Sudan University of Science and Technology
College of Graduate Studies

# Geometric Mean for Toeplitz-Block and Toeplitz Lemma in Geodesic Metric Space

# الوسط الهندسي لكتلة تبوليتز وتمهيدية تبوليتز في الفضاء المتري الجيوديسك

**A Thesis Submitted in Fulfillment of the Requirements for the Degree of Ph.D in Mathematics**

**By**
**Saad Mohammed Eissa Dafalla**

**Supervisor**
**Prof. Dr. Shawgy Hussein AbdAlla**

2022

# Dedication

To my Family.

# Acknowledgements

I would like to thank with all sincerity Allah, and my family for their supports throughout my study. Many thanks are due to my thesis guide, Prof. Dr. Shawgy Hussein AbdAlla Sudan University of Science and Technology.

# Abstract

The effective and computing matrix geometric means satisfying the Ando-Li- Mathias properties are studied. The geometric means of structured matrices and for Toeplitz and Toeplitz -block-block-Toeplitz matrices are also studied. We show the regular operator mappings, monotone and multiplicative geometric means for a class of Toeplitz matrices and the Kähler mean of block – Toeplitz matrices with Toeplitz structured blocks. We explain and develop Beurling-Malliavin theory and geometric significance for Toeplitz Kernels with multipliers between model spaces and improved Toeplitz order. We determine the spectral gaps for sets, measures, two weight inequality for Hilbert transform, higher order multi-dimensional extensions of Cesàro theorem, and law- of large numbers for weighted inductive means in a Hadamard space. The Toeplitz lemma in geodesic metric space, convergence in probability, mean convergence, complete convergence with moment and convergence of inductive means are investigated and obtained.

# الخلاصة

قمنا بدراسة فعالية وحوسبة الأوساط الهندسية للمصفوفة التي تحقق خصائص – اندو- لي – ماثياس. ايضاً درسنا الأوساط الهندسية للمصفوفات المشيدة ولأجل تبوليتز ومصفوفات تبوليتز – كتلة – كتلة – تبوليتز. تم توضيح رواسم المؤثر المنتظمة والأوساط الهندسية الرتيبة والضربية لأجل عائلة مصفوفات تبوليتز ومتوسط كاهلر لمصفوفات كتلة – تبوليتز مع كتل تشييد تبوليتز. تم شرح وتطوير نظرية بيرلينغ- ماليافين والذاتية الهندسية لأجل نويات تبوليتز مع المضاعفات بين فضاءات النموذج ورتبة تبوليتز المحسنة. قمنا بتحديد الفجوات الطيفية لأجل الفئات والقياسات ومتباينة المرجحين لتحويل هلبرت وتمديدات متعدد بعد الرتبة العليا لمبرهنة سيزارو والقانون للأعداد الكبيرة لاجل أوساط الإستنتاج المرجحة في فضاء هادامارد. تمت مناقشة والحصول على تمهيدية تبوليتز في الفضاء المتري الجيوديسك وتقارب الإحتمال والتقارب المتوسط والتقارب التام مع العزم والتقارب للأوساط الأستنتاجية.

# Introduction

We propose a new matrix geometric mean` satisfying the ten properties given by Ando, Li and Mathias. This mean is the limit of a sequence which converges superlinearly with convergence of order 3 whereas the mean introduced by Ando, Li and Mathias is the limit of a sequence having order of convergence 1. A new definition is introduced for the matrix geometric mean of a set of $k$ positive definite $n \times n$ matrices together with an iterative method for its computation. The iterative method is locally convergent with cubic convergence and requires $O(n^3k^2)$ arithmetic operations per step whereas the methods based on the symmetrization technique of Ando, Li and Mathias have complexity $O(n^3k!\,2^k)$. The geometric mean of positive definite matrices is usually identified with the Karcher mean, which possesses all properties – generalized from the scalar case– a geometric mean is expected to satisfy. Unfortunately, the Kärcher mean is typically not structure preserving, and destroys, e.g., Toeplitz and band structures, which emerge in many applications. The Kärcher mean is not always recommended for modeling averages of structured matrices.

We consider the family of Toeplitz operators $T_{J\bar{S}a}$ acting in the Hardy space $H^2$ in the upper halfplane; $J$ and $S$ are given meromorphic inner functions, and a is a real parameter. If $X$ is a closed subset of the real line, denote by $G_X$ the supremum of the size of the gap in the Fourier spectrum, taken over all non-trivial finite complex measures supported on $X$. Let σ and w be locally finite positive Borel measures on ℝ which do not share a common point mass. Assume that the pair of weights satisfy a Poisson $A_2$ condition, and satisfy the testing conditions below, for the Hilbert transform $H$, $\int_I \; H(\sigma 1_I)^2 dw \leq \sigma(I)$, $\int_I \; H(w1_I)^2 \, d\sigma \leq w(I)$, with constants independent of the choice of interval I. Then $H(\sigma \cdot )\; maps\; L^2\; (\sigma)$ to $L^2(w)$, verifying a conjecture of Nazarov–Treil–Volberg.

Three examples are provided which demonstrate that “convergence in probability” versions of the Toeplitz lemma, the Cesàro mean convergence theorem, and the Kronecker lemma can fail. We introduce the notion of regular operator mappings of several variables generalising the notion of spectral function. This setting is convenient for studying maps more general than what can be obtained from the functional calculus, and it allows for Jensen type inequalities and multivariate non-commutative perspectives. As a main application of the theory we consider geometric means of k operator variables extending the geometric mean of k commuting operators and the geometric mean of two arbitrary positive definite matrices. We study the Toeplitz lemma, the

Cesàro mean convergence theorem and the Kronecker lemma. We study "complete convergence" versions of the Toeplitz lemma, the Cesàro mean convergence theorem and the Kronecker lemma. Two counterexamples show that they can fail in general and some sufficient conditions for "complete convergence" version of the Cesàro mean convergence theorem are given.

We introduce using Laurent operators and Fourier coefficients of their symbol functions, a geometric mean for a large class of $n \times n$ positive semi-definite Toeplitz matrices which satisfies the monotonicity property. When one computes an average of positive definite (PD) matrices, the preservation of additional matrix structure is desirable for interpretations in applications. An interesting and widely present structure is that of PD Toeplitz matrices, which we endow with a geometry originating in signal processing theory. As an averaging operation, we consider the barycenter, or minimizer of the sum of squared intrinsic distances. The resulting barycenter, the Kähler mean, is discussed along with its origin. Using the symbol functions and their associated Fourier series, we introduce a new definition of geometric mean for all positive semi-definite Toeplitz matrices and positive semi-definite block-Toeplitz matrices with Toeplitz structured blocks (TBBT).

Let $L^2$ be the Lebesgue space of square-integrable functions on the unit circle. We show that the injectivity problem for Toeplitz operators is linked to the existence of geodesics in the Grassmann manifold of $L^2$ . A new approach to problems of the Uncertainty Principle in Harmonic Analysis, based on the use of Toeplitz operators, has brought progress to some of the classical problems in the area. We develop and systematize the function theoretic component of the Toeplitz approach by introducing a partial order on the set of inner functions induced by the action of Toeplitz operators.

The Cesàro theorem is extended to the cases: higher order Ces`aro mean for sequence (discrete case); and higher order, multi-dimensional and continuous Ces`aro mean for functions. We first study the law of large numbers for weighted inductive means of independent identically distributed random variables taking values in a Hadamard space. We study the Toeplitz lemma for inductive means in a geodesic metric space and by using the Toeplitz lemma, we prove the Cesaro theorem for inductive means.

# The Contents

# Chapter 1
# An Effective Matrix Geometric Mean

We make the new mean very easily computable. We provide a geometric interpretation and a generalization which includes as special cases the mean and the Ando-Li-Mathias mean. We show that the new mean is obtained from the properties of the centroid of a triangle rephrased in terms of geodesics in a suitable Riemannian geometry on the set of positive definite matrices. It satisfies most part of the 10 properties stated by Ando, Li and Mathias; a counterexample shows that monotonicity is not fulfilled. We show a new definition of a geometric mean for structured matrices, its properties are outlined, algorithms for its computation, and numerical experiments are provided. In the Toeplitz case an existing mean based on the Kähler metric is analyzed for comparison.

## Section (1.1): Geometric Mean Satisfying the Ando–Li–Mathias Properties

It is natural to generalize the geometric mean of two positive real numbers $a\#b:=\sqrt{ab}$ to real symmetric positive definite $n\times n$ matrices as

$$A\#B:=A(A^{-1}B)^{\frac{1}{2}}=A^{\frac{1}{2}}\left(A^{-\frac{1}{2}}BA^{-\frac{1}{2}}\right)^{\frac{1}{2}}A^{\frac{1}{2}}. \qquad (1)$$

Several, e.g. [4], [5], [10], and [3] are devoted to studying the geometry of the cone of positive definite matrices $\mathbb{P}^n$ endowed with the Riemannian metric defined by

$$ds=\|A^{-1/2}dAA^{-1/2}\|,$$

where $\| B \|=\sqrt{\sum_{i,j}\left|b_{i,j}\right|^2}$ denotes the Frobenius norm. The distance induced by this metric is

$$d(A,B)=\left\|\log\left(A^{-\frac{1}{2}}BA^{-\frac{1}{2}}\right)\right\|. \qquad (2)$$

It turns out that on this manifold the geodesic joining $X$ and $Y$ has the equation

$$\gamma(t)=X^{1/2}\left(X^{-1/2}YX^{-1/2}\right)^tX^{1/2}=X(X^{-1}Y)^t=:X\#_tY,t\in[0,1],$$

and thus $A\#B$ is the midpoint of the geodesic joining $A$ and $B$. An analysis of numerical methods for computing the geometric mean of two matrices is carried out in [8].

It is less clear how to define the geometric mean of more than two matrices. In [2], Ando, Li and Mathias list ten properties that a "good" matrix geometric mean should satisfy, and they show that several natural approaches based on a generalization of formulas working for the scalar case, or for the case of two matrices, do not work well. They propose a new definition for the mean of $k$ matrices satisfying all the requested properties. We refer to this mean as to the Ando-Li-Mathias mean, or the ALM-mean, for short.

The ALM-mean is the limit of a recursive iteration process where at each step of the iteration $k$ geometric means of $k-1$ matrices must be computed. One of the main drawbacks of this iteration is its linear convergence. In fact, the large number of iterations needed to approximate each geometric mean at all the recursive steps makes it quite expensive to actually compute the ALM-mean with this algorithm. No other algorithms endowed with a higher efficiency are known.

A class of geometric means satisfying the Ando, Li, Mathias requirements has been introduced in [9]. These means are defined in terms of the solution of certain matrix equations. This approach provides interesting theoretical properties concerning the means but no effective tools for their computation.

We propose a new matrix geometric mean satisfying the ten properties of Ando, Li and Mathias. Similar to the ALM-mean, our mean is defined as the limit of an iteration process with the relevant difference that convergence is superlinear with order of convergence at least three. This property makes it much less expensive to compute this geometric mean since the number of iterations required to reach a high accuracy is dropped down to just a few.

The iteration on which our mean is based has a simple geometrical interpretation. In the case $k = 3$, given the positive definite matrices $A_1, A_2, A_3$, we generate three matrix sequences $A_1^{(m)}, A_2^{(m)}, A_3^{(m)}$ starting from $A_i^{(0)} = A_i, i = 1,2,3$. At the step $m + 1$, the matrix $A_1^{(m+1)}$ is chosen along the geodesic which connects $A_1^{(m)}$ with the midpoint of the geodesic connecting $A_2^{(m)}$ to $A_3^{(m)}$ at distance $2/3$ from $A_1^{(m)}$. The matrices $A_2^{(m+1)}$ and $A_3^{(m+1)}$ are similarly defined. In the case of Euclidean geometry, just one step of the iteration provides the value of the limit, i.e., the centroid of the triangle with vertices $A_1^{(m)}, A_2^{(m)}, A_3^{(m)}$. In fact, the medians in a triangle intersect each other at $2/3$ of their length. In the different geometry of the cone of positive definite matrices, the geodesics which play the role of the medians might not even intersect each other.

In the case of $k$ matrices $A_1, A_2, \ldots, A_k$, the matrix $A_i^{(m+1)}$ is chosen along the geodesic which connects $A_i^{(m)}$ with the geometric mean of the remaining matrices, at distance $k/(k + 1)$ from $A_i^{(m)}$. In the customary geometry, this point is the common intersection point of all the "medians" of the $k$-dimensional simplex formed by all the matrices $A_i^{(m)}, i = 1, \ldots, k$. We prove that the sequences $\left(A_i^{(m)}\right)_{m=1}^{\infty}$, $i = 1, \ldots, k$, converge to a common limit $\bar{A}$ with order of convergence at least 3. The limit $\bar{A}$ is our definition of the geometric mean of $A_1, \ldots, A_k$.

It is interesting to point out that our mean and the ALM-mean of $k$ matrices can be viewed as two specific instances of a class of more general means depending on $k - 1$ parameters $s_i \in [0,1], i = 1, \ldots, k - 1$. All the means of this class satisfy the requirements of Ando, Li and Mathias; moreover, the ALM-mean is obtained with $s = (1,1, \ldots ,1,1/2)$, for $s = (s_i)$, while our mean is obtained with $\boldsymbol{s} = ((k - 1)/k, (k - 2)/(k - 1), \ldots ,1/2)$. The new mean is the only one in this class for which the matrix sequences generated at each recursive step converge superlinearly.

We present the ten Ando-Li-Mathias properties and briefly describe the ALM-mean; then, we propose our new definition of a matrix geometric mean and prove some of its properties by also giving a geometrical interpretation; we provide a generalization which includes the ALM-mean and our mean as two special cases. We present some numerical experiments of explicit computations involving this means concerning some problems from physics. It turns out that, in the case of six matrices, the increased speed reached by our approach with respect to the ALM-mean is by a factor greater than 200. We also experimentally demonstrate that the ALM-mean is different, even though very close, from our mean. Finally, for $k = 3$ we provide a pictorial description of the parametric family of geometric means.

We use the positive semidefinite ordering defined by $A \geq B$ if $A - B$ is positive semidefinite. We denote by $A^*$ the conjugate transpose of $A$.

Ando, Li and Mathias [2] proposed the following list of properties that a "good" geometric mean $G(\cdot)$ of three matrices should satisfy.

**P1:** Consistency with scalars. If $A, B, C$ commute, then $(A, B, C) = (ABC)^{1/3}$.

**P2:** Joint homogeneity. $G(\alpha A, \beta B, \gamma C) = (\alpha\beta\gamma)^{1/3} G(A, B, C)$.

**P3:** Permutation invariance. For any permutation $\pi(A, B, C)$ of $A, B, C$ it follows that $G(A, B, C) = G(\pi(A, B, C))$.

**P4:** Monotonicity. If $A \geq A', B \geq B', C \geq C'$, then $G(A, B, C) \geq G(A', B', C')$.

**P5:** Continuity from above. If $A_n, B_n, C_n$ are monotonically decreasing sequences converging to $A, B, C$, respectively, then $G(A_n, B_n, C_n)$ converges to $G(A, B, C)$.

**P6:** Congruence invariance. For any nonsingular , $G(S^*AS, S^*BS, S^*CS) = S^*G(A, B, C)S$.

**P7:** Joint concavity. If $= \lambda A_1 + (1-\lambda)A_2, B = \lambda B_1 + (1-\lambda)B_2, C = \lambda C_1 + (1-\lambda)C_2$, then $G(A, B, C) \geq \lambda G(A_1, B_1, C_1) + (1-\lambda)G(A_2, B_2, C_2)$.

**P8:** Self-duality. $G(A, B, C)^{-1} = G(A^{-1}, B^{-1}, C^{-1})$.

**P9:** Determinant identity. $\det G(A, B, C) = (\det A \det B \det C)^{1/3}$.

**P10:** Arithmetic-geometric-harmonic mean inequality:

$$\frac{A + B + C}{3} \geq G(A, B, C) \geq \left(\frac{A^{-1} + B^{-1} + C^{-1}}{3}\right)^{-1}.$$

It is proved in [2] that P5 and P10 are consequences of the others. All these properties can be easily generalized to the mean of any number of matrices. We will call a geometric mean of three or more matrices any map $G(\cdot)$ satisfying P1-P10 or their analogues for a number $k \geq 3$ of entries.

We denote by $G_2(A, B)$ the usual geometric mean $A\#B$ and, given the $k$-tuple $(A_1, \dots, A_k)$, we define

$$\mathcal{Z}_i(A_1, \dots, A_k) = (A_1, \dots, A_{i-1}, A_{i+1}, \dots, A_k), i = 1, \dots, k,$$

that is, the $k$-tuple where the $i$-th term has been dropped out.

In [2], Ando, Li and Mathias note that the previously proposed definitions of means of more than two matrices do not satisfy all the properties P1-P10, and they propose a new definition that fulfills all of them. Their mean is defined inductively on the number of arguments $k$.

Given $A_1, \dots, A_k$ positive definite, and given the definition of a geometric mean $G_{k-1}(\cdot)$ of $k-1$ matrices, they set $A_i^{(0)} = A_i, i = 1, \dots, k$, and define for $r \geq 0$

$$A_i^{(r+1)} := G_{k-1}\left(\mathcal{Z}_i\left(A_1^{(r)}, \dots, A_k^{(r)}\right)\right), i = 1, \dots, k. \quad (3)$$

For $k = 3$, the iteration reads

$$\begin{bmatrix} A^{(r+1)} \\ B^{(r+1)} \\ C^{(r+1)} \end{bmatrix} = \begin{bmatrix} G_2\left(B^{(r)}, C^{(r)}\right) \\ G_2\left(A^{(r)}, C^{(r)}\right) \\ G_2\left(A^{(r)}, B^{(r)}\right) \end{bmatrix}.$$

Ando, Li and Mathias show that the $k$ sequences $\left(A_i^{(r)}\right)_{r=1}^{\infty}$ converge to the same matrix $\tilde{A}$, and finally they define $G_k(A_1, \dots, A_k) = \tilde{A}$. In the following, we shall denote by $G(\cdot)$ the Ando-Li-Mathias mean, dropping the subscript $k$ when not essential.

An additional property of the Ando-Li-Mathias mean which will turn out to be important in the convergence proof is the following. Recall that $\rho(X)$ denotes the spectral radius of $X$, and let

$$R(A, B) := \max\ \left(\rho(A^{-1}B), \rho(B^{-1}A)\right).$$

This function is a multiplicative metric; that is, we have $R(A,B) \geq 1$ with equality iff $A = B$, and

$$R(A,C) \leq R(A,B)R(B,C).$$

The additional property is

**P11:** For each $k \geq 2$, and for each pair of sequences $(A_1, \dots, A_k), (B_1, \dots, B_k)$, we require that

$$R\big(G(A_1, \dots, A_k), G(B_1, \dots, B_k)\big) \leq \left(\prod_{i=1}^{k} R(A_i, B_i)\right)^{1/k}$$

**Definition (1.1.1)[1]:** We are going to define for each $k \geq 2$ a new mean $\bar{G}_k(\cdot)$ of $k$ matrices satisfying $\mathrm{P1} - \mathrm{P11}$. Let $\bar{G}_2(A,B) = A\#B$, and suppose that the mean has already been defined for up to $k-1$ matrices. Let us denote, for short, $T_i^{(r)} = \bar{G}_{k-1}\left(\mathcal{Z}_i\left(\bar{A}_1^{(r)}, \dots, \bar{A}_k^{(r)}\right)\right)$ and define $\bar{A}_i^{(r+1)}$ for $i = 1, \dots, k$ as

$$\bar{A}_i^{(r+1)} := \bar{G}_k(\bar{A}_i^{(r)}, \underbrace{T_i^{(r)}, T_i^{(r)}, \dots, T_i^{(r)}}_{k-1 \text{ times}}), \tag{4}$$

with $\bar{A}_i^{(0)} = A_i$ for all $i$. Notice that apparently this requires the mean $\bar{G}_k(\cdot)$ to already be defined; in fact, in the special case in which $k-1$ of the $k$ arguments are coincident, the properties P1 and P6 alone allow one to determine the common value of any geometric mean:

$$\begin{aligned} G(X,Y,Y,\dots,Y) &= X^{1/2} G\big(I, X^{-1/2} Y X^{-1/2}, \dots, X^{-1/2} Y X^{-1/2}\big) X^{1/2} \\ &= X^{1/2}\big(X^{-1/2} Y X^{-1/2}\big)^{\frac{k-1}{k}} X^{1/2} = X \#_{\frac{k-1}{k}} Y. \end{aligned}$$

Thus we can use this simpler expression directly in (4) and set

$$\bar{A}_i^{(r+1)} = \bar{A}_i^{(r)} \#_{\frac{k-1}{k}} T_i^{(r)}. \tag{5}$$

We prove that the $k$ sequences $\left(\bar{A}_i^{(r)}\right)_{r=1}^{\infty}$ converge to a common limit $\bar{A}$ with order of convergence at least three, and this will enable us to define $\bar{G}_k(A_1, \dots, A_k) := \bar{A}$. In the following, we will drop the index $k$ from $\bar{G}_k(\cdot)$.

In [4], an interesting geometrical interpretation of the Ando-Li-Mathias mean is proposed for $k = 3$. We propose an interpretation of the new mean $\bar{G}(\cdot)$ in the same spirit. For $k = 3$, the iteration defining $\bar{G}(\cdot)$ reads

$$\begin{bmatrix} \bar{A}^{(r+1)} \\ \bar{B}^{(r+1)} \\ \bar{C}^{(r+1)} \end{bmatrix} = \begin{bmatrix} \bar{A}^{(r)} \#_{\frac{2}{3}} \big(\bar{B}^{(r)} \# \bar{C}^{(r)}\big) \\ \bar{B}^{(r)} \#_{2} \big(\bar{A}^{(r)} \# \bar{C}^{(r)}\big) \\ \bar{C}^{(r)} \#_{\frac{2}{3}} \big(\bar{A}^{(r)} \# \bar{B}^{(r)}\big) \end{bmatrix}.$$

We can interpret this iteration as a geometrical construction in the following way. To find e.g. $\bar{A}^{(r+1)}$, the algorithm is:

(i) Draw the geodesic joining $\bar{B}^{(r)}$ and $\bar{C}^{(r)}$, and take its midpoint $M^{(r)}$.

(ii) Draw the geodesic joining $\bar{A}^{(r)}$ and $M^{(r)}$, and take the point lying at $2/3$ of its length: this is $\bar{A}^{(r+1)}$.

If we execute the same algorithm on the Euclidean plane, replacing the word "geodesic" with "straight line segment", it turns out that $\bar{A}^{(1)}, \bar{B}^{(1)}$, and $\bar{C}^{(1)}$ coincide in the centroid of the triangle with vertices $A, B, C$. Thus, unlike the Euclidean counterpart of the Ando-Li-Mathias mean, this process converges in one step on the plane. When $A, B$ and $C$ are very close to each other, we can approximate (in some intuitive sense) the geometry on the Riemannian manifold $\mathbb{P}^n$ with the geometry on the Euclidean plane: since this construction to find the centroid of a plane triangle converges faster than the Ando-Li-Mathias one, we can expect that also the convergence speed of the resulting algorithm is faster. This is indeed what will result after a more accurate convergence analysis.

In order to prove that the iteration (5) is convergent (and thus that $\bar{G}(\cdot)$ is well defined), we adapt a part of the proof of Theorem (1.1.3) of [2] (namely, Argument 1).

**Theorem (1.1.2)[1]:** Let $A_1, \dots, A_k$, be positive definite.

(i) All the sequences $\left(\bar{A}_i^{(r)}\right)_{r=1}^{\infty}$ converge for $r \to \infty$ to a common limit $\bar{A}$.

(ii) The function $\bar{G}_k(A_1, \dots, A_k)$ satisfies P1-P11.

**Proof.** We work by induction on $k$. For $k = 2$, our mean coincides with the ALMmean, so all the required work has been done in [2]. Let us now suppose that the thesis holds for all $k' \le k - 1$. We have

$$\bar{A}_i^{(r+1)} \le \frac{1}{k}\left(\bar{A}_i^{(r)} + (k-1)T_i^{(r)}\right) \le \frac{1}{k}\sum_{i=1}^{k} \bar{A}_i^{(r)},$$

where the first inequality follows from P10 for the ALM-mean $G_k(\cdot)$ (remember that in the special case in which $k - 1$ of the arguments coincide, $G_k(\cdot) = \bar{G}_k(\cdot)$), and the second follows from P10 for $\bar{G}_{k-1}(\cdot)$. Thus,

$$\sum_{i=1}^{k} \bar{A}_i^{(r+1)} \le \sum_{i=1}^{k} \bar{A}_i^{(r)} \le \sum_{i=1}^{k} A_i. \tag{6}$$

Therefore, the sequence $\left(\bar{A}_1^{(r)}, \dots, \bar{A}_k^{(r)}\right)_{r=1}^{\infty}$ is bounded, and there must be a converging subsequence, say, converging to $(\bar{A}_1, \dots, \bar{A}_k)$.

Moreover, for each $p, q \in \{1, \dots, k\}$ we have

$$\begin{aligned} R\left(\bar{A}_p^{(r+1)}, \bar{A}_q^{(r+1)}\right) &\le R\left(\bar{A}_p^{(r)}, \bar{A}_q^{(r)}\right)^{1/k} R\left(T_p^{(r)}, T_q^{(r)}\right)^{\frac{k-1}{k}} \\ &\le R\left(\bar{A}_p^{(r)}, \bar{A}_q^{(r)}\right)^{1/k} \left(R\left(\bar{A}_q^{(r)}, \bar{A}_p^{(r)}\right)^{\frac{1}{k-1}}\right)^{\frac{k-1}{k}} = R\left(\bar{A}_p^{(r)}, \bar{A}_q^{(r)}\right)^{2/k}, \end{aligned}$$

where the first inequality follows from P11 in the special case, and the second follows from P11 in the inductive hypothesis. Passing to the limit of the converging subsequence, one can verify that

$$R\left(\bar{A}_p, \bar{A}_q\right) \le R\left(\bar{A}_p, \bar{A}_q\right)^{2/k},$$

from which we get $R\left(\bar{A}_p, \bar{A}_q\right) \le 1$, that is, $\bar{A}_p = \bar{A}_q$, because of the properties of $R$; i.e., the limit of the subsequence is in the form $(\bar{A}, \bar{A}, \dots, \bar{A})$. Suppose there is another subsequence converging to $(\bar{B}, \bar{B}, \dots, \bar{B})$; then, by (6), we have both $k\bar{A} \le k\bar{B}$ and $k\bar{B} \le k\bar{A}$, that is, $\bar{A} = \bar{B}$. Therefore, the sequence has only one limit point; thus it is convergent. This proves the first point of the theorem.

We now turn to show that P11 holds for our mean $\bar{G}_k(\cdot)$. Consider the $k$-tuples $A_1, \dots, A_k$ and $B_1, \dots, B_k$, and let $\bar{B}_i^{(r)}$ be defined as $\bar{A}_i^{(r)}$ but starting the iteration from the $k$-tuple $(B_i)$ instead of $(A_i)$. We have for each $i$,

$$\begin{aligned} R\left(\bar{A}_i^{(r+1)},\quad \bar{B}_i^{(r+1)}\right) & \\ &\leq R\left(\bar{A}_i^{(r)}, \bar{B}_i^{(r)}\right)^{1/k} R\left(\bar{G}\left(\mathcal{Z}_i\left(\bar{A}_1^{(r)}, \dots, \bar{A}_k^{(r)}\right)\right), \bar{G}\left(\mathcal{Z}_i\left(\bar{B}_1^{(r)}, \dots, \bar{B}_k^{(r)}\right)\right)\right)^{\frac{k-1}{k}} \\ &\leq R\left(\bar{A}_i^{(r)}, \bar{B}_i^{(r)}\right)^{1/k} \left(\prod_{j\neq i} R\left(\bar{A}_j^{(r)}, \bar{B}_j^{(r)}\right)^{\frac{1}{k-1}}\right)^{\frac{k-1}{k}} \\ &= \prod_{j=1}^{k} R\left(\bar{A}_j^{(r)}, \bar{B}_j^{(r)}\right)^{1/k}. \end{aligned}$$

This yields

$$\prod_{i=1}^{k} R\left(\bar{A}_i^{(r+1)}, \bar{B}_i^{(r+1)}\right) \leq \prod_{i=1}^{k} R\left(\bar{A}_i^{(r)}, \bar{B}_i^{(r)}\right);$$

chaining together these inequalities for successive values of $r$ and passing to the limit, we get

$$R\big(G(A_1, \dots, A_k), G(B_1, \dots, B_k)\big)^k \leq \prod_{i=1}^{k} R(A_i, B_i),$$

which is P11.

The other properties P1-P4 and P6-P9 (remember that P5 and P10 are consequences of these) are not difficult to prove. All the proofs are quite similar, and can be established by induction, using also the fact that since they hold for the ALM-mean, they can be applied to the mean $\bar{G}(\cdot)$ appearing in (5) (since we just proved that all possible geometric means take the same value if applied with $k-1$ equal arguments). We provide only the proof for three of these properties.

**P1:** We need to prove that if the $A_i$ commute, then $\bar{G}(A_1, \dots, A_k) = (A_1 \cdots A_k)^{1/k}$. Using the inductive hypothesis, we have $T_i^{(1)} = \prod_{j\neq i} \bar{A}_i^{\frac{1}{k-1}}$. Using the fact that P1 holds for the ALM-mean, we have

$$\bar{A}_i^{(1)} = A_i^{1/k} \left(\prod_{j\neq i} A_j^{\frac{1}{k-1}}\right)^{\frac{k-1}{k}} = \prod_{i=1}^{k} A_i^{1/k},$$

as needed. So, from the second iteration on, we have $\bar{A}_1^{(r)} = \bar{A}_2^{(r)} = \dots = \bar{A}_k^{(r)} = \prod_{i=1}^{k} A_i^{1/k}$.

**P4:** Let $T_i'^{(r)}$ and $\bar{A}_i'^{(r)}$ be defined as $T_i^{(r)}$ and $\bar{A}_i^{(r)}$ but starting from $A_i' \leq A_i$.

Using monotonicity in the inductive case and in the ALM-mean, we have for each $s \leq 1$ and for each $i$,

$$T_i^{(r+1)} \leq T_i'^{(r+1)}$$

and thus

$$\bar{A}_i^{(r+1)} \leq \bar{A}_i'^{(r+1)}.$$

Passing to the limit for $r \to \infty$, we obtain P4.

**P7:** Suppose $A_i = \lambda A_i' + (1-\lambda)A_i''$, and let $T_i'^{(r)}$ (resp. $T_i''^{(r)}$) and $\bar{A}_i'^{(r)}$ (resp. $\bar{A}_i''^{(r)}$ ) be defined as $T_i^{(r)}$ and $\bar{A}_i^{(r)}$ but starting from $A_i'$ (resp. $A_i''$ ). Suppose that for some $r$ we have $\bar{A}_i^{(r)} \geq \lambda\bar{A}_i'^{(r)} + (1-\lambda)\bar{A}_i''^{(r)}$ for all $i$. Then by joint concavity and monotonicity in the inductive case we have

$$\begin{aligned} T_i^{(r+1)} &= \bar{G}\left(\mathcal{Z}_i\left(\bar{A}_1^{(r)}, \ldots, \bar{A}_k^{(r)}\right)\right) \\ &\geq \bar{G}\left(\mathcal{Z}_i\left(\lambda\bar{A}_1'^{(r)} + (1-\lambda)\bar{A}_1''^{(r)}, \ldots, \lambda\bar{A}_k'^{(r)} + (1-\lambda)\bar{A}_k''^{(r)}\right)\right) \\ &\geq \lambda T_i'^{(r)} + (1-\lambda)T_i''^{(r)}, \end{aligned}$$

and by joint concavity and monotonicity of the Ando-Li-Mathias mean we have

$$\begin{aligned} \bar{A}_i^{(r+1)} &= \bar{A}_i^{(r)} \# \frac{k-1}{k} T_i^{(r)} \\ &\geq \left(\lambda\bar{A}_i'^{(r)} + (1-\lambda)\bar{A}_i''^{(r)}\right) \#_{\frac{k-1}{k}} \left(\lambda T_i'^{(r)} + (1-\lambda)T_i''^{(r)}\right) \\ &\geq \lambda\bar{A}_i'^{(r+1)} + (1-\lambda)\bar{A}_i''^{(r+1)}. \end{aligned}$$

Passing to the limit for $r \to \infty$, we obtain P7.

We will use the big-O notation in the norm sense; that is, we will write $X = Y + O(\varepsilon^h)$ to denote that there are universal positive constants $\varepsilon_0 < 1$ and $\theta$ such that for each $0 < \varepsilon < \varepsilon_0$ it follows that $\| X - Y \| \leq \theta\varepsilon^h$. The usual arithmetic rules involving this notation hold. These constants may depend on $k$, but not on the specific choice of the matrices involved in the formulas.

**Theorem (1.1.3)[1]:** Let $0 < \varepsilon < 1, M$ and $\bar{A}_i^{(0)} = A_i, i = 1, \ldots, k$, be positive definite $n \times n$ matrices, and $E_i := M^{-1}A_i - I$. Suppose that $\|E_i\| \leq \varepsilon$ for all $i = 1, \ldots, k$. Then, for the matrices $\bar{A}_i^{(1)}$ defined in (5) the following hold.

**C1:** We have

$$M^{-1}\bar{A}_i^{(1)} - I = T_k + O(\varepsilon^3), \tag{7}$$

where

$$T_k := \frac{1}{k}\sum_{j=1}^{k} E_j - \frac{1}{4k^2}\sum_{i,j=1}^{k} \left(E_i - E_j\right)^2.$$

**C2:** There are positive constants $\theta, \sigma$ and $\bar{\varepsilon} < 1$ (all of which may depend on k) such that for all $\varepsilon \leq \bar{\varepsilon}$,

$$\left\|M_1^{-1}\bar{A}_i^{(1)} - I\right\| \leq \theta\varepsilon^3$$

for a suitable matrix $M_1$ satisfying $\|M^{-1}M_1 - I\| \leq \sigma\varepsilon$.

**C3:** The iteration (5) converges at least cubically.

**C4:** We have

$$M_1^{-1}\bar{G}(A_1, \ldots, A_k) - I = O(\varepsilon^3). \tag{8}$$

**Proof.** Let us first find a local expansion of a generic point on the geodesic $\#_t B$ : let $M^{-1}A = I + F_1$ and $M^{-1}B = I + F_2$ with $\|F_1\| \leq \delta, \|F_2\| \leq \delta, 0 < \delta < 1$. Then we have

$$
\begin{aligned}
M^{-1}(A\#_t B) = \quad & M^{-1}A(A^{-1}B)^t = (I+F_1)\big((I+F_1)^{-1}(I+F_2)\big)^t \\
= \quad & (I+F_1)\Big(\big(I-F_1+F_1^2+O(\delta^3)\big)(I+F_2)\Big)^t \\
= \quad & (I+F_1)\big(I+F_2-F_1-F_1F_2+F_1^2+O(\delta^3)\big)^t \qquad (9)\\
= \quad & (I+F_1)(I+t(F_2-F_1-F_1F_2+F_1^2) \\
& +\frac{t(t-1)}{2}(F_2-F_1)^2+O(\delta^3)\Bigg) \\
= I+(1-t)F_1+tF_2+\frac{t(t-1)}{2}(F_2-F_1)^2+O(\delta^3),
\end{aligned}
$$

where we have made use of the matrix series expansion $(I+X)^t = I+tX+\frac{t(t-1)}{2}X^2+O(X^3)$. Now, we prove the theorem by induction on $k$ in the following way. Let $\mathrm{C}i_k$ denote the assertion $\mathrm{C}i$ of the theorem (for $i=1,\dots,4$) for a given value of $k$. We show that

(i) $\mathrm{C1}_2$ holds;

(ii) $\mathrm{C1}_k \Longrightarrow \mathrm{C2}_k$;

(iii) $\mathrm{C2}_k \Longrightarrow \mathrm{C3}_k, \mathrm{C4}_k$;

(iv) $\mathrm{C4}_k \Longrightarrow \mathrm{C1}_{k+1}$.

It is clear that these claims imply that the results $\mathrm{C1}-\mathrm{C4}$ hold for all $k \geq 2$ by induction; we will now turn to prove them one by one.

(i) This is simply equation (9) for $t=\frac{1}{2}$.

(ii) It is obvious that $T_k = O(\varepsilon)$; thus, choosing $M_1 := M(I+T_k)$ one has

$$\bar{A}_i^{(1)} = M\big(I+T_k+O(\varepsilon^3)\big) = M_1\big(I+(I+T_k)^{-1}O(\varepsilon^3)\big) = M_1\big(I+O(\varepsilon^3)\big). \quad (10)$$

Using explicit constants in the big-O estimates, we get

$$\left\|M_1^{-1}\bar{A}_i^{(1)}-I\right\| \leq \theta\varepsilon^3,\ \|M^{-1}M_1-I\| \leq \sigma\varepsilon$$

for suitable constants $\theta$ and $\sigma$.

(iii) Suppose $\varepsilon$ is small enough to have $\theta\varepsilon^3 \leq \varepsilon$. We shall apply C2 with initial matrices $\bar{A}_i^{(1)}$, with $\varepsilon_1 = \theta\varepsilon^3$ in lieu of $\varepsilon$ and $M_1$ in lieu of $M$, getting

$$\left\|M_2^{-1}\bar{A}_i^{(2)}-I\right\| \leq \theta\varepsilon_1^3,\ \|M_1^{-1}M_2-I\| \leq \sigma\varepsilon_1.$$

Repeating again for all the steps of our iterative process, we get for all $s=1,2,\dots$,

$$\left\|M_s^{-1}\bar{A}_i^{(s)}-I\right\| \leq \theta\varepsilon_{s-1}^3 = \varepsilon_s,\ \|M_s^{-1}M_{s+1}-I\| \leq \sigma\varepsilon_s \qquad (11)$$

with $\varepsilon_{s+1} := \theta\varepsilon_s^3$ and $M_0 := M$.

We introduce the notation

$$d(X,Y) := \|X^{-1}Y-I\|$$

for any two $n\times n$ symmetric positive definite matrices $X$ and $Y$. It will be useful to notice that $\| X-Y \| \leq \| X \| \, \|X^{-1}Y-I\| \leq \| X \| \, d(X,Y)$ and

$$
\begin{aligned}
d(X,Z) \quad & = \| (X^{-1}Y-I)(Y^{-1}Z-I)+X^{-1}Y-I \\
& \leq d(X,Y)d(Y,Z)+d(X,Y)+d(Y,Z).
\end{aligned} \qquad (12)
$$

With this notation, we can restate (11) as

$$d\left(M_s,\bar{A}_i^{(s)}\right) \leq \varepsilon_s,\ d(M_s,M_{s+1}) \leq \sigma\varepsilon_s.$$

We will now prove by induction that, for $\varepsilon$ smaller than a fixed constant, it follows that

$$d(M_s,M_{s+t}) \leq \left(2-\frac{1}{2^t}\right)\sigma\varepsilon_s. \qquad (13)$$

First of all, for all $t \geq 1$,

$$\varepsilon_{s+t} = \theta^{\frac{3^t-1}{2}} \varepsilon^{3^t},$$

which, for $\varepsilon$ smaller than $\min(1/8, \theta^{-1})$, implies $\frac{\varepsilon_{s+t}}{\varepsilon_s} \leq \varepsilon_s^{\frac{3^t-1}{2}} \leq \varepsilon_s^t \leq \frac{1}{2^{t+2}}$.

Now, using (12), and supposing additionally that $\varepsilon \leq \sigma^{-1}$, we have

$$\begin{aligned} d(M_s, M_{s+t+1}) \leq & \ d(M_s, M_{s+t}) d(M_{s+t}, M_{s+t+1}) \\ & + d(M_s, M_{s+t}) + d(M_{s+t}, M_{s+t+1}) \\ \leq & \ \left(2 - \frac{1}{2^t}\right) \sigma\varepsilon_s + \sigma\varepsilon_s \left(\sigma\varepsilon_{s+t} + \frac{\varepsilon_{s+t}}{\varepsilon_s}\right) \\ \leq & \ \left(2 - \frac{1}{2^t}\right) \sigma\varepsilon_s + \sigma\varepsilon_s \left(2 \frac{\varepsilon_{s+t}}{\varepsilon_s}\right) \\ \leq & \ \left(2 - \frac{1}{2^t}\right) \sigma\varepsilon_s + \sigma\varepsilon_s \frac{1}{2^{t+1}} = \left(2 - \frac{1}{2^{t+1}}\right) \sigma\varepsilon_s \end{aligned}$$

Thus, we have for each $t$,

$$\|M_t - M\| \leq \| M \| \|M^{-1} M_t - I\| \leq 2\sigma \| M \| \varepsilon,$$

which implies $\|M_t\| \leq 2 \| M \|$ for all $t$. By a similar argument,

$$\|M_{s+t} - M_s\| \leq \|M_s\| d(M_{s+t}, M_s) \leq 2\sigma \| M \| \varepsilon_s. \tag{14}$$

Due to the bounds already imposed on $\varepsilon$, the sequence $\varepsilon_s$ tends monotonically to zero with a cubic convergence rate; thus $(M_t)$ is a Cauchy sequence and therefore converges. In the following, let $M^*$ be its limit. The convergence rate is cubic, since passing to the limit (14) we get

$$\|M^* - M_s\| \leq 2\sigma \| M \| \varepsilon_s.$$

Now, using the other relation in (11), we get

$$\begin{aligned} \left\|\bar{A}_i^{(s)} - M^*\right\| & \leq \left\|\bar{A}_i^{(s)} - M_s\right\| + \|M^* - M_s\| \\ & \leq 2 \| M \| d\left(M_s, \bar{A}_i^{(s)}\right) + 2\sigma \| M \| \varepsilon_s \\ & \leq (2\sigma + 2) \| M \| \varepsilon_s; \end{aligned}$$

that is, $\bar{A}_i^{(s)}$ converges with cubic convergence rate to $M^*$. Thus C3 is proved. By (12), (13), and (11), we have

$$\begin{aligned} d\left(M_1, \bar{A}_i^{(t)}\right) & \leq d(M_1, M_t) d\left(M_t, \bar{A}_i^{(t)}\right) + d(M_1, M_t) + d\left(M_t, \bar{A}_i^{(t)}\right) \\ & \leq 2\sigma\varepsilon_1\varepsilon_t + 2\sigma\varepsilon_1 + \varepsilon_t \leq (4\sigma + 1)\varepsilon_1 = O(\varepsilon^3), \end{aligned}$$

which is C4.

(iv) Using C4 $_k$ and (9) with $F_1 = E_{k+1}, F_2 = M^{-1}\bar{G}(A_1, \dots, A_k) = T_k + O(\varepsilon^3), \delta = 2k\varepsilon$, we have

$$\begin{aligned} M^{-1}\bar{A}_{k+1}^{(1)} & = M^{-1}\left(A_{k+1} \# \frac{k}{k+1} \bar{G}(A_1, \dots, A_k)\right) \\ & = I + \frac{1}{k+1} E_{k+1} + \frac{k}{k+1} T_k \\ & - \frac{k}{2(k+1)^2} \left(E_{k+1} - \frac{1}{k} \sum_{i=1}^{k} E_i\right)^2 + O(\varepsilon^3). \end{aligned} \tag{15}$$

Observe that

$$T_k = \frac{1}{k} S_k + \frac{P_k - (k-1)Q_k}{2k^2},$$

where $S_k = \sum_{i=1}^{k} E_i, Q_k = \sum_{i=1}^{k} E_i^2, P_k = \sum_{i,j=1,i\neq j}^{k} E_i E_j$. Since $S_k^2 = P_k + Q_k$ and $S_{k+1} = S_k + E_{k+1}, Q_{k+1} = Q_k + E_{k+1}^2, P_{k+1} = P_k + E_{k+1}S_k + S_k E_{k+1}$, from (15) one finds that

$$\begin{aligned} M^{-1}\bar{A}_{k+1}^{(1)} &= I + \frac{1}{k+1}S_{k+1} - \frac{k}{2(k+1)^2}Q_{k+1} + \frac{1}{2(k+1)^2}P_{k+1} + O(\varepsilon^3) \\ &= I + T_{k+1} + O(\varepsilon^3). \end{aligned}$$

Since the expression we found is symmetric with respect to the $E_i$, it follows that $\bar{A}_j^{(1)}$ has the same expansion for any $j$.

Observe that Theorems (1.1.2) and (1.1.3) imply that the iteration (5) is globally convergent with order of convergence at least 3.

In the case where the matrices $A_i, i = 1, \dots, A_k$, commute with each other, the iteration (5) converges in just one step, i.e., $\bar{A}_i^{(1)} = \bar{A}$ for any $i$. In the noncommutative general case, one has $\det\left(\bar{A}_i^{(s)}\right) = \det(\bar{A})$ for any $i$ and for any $s \geq 1$; i.e., the determinant converges in one single step to the determinant of the matrix mean.

Our mean is different from the ALM-mean, as we will show with some numerical experiments. We prove that our mean and the ALM-mean belong to a general class of matrix geometric means, which depends on a set of $k-1$ parameters.

We introduce a new class of matrix means depending on a set of parameters $s_1, \dots, s_{k-1}$ and show that the ALM-mean and our mean are two specific instances of this class.

We describe this generalization in the case of $k = 3$ matrices $A, B, C$. The case $k > 3$ is outlined. Here, the distance between two matrices is defined in (2).

For $k = 3$, the algorithm presented replaces the triple $A, B, C$ with $A', B', C'$ where $A'$ is chosen in the geodesic connecting $A$ with the midpoint of the geodesic connecting $B$ and $C$, at distance $2/3$ from $A$, and a similar choice is made for $B'$ and $C'$. In our generalization we use two parameters $s, t \in [0,1]$. We consider the point $P_t = B\#_t C$ in the geodesic connecting $B$ to $C$ at distance $t$ from $B$. Then we consider the geodesic connecting $A$ to $P_t$ and define $A'$ to be the matrix on this geodesic at a distance $s$ from $A$. That is, we set $A' = A\#_s(B\#_t C)$.

We do a similar step with $B$ and $C$. This transformation is recursively repeated so that the matrix sequences $A^{(r)}, B^{(r)}, C^{(r)}$ are generated by means of

$$\begin{aligned} A^{(r+1)} &= A^{(r)}\#_s\left(B^{(r)}\#_t C^{(r)}\right) \\ B^{(r+1)} &= B^{(r)}\#_s\left(C^{(r)}\#_t A^{(r)}\right), \ r = 0,1,\dots, \\ C^{(r+1)} &= C^{(r)}\#_s\left(A^{(r)}\#_t B^{(r)}\right), \end{aligned} \qquad (16)$$

starting with $A^{(0)} = A, B^{(0)} = B, C^{(0)} = C$.

By following the same arguments, it can be shown that the three sequences have a common limit $G_{s,t}$ for any $s, t \in [0,1], s \neq 0, (s,t) \neq (1,0), (1,1)$.

Moreover, for $s = 1, t = 1/2$ one obtains the ALM-mean, i.e., $G = G_{1,\frac{1}{2}}$, while for $s = 2/3, t = 1/2$ the limit coincides with our mean, i.e., $\bar{G} = G_{\frac{2}{3},\frac{1}{2}}$. Moreover, it is possible to prove that for any $s, t \in [0,1], s \neq 0, (s,t) \neq (1,0), (1,1)$ the limit satisfies the conditions P1-P11 so that it can be considered a good geometric mean.

Concerning the convergence speed of the sequence generated by (16) we may perform a more accurate analysis. Assume that $A = M(I + E_1), B = M(I + E_2), C = M$ yields

$$A' \doteq M\left(I + (1-s)E_1 + s(1-t)E_2 + stE_3 + \frac{st(t-1)}{2}H_2^2 + \frac{s(s-1)}{2}(H_1 + tH_2)^2\right),$$
$$B' \doteq M\left(I + (1-s)E_2 + s(1-t)E_3 + stE_1 + \frac{st(t-1)}{2}H_3^2 + \frac{s(s-1)}{2}(H_2 + tH_3)^2\right),$$
$$C' \doteq M\left(I + (1-s)E_3 + s(1-t)E_1 + stE_2 + \frac{st(t-1)}{2}H_1^2 + \frac{s(s-1)}{2}(H_3 + tH_1)^2\right),$$
where $\doteq$ denotes equality up to $O(\varepsilon^3)$ terms, with $H_1 = E_1 - E_2, H_2 = E_2 - E_3, H_3 = E_3 - E_1$. Hence we have $A' = M(I + E_1'), B' = M(I + E_2'), C' = M(I + E_3')$, with
$$\begin{bmatrix} E_1' \\ E_2' \\ E_3' \end{bmatrix} \doteq C(s,t)\begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix} + \frac{st(t-1)}{2}\begin{bmatrix} H_2^2 \\ H_3^2 \\ H_1^2 \end{bmatrix} + \frac{s(s-1)}{2}\begin{bmatrix} (H_1 - tH_2)^2 \\ (H_2 - tH_3)^2 \\ (H_3 - tH_1)^2 \end{bmatrix},$$
where
$$C(s,t) = \begin{bmatrix} (1-s)I & s(1-t)I & stI \\ stI & (1-s)I & s(1-t)I \\ s(1-t)I & stI & (1-s)I \end{bmatrix}.$$
Observe that the block circulant matrix $C(s,t)$ has eigenvalues $\lambda_1 = 1, \lambda_2 = \left(1 - \frac{3}{2}s\right) + i\frac{\sqrt{3}}{2}s(2t-1)$, and $\lambda_3 = \bar{\lambda}_2$, with multiplicity $n$, where $i^2 = -1$. Moreover, the pair $(s,t) = (2/3, 1/2)$ is the only one which yields $\lambda_2 = \lambda_3 = 0$. In fact $(2/3, 1/2)$ is the only pair which provides superlinear convergence. For the ALMmean, where $t = 1/2$ and $s = 1$, it follows that $|\lambda_2| = |\lambda_3| = 1/2$, which is the rate of convergence of the ALM iteration [2].
In the case of $k > 3$ matrices, given the $(k-1)$-tuple $(s_1, s_2, \dots, s_{k-1})$ we may recursively define $G_{s_1,\dots,s_{k-1}}(A_1, \dots, A_k)$ as the common limit of the sequences generated by
$$A_i^{(r+1)} = A_i^{(r)} \#_{s_1} G_{s_2,\dots,s_{k-1}}\left(\mathcal{Z}_i\left(A_1^{(r)}, \dots, A_k^{(r)}\right)\right), i = 1, \dots, k.$$
Observe that with $(s_1, \dots, s_{k-1}) = (1, 1, \dots, 1, 1/2)$ one obtains the ALM-mean, while with $(s_1, \dots, s_{k-1}) = ((k-1)/k, (k-2)/(k-1), \dots, 1/2)$ one obtains the new mean introduced.
We have implemented the two iterations converging to the ALM-mean and to the newly defined geometric mean in Matlab, and we have run some numerical experiments on a quad-Xeon 2.8Ghz computer. To compute matrix square roots we used Matlab's built-in sqrtm function, while for $p$-th roots with $p > 2$ we used the rootm function in Nicholas Higham's Matrix Computation Toolbox [7]. To counter the loss of symmetry due to the accumulation of computational errors, we chose to discard the imaginary part of the computed roots.
The experiments have been performed on the same data set as [10]. It consists of five sets, each composed of four to six $6 \times 6$ positive definite matrices, corresponding to physical data from elasticity experiments conducted by Hearmon [6]. The matrices are composed of smaller diagonal blocks of sizes $1 \times 1$ up to $4 \times 4$, depending on the symmetries of the involved materials. Two to three significant digits are reported for each experiment.
We have computed both the ALM-mean and the newly defined mean of these sets; as a stopping criterion for each computed mean, we chose
$$\max_i \left|A_i^{(r+1)} - A_i^{(r)}\right| < \varepsilon,$$

where $|X| := \max_{i,j} |X_{ij}|$, with $\varepsilon = 10^{-10}$. The CPU times, in seconds, are reported in Table 1. For four matrices, the speed gain is a factor of 20 , and it increases even more for more than four matrices.
We then focused on Hearmon's second data set (ammonium dihydrogen phosphate), composed of four matrices. In Table 2, we reported the number of outer ($k = 4$) iterations needed and the average number of iterations needed to reach convergence in the inner ($k = 3$) iterations (remember that the computation of a mean of four matrices requires the computation of three means of three matrices at each of its steps). Moreover, we measured the number of square and $p$-th roots needed by the two algorithms, since they are the most expensive operation in the algorithm. From the results, it is evident that the speed gain in the new

TABLE 1. CPU times in seconds for the Hearmon elasticity data

| Data set (number of matrices) | ALM-mean | New mean |
|---|---|---|
| $NaClO_3$ (5) | 230.0 | 1.30 |
| Ammonium dihydrogen phosphate (4) | 9.9 | 0.39 |
| Potassium dihydrogen phosphate (4) | 9.7 | 0.38 |
| Quartz (6) | 6700.0 | 30.00 |
| Rochelle salt (4) | 10.0 | 0.53 |

TABLE 2. Number of inner and outer iterations needed, and number of matrix roots needed

| | ALM-mean | New mean |
|---|---|---|
| Outer iterations | 23 | 3 |
| Avg. inner iterations | 18.3 | 2 |
| Matrix square roots (`sqrtm`) | 5052 | 72 |
| Matrix $p$-th roots (`rootm`) | 0 | 84 |

mean is due not only to the reduction of the number of outer iterations, but also of the number of inner iterations needed to get convergence at each step of the inner mean calculations. When the number of involved matrices becomes larger, these increased speeds add up at each level.
Hearmon's elasticity data are not suitable for measuring the accuracy of the algorithm, since the results to be obtained are not known. To measure the accuracy of the computed results, we computed instead $|G(A^4, I, I, I) - A|$, which should yield zero in exact arithmetic (due to P1), and its analogue with the new mean. We chose $A$ to be the first matrix in Hearmon's second data set. Moreover, in order to obtain results closer to machine precision, in this experiment we changed the stopping criterion by choosing $= 10^{-13}$ :

| Operation | Result |
|---|---|
| $\|G(A^4, I, I, I) - A\|$ | $3.6\text{E} - 13$ |
| $\|\bar{G}(A^4, I, I, I) - A\|$ | $1.8\text{E} - 14$ |

The results are well within the errors permitted by the stopping criterion, and they show that both algorithms can reach a satisfying precision.
The following examples provide an experimental proof that our mean is different from the ALM-mean.
Consider the following matrices:

$$A = \begin{bmatrix} a & b \\ b & a \end{bmatrix}, B = \begin{bmatrix} a & -b \\ -b & a \end{bmatrix}, C = \begin{bmatrix} 1 & 0 \\ 0 & c \end{bmatrix}.$$

Observe that the triple $(A, B, C)$ is transformed into $(B, A, C)$ under the map $X \to S^{-1}XS$, for $S = \text{diag}(1, -1)$. In this way, any matrix mean $G(A, B, C)$ satisfying condition P3 is such that $G = S^{-1}GS$; that is, the off-diagonal entries of $G$ are zero, whence $G$ must be diagonal. With $a = 2, b = 1, c = 24$, for the ALM-mean $G$ and our mean $\bar{G}$ one finds that

$$\bar{G} = \begin{bmatrix} 1.487443626 & 0 \\ 0 & 4.033766318 \end{bmatrix}, G = \begin{bmatrix} 1.485347837 & 0 \\ 0 & 4.039457861 \end{bmatrix},$$

where we reported the first 10 digits. Observe that the determinant of both the matrices is 6 , that is, the geometric mean of $\det A, \det B, \det C$; moreover, $\rho(\bar{G}) < \rho(G)$.

For the matrices

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 3 & -2 \\ 0 & -2 & 2 \end{bmatrix}, B = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 3 & 2 \\ 0 & 2 & 2 \end{bmatrix},$$

$$C = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 10 & 0 \\ 1 & 0 & 50 \end{bmatrix}, D = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 10 & 0 \\ -1 & 0 & 50 \end{bmatrix},$$

one has

$$\bar{G} = \begin{bmatrix} 1.3481 & 0 & -0.3016 \\ 0 & 3.8452 & 0 \\ -0.3016 & 0 & 6.1068 \end{bmatrix}, G = \begin{bmatrix} 1.3472 & 0 & -0.3106 \\ 0 & 3.8796 & 0 \\ -0.3106 & 0 & 6.0611 \end{bmatrix}.$$

Their eigenvalues are (6.1258,3.8452,1.3290), and (6.0815,3.8796,1.3268), respectively. Observe that, unlike in the previous example, it follows that $\rho(\bar{G}) > \rho(G)$.

In order to illustrate the properties of the set

$$\{G_{s,t}\colon (s,t) \in (0,1] \times (0,1)\},$$

where $G_{s,t}$ is the mean of three matrices defined, we considered the intervals $[1/15,1], [1/15,14/15]$ and discretized them into two sets $\mathcal{S}, \mathcal{T}$ of 15 equidistant points $\{1/15 = s_1 < s_2 < \cdots < s_{15} = 1\}, \{1/15 = t_1 < t_2 < \cdots < t_{15} = 14/15\}$, respectively. For each pair $(s_i, t_j) \in \mathcal{S} \times \mathcal{T}, i,j = 1, \dots, 15$, we computed $G_{s_i,t_j}$ and the orthogonal projection $(x(i,j), y(i,j), z(i,j))$ of the matrix $G_{s_i,t_j} - G_{\frac{2}{3},\frac{1}{2}}$, over a three-dimensional fixed randomly generated subspace. The set

$$\mathcal{V} = \{(x(i,j), y(i,j), z(i,j)) \in \mathbb{R}^3, i,j = 1, \dots, 15\}$$

has been plotted with the Matlab command mesh (x, y, z) which connects each point with coordinates $(x(i,j), y(i,j), z(i,j))$ to its four neighbors with coordinates $(x(i+\delta, j+\gamma), y(i+\delta, j+\gamma), z(i+\delta, j+\gamma))$ for $\delta, \gamma \in \{1, -1\}$.

Figure 1 displays the set $\mathcal{V}$ from six different points of view, where the matrices $A, B$ and $C$ of size 3 have been randomly generated. The set appears to be a flat surface with part of the edge tightly folded on itself. The geometric mean $G_{\frac{2}{3},\frac{1}{2}}$ corresponds to the point with coordinates (0,0,0), which is denoted by a small circle and seems to be located in the central part of the figure. These properties, reported for only one triple $(A, B, C)$, are maintained with very light differences in all the plots that we have performed.

The software concerning our experiments can be delivered upon request.

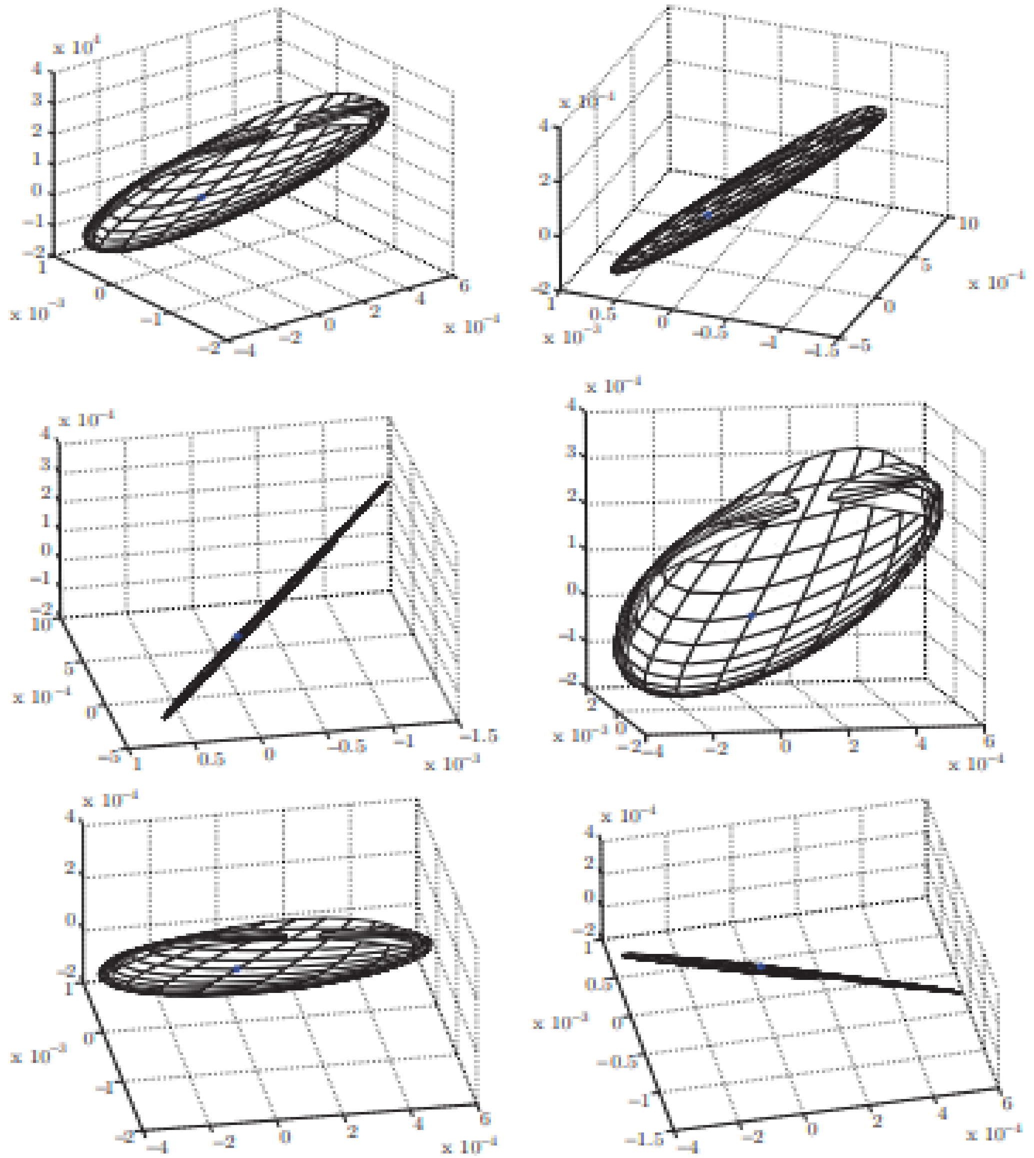


FIGURE 1. Plot of the set $\mathcal{V}$. The small circle corresponds to $G_{2/3,1/2}$.

**Section (1.2): Computing Matrix Geometric Means**

In certain physical applications one has to represent through a single average matrix $G$ the results of several experiments made up by a set of many positive definite $n \times n$ matrices $A_1, A_2, \ldots, A_k$. The arithmetic mean $\frac{1}{k}\sum_{i=1}^{k} A_i$ is not well-suited to represent the needed quantity since for physical reasons one of the required properties is that the average of $A_1^{-1}, \ldots, A_k^{-1}$ as well, must coincide with $G^{-1}$ (see [17], [10] ). Among the classical means of positive real numbers $a_1, \ldots, a_k$, this property is satisfied by the geometric mean $g = \left(\prod_{i=1}^{k} a_i\right)^{1/k}$.

There is large agreement on what is the right definition of the geometric mean $G = A\#B$ of two positive definite matrices $A$ and $B$, namely $G := A(A^{-1}B)^{1/2}$ (see [3] for a concise treatment of the topic), where given a square matrix $M$ having no nonpositive real eigenvalues, $M^{1/2}$ denotes the unique solution of the equation $X^2 = M$ whose eigenvalues lie in the right half plane. That definition was given in the seventies by Pusz and Woronowicz [19], but there are many other equivalent characterizations, the most notable of which has been provided recently in [15], [17] and is related to the Riemannian geometry

obtained endowing the set $\mathbb{P}_n$ of positive definite matrices of size $n$ with the scalar product $g(M,N) = \mathrm{tr}(A^{-1}MA^{-1}N)$ in the tangent space $T_A\mathbb{P}_n$ at $A$.
The link to the geometric mean is through geodesics, in fact it can be proved that there exists a unique geodesic joining two positive definite matrices $A$ and $B$ whose parameterization is

$$A\#_t B := A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^t A^{1/2},\ t \in [0,1],$$

and $A\#B = A\#_{1/2}B$ is its midpoint.

We will use the symbols $\log(A), \exp(A), A^t := \exp(t\log(A))$ to denote the usual functions of a square matrix. If $A$ is diagonalizable, namely if there exists an invertible matrix $M$ and a diagonal matrix $D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$ such that $A = MDM^{-1}$, then $f(A) := Mf(D)M^{-1}$, where $f(D) := \mathrm{diag}\big(f(\lambda_1), \dots, f(\lambda_n)\big)$. The above definition of $A^{1/2}$ coincides with this one (see [14]).
We briefly recall some properties of the matrix exponential and logarithm which will be useful in the following, the proofs can be found in [14].
**Theorem (1.2.1)[11]:** The following properties hold:
- (i) $\log(\alpha I) = \log(\alpha)I$ for any positive constant $\alpha$, in particular $\log I = 0$;
- (ii) if $M$ and $N$ commute and have real positive eigenvalues then $\log(MN) = \log(M) + \log(N)$;
- (iii) for any invertible matrix $M, f(MAM^{-1}) = Mf(A)M^{-1}$, in particular $\exp(MAM^{-1})$ $M\exp(A)M^{-1}$ and $\log(MAM^{-1}) = M\log(A)M^{-1}$;
- (iv) $\det(\exp(A+B)) = \det(\exp(A))\det(\exp(B))$;
- (v) $\exp(-X) = \exp(X)^{-1}$.

In the setting of matrix functions, it is often easy to prove general results in an elegant way. For example the following result holds.
**Theorem (1.2.2)[11]:** Let $A$ and $B$ be positive definite matrices and let $f$ be a function defined on the eigenvalues of $A^{-1}B$, then $Af(A^{-1}B) = A^{1/2}f\left(A^{-1/2}BA^{-1/2}\right)A^{1/2}$
**Proof.** First, observe that the matrix $A^{-1}B$ is diagonalizable. From the above definition of matrix function it follows that for any diagonalizable matrix $A$ one has $f(N^{-1}AN) = N^{-1}f(A)N$, thus

$$Af(A^{-1}B) = Af\left(A^{-1/2}A^{-1/2}BA^{-1/2}A^{1/2}\right) = A^{1/2}f\left(A^{-1/2}BA^{-1/2}\right)A^{1/2}.$$

Theorem (1.2.2) explains why $A\#_{1/2}B = A(A^{-1}B)^{1/2}$.

The generalization of the definition of geometric mean to more than two positive definite matrices seems to be considerably more difficult.
Ando, Li and Mathias [2] proposed a list of ten properties (the ALM properties) that a "good" geometric mean $G(\cdot)$ of $k$ matrices should satisfy. Here, for simplicity we report this list in the case $k = 3$ where we write $A > B$ if $A - B$ is positive definite and $A \geqslant B$ if $A - B$ is positive semi-definite.
P1 Consistency with scalars. If $A, B, C$ commute then $G(A,B,C) = (ABC)^{1/3}$.
P2 Joint homogeneity. $G(\alpha A, \beta B, \gamma C) = (\alpha\beta\gamma)^{1/3}G(A,B,C)$, for $\alpha, \beta, \gamma > 0$.
P3 Permutation invariance. For any permutation $\pi(A,B,C)$ of $A,B,C$, it holds that $G(A,B,C) = G(\pi(A,B,C))$.
P4 Monotonicity. If $A \geqslant A', B \geqslant B', C \geqslant C'$, then $G(A,B,C) \geqslant G(A',B',C')$.
P5 Continuity from above. If $A_n, B_n, C_n$ are monotonic decreasing sequences converging to $A, B, C$, respectively, then $G(A_n, B_n, C_n)$ converges to $G(A,B,C)$.
P6 Congruence invariance. For any nonsingular $S$, it holds that $S^*G(A,B,C)S = G(S^*AS, S^*BS, S^*CS)$.

P7 Joint concavity. If $A = \lambda A_1 + (1-\lambda)A_2, B = \lambda B_1 + (1-\lambda)B_2, C = \lambda C_1 + (1-\lambda)C_2$, then $G(A,B,C) \geqslant \lambda G(A_1,B_1,C_1) + (1-\lambda)G(A_2,B_2,C_2)$ for $0 < \lambda < 1$.
P8 Self-duality. $G(A,B,C)^{-1} = G(A^{-1},B^{-1},C^{-1})$.
P9 Determinant identity. $\det G(A,B,C) = (\det A \det B \det C)^{1/3}$.
P10 Arithmetic-geometric-harmonic mean inequality.

$$\frac{A+B+C}{3} \geqslant G(A,B,C) \geqslant \left(\frac{A^{-1}+B^{-1}+C^{-1}}{3}\right)^{-1}.$$

It has been proved in [2] that P5 and P10 are consequences of the others. Notice that all these properties can be easily generalized to the mean of any number of matrices. For $k=2$ this list uniquely defines $G = A\#B = A(A^{-1}B)^{1/2}$. In the case $k>2$ there are infinitely many means satisfying the ALM properties.
In [2] Ando, Li and Mathias propose a numerical scheme for computing a mean of $k$ matrices which satisfies the ALM properties. For $k=3$ they show that the sequences

$$\begin{aligned} A^{(\nu+1)} &= B^{(\nu)}\#C^{(\nu)} \\ B^{(\nu+1)} &= C^{(\nu)}\#A^{(\nu)},\ \nu = 0,1,\ldots, \\ C^{(\nu+1)} &= A^{(\nu)}\#B^{(\nu)} \end{aligned} \quad (17)$$

obtained with $A^{(0)} = A, B^{(0)} = B, C^{(0)} = C$, converge to a common limit $G$ satisfying the ALM properties. For a set $A_1,\ldots,A_k$ of $k>3$ matrices these sequences can be defined as

$$A_i^{(\nu+1)} = G_{k-1}\left(A_1^{(\nu)},\ldots,A_{i-1}^{(\nu)},A_{i+1}^{(\nu)},\ldots,A_k^{(\nu)}\right),\ i = 1,\ldots,k, \quad (18)$$

where $G_{k-1}$ denotes the mean of $k-1$ matrices recursively defined by means of the same relations. Indeed, also these sequences converge to a common limit which satisfies the ALM properties. See this limit as the ALM mean. It is proved that convergence is linear with convergence factor $1/2$. It is easy to find out that the computational cost of this scheme for general $k$ is $O\left(n^3 k! \prod_{i=3}^{k} p_i\right)$ where $n$ is the matrix size and $p_i$ is the number of iterations needed in the computation of the means of $i$ matrices.
A substantial improvement has been achieved in [1] relying on these observations: in the sequences (17) converging to the ALM mean, $A^{(\nu+1)}$ is the midpoint of the geodesics joining the matrix $B^{(\nu)}$ with $C^{(\nu)}$; in the Euclidean geometry the limit of this sequence is the centroid of the triangle $ABC$; the centroid is also located in the median which connects $A$ with the midpoint of the edge $BC$ at distance $2/3$ from $A$, that is $A\#_{2/3}\left(B\#_{1/2}C\right)$; the three medians have the centroid as common point. Due to the negative curvature of $\mathbb{P}_n$ the three points $A\#_{\frac{2}{3}}(B\#_{1/2}C), B\#2/3(C\#_{1/2}A), C\#_{2/3}$ trices
Therefore the iteration is given by

$$\begin{aligned} A^{(\nu+1)} &= A^{(\nu)}\#_{2/3}\left(B^{(\nu)}\#C^{(\nu)}\right) \\ B^{(\nu+1)} &= B^{(\nu)}\#_{2/3}\left(C^{(\nu)}\#A^{(\nu)}\right),\ \nu = 0,1,\ldots \\ C^{(\nu+1)} &= C^{(\nu)}\#_{2/3}\left(A^{(\nu)}\#B^{(\nu)}\right) \end{aligned}$$

It is proved that the three matrix sequences have a common limit, different from the ALM mean, which satisfies the ALM properties, and the convergence is cubic. See this mean as the BMP mean. The same iteration can be generalized to the case of $k>3$ matrices.
The computational cost is the same as the ALM scheme, however, the number $p_i$ of iterations is reduced by relying on a numerical scheme having cubic convergence so the acceleration in certain applications is dramatic. Unfortunately, the growth of the

computational cost with $k$ is still exponential; therefore, for moderate values of $k$ also this iteration is infeasible.
The idea of [2], [1] can be generalized by considering new means obtained by assembling existing ones through a recursive procedure. Unfortunately, it has been proved that no such definition could produce a mean whose computational cost is polynomial with respect to $k$ [18]. We follow a different direction.

By relying on the geometric interpretation given in terms of geodesics in the Riemannian geometry on the variety $\mathbb{P}_n$, we introduce a new iteration for computing a geometric mean of $k$ matrices with the following features: unlike the known methods, the computation of the mean of $k$ matrices does not require computing the mean of $k-1$ matrices and no recursive process is needed; the convergence speed of the new iteration is cubic; its computational cost is polynomial, namely $O(n^3k^2p_k)$, where $p_k$ is the number of iterations needed by the method (typically just a few); for $k=2$ the limit is $A\#B$, so the proposed mean generalizes the geometric mean of two matrices; the limit of $k$ sequences satisfies the ALM properties P1-P3, P6, P8 and P9; we provide a counterexample where P4 is not satisfied. The counterexample requires that the matrices be very far from each other; counterexamples where the matrices $A_i, i=1,\dots,k$ are in a relatively small neighborhood of their mean are not known. We refer to this new mean as the Cheap Mean.
The idea on which this iteration is based relies once again on the geometric interpretation of the centroid $G$ of a triangle $ABC$. In the Euclidean geometry the centroid G satisfies the equations

$$G = A + \frac{1}{3}\big((B-A)+(C-A)+(A-A)\big),$$

that is, it lies in the geodesic passing through $A$ and tangent in $A$ to the arithmetic mean of the tangent vectors in $A$ of the three geodesics connecting $A$ with $B, C$ and $A$, respectively. Obviously, the third vector is zero. Similar expressions are obtained starting from $B$ and $C$, respectively.
In the Riemannian manifold $\mathbb{P}_n$ this procedure gives three different points $A', B'$ and $C'$, and can be viewed as a step of an iterative procedure converging to a possible mean. Observe that the mean of the tangent vectors is done in the tangent space at a point which is Euclidean, where it is natural to choose the arithmetic mean.
This procedure can be easily generalized to $k \geqslant 3$. Given $A_1,\dots,A_k$, it is enough to consider, for each $i$, the geodesic starting at $A_i$ and whose tangent vector is the arithmetic mean of the $k$ tangent vectors at $A_i$ to the geodesic joining $A_i$ with $A_j$ (where if $i=j$ the vector is 0 ). Then $A_i'$ will be the point of that geodesic for $t=1$.
Since the tangent vector at $A$ to the geodesic joining $A$ and $B$ is the symmetric matrix $A\log(A^{-1}B)$, one obtains the following iteration

$$A_i^{(\nu+1)} = A_i^{(\nu)} \exp\left(\frac{1}{k}\sum_{j=1, j\neq i}^{k} \log\left(\left(A_i^{(\nu)}\right)^{-1} A_j^{(\nu)}\right)\right),\ i=1,\dots,k, \quad (19)$$

with $A_i^{(0)} = A_i, i=1,\dots,k$. Observe that, by Theorem (1.2.2), (19) can be equivalently rewritten as

$$A_i^{(\nu+1)} = \left(A_i^{(\nu)}\right)^{\frac{1}{2}} \exp\left(\frac{1}{k}\sum_{j=1, j\neq i}^{k} \log\left(\left(A_i^{(\nu)}\right)^{-\frac{1}{2}} A_j^{(\nu)} \left(A_i^{(\nu)}\right)^{-\frac{1}{2}}\right)\right)\left(A_i^{(\nu)}\right)^{\frac{1}{2}}. \quad (20)$$

This equation shows that the sequences $\left\{A_i^{(\nu)}\right\}_\nu$ are formed by symmetric positive definite matrices.
For $k = 2$ the first step of the iteration yields $A_1^{(1)} = A\exp\left(\frac{1}{2}\left(\log(A^{-1}B)\right)\right) = A\#B$ and similarly $B_1^{(1)} = A\#B$. Thus, in the case of two matrices the iteration yields the geometric mean since the first step.
We prove that if the sequences $\left\{A_i^{(\nu)}\right\}_\nu$ converge to the same limit for $i = 1, \dots, k$, then the convergence is cubic. Moreover we give conditions under which convergence occurs. Even though the local convergence condition may appear rather restrictive, from the many numerical experiments that we have performed we never encountered failure of convergence.
We have implemented the computation of the Cheap Mean in the Matrix Mean Toolbox, available for Matlab and Octave [13], and performed some numerical tests.
In particular, we have compared the Cheap mean with the "least square geometric mean" [5], also called "Riemannian geometric mean" [17], or Karcher mean [16], that is, the unique positive definite solution of the matrix equation

$$\sum_{i=1}^{k} X^{1/2} A_i^{-1} X^{1/2} = 0. \tag{21}$$

It is known that this mean satisfies all the ALM properties of a geometric mean except perhaps the monotonicity property, for which no counterexample is known so far. By means of numerical experiments we show that the Cheap mean is much faster to compute than the Karcher mean (if for the latter, the algorithms of [16], [17] or a gradient algorithm applied to (21) are used). In fact, in all the experiments performed so far, iteration (19) converges to the Cheap mean in at most 5 iterations with a relative error of the order of $10^{-15}$ independently of the condition number, whereas for the Karcher mean, the iterations of [16], [17] do not converge in certain cases and in the other cases require a larger computational cost. The gradient methods require always a larger computational cost.
We wish to point out that the iteration of [16] is given by

$$X^{(\nu+1)} = X^{(\nu)} \exp\left(\frac{1}{k}\sum_{i=1}^{k} \log\left(\left(X^{(\nu)}\right)^{-1} A_i\right)\right),\ X_0 = A_1, \tag{22}$$

which is very similar to our iteration (19). In fact, each step of (19) can be viewed as $k$ first steps of iteration (22) with $A_i = A_i^{(\nu)}, i = 1, \dots, k$ and $X_0 = A_i^{(\nu)}$. In [16] the convergence of (22) has been proved in the special orthogonal group provided that the matrices $A_i$ are sufficiently close to each other. The numerical tests show that iteration (22) does not converge if the matrices $A_i$ are positive definite and not close each other and that when convergence occurs it is linear.
Another comparison that we have performed concerns the definition of geometric mean by

$$G = \exp\left(\frac{1}{k}\sum_{i=1}^{k} \log A_i\right).$$

This mean, referred to as ExpLog mean, is studied in [2], and can be computed with a cost of $O(n^3 k)$ ops. However, its properties are poorer than the properties of the Cheap mean. First, the ExpLog mean of two matrices is different from $A\#B$. Second, it is not congruence invariant as shown in the numerical experiments. Third, the ExpLog mean looses the monotonicity property in a very large part of cases. In fact, from a wide numerical experimentation it turns out that even though the matrices $A_i$ are tightly close to each other and have a moderate condition number, this mean fails to be monotone. Whereas, the Cheap mean fails to be monotone only when the matrices are severely ill conditioned and they are not tightly close to each other. Finally, for modeling reasons, any practical definitions of geometric mean should lie in a small neighborhood; from numerical experiments it turns out that the ALM, BMP, and Cheap means form a very tight cluster while the ExpLog mean lies very far from this cluster.

We prove the local cubic convergence of iteration (20), we prove most of the ALM properties for the Cheap mean and provide a counterexample for the monotonicity. We discuss the results of the numerical experiments.

We consider a single step of iteration (20) and for notational simplicity we write

$$A_i' = A_i^{\frac{1}{2}} \exp\left(\frac{1}{k}\sum_{j=1}^{k} \log\left(A_i^{-\frac{1}{2}} A_j A_i^{-\frac{1}{2}}\right)\right) A_i^{\frac{1}{2}}. \tag{23}$$

Observe that the condition $i \neq j$ is not needed in (23) since the term obtained for $i = j$ is zero.

We introduce the following notation

$$A_j^{-1/2} A_i A_j^{-1/2} = I + A_j^{-1/2}(A_i - A_j) A_j^{-1/2} = I + X_{i,j},$$
$$X_{i,j} = A_j^{-1/2} E_{i,j} A_j^{-1/2},\ E_{i,j} = A_i - A_j,$$

so that equation (23) can be rewritten as

$$A_i' = A_i^{\frac{1}{2}} \exp\left(\frac{1}{k}\sum_{j=1}^{k} \log\left(I + X_{i,j}\right)\right) A_i^{\frac{1}{2}}. \tag{24}$$

We recall that if $\rho(X) < 1$ then

$$\begin{aligned} \log(I + X) &= X - \frac{1}{2}X^2 + \frac{1}{3}X^3 - \cdots \doteq X - \frac{1}{2}X^2 \\ \exp(W) &= I + W + \frac{1}{2}W^2 + \frac{1}{3!}W^3 + \cdots \doteq I + W + \frac{1}{2}W^2 \end{aligned} \tag{25}$$

where $\doteq$ denotes equality up to terms of the third order in $X$ or in $W$.

Here we assume that the matrices are close enough to each other, we assume that

$$\left\| A_j^{-1/2} E_{i,j} A_j^{-1/2} \right\| \leqslant \varepsilon < 1,\ i,j = 1, \dots, k,$$

for $\varepsilon > 0$ small enough, where $\|\cdot\|$ denotes the spectral norm.

Since $\|X_{i,j}\| < 1$, applying (25) with $X = X_{i,j}$ in (24) yields

$$A_j' \doteq A_j^{1/2}\left[I + Z_j + \frac{1}{2}Z_j^2\right] A_j^{1/2},$$

$$Z_j \doteq \frac{1}{k}\sum_{i=1}^{k} \left(X_{i,j} - \frac{1}{2}X_{i,j}^2\right),$$

whence

$$A_j' \doteq A_j + \frac{1}{k}\sum_{i=1}^{k} E_{i,j} - \frac{1}{2k}\sum_{i=1}^{k} E_{i,j} A_j^{-1} E_{i,j} + \frac{1}{2k^2}\sum_{r,s=1}^{k} E_{r,j} A_j^{-1} E_{s,j}.$$

Writing down the same equation for $A'_h$ and subtracting the two expressions yields the equation which relates $E'_{h,j} = A'_h - A'_j$ to $E_{i,j}$ :

$$\begin{aligned} E'_{h,j} \doteq \quad & E_{h,j} + \frac{1}{k}\sum_{i=1}^{k} \left(E_{i,h} - E_{i,j}\right) - \frac{1}{2k}\sum_{i=1}^{k} \left(E_{i,h}A_h^{-1}E_{i,h} - E_{i,j}A_j^{-1}E_{i,j}\right) \\ & + \frac{1}{2k^2}\sum_{r,s=1}^{k} \left(E_{r,h}A_h^{-1}E_{s,h} - E_{r,j}A_j^{-1}E_{s,j}\right). \end{aligned}$$

Now, since $E_{h,j} + \frac{1}{k}\sum_{i=1}^{k}\left(E_{i,h} - E_{i,j}\right) = A_h - A_j + \frac{1}{k}\sum_{i=1}^{k}\left(A_j - A_h\right) = 0$, one has

$$\begin{aligned} E'_{h,j} \doteq \quad & -\frac{1}{2k}\sum_{i=1}^{k} \left(E_{i,h}A_h^{-1}E_{i,h} - E_{i,j}A_j^{-1}E_{i,j}\right) \\ & + \frac{1}{2k^2}\sum_{r,s=1}^{k} \left(E_{r,h}A_h^{-1}E_{s,h} - E_{r,j}A_j^{-1}E_{s,j}\right). \end{aligned} \qquad (26)$$

This implies that there exists a constant $\sigma$, depending only on the matrices $A_1, \ldots, A_k$ such that $\max_{i,j}\|E'_{i,j}\| < \sigma \max_{i,j}\|E_{i,j}\|^2$, so that if the sequence $\left\{E_{h,j}^{(\nu)}\right\}_\nu$ converges to zero the convergence is at least quadratic.

We can prove more by observing that

$$\frac{1}{k}\sum_{i=1}^{k} E_{i,h}A_h^{-1}E_{i,h} = \frac{1}{k}\sum_{i=1}^{k} A_iA_h^{-1}A_i - 2M + A_h,$$

$$\frac{1}{k^2}\sum_{i,j=1}^{k} E_{i,h}A_h^{-1}E_{j,h} = MA_h^{-1}M - 2M + A_h,$$

where we set $M = \frac{1}{k}\sum_{i=1}^{k} A_i$. Replacing the latter equations in (26) one obtains

$$E'_{h,j} = -\frac{1}{2}\left(M\left(A_h^{-1} - A_j^{-1}\right)M - \frac{1}{k}\sum_{i=1}^{k} A_i\left(A_h^{-1} - A_j^{-1}\right)A_i\right).$$

Since $M = \frac{1}{k}\sum_{i=1}^{k} A_i$, formally the latter expression is a quadratic form in $A_1, \ldots, A_k$, namely,

$$E'_{h,j} = \sum_{r,s=1}^{k} \eta_{r,s}A_r\left(A_h^{-1} - A_j^{-1}\right)A_s, \ \frac{1}{2k^2}\left(kI - ee^T\right) = \left(\eta_{r,s}\right),$$

where $e = (1, \ldots, 1)^T$, that is, the matrix associated with this quadratic form is

$$Q_{h,j} = \frac{1}{2k^2}\left(kI - ee^T\right) \otimes \left(A_h^{-1} - A_j^{-1}\right),$$

where $\otimes$ denotes the Kronecker product.

Now, the matrix $kI - ee^T$ can be rewritten as

$$kI - ee^T = kUT^{-1}U^T, \ T = U^TU \qquad (27)$$

where $U \in \mathbb{R}^{k\times(k-1)}$, $Ue_i = e_i - e_{(i-1)\mathrm{mod}k}$, for $i = 1, \ldots, k-1$, and $T = U^TU$ is the $(n-1)\times(n-1)$ symmetric tridiagonal matrix having diagonal entries equal to 2 and super-diagonal entries equal to $-1$. In fact, the two matrices in the left-hand and in the right-hand side of (27) have the vector $e$ in their kernels and thus coincide in the linear space orthogonal to $e$ spanned by the columns of $U$. Therefore

$$Q_{h,j} = \frac{1}{2k}[U \otimes I]\left[T^{-1} \otimes \left(A_h^{-1} - A_j^{-1}\right)\right][U^T \otimes I]$$

so that we may write

$$\begin{aligned} E'_{h,j} &= [A_1 \quad \cdots \quad A_k] Q_{h,j} \begin{bmatrix} A_1 \\ \vdots \\ A_k \end{bmatrix} \\ &= [E_{1,2} \quad \cdots \quad E_{k-1,k}] \left( \frac{1}{2k} T^{-1} \otimes \left(A_h^{-1} E_{h,j} A_j^{-1}\right)\right) \begin{bmatrix} E_{1,2} \\ \vdots \\ E_{k-1,k} \end{bmatrix} \\ &= \frac{1}{2k} \sum_{r=1}^{k-1} \sum_{s=1}^{k-1} \gamma_{r,s} E_{r,r+1} A_h^{-1} E_{h,j} A_j^{-1} E_{s,s+1}, \end{aligned}$$

with $T^{-1} = \left(\gamma_{r,s}\right)$. Denoting $\gamma = \gamma(k) = \frac{1}{2k}\sum_{r,s}\gamma_{r,s}$ one has

$$\|E'_{h,j}\| \leqslant \|E_{h,j}\| \gamma(k) \max_r \|E_{r,r+1}\|^2 \|A_h^{-1}\| \cdot \|A_j^{-1}\|.$$

Therefore,

$$\max_{i,j} \|E'_{i,j}\| \leqslant \gamma(k) \max_{i,j} \|E_{i,j}\|^3 \cdot \max_j \|A_j^{-1}\|^2. \tag{28}$$

We synthesize the above discussion in the following result, where we give also a condition such that iteration (19) converges.

**Theorem (1.2.3)[11]:** If the sequences $A_i^{(\nu)}$ generated by (19) have a common limit $G$, then there exists a constant $\gamma$ such that $\left\|A_i^{(\nu+1)} - A_j^{(\nu+1)}\right\| \leqslant \gamma \left\|A_i^{(\nu)} - A_j^{(\nu)}\right\|^3$, for any $i,j = 1, \dots, k$, i.e., convergence has order at least 3. If $\max_j \|A_j^{-1}\| \cdot \max_{i,j} \| A_i - A_j \| < \varepsilon$ for $i,j = 1, \dots, k$, where $0 < \varepsilon < 1/3$ then $\max_j \left\|\left(A_j^{(\nu)}\right)^{-1}\right\| \cdot \max_{i,j} \| A_i^{(\nu)} - A_j^{(\nu)} \| < \varepsilon$ for $i,j = 1, \dots, k$, for any $\nu$, moreover $\max_{i,j} \left\|A_i^{(\nu)} - A_j^{(\nu)}\right\| \leqslant (2\varepsilon/(1-\varepsilon))^\nu$, and the sequences $A_i^{(\nu)}$ converge to the same limit $G$.

**Proof.** The first part of the theorem follows from (28). Concerning the second part, denote $= \max_{i,j} \|A_i - A_j\|, \delta' = \max_{i,j} \|A'_i - A'_j\|, f = \max_i \|A_i^{-1}\|, f' = \max_i \|A_i'^{-1}\|$, and observe that $\|A_j^{-1}\left(A_i - A_j\right)\| \leqslant \delta f$. Let us prove that if $\delta f < \varepsilon$ with $\varepsilon$ sufficiently small, then also $\delta' f' \leqslant \varepsilon$ so that $\|A_j'^{-1}\left(A'_i - A'_j\right)\| \leqslant \varepsilon$ as well. From (26) one finds that

$$\delta' \leqslant 2\delta^2 \max_i \|A_i^{-1}\| = 2\delta^2 f. \tag{29}$$

Now we provide an upper bound to $f'$ by proving that

$$f' \leqslant \frac{f}{1 - \delta f}. \tag{30}$$

We rely on the following inequalities which derive directly from the definition of the matrix functions exp and log by taking the norms of both sides of (25):

$$\begin{aligned} &\| \exp(X) \| \leqslant \exp(\| X \|) \\ &\| \log(I + X) \| \leqslant -\log(1 - \| X \|), \text{ if } \| X \| < 1. \end{aligned} \tag{31}$$

We note

$$\|A_i'^{-1}\| \leqslant \left\|\exp\left(-\frac{1}{k}\sum_{j=1}^{k} \log\left(A_i^{-1}A_j\right)\right)\right\| \cdot \|A_i^{-1}\| \leqslant \left\|\exp\left(-\frac{1}{k}\sum_{j=1}^{k} \log\left(A_i^{-1}A_j\right)\right)\right\| f. \tag{32}$$

By using (31) one finds that

$$\left\|\exp\left(-\frac{1}{k}\sum_{j=1}^{k}\log\left(A_i^{-1}A_j\right)\right)\right\| \leqslant \exp\left(\left\|\frac{1}{k}\sum_{j=1}^{k}\log\left(A_i^{-1}A_j\right)\right\|\right)$$
$$\leqslant \exp\left(\frac{1}{k}\sum_{j=1}^{k}\left\|\log\left(A_i^{-1}A_j\right)\right\|\right)$$
$$= \exp\left(\frac{1}{k}\sum_{j=1}^{k}\left\|\log\left(I + A_i^{-1}E_{j,i}\right)\right\|\right)$$

Now, since $\|A_i^{-1}E_{j,i}\| \leqslant \delta f \leqslant \varepsilon < 1$ we may apply (31) and get

$$\left\|\exp\left(-\frac{1}{k}\sum_{j=1}^{k}\log\left(A_i^{-1}A_j\right)\right)\right\| \leqslant \exp\left(-\frac{1}{k}\sum_{j=1}^{k}\log\left(1 - \|A_i^{-1}E_{j,i}\|\right)\right)$$
$$= \left(\prod_{j=1}^{k}\left(1 - \|A_i^{-1}E_{j,i}\|\right)\right)^{-1/k}$$
$$\leqslant (1-\delta f)^{-1}$$

which in the view of $(32)$ yields $(30)$. Now we are ready to prove that if the condition $\delta f \leqslant \varepsilon$ is satisfied then $\delta' f' \leqslant \varepsilon$ as well. Combining (29) and (30) yields

$$\delta' f' \leqslant (\delta f)^2 \frac{2}{1-\delta f}.$$

Clearly, if $\varepsilon < 1/3$ then $\delta' f' < \varepsilon$ and from (29) one deduces that

$$\delta' \leqslant \frac{2}{3}\delta.$$

An inductive process completes the convergence proof.
Proving global convergence is still an open problem. From the many numerical experiments that we have performed we have always observed convergence. It is interesting to point out that if the matrices $A_i$ pairwise commute then convergence occurs in just one step for any $k$-tuple of positive definite matrices $A_1, \dots, A_k$.

A large number of the ALM properties are satisfied by the Cheap mean. We give a formal proof for the properties P1, P2, P3, P6, P8, and P9, while for P4 we provide a counterexample which shows that monotonicity is not fulfilled by our mean. The proof of validity of P5, P7 and P10 is usually performed relying on P4. We have no counterexample for P5, P7 and P10.

We provide the proofs in the case $k = 3$. The generalization to any $k$ is straightforward. We show that starting with $A_0 = A$, $B_0 = B$ and $C_0 = C$, properties P1, P2, P6, P8 and P9, are held by $A_1$ itself (and $B_1$ and $C_1$ ). This fact can be used in an induction argument, proving that the same properties hold for $A_k, B_k$ and $C_k$, for each $k > 0$ and thus for the limit.
P1 Consistency with scalars. If $A, B, C$ commute, then

$$A_1 = A\exp\left(\frac{1}{3}\left(\log(A^{-1}B) + \log(A^{-1}C)\right)\right) \quad = A\exp\left(\frac{1}{3}\left(\log(A^{-1}BA^{-1}C\right)\right)$$
$$= A(A^{-2}BC)^{1/3} = (ABC)^{1/3}.$$

where we have used Property 2 of Theorem (1.2.1). The same holds for $B_1$ and $C_1$.

P2 Joint homogeneity. If $\hat{A} = \alpha A, \hat{B} = \beta B$ and $\hat{C} = \gamma C$, with $\alpha, \beta, \gamma > 0$, then

$$\begin{aligned}
\hat{A}_1 \quad &= \alpha A\exp\left(\frac{1}{3}\left(\log\left(A^{-1}B\frac{\beta}{\alpha}\right) + \log\left(A^{-1}C\frac{\gamma}{\alpha}\right)\right)\right) \\
&= \alpha A\exp\left(\frac{1}{3}\left(\log(A^{-1}B) + \log(A^{-1}C) + \log\left(\frac{\beta\gamma}{\alpha^2}I\right)\right)\right) \\
&= \alpha A\exp\left(\frac{1}{3}\left(\log(A^{-1}B) + \log(A^{-1}C)\right)\right)\exp\log\left(\left(\frac{\beta\gamma}{\alpha^2}\right)^{1/3}\right) = (\alpha\beta\gamma)^{1/3}A_1,
\end{aligned}$$

where we have used Properties 1 and 2 of Theorem (1.2.1). The same holds for $B_1$ and $C_1$.

P3 Permutation invariance. It follows immediately from the definition.

P4 Monotonicity. This property is not satisfied in general as it is shown by the following numerical counterexample.

Let

$$A = I,\ B = \begin{bmatrix} \varepsilon & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},\ C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \varepsilon & 0 \\ 0 & 0 & 1 \end{bmatrix},\ \tilde{A} = A + hee^T.$$

For $\varepsilon = 0.0001$ and $0 < h \leqslant 3$ it holds that $\tilde{A}(h) \geqslant A$ and the matrix $G(\tilde{A}, B, C) - G(A, B, C)$ has a negative eigenvalue. For instance, for $h = 1$ the eigenvalues are $-2.4131\mathrm{e}-3, 2.2853\mathrm{e}-2, 1.0826\mathrm{e}-1$.

P6 Congruence invariance. Observe that starting from $\hat{A} = S^*AS, \hat{B} = S^*BS, \hat{C} = S^*CS$ one has

$$\begin{aligned}
\hat{A}_1 \quad &= \hat{A}\exp\left(\frac{1}{3}\left(\log(\hat{A}^{-1}\hat{B}) + \log(\hat{A}^{-1}\hat{C})\right)\right) \\
&= S^*AS\exp\left(\frac{1}{3}\left(\log(S^{-1}A^{-1}S^{-*}S^*BS) + \log(S^{-1}A^{-1}S^{-*}S^*BS)\right)\right) \\
&= S^*ASS^{-1}\exp\left(\frac{1}{3}(\log(A^{-1}B) + \log(A^{-1}C))\right)S = S^*A_1S,
\end{aligned}$$

where Property 3 of Theorem (1.2.1) has been used. The same holds for $\hat{B}_1$ and $\hat{C}_1$.

P8 Self duality. Observe that

$$\begin{aligned}
A_1^{-1} \quad &= \exp\left(-\frac{1}{3}\left(\log(A^{-1}B) + \log(A^{-1}C)\right)\right)A^{-1} \\
&= \exp\left(\log(B^{-1}A)^{1/3} + \log(C^{-1}A)^{1/3}\right)A^{-1} \\
&= A^{-1}\exp\left(\log(AB^{-1})^{1/3} + \log(AC^{-1})^{1/3}\right) = \hat{A}_1,
\end{aligned}$$

where $\hat{A}_1$ is obtained from $\hat{A} = A^{-1}, \hat{B} = B^{-1}$ and $\hat{C} = C^{-1}$, thus the self-duality holds for $A_1$. The same holds for $B_1$ and $C_1$.

P9 Determinant identity. The identity follows from $\det(e^{A+B}) = \det(e^Ae^B)$, in fact for $A, B$ and $C$

$$\det A_1 = \det A \det\left(\exp\left(\log(A^{-1}B)^{1/3}\right)\right)\det\left(\exp\left(\log(A^{-1}C)^{1/3}\right)\right)$$
$$= \det A \det(A^{-1}B)^{1/3}\det(A^{-1}C)^{1/3} = (\det A \det B \det C)^{1/3},$$

where Property 4 of Theorem (1.2.1) has been used. The same holds for $\det B_1$ and $\det C_1$. Observe that in the counterexample concerning monotonicity the matrices $A, B$ and $C$ are quite far from each other and do not satisfy the convergence conditions of Theorem (1.2.3). We do not have any counterexample to the monotonicity where the matrices $A_i$ satisfy the convergence conditions of Theorem (1.2.3), and we believe that monotonicity is satisfied "locally", i.e., if the set of matrices $A_i, i = 1, \ldots, k$ lie in a neighborhood of their mean. Observe, moreover, that $A_1$ verifies properties P1, P2, P6, P8 and P9, thus, it can be viewed as a rough mean.

We have implemented the computation of the Cheap mean together with other algorithms for matrix means in the Matrix Means Toolbox [13] available for Matlab and Octave. Here we report part of the many numerical experiments that we have performed.
In the first set of tests we compare the execution times of computing the Cheap mean and the mean of [1], in the following BMP mean, which among the ALM means is the fastest available.
The test matrices are generated randomly with different values of their condition numbers according to the following Matlab commands:

$$\mathrm{n} = 10; \mathrm{W} = \mathrm{rand(n)} - \mathrm{rand(n)}; \mathrm{X} = \mathrm{W}' * \mathrm{W}; \mathrm{X} = \mathrm{X} - \mathrm{eye(n)} * \mathrm{m}\ \ (\mathrm{eig(X)})$$
$$\mathrm{X} = \mathrm{X/norm(X)}; \mathrm{X} = \mathrm{X} + \mathrm{eye(n)/(cnd} - 1); \mathrm{X} = \mathrm{X/norm(X)};$$

so that the parameter cnd coincides with the condition number of $X$.
For various values of the condition number cnd, for $n = 4$ and $k = 3{:}\,10$, in Table 4 we report the CPU time required to compute the Cheap mean and the BMP mean together with the Euclidean distance of the two means. A "*" denotes a CPU time larger than $10^4$ seconds. The number of iterations required to compute the Cheap mean as well as the number of outer iterations in the recursive process to compute the BMP mean has been between 4 and 5.
The exponential growth with $k$ of the complexity of the BMP mean is evident, while the polynomial complexity of the Cheap mean makes the computation possible even for much larger values of $k$. It is interesting to observe that the Cheap mean and the BMP mean are not so far from each other.
The second bunch of tests compares the Cheap mean with the mean

$$G = \exp\left(\frac{1}{k}\sum_{i=1}^{k} \log(A_i)\right)$$

which, for simplicity we call ExpLog mean, in order to find out the cases where the monotonicity property is not satisfied. To this end, we consider a $3 \times 3$ diagonal matrix $A_1$ with diagonal entries $1, \delta, \delta^2$, for $0 < \delta < 1$ so that $\|A_1\| = 1$ and its condition

| | cnd= 1.e2 | | | cnd= 1.e4 | | | cnd= 1.e8 | | |
|---|---|---|---|---|---|---|---|---|---|
| $k$ | Cheap | BMP | Dist. | Cheap | BMP | Dist. | Cheap | BMP | Dist. |
| 3 | 1.e-2 | 1.e-2 | 5.e-3 | 1.e-2 | 1.e-2 | 3.e-2 | 1.e-2 | 1.e-2 | 3.e-2 |
| 4 | 2.e-2 | 2.e-1 | 6.e-3 | 2.e-2 | 2.e-1 | 2.e-2 | 2.e-2 | 2.e-2 | 8.e-2 |
| 5 | 2.e-2 | 1.e0 | 7.e-3 | 3.e-2 | 2.e0 | 4.e-2 | 3.e-1 | 2.e0 | 5.e-2 |
| 6 | 3.e-2 | 1.e+1 | 5.e-2 | 4.e-2 | 3.e+1 | 2.e-2 | 4.e-2 | 3.e+1 | 5.e-2 |
| 7 | 3.e-2 | 2.e+2 | 8.e-3 | 5.e-3 | 4.e+2 | 2.e-2 | 5.e-2 | 4.e+2 | 1.e-2 |
| 8 | 4.e-2 | 2.e+3 | 1.e-2 | 6.e-2 | 5.e+3 | 2.e-2 | 7.e-2 | 5.e+3 | 3.e-2 |
| 9 | 4.e-2 | * | – | 7.e-2 | * | – | 7.e-2 | * | – |
| 10 | 5.e-2 | * | – | 9.e-2 | * | – | 1.e-1 | * | – |

**Table (1)[]:**

CPU times in seconds, rounded to one digit, required to compute the BMP mean $G_1$ and the Cheap mean $G_2$, together with the distances $\|G_1 - G_2\|_2$. $A$ " $*$ " denotes a CPU time larger than $10^4$ seconds.

| $\varepsilon$ | 1e-6 | 1e-5 | 1e-4 | 1e-3 | 1e-2 | 1e-1 | 1 | 1e1 |
|---|---|---|---|---|---|---|---|---|
| cnd=1e4 | 0 | 0 | 0 | 0 | 0 | 5 | 34 | 64 |
| cnd=1e6 | 0 | 0 | 0 | 7 | 28 | 24 | 58 | 78 |
| cnd=1e8 | 0 | 0 | 21 | 45 | 16 | 20 | 71 | 77 |
| cnd=1e10 | 0 | 36 | 41 | 24 | 6 | 22 | 77 | 82 |
| cnd=1e12 | 39 | 56 | 14 | 5 | 3 | 40 | 85 | 85 |

**Table (2)[]:**

Percentage of cases where the ExpLog mean of $A_1, A_2, A_3$ fails to be monotone where $A_2$ and $A_3$ are chosen in a neighborhood of A1 of radius " and A1 has condition number cnd number is $1/\delta^2$, and define $A_2 = A_1 + \varepsilon U_1, A_3 = A_1 + \varepsilon U_2$, where $U_1, U_2$ are positive definite random matrices with norm 1 , generated as follows:

$$\mathrm{W = rand(3) - rand(3); W = W * \ W'; U = W/norm(W);}$$

In this way the matrices $A_2, A_3$ stay in the sphere of center $A_1$ and radius $\varepsilon$. We have generated 100 random values and computed the number of cases where the matrix $G(A_1 + 0.01 * A_2, A_2, A_3) - G(A_1, A_2, A_3)$ is not positive definite. Tables (2) and (3) report these values according to the conditioning of $A_1$ and to the radius of the neighborhood of $A_1$. It is evident that the ExpLog mean fails to be monotone even for moderate values of the condition number and for relatively small neighborhoods of $A_1$, whereas the Cheap mean seems to be more robust.

It is not difficult to construct numerical examples showing that the ExpLog mean is not congruence invariant, for instance if $A = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$ and $S = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$

$$S^* \exp\left(\frac{1}{2}(\log(A) + \log(I))\right) S \quad = \begin{bmatrix} 2 & 2 \\ 2 & 8 \end{bmatrix},$$

$$\exp\left(\frac{1}{2}(\log(S^*AS) + \log(S^*S))\right) \quad \approx \begin{bmatrix} 3.0 & 5.4 \\ 5.4 & 13.5 \end{bmatrix}.$$

The last bunch of tests, taken from [12], reports the number of iterations needed and the Cheap mean.

| $\varepsilon$ | 1e-6 | 1e-5 | 1e-4 | 1e-3 | 1e-2 | 1e-1 | 1 | 1e1 |
|---|---|---|---|---|---|---|---|---|
| cnd=1e4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| cnd=1e6 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| cnd=1e8 | 0 | 0 | 0 | 0 | 0 | 4 | 33 | 23 |
| cnd=1e10 | 0 | 0 | 0 | 0 | 0 | 4 | 33 | 23 |
| cnd=1e12 | 0 | 0 | 0 | 1 | 1 | 18 | 70 | 67 |

TABLE 4.3

*Percentage of cases where the Cheap mean of $A_1, A_2, A_3$ fails to be monotone where $A_2$ and $A_3$ are chosen in a neighborhood of $A_1$ of radius $\varepsilon$ and $A_1$ has condition number* cnd

| | cond=1.e2 | | | cond=1.e4 | | |
|---|---|---|---|---|---|---|
| $k\backslash X_0$ | $I$ | AM | Cheap | $I$ | AM | Cheap |
| 3 | 74 | 26 | 17 | 114 | 89 | 41 |
| 4 | 66 | 21 | 17 | 82 | 59 | 37 |
| 5 | 65 | 19 | 16 | 87 | 58 | 35 |
| 6 | 62 | 20 | 16 | 81 | 54 | 31 |
| 7 | 61 | 21 | 15 | 83 | 63 | 29 |
| 8 | 61 | 20 | 15 | 93 | 55 | 29 |
| 9 | 58 | 19 | 14 | 89 | 50 | 29 |
| 10 | 56 | 19 | 14 | 94 | 47 | 28 |

TABLE 4.4

*Number of iterations needed to approximate the Karcher mean up to the error* 1.e-11 *by means of the iteration (4.1) (a), starting with the identity matrix, the arithmetic mean*

for approximating the Karcher mean relying on the iteration

$$X_{\nu+1} = g(X_\nu),\ \nu = 0,1,\dots,$$
$$g(X) = X - \vartheta X^{1/2} \sum_{i=1}^{k} \log\left(X^{1/2} A_i^{-1} X^{1/2}\right) X^{1/2}, \qquad (33)$$

starting with $X_0$ equal to the identity matrix, the arithmetic mean and the Cheap mean. Here we have choosen the value of $\vartheta$ which minimizes the number of iterations. It is interesting to point out that the number of iterations required is much larger than the number of iterations needed to approximate the Cheap mean which in all the treated cases is less than or equal to 5. Moreover, choosing as starting approximation the Cheap mean yields a faster convergence.

We conclude with an example showing the mutual distance of most of the means of interest. We consider the following matrices

$$A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix},\ B = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix},\ C = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \qquad (34)$$

and we compute the ALM, BMP, Cheap, ExpLog and the Karcher means of them. Moreover, we compute the arithmetic-harmonic-geometric (AHG) mean, that is the geometric mean of the arithmetic mean and the harmonic mean. The latter does not satisfy most of the ALM properties, but it is easy to compute. In Figure 4 we have plotted the corresponding points in the three dimensional space of $2 \times 2$ symmetric matrices. One can observe that the ALM, BMP, Cheap and Karcher means are very

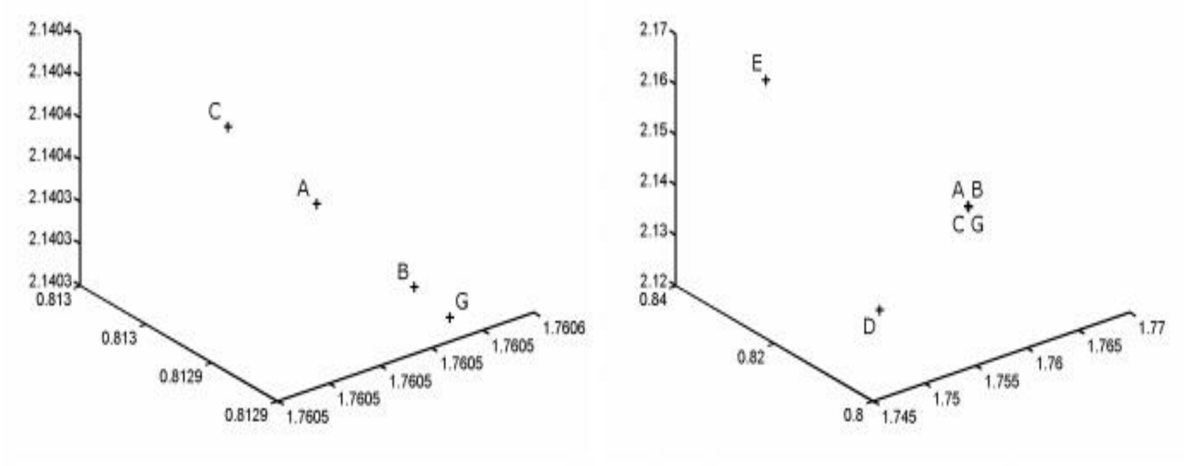



FIG. 4.1. *Localization of the ALM (A), BMP (B), Cheap (C), Arithmetic-Harmonic-Geometric (D), ExpLog (E) and Karcher (G) mean for A, B and C as in (4.2)*

near to each other, while ExpLog and AHG means are relatively far from the others. This is a typical situation that makes the Cheap mean preferable with respect to the ExpLog and AHG means.

We have introduced a new definition of geometric mean which, unlike the ALM means, can be computed with low computational effort even for a large number of input matrices (Cheap mean). We have proved its local convergence and that it fulfills most of the ALM properties.

A proof of global convergence of the iteration for the Cheap Mean is missing; concerning the lack of monotonicity, it would be interesting to find out under which conditions on the matrices $A_i$ the Cheap mean keeps monotonicity. For instance, it seems reasonable that if the matrices $A_i$ are close enough to each other then monotonicity should be satisfied.

**Section (1.3): Geometric Means of Structured Matrices**

We generalize the concept of the geometric mean to positive definite (positive for short) matrices and, on the other hand, the need to average quantities expressed by positive matrices in certain applications have led to the definition and the study of the Karcher mean [3], [5], [17].

We consider the set of positive Hermitian $n \times n$ matrices, denoted by $\mathcal{P}_n$, as a manifold [21], in particular, there is a diffeomorphism from $\mathcal{P}_n$ to $\mathbb{R}^{n^2}$. In each point of $A \in \mathcal{P}_n$ one can define the tangent space $T_A\mathcal{P}_n$, which can be identified with the space of Hermitian matrices. The Karcher mean can now be defined in terms of a Riemannian geometry defined on $\mathcal{P}_n$ and induced by the inner product

$$g_A(X, Y) := \operatorname{tr}(A^{-1}XA^{-1}Y), \; X, Y \in T_A\mathcal{P}_n \tag{35}$$

on the tangent space $T_A\mathcal{P}_n$. This inner product $g_A$ makes $\mathcal{P}_n$ a complete Riemannian manifold with non-positive curvature and yields the following distance between two matrices $A, B \in \mathcal{P}_n$ :

$$\delta(A, B) = \left(\sum_{k=1}^{n} \log^2 \lambda_k\right)^{\frac{1}{2}}, \tag{36}$$

where $\lambda_1, \ldots, \lambda_n$, are the eigenvalues of $A^{-1}B$, which are positive numbers (for all the proofs see [3]). The Karcher mean of a set of $m$ positive matrices, $A_1, \ldots, A_m \in \mathcal{P}_n$, is defined as the unique positive minimizer $G(A_1, \ldots, A_m)$ of the function

$$f(X; A_1, \ldots, A_m) := \sum_{j=1}^{m} \delta^2(X, A_j). \tag{37}$$

Since this mean minimizes the sum of squared intrinsic distances to each of the matrices $A_j$ it is a barycenter of these matrices with respect to the aforementioned metric.
An important feature of the Karcher mean is that it possesses all the properties desired by a geometric mean, like the ten Ando-Li-Mathias (ALM) axioms [2]. For this reason, it is a viable tool in applications requiring some of these properties [24], [10]. A geometric mean should for instance be: permutation invariant, monotone, joint concave, and should satisfy the arithmetic-geometric-harmonic inequality (see [2] for the precise statements of the properties). In particular, one of the most characteristic properties of a geometric mean is its invariance under inversion:

$$G(A_1^{-1}, \ldots, A_m^{-1}) = G(A_1, \ldots, A_m)^{-1}. \tag{38}$$

Prior to having the proofs of all the properties of the Karcher mean, some of which are very elusive [41], other definitions of a matrix geometric mean had been proposed [2], [1], [11], [43], even if nowadays there is large agreement in considering the Karcher mean as the "right" matrix geometric mean.

In certain applications, however, besides the positive definiteness, the data matrices have some further structure in the sense that they belong to some special subset $\mathcal{S}$, say a linear space. For instance, in the design and analysis of certain radar systems, the matrices to be averaged are correlation matrices, which are positive Toeplitz matrices [40]. In these cases, one would like the geometric mean to belong to the same class $\mathcal{S}$ as the data. Unfortunately, the Karcher mean does not preserve many structures, in particular the Karcher mean of Toeplitz and/or band matrices is typically not of Toeplitz and/or band form anymore, as illustrated by the following simple example.

**Example (1.3.1)[20]:** Let $\mathcal{S}$ be the set of tridiagonal Toeplitz matrices and choose $A_1, A_2 \in \mathcal{S}$ where $A_1 = I$, and $A_2 = \text{tridiag}(1,2,1)$ is the matrix with 2 's on the main, and 1 's appearing on sub- and superdiagonals. We have $A_1 A_2 = A_2 A_1$, thus the Karcher mean equals $(A_1 A_2)^{1/2}$. For $n = 3$ we get

$$(A_1 A_2)^{1/2} = \frac{\sqrt{2}}{4} \begin{bmatrix} \sqrt{2+\sqrt{2}}+2 & \sqrt{2}\sqrt{2-\sqrt{2}} & \sqrt{2+\sqrt{2}}-2 \\ \sqrt{2}\sqrt{2-\sqrt{2}} & \sqrt{2+\sqrt{2}} & \sqrt{2}\sqrt{2-\sqrt{2}} \\ \sqrt{2+\sqrt{2}}-2 & \sqrt{2}\sqrt{2-\sqrt{2}} & \sqrt{2+\sqrt{2}}+2 \end{bmatrix} \tag{39}$$

which is neither tridiagonal nor Toeplitz.

We introduce the concept of a structured geometric mean of positive matrices in such a way that if $A_1, \ldots, A_m \in \mathcal{S}$ their mean also belongs to $\mathcal{S}$. Given a subset $\mathcal{S}$ of $\mathcal{P}_n$ and matrices $A_1, \ldots, A_m \in \mathcal{S}$, we say that $G \in \mathcal{S}$ is a structured geometric mean with respect to $\mathcal{S}$ of $A_1, \ldots, A_n$ if the function $f(X; A_1, \ldots, A_m)$ of (37) takes its minimum value over $\mathcal{S}$ at $G$. The set of all structured geometric means of $A_1, \ldots, A_m$ with respect to $\mathcal{S}$ is denoted by $G_{\mathcal{S}} = G_{\mathcal{S}}(A_1, \ldots, A_m)$.

We show that if $\mathcal{S}$ is closed (and nonempty) then $G_{\mathcal{S}}$ is nonempty and the matrices $G \in G_{\mathcal{S}}$ satisfy most of the ALM axioms in a suitably adjusted form. For instance, the invariance under inversion property (38) turns into

$$G_{\mathcal{S}}(A_1, \dots, A_m) = G_{\mathcal{S}^{-1}}(A_1^{-1}, \dots, A_m^{-1})^{-1},$$

where for a set $\mathcal{U} \subseteq \mathcal{P}_n$ we denote $\mathcal{U}^{-1} = \{X^{-1}: X \in \mathcal{U}\}$. That is, the inverse of any structured geometric mean of the matrices $A_1, \dots, A_m \in \mathcal{S}$ with respect to $\mathcal{S}$ coincides with a structured mean of the inverses $A_1^{-1}, \dots, A_m^{-1}$ with respect to the set $\mathcal{S}^{-1}$ where these inverses reside.

We show that, in many interesting cases, structured geometric means can be characterized in terms of the positive solutions of a suitable vector equation and provide algorithms for their computation.

In the Toeplitz case we also consider a different approach, where the mean is defined as a barycenter for a suitable metric on the manifold [23]. We analyze this barycenter and its properties in detail, obtaining an explicit expression in the real case and a quick algorithm in the complex case.

The cost function (37) is examined with special focus on the existence of the minimizer over a closed set. The structured matrix mean itself is the subject of study, where the theoretical properties it should satisfy are examined. We propose two algorithms for computing a structured mean $G$ in a linear space together with their convergence analysis. For one algorithm, it is shown that the convergence speed is independent of the condition number of the mean and is faster when the condition numbers of the matrices $A_i^{-1/2} G A_i^{-1/2}$ are smaller, for $i = 1, \dots, n$. Because of its nature and its convergence properties, this algorithm can be viewed as the natural extension to the structured case of the Richardson-like algorithm introduced and analyzed in [26] for the computation of the Karcher mean of unstructured matrices. For Toeplitz matrices, a different structured matrix mean [23] as a barycenter is considered, and an algorithm for computing it is developed. We show numerical experiments related to accuracy and speed for computing the structured matrix mean.

Given a matrix $A$, we define $\sigma(A)$ the spectrum of $A$, that is, the set of all the eigenvalues of $A$, and $\rho(A) = \max_{\lambda \in \sigma(A)} |\lambda|$ the spectral radius of $A$. Moreover we denote by $\| A \|_F := \left(\operatorname{trace}(A^* A)\right)^{1/2} = \left(\sum_{i,j} |a_{ij}|^2\right)^{1/2}$ the Euclidean (Frobenius) norm of $A$, and $\| A \|_2 = \rho(A^* A)^{1/2}$ is the spectral norm. By $A^*$ we denote the transposed conjugate of $A$. Recall that for a positive matrix $A$ there exists a unique positive solution to the equation $X^2 = A$. This solution, denoted by $A^{1/2}$, is called the square root of $A$[3]. Given a matrix $A \in \mathbb{C}^{n \times n}$, we use the vec-operator to build $\operatorname{vec}(A) \in \mathbb{C}^{n^2}$, a long vector obtained by stacking the columns of $A$. We will use the Kronecker product $\otimes$ such that $A \otimes B$ is the block matrix whose $(i, j)$ th block is defined as $a_{ij} B$. The vec operator and the Kronecker product interplay in the following way [34]

$$\operatorname{vec}(ABC) = (C^T \otimes A) \operatorname{vec}(B). \tag{40}$$

Finally, we recall a natural partial order in $\mathcal{P}_n$ that will be used in the following: let $A$ and $B$ be positive, we write $A \geqslant B$ if the matrix $A - B$ is semidefinite positive.

The existence of a structured geometric mean and its relation to the classical Karcher mean is studied. First some necessities are repeated.

The Riemannian geometry on $\mathcal{P}_n$ given by the inner product (35) turns out to be complete and a parametrization of the geodesic joining two positive matrices $A$ and $B$ is known to be [3]

$$A\#_t B = A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^t A^{1/2} = A(A^{-1}B)^t,\ t \in [0,1], \quad (41)$$

where the midpoint $A\#_{1/2}B$ coincides with the geometric mean of the two matrices [35], [15].

Given a set of matrices $A_1, \dots, A_m \in \mathcal{P}_n$, the function $f(X) = f(X; A_1, \dots, A_m)$ in (37) is strictly geodesically convex, which means that for any two different matrices $X, Y \in \mathcal{P}_n$, we have

$$f(X\#_t Y) < (1-t)f(X) + tf(Y),\ 0 < t < 1. \quad (42)$$

This property follows from [3], where it is stated that for $m = 1$ the function $f(X)$ is strictly geodesically convex. The case $m > 1$ follows by summing up the $m$ inequalities obtained by applying (42) to the functions $f(X) = f(X; A_i)$, for $i = 1, \dots, m$, respectively.
Geodesical convexity is a key ingredient for the proof of the existence of a unique minimizer of $f$ over $\mathcal{P}_n$ given in [3]. A different proof is obtained using the fact that $\mathcal{P}_n$, with the inner product (35), forms a Cartan-Hadamard manifold [29], [15], [42], which is a Riemannian manifold, complete, simply connected and with non-positive sectional curvature everywhere. On such a Cartan-Hadamard manifold the Karcher mean (the so-called center-of-mass) exists and is unique [17], [38], [39].
The notion of geodesical convexity in $\mathcal{P}_n$ is different from the customary convexity in the Euclidean space where one requires that

$$f\left((1-t)X + tY\right) \leqslant (1-t)f(X) + tf(Y),\ t \in [0,1].$$

In fact, the function $f$ is not convex in the traditional sense as the following example shows.
**Example (1.3.2)[20]:** Consider the set made of the unique matrix $A = 1$, and $x, y \in \mathbb{R}_+^* = \mathcal{P}_1$. We have $f(x) = \delta^2(x, A) = \log^2(x)$ which is not convex. On the other hand the function $\log^2(x)$ is strictly geodesically convex and this can be shown by an elementary argument: in fact, it is continuous and

$$\begin{aligned}\delta^2(\sqrt{xy}, 1) = \log^2(\sqrt{xy}) &= \frac{1}{4}(\log^2 x + \log^2 y + 2\log x \log y) \\ &= \frac{1}{2}(\log^2 x + \log^2 y) - \frac{1}{4}(\log x - \log y)^2 \\ &< \frac{1}{2}(\log^2 x + \log^2 y) = \frac{1}{2}(\delta^2(x,1) + \delta^2(y,1)).\end{aligned}$$

Iterative selection of midpoints, by using midpoints and a continuity argument completes the proof.
Since $f$ is strictly geodesically convex, it can be proved that it has a unique minimizer over any closed, geodesically convex subset $\mathcal{S}$ of $\mathcal{P}_n$, where we say that a subset $\mathcal{S} \subseteq \mathcal{P}_n$ is geodesically convex if for any $X, Y \in \mathcal{S}$, the entire geodesic $X\#_t Y, t \in [0,1]$ belongs to $\mathcal{S}$. Indeed, if $X_1$ and $X_2$ were two different matrices in $\mathcal{S}$ where $f$ takes its minimum, then from (42) it would follow that $f(X_1\#_t X_2) < f(X_1) = f(X_2)$ for any $0 < t < 1$ which contradicts the assumption.

For a generic closed subset $\mathcal{U}$ of $\mathcal{P}_n$, which is not necessarily geodesically convex, we can prove the existence of a minimum point by using the fact that $f(X)$ is continuous. In order to prove this, we first give a couple of preliminary results.
**Lemma (1.3.3)[20]:** Let $A, X, Y \in \mathcal{P}_n$ be such that $Y = A^{-1/2}XA^{-1/2}$. Then for any operator norm,

$$\| Y \| \geqslant \| X \| / \|A^{1/2}\|^2, \ \|Y^{-1}\| \geqslant \|X^{-1}\| / \|A^{-1/2}\|^2.$$

**Proof.** The condition $Y = A^{-1/2}XA^{-1/2}$ can be rewritten as $X = A^{1/2}YA^{1/2}$. Taking norms yields $\| X \| \leqslant \|A^{1/2}\|^2 \| Y \|$ from which the first inequality follows. The second inequality holds similarly starting from $Y^{-1} = A^{1/2}X^{-1}A^{1/2}$.

**Lemma (1.3.4)[20]:** For the function $\delta^2(X, A)$ we have

$$\delta^2(X, A) \geqslant \log^2 s$$

where $s = \max\{\rho(A^{-1/2}XA^{-1/2}), \rho(A^{1/2}X^{-1}A^{1/2})\}$.

**Proof.** This follows from the equation

$$\delta^2(X, A) = \sum_i \log^2 \lambda_i(A^{-1/2}XA^{-1/2})$$

and the fact that all terms are positive, implying that $\sum_i \log^2 \lambda_i(A^{-1/2}XA^{-1/2})$ is greater than any single term in the summation, in particular those given by the extreme eigenvalues of $A^{-1/2}XA^{-1/2}$, that is, the spectral radius $\rho(A^{-1/2}XA^{-1/2})$ and its inverse $\rho(A^{1/2}X^{-1}A^{1/2})$.

**Theorem (1.3.5)[20]:** Let $\mathcal{U} \subseteq \mathcal{P}_n$ be a closed subset. Then for any $A_1, \ldots, A_m \in \mathcal{P}_n$ the function $f(X) = f(X; A_1, \ldots, A_m)$ has a minimum in $\mathcal{U}$.

**Proof.** If $\mathcal{U}$ is bounded, then it is compact and the continuous function $f(X)$ has a minimum in it, so we may assume that $\mathcal{U}$ is unbounded. Let $t > 0$ and $\mathcal{A}_t = \{X \in \mathcal{P}_n\colon \ \| X \|_2 \leqslant t, \|X^{-1}\|_2 \leqslant t\}$, such that $A_t$ is closed and bounded. We claim that there exists a sufficiently large value $t$ such that outside the set $\mathcal{U} \cap \mathcal{A}_t$ the function $f(X)$ takes values larger than $\gamma = \inf_{X \in \mathcal{U}} f(X)$. In this way, the set where we minimize the function can be restricted to $\mathcal{U} \cap \mathcal{A}_t$ which is compact and hence $f(X)$ takes its minimum over it. For simplicity, we prove the existence of $t$ for $m = 1$. The case $m > 1$ can be obtained by using the same arguments.

Combining Lemma (1.3.3) with $\|\cdot\| = \|\cdot\|_2$, Lemma (1.3.4), and using the properties of the spectral norm, we find that there exist positive constants $\alpha$ and $\beta$ (depending on $A$ ) such that

$$\delta^2(X, A) \geqslant \max \ \{\log^2(\alpha \| X \|_2), \log^2(\beta \|X^{-1}\|_2)\} \qquad (43)$$

for any $X \in \mathcal{P}_n$. Choosing $t$ sufficiently large in such a way that $\log^2(\alpha t), \log^2(\beta t) > \gamma$, it follows from (43) that $\delta^2(X, A) > \gamma$ for any $X$ having $\| X \|_2 > t$ or $\|X^{-1}\|_2 > t$. This completes the proof of the existence of a minimum of $f(X, A)$. Considering the summation in (37) this generalizes to an arbitrary $f(X)$.

In general, uniqueness of the point where $f(X)$ takes its minimum cannot be guaranteed. For instance, if both $A$ and $A^{-1}$ belong to $\mathcal{U}$ while $I = A\#_{1/2}A^{-1}$ does not, then the function $f_1(X) := \delta^2(X, A) + \delta^2(X, A^{-1})$ reaches its minimum at a point $G \neq I \in \mathcal{U}$. Clearly, $f_1(G^{-1}) = f_1(G)$ and if $G^{-1} \neq G$ belongs to $\mathcal{U}$, then we have at the following.

**Example (1.3.6)[20]:** Consider the $2 \times 2$ matrices $A = I$ and $B = \begin{bmatrix} a & 0 \\ 0 & a^{-1} \end{bmatrix}$, where $a > 1$ Define the segment $\mathcal{U} = \{G(t) = A + t(B - A), t \in [0,1]\}$, which is closed and convex, but not geodesically convex. The function $f(t) = \delta^2(G(t), A) + \delta^2(G(t), B)$ takes the form $f(t) = \log^2((1 - t)/a + t) + \log^2(a(1 - t) + t) + \log^2((1 - t) + t/a) + \log^2((1 - t) + at)$ and is symmetric with respect to $t = 1/2$. For $a = 200$ the function has the graph shown in Figure 1 with a local maximum at $t = 1/2$ and two global minima close to the edges of the segment.

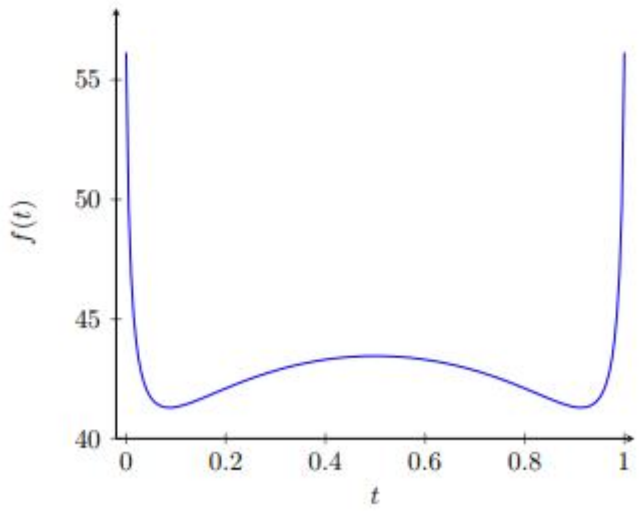



FIGURE 1. Graph of $f(t) = \delta^2(G(t), A) + \delta^2(G(t), B)$ for $G(t) = A + t(B - A)$ with $A = I$ and $B = \mathrm{diag}(200, 1/200)$.

We discuss the relation between the structured and generic geometric mean, together with the adaptation of the generic properties to the structured setting. We will discuss just the real case, the set $\mathcal{P}_n$ stands for the manifold of real positive definite matrices whose tangent space is the set of real and symmetric matrices.

The properties shown imply that a structured geometric mean with respect to $\mathcal{U}$, as defined always exists for any closed subset $\mathcal{U}$ of $\mathcal{P}_n$. In particular, this holds in the cases where $\mathcal{U} = \mathcal{S} \cap \mathcal{P}_n$ for any linear space $\mathcal{S}$ of matrices $\mathcal{S}$ and also for $\mathcal{U}^{-1} := \mathcal{S}^{-1} \cap \mathcal{P}_n$ where $\mathcal{S}^{-1} = \{A^{-1} : A \in \mathcal{S}, \det A \neq 0\}$. This captures a wide class of interesting structures emerging in applications, e.g., Toeplitz and band matrices, as well as their inverses. For simplicity we will restrict our analysis in the remainder to the real case.

More general structures are given in terms of a parametrization $\sigma(t): \mathcal{V} \to \mathbb{R}^{n\times n}$, with $\sigma$ a differentiable function defined in the open subset $\mathcal{V}$ of $\mathbb{R}^q$, which we will call the parameter space. The set $\mathcal{T} = \sigma(\mathbb{R}^q)$ is the structure determined by $\sigma$. If $\sigma$ is linear and $\mathcal{V} = \mathbb{R}^q$, then $\mathcal{T}$ is a linear space. Examples of sets $\mathcal{T}$ of interest which generally do not form a linear space are the set of matrices with a given displacement rank [25], the set of semiseparable [47], and quasiseparable matrices [17]. For an $n \times n$ symmetric Toeplitz matrix, a possible parametrization is given by

$$\sigma(t) = \sigma([t_0, t_1, \dots, t_{n-1}]) = \begin{bmatrix} t_0 & t_1 & \cdots & t_{n-1} \\ t_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_1 \\ t_{n-1} & \cdots & t_1 & t_0 \end{bmatrix}. \qquad (44)$$

For a band matrix, one can, e.g., just store the nonzero-elements in a long vector and map them onto their exact locations. In the following, given a closed set $\mathcal{T}$ we let $\mathcal{U} = \mathcal{T} \cap \mathcal{P}_n$.

In Example (1.3.6) we illustrated that the minimum of the cost function restricted to a closed subset $\mathcal{U} \subseteq \mathcal{P}_n$ is not necessarily unique. For this reason, we consider the structured geometric mean $G_U = G_{\mathcal{U}}(A_1, \dots, A_m)$ of $A_1, \dots, A_m \in \mathcal{U}$ as the set of matrices in $\mathcal{U}$ where the function $f(X)$ attains its minimum. Formally speaking, for $A_1, \dots, A_m \in \mathcal{U}$, let $g \in \mathbb{R}^q$ be such that $\hat{G} = \sigma(g) \in G_{\mathcal{U}}(A_1, \dots, A_m)$, then

$$f(\sigma(g); A_1, \dots, A_m) = \min_{t\in\mathbb{R}^q} f(\sigma(t); A_1, \dots, A_m).$$

Since $\mathcal{U} \subseteq \mathcal{P}_n$, the minimum over $\mathcal{P}_n$ is less than or equal to the minimum over $\mathcal{U}$. In general it will often happen that $\hat{G} \neq G(A_1, \dots, A_m)$ like in (39).

Some desired properties for a matrix geometric mean were stated by Ando, Li and Mathias in [2], of which the most noticeable are enlisted here.

Consistency with scalars: If $A_1, \dots, A_m$ commute, then

$$G(A_1, \dots, A_m) = (A_1 \cdots A_m)^{1/m}.$$

Permutation invariance: For any permutation $\pi$ of $\{1, \dots, k\}$,

$$G(A_1, \dots, A_m) = G\big(A_{\pi(1)}, \dots, A_{\pi(m)}\big).$$

Joint homogeneity:

$$G(\alpha_1 A_1, \alpha_2 A_2, \dots, \alpha_m A_m) = (\alpha_1 \cdots \alpha_m)^{1/m} G(A_1, \dots, A_m).$$

Monotonicity: If $A_i \geqslant A_i'$, for $i = 1, \dots, k$, then

$$G(A_1, \dots, A_m) \geqslant G(A_1', \dots, A_m').$$

Invariance under congruence: For any nonsingular $M$,

$$G(M^* A_1 M, \dots, M^* A_m M) = M^* G(A_1, \dots, A_m) M.$$

Invariance under inversion:

$$G(A_1, \dots, A_m)^{-1} = G(A_1^{-1}, \dots, A_m^{-1}).$$

Arithmetic-geometric-harmonic mean inequality:

$$\frac{1}{m}(A_1 + \cdots + A_m) \geqslant G(A_1, \dots, A_m) \geqslant m(A_1^{-1} + \cdots + A_m^{-1})^{-1}.$$

Yet another property naturally desired of a geometric mean, but not required in the list of Ando, Li and Mathias, is the repetition invariance, that is, for any set of positive matrices $A_1, \dots, A_m \in \mathcal{P}_n$,

$$G(A_1, \dots, A_m, A_1, \dots, A_m) = G(A_1, \dots, A_m). \qquad (45)$$

Now, we consider the properties of the structured geometric mean. Some properties such as the permutation invariance trivially hold, others should be restated. In fact, in the generic case the structures we consider are neither invariant under inversion nor under congruence. That is because if $A \in \mathcal{U}$ then it is not necessarily true that $A^{-1} \in \mathcal{U}$ or $M^* A M \in \mathcal{U}$.

We start with the invariance under inversion as this is one of the most characteristic properties of the geometric mean. To this end we consider the set $\mathcal{T}^{-1} = \{T^{-1} : T \in \mathcal{T}, \det T \neq 0\}$ parametrized with the function $\sigma_{-1}(t) := \sigma(t)^{-1}$. Clearly, the intersection $\mathcal{U}$ of $\mathcal{T}$ with $\mathcal{P}_n$ yields always invertible matrices, so that $\mathcal{T}^{-1} \cap \mathcal{P}_n = \mathcal{U}^{-1}$.

According to our definition, the structured geometric mean of $A_1^{-1}, \dots, A_m^{-1} \in \mathcal{U}^{-1}$ is given by the set $G_{\mathcal{U}^{-1}}(A_1^{-1}, \dots, A_m^{-1})$. For any $\tilde{G} \in G_{\mathcal{U}^{-1}}$, we have $\tilde{G} = \sigma(\tilde{g})^{-1}$ such that

$$f(\sigma(\tilde{g})^{-1}; A_1^{-1}, \dots, A_m^{-1}) = \min_{t \in \mathbb{R}^q} f(\sigma(t)^{-1}; A_1^{-1}, \dots, A_m^{-1}).$$

Since $\delta(A, B) = \delta(A^{-1}, B^{-1})$, one gets $f(X; A_1, \dots, A_m) = f(X^{-1}; A_1^{-1}, \dots, A_m^{-1})$ so that

$$f(\sigma(\tilde{g}); A_1, \dots, A_m) = \min_{t \in \mathbb{R}^q} f(\sigma(t); A_1, \dots, A_m)$$

and thus $\tilde{G}^{-1} \in G_{\mathcal{U}}(A_1, \dots, A_m)$. Since $\tilde{G}$ was chosen arbitrarily, and since $\mathcal{U}$ can be interchanged with $\mathcal{U}^{-1}$, we have the analogue of the invariance under inversion for the structured geometric mean:

$$G_{\mathcal{U}}(A_1, \dots, A_m)^{-1} = G_{\mathcal{U}^{-1}}(A_1^{-1}, \dots, A_m^{-1}). \qquad (46)$$

In a similar manner we can restate the invariance under congruence in a structured style by defining, for any nonsingular $M$, the set $\mathcal{U}_M := M^* \mathcal{U} M = \{M^* T M : T \in \mathcal{U}\}$. The invariance under congruence is then understood as

$$G\mathcal{U}_M(M^* A_1 M, \dots, M^* A_m M) = M^* G_{\mathcal{U}}(A_1, \dots, A_m) M.$$

Joint homogeneity, in order to be defined, requires that the set $\mathcal{T}$ satisfies the following property:

$$A \in \mathcal{T} \Rightarrow \alpha A \in \mathcal{T}$$

for any scalar $\alpha > 0$. This property clearly holds if $\mathcal{T}$ is a linear space or the set formed by the inverses of the nonsingular matrices of a linear space. For these sets, the joint homogeneity holds.
Repetition invariance holds true as well by (37), since

$$f(X; A_1, \dots, A_m, A_1, \dots, A_m) = 2f(X; A_1, \dots, A_m),$$

so the minimizers (over a subset) of the functions $f(X; A_1, \dots, A_m, A_1, \dots, A_m)$ and $f(X; A_1, \dots, A_m)$ are the same.
Regarding the remaining properties, we observe that the consistency with scalars is violated, as Example (1.3.1) shows. Weaker consistency properties hold, such as idempotency, namely $G_{\mathcal{U}}(A, A, \dots, A) = A$ for each structure $\mathcal{U}$ and $A \in \mathcal{U}$.
Moreover, if the set $\mathcal{U}$ is closed and geodesically convex then

$$G_{\mathcal{U}}(A_1, \dots, A_m) = G(A_1, \dots, A_m),$$

so the geometric and structured geometric mean coincide. An interesting case of a geodesically convex set is given by $\mathcal{U} = \mathcal{T} \cap \mathcal{P}_n$, when $\mathcal{T}$ is an algebra, i.e., a linear space closed under multiplication and inversion.
Finally, the properties related to the ordering of positive matrices such as monotonicity are not true as shown by the following numerical example.
**Example (1.3.7)[20]:** We consider the four Toeplitz matrices

$$T_1 = \begin{bmatrix} 1 & 1/2 & 1/2 \\ 1/2 & 1 & 1/2 \\ 1/2 & 1/2 & 1 \end{bmatrix}, \ T_2 = T_1, \ T_3 = \begin{bmatrix} 3/4 & 1/2 & 0 \\ 1/2 & 3/4 & 1/2 \\ 0 & 1/2 & 3/4 \end{bmatrix}, \ S = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix},$$

and, using the algorithms presented, we compute a structured geometric mean $G_\varepsilon$ of the three matrices $T_1, T_2$ and $T_3 + \varepsilon S$ for various $\varepsilon \geqslant 0$. The norm of $G_\varepsilon - G_0$ becomes small as $\varepsilon$ tends to 0 and we observe that $G_\varepsilon - G_0$ is not positive (semi)definite, while $T_3 + \varepsilon S \geqslant T_3$. This gives numerical evidence of the lack of monotonicity of a structured geometric mean. On the other hand, computing the arithmetic mean $A$ of $T_1, T_2$ and $T_3$, one observes also that the expected inequality $A \geqslant G_0$ does not hold in this case.

We start from the Karcher mean, which is obtained as the unique solution in $\mathcal{P}_n$ of the matrix equation

$$\sum_{i=1}^{m} \log(XA_i^{-1}) = 0. \tag{47}$$

Equation (47) is obtained using the fact that $f$ is differentiable and has a minimum at the Karcher mean. Thus the Karcher mean satisfies the condition $\nabla f_X = 0$, where $\nabla f_X = 2X^{-1}\sum_{i=1}^{m} \log(XA_i^{-1})$ denotes the (Euclidean) gradient of $f$ with respect to $X$ (see [37], [17]). In the general case, the restriction of $f$ to a structure given by $\sigma(t)$ is investigated. For any minimum $g$ (with corresponding $(g)$ ) not located at the boundary of the parameter space, the gradient $\nabla(f \circ \sigma)_t$ of the function with respect to $t$ must be zero, so we are interested in the solutions of the vector equation $\nabla(f \circ \sigma)_t = 0$.
From the chain rule of derivation, one obtains that

$$\nabla(f \circ \sigma)_t = \left( \sum_{i,j} \frac{\partial f(\sigma(t))}{\partial x_{i,j}} \frac{d\sigma_{i,j}(t)}{dt_s} \right)_{s=1,\dots,q} = 0$$

which leads to the vector equation

$$\sum_{i,j} (\Gamma(\sigma(t)))_{i,j} \frac{d\sigma_{i,j}(t)}{dt_s} = 0, s = 1, \dots, q, \tag{48}$$

where $\Gamma(X) := \frac{1}{2}\nabla f_X$.

In the case where $\mathcal{T}$ is a linear space, the function $\sigma(t)$ is linear and can be written in matrix form as

$$\text{vec}(\sigma(t)) = Ut, U \in \mathbb{R}^{n^2 \times q},$$

so that equation (48) turns into

$$U^T \text{vec}(\Gamma(\sigma(t))) = 0,\ \Gamma(X) = X^{-1} \sum_{i=1}^{m} \log(XA_i^{-1}). \tag{49}$$

If $\sigma(t)$ is chosen to be orthogonal, i.e. such that $U^T U = I$, then $U^T$ coincides with the Moore-Penrose inverse of $U$.

When $\mathcal{T}$ denotes the set of symmetric Toeplitz matrices, the parametrization (44) leads to a matrix $U$ having orthogonal columns. In fact one has $U^T U = D = \text{diag}(n, 2(n-1), 2(n-2), \dots, 2)$. In particular, for $n = 3$ one has

$$U^T = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

For $\mathcal{T}$ being the set of symmetric tridiagonal matrices the parametrization

$$\sigma(t) = \begin{bmatrix} t_1 & t_{n+1} & & & \\ t_{n+1} & t_2 & t_{n+2} & & \\ & \ddots & \ddots & \ddots & \\ & & t_{2n-2} & t_{n-1} & t_{2n-1} \\ & & & t_{2n-1} & t_n \end{bmatrix}$$

also leads to a matrix $U$ having orthogonal columns. Moreover, $U^T U = \text{diag}(I_n, 2I_{n-1})$. For $n = 3$, e.g., one has

$$U^T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}.$$

We will give two algorithms for computing structured geometric means when they are characterized in terms of the solutions $g$ of a vector equation, as, for instance, in the linear case.

We first provide a general definition of a class of algorithms based on preconditioned functional iteration, then we specialize to two algorithms given by two different preconditioners.

The first, provided is derived by relying on the projection of the gradient with respect to the Euclidean scalar product. The second, presented, is obtained through projection with respect to the Riemannian metric of $\mathcal{P}_n$ decribed.

We assume that $A_1, \dots, A_m \in \mathcal{U}$, where $\mathcal{U} = \mathcal{T} \cap \mathcal{P}_n$ and $\mathcal{T}$ is a linear space with a parametrization $\sigma(t)$ such that vec $(\sigma(t)) = Ut$, and $D = U^T U$.

The structured geometric mean $G_{\mathcal{U}}$ is the set of minimizers of the function $f(X; A_1, \dots, A_m)$ over $\mathcal{U}$. These minimizers must be sought among the stationary points of the function $f$, that is, among the solutions to the vector equation (49).

Therefore, a way to design algorithms for computing structured means $G_{\mathcal{U}}$ is to apply numerical techniques to solve the vector equation (49). We consider a preconditioned Richardson-like iteration constructed in the spirit of [37]. Let $V(X)$ be a nonsingular and sufficiently differentiable matrix function and define

$$\varphi(t) = t - \theta S(t),\ S(t) = V(\sigma(t))^{-1} U^T \mathrm{vec}(\Gamma(\sigma(t))),$$
$$t_{\nu+1} = \varphi(t_\nu),\ \nu = 0,1,\dots, \tag{50}$$

where $\theta$ is a parameter introduced to enhance convergence, $V(\sigma(t))$ is a preconditioner and $t_0$ is a given vector such that $\sigma(t_0)$ is positive. Observe that the fixed points of $\varphi(t)$ are the solutions of the vector equation (49) and if convergent, the sequence $t_\nu$ converges to a solution of the vector equation (49).

Given a matrix function $f(X)$, where $X = (x_{i,j})$ and $f(X)$ are $n \times n$ matrices, we denote by $J_f(G)$ the $n^2 \times n^2$ Jacobian matrix of vec $(f(X))$ with respect to the variable vec $(X)$ computed at $X = G$, similarly we denote $J_{f\circ\sigma}(t_G)$ the $n^2 \times q$ Jacobian of the composed function vec $(f(\sigma(t)))$ with respect to the variables $(t_1, \dots, t_q)$ at $t = t_G$. In this notation, the function in the subscript as well as the variable between parentheses specify if the derivatives are taken w.r.t. the matrix variable $X$ or the vector variable $t$.

Observe that if $V(\sigma(t))$ is chosen as the Jacobian of $U^T\mathrm{vec}(\Gamma(\sigma(t)))$, then (50) coincides with Newton's iteration.

If $t_G$ is a solution of (49) and if $t_\nu$ is sufficiently near to $t_G$, then

$$t_{\nu+1} - t_G = J_\varphi(t_G)(t_\nu - t_G) + O(\|t_\nu - t_G\|^2),$$

so that in order to study the local convergence of this sequence it is sufficient to estimate the spectral radius $\rho$ or any induced norm of $J_\varphi(t_G)$ and determine $\theta$ in such a way that $\rho\left(J_\varphi(t_G)\right) < 1$. Notice that the Jacobian of $\varphi(t)$ at $t = t_G$ is given by $I - \theta K$ where $K = J_S(t_G)$ is the Jacobian of $S(t)$ at $t = t_G$. Therefore, if we can find a preconditioner $V(t)$ such that $K$ has real positive eigenvalues with minimum and maximum eigenvalues $\kappa_{\min}$ and $\kappa_{\max}$ respectively, then the choice $\theta = 2/(\kappa_{\min} + \kappa_{\max})$ insures local convergence and provides the minimum spectral radius of $J_\varphi(t_G)$ given by

$$\rho\left(J_\varphi(t_G)\right) = \frac{\kappa_{\max} - \kappa_{\min}}{\kappa_{\max} + \kappa_{\min}} = \frac{\mu - 1}{\mu + 1} < 1,\ \mu = \kappa_{\max}/\kappa_{\min}.$$

Moreover, any values $\hat{\kappa}_{\min} \leqslant \hat{\kappa}_{\max}$ such that $\hat{\kappa}_{\min} \leqslant \kappa_{\min} \leqslant \kappa_{\max} \leqslant \hat{\kappa}_{\max}$ can be used instead of $\kappa_{\min}$ and $\kappa_{\max}$ to determine a value $\hat{\theta} = 2/(\hat{\kappa}_{\min} + \hat{\kappa}_{\max})$ which insures convergence. Also notice that the closer $\mu$ is to 1 the faster is the convergence of the iteration.

We perform a spectral analysis of $K$ and to find an upper bound to the ratio $\mu = \kappa_{\max}/\kappa_{\min}$, assuming that all the eigenvalues of $K$ are real positive. From the composition rule of derivatives one finds that

$$K = V(\sigma(t_G))^{-1} U^T J_\Gamma(G) U + J_{V(\sigma(t_G))^{-1}}(\sigma(t_G)) U^T \mathrm{vec}\left(\Gamma(\sigma(t_G))\right)$$

and since $U^T\mathrm{vec}\left(\Gamma(\sigma(t_G))\right) = 0$, it follows that

$$K = V(\sigma(t_G))^{-1} U^T J_\Gamma(G) U. \tag{51}$$

To evaluate $J_\Gamma(G)$, we recall that $\Gamma(X) = \sum_{i=1}^m X^{-1}\log(XA_i^{-1})$, so that it is sufficient to determine the formal expression of $J_\psi(G)$ for $\psi(G, A) = G^{-1}\log(GA^{-1})$ for a generic $A$ and then to write $J_\Gamma(G) = \sum_{i=1}^m J_{\psi(G,A_i)}(G)$. In order to evaluate $J_\psi(G)$, we rely on the definition of Fréchet derivative of a matrix function $f(X)$ at $X$ in the direction $E$

$$Df_X[E] = \lim_{t\to 0}\frac{f(X+tE)-f(X)}{t} = \frac{d}{dt}\bigg|_{t=0} f(X+tE).$$

In fact, the $n^2 \times n^2$ Jacobian matrix $J_f(X)$ of the vector function vec of o vec $^{-1}$ at $\mathrm{vec}(X)$ is related to the Fréchet derivative by the equation

$$\mathrm{vec}(Df_X[E]) = J_f(X)\,\mathrm{vec}(E). \qquad (52)$$

We recall also the following properties of the Fréchet derivative [14] where $f, g$ are given matrix functions and $(X) = X^{-1}$ :

$$\begin{aligned} &D(fg)_X[E] = Df_X[E]g(X) + f(X)Dg_X[E], && \text{product rule,}\\ &D(f\circ g)_X[E] = Df_{g(X)}\big[Dg_X[E]\big], && \text{chain rule,} \qquad (53)\\ &D\varphi_X[E] = -X^{-1}EX^{-1}, && \text{inversion.}\end{aligned}$$

For the derivative of the exponential function we have (see [14])

$$J_{\exp}(Y) = (I\otimes \exp Y)\beta(Y^T\otimes I - I\otimes Y),\ \beta(z) = (e^z-1)/z.$$

Therefore, since $J_{\log}(X) = J_{\exp}(Y)^{-1}$ for $Y = \log X$, we find that

$$J_{\log}(X) = \gamma(\log(X^T)\otimes I - I\otimes \log X)(I\otimes X^{-1}),\ \gamma(z) = z/(e^z-1). \qquad (54)$$

We provide an explicit expression of the Fréchet derivative of the function $\psi(X,A) = X^{-1}\log(XA^{-1})$ and of the Jacobian $J_{\psi(X,A)}(X)$.

**Lemma (1.3.8)[20]:** Let $\psi(X) = X^{-1}\log(XA^{-1})$. Assume that $A, X$ are positive. For the matrix $J_\psi(X)$ such that $\mathrm{vec}(D\psi_X[E]) = J_\psi(X)\mathrm{vec}(E)$ we have

$$J_\psi(X) = -X^{-1}\log(XA^{-1})\otimes X^{-1} + (A^{-1}\otimes X^{-1})\gamma(W)(I\otimes AX^{-1}),$$
$$W = \log(XA^{-1})\otimes I - I\otimes\log(XA^{-1}),$$

with $\gamma(z) = z/(e^z-1)$.

**Proof.** Since $h(X) := \log(XA^{-1})$ is the composition of $f(X) = \log(X)$ and $(X) = XA^{-1}$, we get by (53)

$$Dh_X[E] = D\log_{XA^{-1}}[EA^{-1}].$$

As $\psi(X)$ is the product of $f(X) = X^{-1}$ and $h(X)$, (53) gives us

$$D\psi_X[E] = -X^{-1}EX^{-1}\log(XA^{-1}) + X^{-1}Dh_X[E].$$

Combining the latter two equations yields

$$D\psi_X[E] = -X^{-1}EX^{-1}\log(XA^{-1}) + X^{-1}D\log_{XA^{-1}}[EA^{-1}].$$

By using (52) and (40) we find that the matrix $J_\psi(X)$ representing $D\psi_X$ is given by

$$J_\psi(X) = -(X^{-1}\log(XA^{-1}))^T\otimes X^{-1} + (I\otimes X^{-1})J_{\log}(XA^{-1})(A^{-T}\otimes I).$$

Replacing (54) in the equation above and using the fact that $A = A^T, X = X^T$ yields

$$\begin{aligned} J_\psi(X) &= -\log(A^{-1}X)X^{-1}\otimes X^{-1}\\ &\quad +(I\otimes X^{-1})\gamma\big(\log(A^{-1}X)\otimes I - I\otimes\log(XA^{-1})\big)(A^{-1}\otimes AX^{-1}).\end{aligned}$$

Using the fact that $W\log(V)W^{-1} = \log(WVW^{-1})$, the first term can be written as $-X^{-1}\log(XA^{-1})\otimes X^{-1}$. The sécond term can bee writtèn ás $(I\otimes X^{-1})(A^{-1}\otimes \mathrm{I})$ $\gamma\big(\log(XA^{-1})\otimes I - I\otimes\log(XA^{-1})\big)(I\otimes AX^{-1})$, which completes the proof.

Recall that $\Gamma(X) = \sum_{i=1}^m \psi(X, A_i)$ and $G^{-1}\sum_{i=1}^m \log(GA_i^{-1}) = 0$, for $G = \sigma(t_G)$. Then by Lemma (1.3.8), we obtain the following formula for the Jacobian $J_\Gamma(\sigma(t))$:

$$J_\Gamma(G) = (I\otimes G^{-1})H(I\otimes G^{-1}),\ H = \sum_{i=1}^m H_i,$$
$$H_i = (A_i^{-1}\otimes I)\gamma\left(\log(GA_i^{-1})\otimes I - I\otimes\log(GA_i^{-1})\right)(I\otimes A_i).$$

Moreover, by using the properties of the Kronecker product and the fact that $\log(GA^{-1}) = A^{1/2}\log\left(A^{-1/2}GA^{-1/2}\right)A^{-1/2}$, we can write

$$H_i = \left(A_i^{-1/2} \otimes A_i^{1/2}\right)\gamma(\log M_i \otimes I - I \otimes \log M_i)\left(A_i^{-1/2} \otimes A_i^{1/2}\right),$$
$$M_i = A_i^{-1/2}GA_i^{-1/2}.$$

From this expression it turns out that $H_i$ is positive, and from (51) we find that $J_S(t_G)$ is the product of the matrices $V\left(\sigma(t_G)\right)^{-1}$ and the positive matrix $U^T(I \otimes G^{-1})\sum_{i=1}^{m} H_i(I \otimes G^{-1})U$.

Thus we may conclude with the following

**Theorem (1.3.9)[20]:** The Jacobian $K$ of the function $S(t)$ in (51) at $\sigma(t_G) = G$ is given by

$$K = V^{-1}U^T(I \otimes G^{-1})H(I \otimes G^{-1})U,$$
$$H = \sum_{i=1}^{m} H_i,\ H_i = \left(A_i^{-1/2} \otimes A_i^{1/2}\right)\gamma(\log M_i \otimes I - I \otimes \log M_i)\left(A_i^{-1/2} \otimes A_i^{1/2}\right),$$
$$M_i = A_i^{-1/2}GA_i^{-1/2},$$
$$\gamma(z) = z/(e^z - 1).$$

Moreover, the eigenvalues of $K$ are the solutions of the equation

$$\det(\kappa V - U^T(I \otimes G^{-1})H(I \otimes G^{-1})U) = 0.$$

The simplest choice for the preconditioner $V(t)$ in (50) is $V(t) = U^TU = D$. This corresponds to projecting the gradient of the function $f\left(X, A_1, \dots, A_p\right)$ on the set $\mathcal{U}$ according to the Euclidean scalar product. The problem $\det(\kappa I - K) = 0$ turns into the generalized $q$-dimensional symmetric eigenvalue problem

$$\det\left(U^T\left(\kappa I - (I \otimes G^{-1})H(I \otimes G^{-1})\right)U\right) = 0.$$

This problem is the projection on the space spanned by the columns of $U$ of the problem $\det\left(\nu I - (I \otimes G^{-1})H(I \otimes G^{-1})\right) = 0$, which has real positive solutions.

Now we recall the following result, valid for general positive matrices $A, B$, which relates the generalized eigenvalues of the pair $(A, B)$ to the ones of the projected pair $(U^TAU, U^TBU)$.

**Lemma (1.3.10)[20]:** Let $A, B$ be positive $n \times n$ matrices and $U$ an $n \times m$ matrix. Then the generalized eigenvalues of the pair $(U^TAU, U^TBU)$, which solve the equation $\det(U^T(A - \kappa B)U) = 0$, are real positive and lie in between the maximum and minimum eigenvalues $\lambda$ of the pair $(A, B)$, which satisfy $\det(A - \lambda B) = 0$. Moreover, the extreme eigenvalues $\lambda_{\min}, \lambda_{\max}$ of the pair $(A, B)$ are bounded by the inequality $\alpha_{\min}/\beta_{\max} \leqslant \lambda_{\min} \leqslant \lambda_{\max} \leqslant \alpha_{\max}/\beta_{\min}$, where $\alpha_{\min}, \alpha_{\max}, \beta_{\min}, \beta_{\max}$ are the minimum and maximum eigenvalues of the matrices $A$ and $B$, respectively.

**Proof.** The condition $\det(\lambda B - A) = 0$ is equivalent to $\det\left(\lambda I - B^{-1/2}AB^{-1/2}\right) = 0$, which has real positive solutions since $B^{-1/2}AB^{-1/2}$ is positive. The remaining part of the lemma follows from the fact that maximum and minimum eigenvalues of the larger and smaller problems coincide with maximum and minimum value of the Rayleigh quotient $x^TAx/x^TBx$ for $x \in \mathbb{R}^n$, and for $x \in \operatorname{span}(U)$, respectively.

A first consequence of the above lemma is that the extreme eigenvalues $\kappa_{\min}$ and $\kappa_{\max}$ of $K$ are in between the maximum and the minimum eigenvalue of the $n^2$-dimensional symmetric matrix $Y = (I \otimes G^{-1})H(I \otimes G^{-1})$, so that the ratio $\mu$ between the maximum and minimum eigenvalue of $K$ is less than or equal to the condition number $\mu(Y)$ of the symmetric matrix $Y$. Moreover since $Y = \sum_{i=1}^{m} Y_i$ with

$$Y_i = \left(A_i^{-1/2} \otimes A_i^{-1/2}\right)(I \otimes M_i^{-1})\gamma(\log M_i \otimes I - I \otimes \log M_i)(I \otimes M_i^{-1})\left(A_i^{-1/2} \otimes A_i^{-1/2}\right),$$

one finds that $\hat{k}_{\min} := \sum_{i=1}^{m} \lambda_{\min}^{(i)} \leqslant \kappa_{\min}$ and $\hat{k}_{\max} := \sum_{i=1}^{m} \lambda_{\max}^{(i)} \geqslant \kappa_{\max}$, where $\lambda_{\min}^{(i)}$ and $\lambda_{\max}^{(i)}$ are the minimum and the maximum eigenvalues of $Y_i$. Moreover, from Lemma (1.3.10) and from the expression above for $Y_i$ it follows that $\lambda_{\min}^{(i)} \geqslant \gamma_{\min}^{(i)} / \left(\alpha_{\max}^{(i)}\right)^2, \lambda_{\max}^{(i)} \leqslant \gamma_{\max}^{(i)} / \left(\alpha_{\min}^{(i)}\right)^2$, where $\alpha_{\min}^{(i)}, \alpha_{\max}^{(i)}$ are the minimum and the maximum eigenvalues of $A_i$, respectively, while $\gamma_{\min}^{(i)}$ and $\gamma_{\max}^{(i)}$ are the minimum and maximum eigenvalues of $(I \otimes M_i^{-1})\gamma(\log M_i \otimes I - I \otimes \log M_i)(I \otimes M_i^{-1})$, respectively. maximum eigenvalues of $(I \otimes M_i^{-1})\gamma(\log M_i \otimes I - I \otimes \log M_i)(I \otimes M_i^{-1})$, respectively. From the properties of the matrix function $\gamma(\cdot)$ and from the properties of the Kronecker product one finds that the eigenvalues of the latter matrix can be explicitly given in terms of the eigenvalues $\nu_r^{(i)}$ of the matrix $M_i$. In fact, they coincide with $\frac{1}{\left(\nu_s^{(i)}\right)^2}\left(\log t_{r,s}^{(i)}\right) / \left(t_{r,s}^{(i)} - 1\right)$ where $t_{r,s}^{(i)} = \frac{\nu_r^{(i)}}{\nu_s^{(i)}}$.

Since the function $(\log t)/(t-1)$ is monotonically decreasing, its minimum and maximum are

$$\eta_{\min}^{(i)} = (\log \mu^{(i)})/(\mu^{(i)} - 1),$$
$$\eta_{\max}^{(i)} = \log(1/\mu^{(i)})/(1/\mu^{(i)} - 1) = \mu^{(i)}(\log \mu^{(i)})/(\mu^{(i)} - 1),$$

for $\mu^{(i)} = \mu(M_i)$ the spectral condition number of $M_i$. Additionally, taking the factor $(\nu^{(i)})^{-2}$ into consideration gives

$$\gamma_{\min}^{(i)} \geqslant \eta_{\min}^{(i)} \left(\nu_{\max}^{(i)}\right)^{-2},$$
$$\gamma_{\max}^{(i)} \leqslant \eta_{\max}^{(i)} \left(\nu_{\min}^{(i)}\right)^{-2} \leqslant \mu^{(i)} \left(\nu_{\min}^{(i)}\right)^{-2},$$

where $\nu_{\min}^{(i)}$ and $\nu_{\max}^{(i)}$ represent respectively the minimum and maximimum eigenvalue of $M_i$.

Therefore, we may conclude that the eigenvalues of $K$ are bounded by $\tilde{\kappa}_{\min} := \sum_{i=1}^{m} \eta_{\min}^{(i)} / \left(\nu_{\max}^{(i)} \alpha_{\max}^{(i)}\right)^2$ and $\tilde{\kappa}_{\max} := \sum_{i=1}^{m} \eta_{\max}^{(i)} / \left(\nu_{\min}^{(i)} \alpha_{\min}^{(i)}\right)^2$.

Observe that this bound gets worse when either some matrix is ill-conditioned or if some matrix $A_i^{-1/2} G A_i^{-1/2}$ is ill-conditioned. The latter case cannot occur if the matrices $A_i$ do not differ much from $G$. The dependence of this bound on the conditioning of $A_i$ makes this algorithm very inefficient as long as some $A_i$ is ill-conditioned. This drawback is overcome, where we design a more effective preconditioner.

The Karcher mean for positive matrices inherits a beautiful interpretation in terms of differential geometry. It can be considered as the center of mass for a well chosen inner product on the manifold of positive matrices. We consider two approaches inspired by this idea. For more information see [36], and [28], [33], [44]-[46].

When considering a manifold optimization approach, the intersection $\mathcal{U}$ of a linear space $\mathcal{T}$ with the manifold of positive matrices $\mathcal{P}_n$ can be viewed as a Riemannian submanifold of $\mathcal{P}_n$ itself, which in turn is called the enveloping space. This entails that the inner product from this enveloping space is induced on the submanifold. An immediate consequence is that the gradient of the cost function for the submanifold is given by the orthogonal projection (with respect to the inner product) of the gradient for the enveloping space. Similar to the space of symmetric matrices, being the tangent space to the manifold of positive matrices, the intersection $\mathcal{V}$ of the linear space $\mathcal{T}$ with the space of symmetric matrices is the tangent space to $\mathcal{U}$.

First consider the manifold of positive matrices endowed with the Euclidean inner product $g_X(A, B) = \operatorname{tr}(AB)$, with $A$ and $B$ symmetric, and $X$ a positive matrix. Note that even though this inner product $g_X$ is independent of $X$, the subscript notation is kept for consistency. In this case, the orthogonal projection of a symmetric matrix $A$ onto $\mathcal{T}$ gives a matrix $T$, with

$$\operatorname{vec}(T) = U(U^T U)^{-1} U^T \operatorname{vec}(A),$$

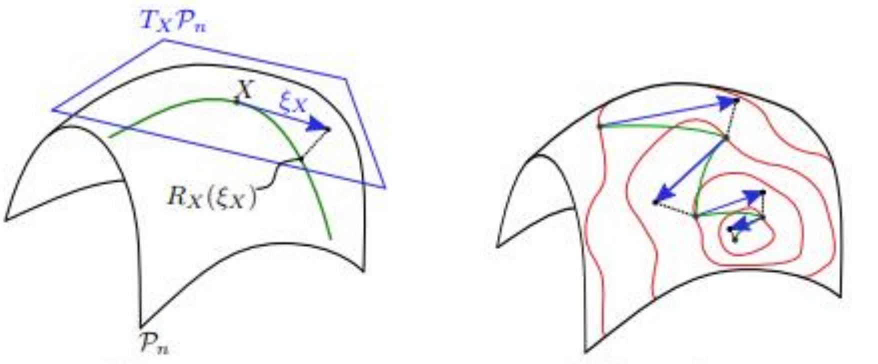
or $\operatorname{vec}(T) = Ut$, with

$$t = (U^T U)^{-1} U^T \operatorname{vec}(A). \tag{55}$$

The expression for the gradient of the Karcher cost function, corresponding to the Euclidean inner product, is known for the manifold of positive matrices and is given by

$$\operatorname{grad}_e f(X; A_1, \dots, A_m) = 2X^{-1} \sum_{i=1}^{m} \log(X A_i^{-1}). \tag{56}$$

The gradient naturally defines the direction of steepest ascent. The gradient lies in the tangent space, and to build an algorithm from this, a practical way is to follow the gradient and then go back to the manifold through a suitable function, called retraction. The precise definition of a retraction, together with general theoretical assumptions it should satisfy, can be found in [21]. Figure 2(a) graphically illustrates the concept of a retraction, where a vector $\xi_X$ in the tangent space $T_X\mathcal{P}_n$ of the positive matrices is retracted to a point $R_X(\xi_X)$ residing on the manifold $\mathcal{P}_n$. On a manifold, the classical steepest descent algorithm is graphically




(A) Retraction. (B) Steepest Descent.

FIGURE 2. Graphical representation of a retraction and steepest descent flow.

depicted in Figure 2(b). The thin red lines depict the contour lines, the blue arrows the gradients, and the green curves the retractions to the manifold.

Observe that for $\mathcal{P}_n$ immersed in the set of symmetric matrices, the tangent space at a point is the whole set of symmetric matrices. So one can consider the basic retraction $R_X(A) = X + A$ for a sufficiently small symmetric matrix $A$.

Entering now the gradient (56) in projection (55) and applying a gradient descent method with the basic retraction $R_X(A) = X + A$, we arrive exactly at the Richardson-like algorithm for finding the fixed points of function (50).
However, since the function $f$ to be minimized is defined through the distance (36), it is more natural to consider the manifold of positive matrices endowed with the inner product $g_X(A,B) = \operatorname{tr}(AX^{-1}BX^{-1})$, with $A, B$ and $X$ as before. In this case, the gradient for the enveloping space is known to be

$$\operatorname{grad}_n f(X; A_1, \ldots, A_m) = 2X \sum_{i=1}^{m} \log(A_i^{-1}X).$$

Note the difference with (56).
The orthogonal projection $T$ onto the intersection $\mathcal{V}$ (of $\mathcal{T}$ and the space of symmetric matrices) of this gradient, with respect to the Riemannian scalar product, can be found as the solution of the equations

$$\operatorname{grad}_n f(X) = T + S,$$
$$g_X(S,K) = \operatorname{tr}(SX^{-1}KX^{-1}) = 0, \ \text{ for every } K \in \mathcal{V}.$$

Writing again $\operatorname{vec}(T) = Ut$, we find in parameter space

$$t = (U^T(X^{-1} \otimes X^{-1})U)^{-1}U^T(X^{-1} \otimes X^{-1}) \operatorname{vec}(\operatorname{grad}_n f(X)). \qquad (57)$$

The factor $U^T(X^{-1} \otimes X^{-1})U$ is recurring and is abbreviated as $D_X$, where the subscript points to the intrinsic variable $X$. Observe that this Riemannian orthogonal projection can be seen as a Euclidean oblique projection where the two bases of the subspace are the columns of $U$ and $(X^{-1} \otimes X^{-1})U$, respectively.
Using this expression, it is possible to define another gradient descent method where we are now searching the fixed points of the function

$$\varphi(t) = t - \theta D_{\sigma(t)}^{-1} U^T(\sigma(t)^{-1} \otimes \sigma(t)^{-1}) \operatorname{vec}\left(\sigma(t) \sum_{\mathrm{i}=1}^{\mathrm{m}} \log\left(A_i^{-1}\sigma(t)\right)\right). \ (58)$$

Relying on (40) to incorporate the Kronecker product into the vectorization, we find that $(\sigma^{-1} \otimes \sigma^{-1})\operatorname{vec}\left(\sigma \sum_{i=1}^{m} \log(A_i^{-1}\sigma)\right) = \operatorname{vec}(\sum_{i=1}^{m} \log(A_i^{-1}\sigma)\sigma^{-1})$. Applying a property of the matrix logarithm we may rewrite the latter expression as $\operatorname{vec}\left(\sigma^{-1}\sum_{i=1}^{m} \log(\sigma A_i^{-1})\right)$.
This way, equation (58) takes the form of (50) with

$$V = U^T(G^{-1} \otimes G^{-1})U.$$

To analyze the convergence of (50) with the choice $V = U^T(G^{-1} \otimes G^{-1})U$, we have to analyze the eigenvalues of the Jacobian $K = J_S(t_G)$ of $S(t)$ in (50) where the equation $\det(\kappa I - K) = 0$ takes the form of the following generalized eigenvalue problem

$$\det\left(U^T\left(\kappa(G^{-1} \otimes G^{-1}) - (I \otimes G^{-1})H(I \otimes G^{-1})\right)U\right) = 0. \ (59)$$

Since the two matrices in equation (59) are positive, in view of Lemma (1.3.10), the solutions of this generalized eigenvalue problem are real positive and are located in between the maximum and the minimum solution of the larger problem

$$\det\left(\lambda(G^{-1} \otimes G^{-1}) - (I \otimes G^{-1})H(I \otimes G^{-1})\right) = 0,$$

which in turn can be rewritten as a standard eigenvalue problem

$$\det\left(\lambda I - \left(G^{1/2} \otimes G^{1/2}\right)(I \otimes G^{-1})H(I \otimes G^{-1})\left(G^{1/2} \otimes G^{1/2}\right)\right) = 0.$$

Since $H = \sum_{i=1}^{m} H_i$, and the matrices $H_i$ are real symmetric, the eigenvalues of this problem are located in between the sum of the minimum and the sum of the maximum eigenvalues of each subproblem

$$\det\left(\lambda I - \left(G^{\frac{1}{2}} \otimes G^{\frac{1}{2}}\right)(I \otimes G^{-1})H_i(I \otimes G^{-1})\left(G^{\frac{1}{2}} \otimes G^{\frac{1}{2}}\right)\right) = 0, (60)$$

that is $\det(\lambda(G^{-1} \otimes G) - H_i) = 0$, or equivalently $\det\big(\lambda I - (G \otimes I)H_i(I \otimes G^{-1})\big) = 0$. The matrix in the latter expression is similar to $\left(A_i^{-1/2} \otimes A_i^{-1/2}\right)(G \otimes I)H_i(I \otimes G^{-1})\left(A_i^{1/2} \otimes A_i^{1/2}\right)$, which, using the expression of $H_i$ provided in Theorem (1.3.9), can be written as

$$(M_i \otimes I)\gamma(\log M_i \otimes I - I \otimes \log M_i)(I \otimes M_i^{-1}).$$

This way, the eigenvalues of (60) can be explicitly given in terms of the eigenvalues $\nu_r^{(i)}$ of the matrix $M_i$. In fact, they coincide with the $t_{r,s}^{(i)}\left(\log t_{r,s}^{(i)}\right)/\left(t_{r,s}^{(i)} - 1\right)$ where $t_{r,s}^{(i)} = \frac{\nu_r^{(i)}}{\nu_s^{(i)}}$.

Since the function $t(\log t)/(t-1)$ is monotone, for the minimum and maximum solution to (60) we have

$$\eta_{\min}^{(i)} = \left(1/\mu^{(i)}\right)\log\left(1/\mu^{(i)}\right)/\left(1/\mu^{(i)} - 1\right) = \left(\log \mu^{(i)}\right)/\left(\mu^{(i)} - 1\right),$$
$$\eta_{\max}^{(i)} = \mu^{(i)}\left(\log \mu^{(i)}\right)/\left(\mu^{(i)} - 1\right),$$

respectively, for $\mu^{(i)} = \mu(M_i)$ the spectral condition number of $M_i$. Therefore, we may conclude that the eigenvalues of $K$ are in between $\sum_{i=1}^{m} \eta_{\min}^{(i)}$ and $\sum_{i=1}^{m} \eta_{\max}^{(i)}$. This way, we find for the optimal value of $\theta$ and for the optimal spectral radius the estimates

$$\theta = \frac{2}{\sum_{i=1}^{m} \frac{\mu^{(i)} + 1}{\mu^{(i)} - 1}\log \mu^{(i)}},$$

$$\rho = \frac{\sum_{i=1}^{m} \log \mu^{(i)}}{\sum_{i=1}^{m} \frac{\mu^{(i)} + 1}{\mu^{(i)} - 1}\log \mu^{(i)}}.$$

It is interesting to point out that in this case the convergence speed is related neither to the condition number of the geometric mean $G$ nor to those of the matrices $A_i$ but is related only to the relative distances of $G$ from each $A_i$ measured by the quantities $\mu^{(i)} = \mu(M_i), M_i = A_i^{-1/2} G A_i^{-1/2}$. The closer they are to 1 the faster is the convergence. Therefore, if the matrices to average are not to far from each other, so that the quantities $\mu(M_i)$ are close to 1 , then the optimal value of $\theta$ is close to $1/m$ and a very fast convergence is expected. This analysis is confirmed by the numerical experiments.

From the computational point of view, at each step of the iteration (50) one has to compute $U^T \mathrm{vec}(\Gamma(\sigma(t)))$ and then to solve a linear system with the matrix $V(\sigma(t))$. The former computation, based on (49), requires $O(\mathrm{mn}^3)$ arithmetic operations (ops), while the cost of the latter depends on the structure of $V(\sigma(t))$.

We examine the case where $\mathcal{U}$ is the class of symmetric Toeplitz matrices and where $\sigma(t)$ associates $t$ with the Toeplitz matrix having as first column $t$. We describe a way to make the algorithm.

Indeed, for the iteration analyzed, $V$ is the diagonal matrix with diagonal entries $(n, 2n-2, \ldots, 2)$ and the cost of solving a system with matrix $V$ amounts to $n$ divisions.

The iteration examined has a higher convergence speed but at each step an $n \times n$ system with $V = U^T(X^{-1} \otimes X^{-1})U$ must be solved, where $X$ is a symmetric positive definite Toeplitz matrix.
We split the computation in two steps. In the first, the $n^2$ entries of $V$ are computed, in the second step a standard $O(n^3)$ ops linear system solver is used. Concerning the first step we discuss two approaches.
In both approaches the inverse of the Toeplitz matrix $X$ needs to be computed, which can be done efficiently using the Gohberg Semencul formula [25]. Here, vectors $v_1, v_2, v_3, v_4$ are determined such that $X^{-1} = L(v_1)L(v_2)^T - L(v_3)L(v_4)^T$, where $L(v)$ is the lower triangular Toeplitz matrix whose first column is $v$. From these, the $n^2$ entries of $X^{-1}$ can be found. The overall cost is $O(n^2)$ ops.
(1) As a first attempt, the entries of $V$ are computed in a straightforward manner using the entries of $X^{-1}$ :

$$V = \begin{bmatrix} \gamma_{1,1} & 2\gamma_{1,2} & \cdots & 2\gamma_{1,n} \\ 2\gamma_{1,2} & 2\gamma_{2,2} & \cdots & 2\gamma_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ 2\gamma_{1,n} & 2\gamma_{2,n} & \cdots & 2\gamma_{n,n} \end{bmatrix},$$

where

$$\begin{aligned} \gamma_{1,j} &= \sum_{i=1}^{n} \sum_{k=1}^{n-j+1} (X^{-1})_{i,k}(X^{-1})_{i,k+j-1}, \\ \gamma_{j,p} &= \sum_{i=1}^{n-j+1} \sum_{k=1}^{n-p+1} \left((X^{-1})_{i,k}(X^{-1})_{i+j-1,k+p-1} + (X^{-1})_{i,k+p-1}(X^{-1})_{i+j-1,k}\right). \end{aligned}$$

The cost of this approach in terms of arithmetic operations is of the order $O(n^4)$.
(2) In the second approach, we show that the cost of this computation can be kept at the level of $O(n^3 \log n)$ ops by combining the Gohberg Semencul formula and the the FFT. For a given $i$, the product vector $w_i = (X^{-1} \otimes X^{-1})Ue_i$, where $e_i$ is the $i$ th vector of the canonical basis, is such that $w_i = \mathrm{vec}(X^{-1}E_iX^{-1})$, with $E_i$ being the symmetric Toeplitz matrix whose first column is $e_i$. Therefore, compute first the columns of $E_iX^{-1}$ by performing $O(n^2)$ additions, and then multiply $X^{-1}$ by these columns, stacking the results to obtain $w_i$. This computation is performed in $O(n^2 \log n)$ operations for each $i$ by using the Goghberg Semencul formula, since the multiplication of a lower triangular Toeplitz matrix and a vector can be performed in $O(n \log n)$ operations by means of the FFT [25]. Therefore the overall computation of this stage for $i = 1, \ldots, n$ is $O(n^3 \log n)$ ops. Finally, compute for any $i$ the vector $U^T w_i$ for the cost of $O(n^2)$ additions.
The performance of these methods will be compared.

The Karcher mean of positive definite matrices has the specific interpretation of being the barycenter of the given matrices for the natural metric (36) on this manifold. Hence there are in a certain sense two possible generalizations. On the one hand, try to generalize the geometric mean concept, or, on the other hand, try to generalize the barycenter concept. Previously we focused on an extension of the geometric mean. Hereafter we focus on the positive definite Toeplitz matrix manifold itself, denoted by $\mathcal{T}_n$, and consider a barycenter in this case. This mean cannot be called a geometric mean in the sense of satisfying all required properties, but through its intuitive definition, many desirable properties could arise.

The concept of a barycenter is not restricted to the specific metric used to define the Karcher mean. For example, when the set $\mathcal{T}_n$ is endowed with the classical Euclidean inner product, the resulting barycenter is nothing else than the arithmetic mean. Using a probabilty argument, in [22], [23] a metric on $\mathcal{T}_n$ is introduced, called the Kähler metric. This metric results in a complete, simply connected manifold with non-positive sectional curvature everywhere, or a Cartan-Hadamard manifold. Thus, by [17], [38], existence and uniqueness are guaranteed for the barycenter with respect to this metric.
We will recall some known facts about the Kähler metric, and then we will give an explicit formula for the barycenter in the real case and a numerical procedure to compute the barycenter in the complex case.
To construct the Kähler metric, a Toeplitz matrix is first transformed to an $n$-tuple $(P_0, \mu_1, \dots, \mu_{n-1})$ in $\mathbb{R}_+^* \times \mathbb{D}^{n-1}$, with $\mathbb{R}_+^*$ the set of strictly positive real numbers and $\mathbb{D}$ the set of complex numbers of modulus less than one. This transformation, denoted as $\zeta(T) = \left[P_T, \mu_{T,1}, \dots, \mu_{T,n-1}\right]^T$, is performed as follows:

$$P_T = t_0,\ \mu_{T,j} = (-1)^j \frac{\det(S_j)}{\det(R_j)},$$

with $t_0$ the main diagonal element of $T, R_j$ the principal submatrix of size $j$ of $T$ (the upper left $j \times j$ submatrix) and $S_j$ obtained by shifting $R_j$ down one row, or equivalently, by removing the first row and last column of $R_{j+1}$ (the inverse transformation can be found in [48]). In what follows, we use this one-to-one relation between the Toeplitz matrices and the corresponding $n$-tuple, and when clear by the context, we will neglect the distinction and identify one with the other.
For $X$ and $Y$ being the transformations of two positive Toeplitz matrices $X = \left[P_X, \mu_{X,1}, \dots, \mu_{X,n-1}\right]^T$ and $Y = \left[P_Y, \mu_{Y,1}, \dots, \mu_{Y,n-1}\right]^T$, the metric is given by

$$\begin{aligned} \mathrm{d}(X,Y) &= \left(n\sigma(P_X,P_Y)^2 + \sum_{j=1}^{n-1} (n-j)\tau\left(\mu_{X,j},\mu_{Y,j}\right)^2\right)^{1/2}, \\ \sigma(P_X,P_Y) &= \left|\log\left(\frac{P_Y}{P_X}\right)\right|,\ \tau\left(\mu_{X,j},\mu_{Y,j}\right) = \operatorname{atanh}\left(\left|\frac{\mu_{Y,j}-\mu_{X,j}}{1-\mu_{X,j}\mu_{Y,j}}\right|\right), \end{aligned}$$

where $\operatorname{atanh}(z) = \frac{1}{2}\log\left(\frac{1+z}{1-z}\right)$.
The barycenter of the positive Toeplitz matrices $T_i$, for $i = 1, \dots, m$, with respect to the Kähler metric will be denoted by $B(T_1, \dots, T_m) = \left[P_B, \mu_{B,1}, \dots, \mu_{B,n-1}\right]^T$. It is obtained in this transformed space by minimizing the function

$$f(X) = \sum_{i=1}^{m} \mathrm{d}^2(X, T_i)$$

over $\mathbb{R}_+^* \times \mathbb{D}^{n-1}$. Notice that the problem of minimizing $f(X)$ can be decoupled into the problems of minimizing $\varphi_0(x) = \sum_{i=1}^m \sigma\left(x, P_{T_i}\right)^2$ over $\mathbb{R}_+^*$, and the $n-1$ scalar functions

$$\varphi_j(z) = \sum_{i=1}^{m} \tau\left(z, \mu_{T_i,j}\right)^2,\ j = 1, \dots, n-1$$

over $\mathbb{D}$. The minimum of $\varphi_0(x)$ is easily obtained as $P_B = \left(P_{T_1} \cdots P_{T_m}\right)^{1/m}$ by solving the equation $\nabla\varphi_0(x) = 0$. The minimum of $\varphi_j(z)$ is nothing else than the barycenter of $\mu_{T_1,j}, \dots, \mu_{T_m,j}$ with respect to the customary Poincaré metric on the unit disk and is the point where the gradient

$$\nabla\varphi_j(z) = 2(|z|^2 - 1)\sum_{\mathrm{i}=1}^{\mathrm{m}} \operatorname{sign}(c_{i,j})\operatorname{atanh}(|c_{i,j}|),\ c_{i,j} = \frac{\mu_{T_i,j} - z}{1 - \bar{z}\mu_{T_i,j}}, \text{equals zero.} \quad (61)$$

In the real case we are able to find an explicit expression for this barycenter as well, since $\operatorname{sign}(c)\operatorname{atanh}(|c|) = \operatorname{atanh}(c)$ and after some manipulations we get

$$\mu_{X,j} = \mathcal{C}\left(\left(\mathcal{C}(\mu_{T_1,j}) \cdots \mathcal{C}(\mu_{T_m,j})\right)^{1/m}\right),$$

where $\mathcal{C}(z) = (1 - z)/(1 + z)$ is the Cayley transform.
In the complex case we were not able to find such an explicit formula but a quick numerical method can be devised using a gradient descent algorithm. We recall that the tangent space to the Poincaré disk can be identified with the complex plane and thus for a sufficiently small tangent vector $v \in \mathbb{C}$, one can consider the retraction $R_z(v) = z + v$, which captures the fact that the manifold is an open subset of the complex plane. The resulting algorithm to find the barycenter of $\mu_1, \dots, \mu_n \in \mathbb{C}$ is given by the iteration

$$z_{k+1} = z_k + t_k v_k,\ v_k = (1 - |z_k|^2)\sum_{i=1}^{n} \operatorname{sign}(c_{i,k})\operatorname{atanh}(|c_{i,k}|),\ c_{i,k} = \frac{\mu_i - z_k}{1 - \bar{z}_k\mu_i}, \quad (62)$$

for a suitable initial value $z_0$ and a sufficiently small steplength $t_k$.
Another possibility is to consider the retraction

$$R_z(v) = \frac{z + e^{i\theta} + (z - e^{i\theta})e^{-s}}{1 + \bar{z}e^{i\theta} + (1 - \bar{z}e^{i\theta})e^{-s}},\ \theta = \arg v,\ s = \frac{2|v|}{1 - |z|^2},$$

which corresponds to moving along the geodesics of the Poincaré disk. The corresponding gradient descent method is

$$z_{k+1} = R_{z_k}(t_k v_k),$$

with the same $v_k$ as (62).

Regarding the properties of this barycenter, it is easily seen that it is permutation invariant, repetition invariant and idempotent (this holds for any barycenter). Moreover, for any $\alpha > 0$, the transformed values of $\alpha T$ are $\left[\alpha P_T, \mu_{T,1}, \dots, \mu_{T,m}\right]^T$ and from the explicit expression of $P_B$ in the real case we get that $B(\alpha T_1, \alpha T_2, \dots, \alpha T_m) = \alpha^{1/m} B(T_1, \dots, T_m)$, that is, homogeneity holds.
Unfortunately, this new barycenter does not possess other properties as shown by the following example.
**Example (1.3.11)[20]:** From the explicit expression for the mean in the real case we get a simple formula for the Kähler barycenter of two $2 \times 2$ matrices

$$T_1 = \begin{bmatrix} x_1 & y_1 \\ y_1 & x_1 \end{bmatrix},\ T_2 = \begin{bmatrix} x_2 & y_2 \\ y_2 & x_2 \end{bmatrix},$$

namely

$$B(T_1, T_2) = \sqrt{x_1 x_2}\begin{bmatrix} 1 & \dfrac{a - b}{a + b} \\ \dfrac{a - b}{a + b} & 1 \end{bmatrix}, \text{with } \begin{cases} a = \sqrt{(x_1 + y_1)(x_2 + y_2)} \\ b = \sqrt{(x_1 - y_1)(x_2 - y_2)} \end{cases}.$$

Now consider the following matrices

$$T_1 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \tilde{T}_1 = \begin{bmatrix} 4 & -1 \\ -1 & 4 \end{bmatrix}, \quad T_2 = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix},$$

with $\tilde{T}_1 \geqslant T_1$. By symbolic computation, one gets that

$$B(\tilde{T}_1, T_2) = \begin{bmatrix} 2\sqrt{2} & \sqrt{2}(\sqrt{5}-3) \\ \sqrt{2}(\sqrt{5}-3) & 2\sqrt{2} \end{bmatrix} \not B(T_1, T_2) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix},$$

in fact one eigenvalue of $B(\tilde{T}_1, T_2) - B(T_1, T_2)$ is $\lambda = \sqrt{10} - 2 - \sqrt{2} < 0$. Thus, we have proved that the Kähler barycenter is not monotonic. Moreover,

$$B(T_1, T_2) \neq (T_1 T_2)^{1/2} = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & \sqrt{3} \end{bmatrix},$$

and hence the Kähler barycenter does not coincide with the Karcher mean for circulant matrices. In particular, it is not a structured geometric mean as defined.

Observe that in the previous example $B(T_1, T_2)$ surprisingly coincides with the arithmetic mean of $T_1$ and $T_2$. It is not difficult to construct examples where it is not true that $B(T_1, T_2) \leqslant (T_1 + T_2)/2$ as it should be for a geometric mean.

The different algorithms proposed and will be compared w.r.t. speed and accuracy. The numerical experiments are confined to Toeplitz matrices, because of applicational interest in computing their structured matrix mean [40]. These matrices are constructed randomly, but with chosen condition number, using the technique described in [10]. Performance, accuracy and computational distance are subjects of the forthcoming investigations. For clarity we remind that the Richardson-iteration corresponds to a projection technique on a manifold, with the classical Euclidean inner product. For all algorithms, the stopping criteria is based on checking the relative size of the gradient and on comparing two consecutive iteration points.

In spite of the lack of the proof of uniqueness for structured geometric mean in the Toeplitz case, for any fixed set of data matrices used in our experiments, any initial value and any algorithm yielded always the same structured geometric mean. This suggests the conjecture that in the Toeplitz case there is a unique structured geometric mean.

We have also compared the structured geometric mean obtained by our algorithms with the Kähler metric mean, getting in most experiments a relative difference of the order $10^{-1}$, which indicates that these two means are relatively far from each other.

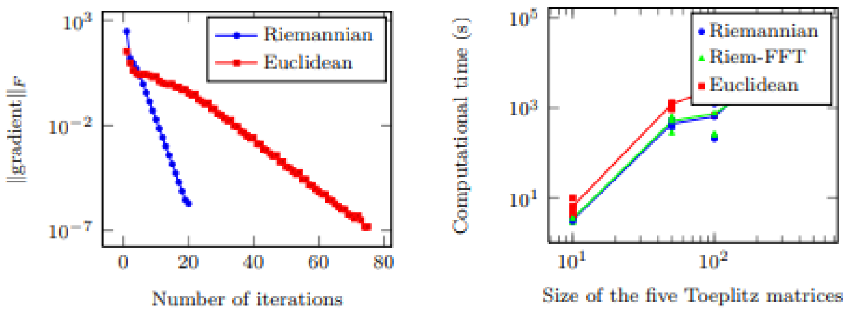
The performance of the projection methods explained can be compared by looking at both the number of iterations the methods require and the total amount of computational time they need.

In Figure $3(\mathrm{a})$, the evolution of the gradient over the iterations is displayed for both techniques (and hence also the number of iterations). Using the projection method introduced gives a faster decrease of the gradient and results in fewer iteration steps. The number of iterations remains almost constant for this method as the size of the matrices increases. For the projection technique on the other hand, this number starts to increase when the matrix size grows.

However, comparing expression (55) and (57), it can be seen that the second one is computationally more expensive and hence the advantage of requiring fewer iterations could be nullified. Therefore, Figure $3(b)$ displays the total computational time of both methods for varying sizes of the matrices. The two methods based maintain an advantage despite their larger computational cost per iteration. Note that for the largest matrix size the computational time of the Euclidean based method appears less than one of the other methods. However, this is caused by the increasing number of iterations required by this

Euclidean method. Consequently, the maximum number of iterations is reached before convergence and the algorithms is terminated prematurely. Concerning the operation count, the advantage of the method based on FFT starts to appear when the matrices become sufficiently large. Accuracy. In order to analyze the accuracy of the projection methods, we implement a high precision version of the first algorithm using the vpa functionality of MATLAB. The relative distance, based on the intrinsic distance (36), between this high precision computation and the result of the actual algorithms is shown in Figure 4 . For small condition numbers, the accuracy of all methods is similar in average, but as the condition of the matrices becomes worse, the accuracy of the projection method based on Euclidean geometry deteriorates much faster than that of the method based on the Riemannian geometry. This first method even fails to converge when the condition number of the matrices becomes significantly large. The accuracy of the two approaches is similar and deteriorates steadily as the condition numbers of the matrices increase.

The Karcher mean for positive definite matrices to structured positive definite matrices was proposed. Besides a theoretical investigation and adaptation of the desired properties of such a mean, algorithms were proposed. In the design of the algorithms, two trajectories were put forward, one




(A) Gradient evolution for five $50 \times 50$ matrices.

(B) Computational time for various matrix sizes. The mean of the samples is connected by a line.

FIGURE 3. Comparison of the projection methods for Toeplitz matrices. In the legends, Euclidean indicates the method of Section 4.2, Riemannian indicates the first approach described in Section 4.4, and Riem-FFT the second.

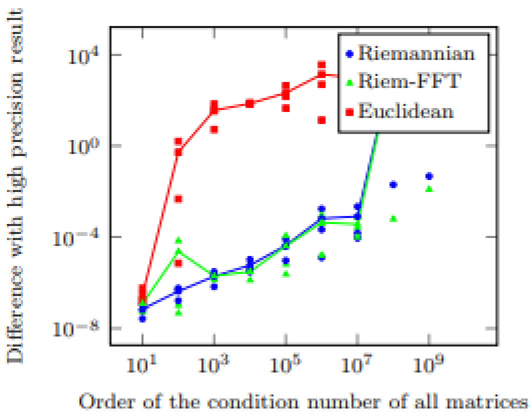



FIGURE 4. Accuracy of the projection methods when compared to a high precision version. The mean of the samples is connected by a line. In the legends, Euclidean indicates the method of Section 4.2, Riemannian indicates the first approach described in Section 4.4, and Riem-FFT the second.

relying mostly on linear algebra, and one based on differential geometry. A convergence analysis has been performed showing the superiority of the algorithm based on differential geometry. Numerical experiments compared the accuracy and speed of the various techniques and confirmed the theoretical analysis.

In the case of Toeplitz matrices, we have considered also the Kähler metric mean [23], whose properties have been investigated, providing an explicit expression in the real case and a quick algorithm in the complex case. For Toeplitz matrices, both the new structured geometric mean and the Kähler metric mean have not completely satisfying properties. In fact they are neither monotone, nor do they satisfy the arithmetic-geometric inequality. We wonder if it is possible to provide a definition of geometric mean for Toep to the ordering of positive matrices.

# Chapter 2
# Toeplitz Kernels and Spectral Gaps and Two Weight Inequality

We show that where the argument of the meromorphic inner function has a power law type behavior on the real line, we compute the critical value $(J,S) = inf\,\{a: ker\,T_{J\bar{S}a} \neq 0\}$ . The formula for $c(J,S)$ generalizes the Beurling-Malliavin theorem on the radius of completeness for a system of exponentials. We attempt to find the supermom of the size of the gap in the Fourier spectrum in terms of metric properties of $X$. We show here two components, a 'global to local' reduction, carried out, and an analysis of the 'local' problem, carried out.

## Section (2.1): Beurling-Malliavin Theory

For $\Lambda \subset \mathbf{C}$ denote

$$\mathcal{E}_\Lambda = \{e^{i\lambda x}: \lambda \in \Lambda\}.$$

By definition, the radius of completeness for the family $\mathcal{E}_\Lambda$ is the number

$$R(\Lambda) = \text{s}\ \{a: \mathcal{E}_\Lambda \text{ is complete in } L^2(-a,a)\}.$$

(A family is complete if finite linear combinations of its elements are dense in the corresponding space.) In [50]-[51], Beurling and Malliavin established a formula for $R(\Lambda)$ in terms a certain density of $\Lambda$ at infinity.

If $\Lambda \subset \mathbb{R}$, then the effective (or Beurling-Malliavin) density $D_{\text{eff}}(\Lambda)$ is the supremum of the set of numbers $a \geq 0$ such that there is a collection of disjoint intervals $\{l_j\}$ satisfying the following two conditions:

$$\sum_j \frac{|l_j|^2}{1+d_j^2} = \infty,\ d_j := \text{dist}(0, l_j),$$

and

$$\forall j,\ \#(\Lambda \cap l_j) \geq a|l_j|.$$

Beurling-Malliavin's "Second Theorem" (BM2 for short) states that if $\Lambda \subset \mathbb{R}$, then

$$R(\Lambda) = \pi D_{\text{eff}}(\Lambda).$$

This formula extends to the general case $\Lambda \subset \mathbf{C}$ as follows. If $\Lambda$ satisfies the Blaschke condition

$$\text{(B)} \sum_{\lambda\in\Lambda} |\Im\lambda^{-1}| < \infty,$$

then

$$R(\Lambda) = \pi D_{\text{eff}}(\Lambda^*),$$

where

$$\Lambda^* = \{\lambda^*: \lambda \in \Lambda, \Re\lambda \neq 0\},\ \lambda^* := [\Re\lambda^{-1}]^{-1};$$

otherwise

$$\Lambda \notin (B) \ \Rightarrow\ R(\Lambda) = \infty.$$

The Beurling-Malliavin Theorem (2.1.2)rowned a long search for a solution of the completeness problem, see [78],[70],[81],[69]. See [80] for historical information; let us only mention that one of the earliest results of the theory was the estimate

$$R(\Lambda) \leq \pi D(\Lambda),\ (\Lambda \subset \mathbb{R}), \tag{1}$$

where $D(\Lambda)$ is the usual upper density of $\Lambda$ at infinity.

The Beurling-Malliavin theory also comprises their "First Theorem" (BM1), a result of considerable independent interest and (so far) a necessary step in the proof of BM2. A detailed exposition of BM theory (including clarification and further improvements of the argument) is presented in [66], [67], [58], see also [63], [55]. New applications and new approaches to various parts of the theory have been suggested; see [52], [59], [73] for some recent developments; also see [64] for a modern overview of the completeness problem for exponential systems.

We generalize BM theory to many other families of special functions. We state our results in the language of Toeplitz kernels referring to [72] for a detailed explanation of how results of this type are related to the completeness problem for families of solutions of general Sturm-Liouville problems.

The completeness radius problem can be restated in terms of Toeplitz operators as follows. Recall that the Toeplitz operator $T_U$ with a symbol $U \in L^\infty(\mathbb{R})$ is the map

$$T_U: H^2 \to H^2,\ F \mapsto P_+(UF),$$

where $P_+$is the orthogonal projection in $L^2(\mathbb{R})$ onto the Hardy space $H^2 = H^2(\mathbf{C}_+)$ in the upper halfplane $\mathbf{C}_+ = \{\Im sz > 0\}$. By duality and the definition of the classical Fourier transform,

$$f(t) \mapsto \hat{f}(z) = \int e^{izt} f(t)dt,$$

the exponential family $\mathcal{E}_\Lambda$ is complete in $L^2(-a,a)$ if and only if there is a non-trivial function $F$ in the Paley-Wiener space

$$\mathrm{PW}_a = \{\hat{f}: f \in L^2(-a,a)\}$$

such that $F = 0$ on $\Lambda$. According to Paley-Wiener's theorem, the Fourier transform isometrically identifies $L^2(0,\infty)$ with $H^2(\mathbf{C}_+)$, and therefore

$$\mathrm{PW}_a = e^{-iaz}\left[H^2 \ominus e^{2aiz} H^2\right].$$

The subspace $H^2 \oplus e^{2aiz}H^2$ is the so called model space of the inner function $e^{2aiz}$. More generally, one defines model spaces

$$K_\Theta = H^2 \ominus \Theta H^2$$

for arbitrary inner functions $\Theta$. The elements of $K_\Theta$ are analytic functions in $\mathbf{C}_+$ but if $\Theta$ has a meromorphic extention to the whole complex plane, (we call such $\Theta$ 's meromorphic inner functions), then the elements of $K_\Theta$ are defined as functions in C. The completeness problem for exponentials is exactly the problem of describing the sets of uniqueness for the model space of $e^{2aiz}$.

Suppose now that $\Lambda$ is a subset of $\mathbf{C}_+$satisfying the Blaschke condition, and let $B_\Lambda$ be the corresponding Blaschke product. A simple argument shows that $\Lambda$ is a set of uniqueness for $K_\Theta$ if and only if the Toeplitz operator with the symbol $U = B_\Lambda \bar{\Theta}$ has a trivial kernel. In particular, we obtain the formula

$$R(\Lambda) = \inf\ \{a: \ker T_{B_\Lambda e^{-2aiz}} \neq 0\}.$$

There is a similar statement in the general case $\Lambda \subset \mathbf{C}$, see [72]. For example, if $\Lambda \subset \mathbb{R}$, then

$$R(\Lambda) = \inf\ \{a: \ker T_{J_\Lambda e^{-2aiz}} \neq 0\},$$

where $J_\Lambda$ denotes some/any meromorphic inner function $J$ such that $\Lambda$ is precisely the level set $\{J = 1\}$.

We should mention that the idea of the Toeplitz operator approach in the study of exponential systems was introduced in the series of [79], [75], [61]. This approach has been

particularly successful for the interpolation and sampling theory in Paley-Wiener spaces, see [71], [77], [82].
We will use the following notation for kernels of Toeplitz operators (or Toeplitz kernels in $H^2$ ):

$$N[U] = \ker T_U.$$

(For example, $N[\bar{\Theta}] = K_\Theta$ if $\Theta$ is an inner function.) We will also consider Toeplitz kernels in the Smirnov-Nevanlinna class $\mathcal{N}^+ = \mathcal{N}^+(\mathbf{C}_+)$,

$$N^+[U] = \{F \in \mathcal{N}^+ \cap L^1_{\text{loc}}(\mathbb{R}) : \bar{U}\bar{F} \in \mathcal{N}^+\},$$

and in the Hardy spaces $H^p = H^p(\mathbf{C}_+)$,

$$N^p[U] = N^+[U] \cap L^p(\mathbb{R}),\ (0 < p \leq \infty).$$

See [65], [56], [76] for general references concerning the Hardy-Nevanlinna theory.

A natural way to generalize the completeness radius problem (and the BM2 theorem) is to ask about the exact value of the infimum

$$\inf\ \{a : \ker T_{J\bar{S}^a} \neq 0\} \qquad (2)$$

for arbitrary meromorphic inner functions $J$ and $S$. We will give an answer in the case where the argument of $S$ has a power law type behavior,

$$(\arg S)'(x) \asymp |x|^\kappa,\ x \to \pm\infty,$$

with $\kappa \geq 0$. (We call the case $\kappa \geq 0$ super-exponential to underline the relation to the classical case $S(x) = e^{iax}$. We will comment on the sub-exponential case $\kappa < 0$.)
As explained in [72], the computation of the "radius" (2) has some immediate consequences for the theory of Sturm-Liouville (SL) operators. The case of SL operators with eigenvalues

$$\lambda_n \asymp n^\nu$$

belongs to the theory with parameter

$$\kappa = \frac{2}{\nu} - 1 \geq 0.$$

If $\kappa > 0$, the SL operators are singular in contrast to the BM case $S(x) = e^{iax}$, which applies to regular operators. In addition to the completeness problem for systems of solutions of SL equations, cf [60], the generalized BM theory applies to certain problems of spectral theory as well as the theory of (Weyl-Titchmarsh) Fourier transforms associated with SL operators and the corresponding (de Branges) spaces of entire functions.
To state our results, we need to introduce the notion of $BM$ intervals. Let $\gamma$ be a continuous function $\mathbb{R} \to \mathbb{R}$ such that $\gamma(\mp\infty) = \pm\infty$. i.e.

$$\lim_{x\to-\infty} \gamma(x) = +\infty,\ \lim_{x\to+\infty} \gamma(x) = -\infty.$$

The family $\mathcal{BM}(\gamma)$ is defined as the collection of the components of the open set

$$\left\{x : \gamma(x) \neq \max_{[x,+\infty)} \gamma\right\}.$$

For an interval $l \in \mathcal{BM}(\gamma)$, we denote its length by $|l|$ or simply by $l$, and we denote the distance to the origin by $d = d(l)$.

If $\kappa \geq 0$, then we say that $\gamma$ is $(\kappa)$-almost decreasing if

$$\gamma(\mp\infty) = \pm\infty,\ \sum_{l \in \mathcal{BM}(\gamma)} d^{\kappa-2} l^2 < \infty, \qquad (3)$$

where the sum is taken over intervals satisfying $d(l) \geq 1$.
The standard terminology in the classical $\kappa = 0$ case is the following: the family $\mathcal{BM}(\gamma)$ is short if $\gamma$ is almost decreasing; otherwise we say that $\mathcal{BM}(\gamma)$ is long.
**Corollary (2.1.1)[49]:** Let $U = e^{i\gamma}$ and $S = e^{i\sigma}$ be such that

$$\gamma'(x) \gtrsim -|x|^\kappa,\ \sigma'(x) \asymp |x|^\kappa,\ (x \to \infty),$$
and let $c = c(U, S; \kappa)$. Then for all $p < 1/3$ we have
$$N^p[U\bar{S}^a] = 0\ (a < c),\ N^p[U\bar{S}^a] \neq 0\ (a > c).$$
Indeed, if $a < c$ then $\gamma_{a+\epsilon}$ is not almost decreasing for some $\epsilon > 0$, and we have
$$N^p[U\bar{S}^a] \subset N^+[U\bar{S}^{a+\epsilon}S^\epsilon] = 0$$
by Theorem (2.1.22), which can be applied because $\gamma_a'(x) \gtrsim -|x|^\kappa$ for all $a$ 's. Similarly, if $a > c$, then $\gamma_{a-\epsilon}$ is almost decreasing for some $\epsilon > 0$, and we have
$$N^p[U\bar{S}^a] = N^p[U\bar{S}^{a-\epsilon}\bar{S}^\epsilon] \neq 0.$$
In the special case where $U$ is an inner function, we can extend the statement of he corollary to all values of $p$, in particular $p = 2$.

It is easy to see that the statement of Theorem (2.1.22) (and Theorem (2.1.23)) can not be extended to the case $\kappa < 0$. For example, the functions
$$\sigma(x) = 2\mathrm{sign}(x)|x|^{1/4},\ \gamma(x) = 2(1+\sqrt{2})1_{\mathbb{R}_-}(x)|x|^{1/4}$$
satisfy the conditions (35) with $\kappa = -3/4$, and of course $\gamma(+\infty) \neq -\infty$.For $U = e^{i\gamma}$ and $S = e^{i\sigma}$ we have
$$N^\infty[US] \neq 0,$$
because
$$US = \bar{f}/f,\ f(z) = \exp\{-(1+i)z^{1/4}\} \in H^\infty(\mathbf{C}_+).$$
(Also note that the sum $\sum d^{\kappa-2}l^2$ in (3) is always finite if $\kappa < 0$.)
The Beurling-Malliavin theory extends to the sub-exponential case in a different fashion. For $\kappa \in (-1,0]$ we consider the weighted (non-linear) Smirnov-Nevanlinna classes
$$\mathcal{N}_\kappa^+ = \left\{F \in \mathcal{N}^+ : \log|F| \in L^1\left(\mathbb{R}, \frac{1}{1+|x|^{2+\kappa}}\right)\right\},$$
and define the corresponding Toeplitz kernels as follows:
$$N_\kappa^+[U] = N^+[U] \cap \mathcal{N}_\kappa^+,\ N_\kappa^p[U] = N^p[U] \cap \mathcal{N}_\kappa^+.$$
**Theorem (2.1.2)[49]:** Let $\kappa \in (-1,0]$, and let $U = e^{i\gamma}$ and $S = e^{i\sigma}$ be smooth unimodular functions such that
$$\gamma'(x) \gtrsim -|x|^\kappa,\ \sigma'(x) \gtrsim |x|^\kappa\ (x \to \infty).$$
(i) If the family $\mathcal{BM}(\gamma)$ is long, then $N_\kappa^+[US^\epsilon] = 0$ for all $\epsilon > 0$.
(ii) If the family $\mathcal{BM}(\gamma)$ is short, then $N_\kappa^p[U\bar{S}^\epsilon] \neq 0$ for all $\epsilon > 0$ and all $p < \frac{1}{3}$.
One can also state a theorem similar to Theorem (2.1.23). Applications of these results to Volterra operators, see [57], and higher order differential operators will be discussed.

The main tool in the proof of the theorems stated above is the one-dimensional Hilbert transform. Let $\Pi$ denote the Poisson measure on $\mathbb{R}$,
$$d\Pi(t) = \frac{dt}{1+t^2}.$$
If $h \in L_\Pi^1 \equiv L^1(\mathbb{R}, \Pi)$ is a real-valued function, and if $\mathcal{S}h$ denotes its Schwarz integral,
$$\mathcal{S}h(z) = \frac{1}{\pi i}\int_{\mathbb{R}}\left[\frac{1}{t-z} - \frac{t}{1+t^2}\right]h(t)dt,\ (z \in \mathbf{C}_+), \qquad (4)$$
then $\tilde{h}$, the Hilbert transform of $h$, is defined a.e. on $\mathbb{R}$ as the angular limit of $\Im[\mathcal{S}h]$. Alternatively, $\tilde{h}$ can be defined as a singular integral:
$$\tilde{h}(x) = \frac{1}{\pi}\ \text{v.p.} \int \left[\frac{1}{x-t} + \frac{t}{1+t^2}\right]h(t)dt,\ (x \in \mathbb{R}).$$

(As a general rule we identify Nevanlinna class functions in the halfplane $\mathbf{C}_+$with their angular boundary values on $\mathbb{R}$; e.g. we may write $\mathcal{S}h = h + i\tilde{h}$. )
The relevance of the Hilbert transform in the theory of Toeplitz kernels can be explained by the following simple observation, see [72].
Suppose $\gamma: \mathbb{R} \to \mathbb{R}$ is a smooth function. Then $N^+[e^{i\gamma}] \neq 0$ if and only if

$$\gamma = -\alpha + \tilde{h} \tag{5}$$

for some smooth increasing function $\alpha$ and some $h \in L^1_{\Pi^-}$. There is a similar criterion for Toeplitz kernels in Hardy spaces: $N^p[e^{i\gamma}] \neq 0$ if and only if $\gamma$ admits a representation (5) with $\alpha$ being the argument of some inner function and with $h \in L^1_\Pi$ such that $e^{-h} \in L^{p/2}(\mathbb{R})$.
We recall some properties of the Hilbert transform. We denote by $L^{(1,\infty)}_\Pi$ the usual weak $L^1$-space with respect to the Poisson measure. Kolmogorov's theorem states that

$$\tilde{h} \in L^{o(1,\infty)}_\Pi,$$

where $L^{o(1,\infty)}_\Pi$ stands for the "little o" subspace of $L^{(1,\infty)}_\Pi$, i.e.

$$\Pi\{|\tilde{h}| > A\} = o\left(\frac{1}{A}\right),\ A \to \infty. \tag{6}$$

For bounded functions we have the following (Smirnov-Kolmogorov) estimate :

$$\| h \|_\infty < \frac{\pi}{2} \ \Rightarrow \ e^{\tilde{h}} \in L^1_\Pi.$$

Together with the criterion (5), this implies

$$\| \gamma \|_\infty < \frac{\pi}{p} \ \Rightarrow \ N^p\left[\bar{b}^{\frac{2}{p}} e^{i\gamma}\right] \neq 0, \tag{7}$$

where $b$ is the Blaschke factor

$$b(z) = \frac{i-z}{i+z},\ z \in \mathbf{C}_+. \tag{8}$$

We prove Theorems (2.1.22) and (2.1.23). We closely follow all the steps in our presentation of the classical Beurling-Malliavin theory in [72], combining them with certain estimates of the Hilbert transform. To make the proof selfcontained, in several places we had to repeat the argument outlined in [72]. To avoid further repetitions we decided to omit the proof of Theorem (2.1.2) because the reasoning in the sub-exponential case is quite similar. The proof of Theorems (2.1.22) and (2.1.23) is organized as follows.
(i) Upper density estimate: $\gamma(\pm\infty) \neq \mp\infty$ implies $N^+[US^\epsilon] = 0$. This statement is analogous to the estimate (i).
(ii) Effective density estimate: $\sum d^{\kappa-2} l^2 = \infty$ implies $N^+[US^\epsilon] = 0$. Together with (i) this generalizes the estimate $R(\Lambda) \leq \pi D_{\text{eff}}(\Lambda)$ in BM2.
(iii) Little multiplier theorem: if $\gamma$ is almost decreasing, then $N^+[U\bar{S}^\epsilon] = 0$.
(iv) BM multiplier theorem: if the weighted Dirichlet norm of $\log W$ is finite, then $W$ belongs to some Hardy space up to a factor from $N^+[\bar{S}^\epsilon]$.
(v) A version of BM1: the logarithm of any outer function in $N^+$-kernel has a finite weighted Dirichlet norm. This is used to show that non-triviality of $N^+$-kernels implies non-triviality of $N^p$-kernels for symbols involving inner functions.
(vi) $L^p$-multipliers: approximation by inner functions and multiplying the elements of $N^p$-kernels down to $H^\infty$.

We discuss various consequences of the weighted one-sided Lipschitz condition

$$\tilde{h}'(x) \lesssim |x|^\kappa,\ x \to \infty,$$

for smooth real-valued functions $h \in L^1_\Pi$. (Here and elsewhere $\tilde{h}'$ means $(\tilde{h})'$.)

**Lemma (2.1.3)[49]:** If $\kappa \ge 0$ and $h \in L^1_\Pi$, then

$$\tilde{h}'(x) \lesssim x^\kappa \;\Rightarrow\; \tilde{h}(x) = o(x^{\kappa+1}) \;\text{ as }\; x \to +\infty.$$

**Proof:** By Kolmogorov's theorem, we have

$$\tilde{h} \in L_\Pi^{o(1,\infty)}. \tag{9}$$

If $x_* \gg 1$ and $\tilde{h}(x_*) \ge c x_*^{\kappa+1}$ for some $c > 0$, then for all $x$ such that $1 \ll x \le x_*$ we have

$$\tilde{h}(x) \ge \tilde{h}(x_*) - a\int_x^{x_*} t^\kappa dt \ge c x_*^{\kappa+1} - a(x_*^{\kappa+1} - x^{\kappa+1}),$$

and it follows that

$$\tilde{h}(x) \gtrsim x_*^{\kappa+1} \ge x_*$$

for all $x$ in some interval $(x_{**}, x_*)$ of length $\asymp x_*$. Since $\Pi(x_{**}, x_*) \asymp 1/x_*$, this contradicts (9), see (6).

If $\tilde{h}(x_*) \le -c x_*^{\kappa+1}$ for some $c > 0$, then by a similar argument we have

$$\tilde{h}(x) \lesssim -x_*^{\kappa+1} \le x_*$$

for all $x$ in some interval $(x_*, x_{**})$ of length $\asymp x_*$, which again contradicts (9).

We will also need the following version of this lemma.

**Lemma (2.1.4)[49]:** Let $h \in L^1_\Pi, \kappa \ge 0$, and $a \in \mathbb{R}$. If

$$\tilde{h}'(x) + a x^{-1}\tilde{h}(x) \le x^\kappa,\; x \gg 1, \tag{10}$$

then

$$\tilde{h}(x) = o(x^{\kappa+1}) \;\text{ and }\; \tilde{h}'(x) \le x^\kappa + o(x^\kappa) \;\text{ as }\; x \to +\infty.$$

**Proof:** Suppose we have $\tilde{h}(x_*) \ge c x_*^{\kappa+1}$ for some $x_* \gg 1$. Let $x_1$ be the smallest positive number such that $\tilde{h}(x_1) = c x_*^{\kappa+1}$, so we have $1 \ll x_1 \le x_*$ and $\tilde{h} \le c x_*^{\kappa+1}$ on $(0, x_1)$. Together with (10), this implies

$$\tilde{h}'(x) \lesssim \frac{x_*^{\kappa+1}}{x_1},\; x \in \left(\frac{x_1}{2}, x_1\right).$$

Arguing as in the previous proof, we see that $\tilde{h} \gtrsim x_*^{\kappa+1} \ge x_1$ on some interval of length $\asymp x_1$, which contradicts the weak $L^1$-estimate (6). The argument in the case $\tilde{h}(x_*) \le -c x_*^{\kappa+1}$ is similar.

**Lemma (2.1.5)[49]:** If $\kappa \in [-1,0)$ and $h \in L^1(|x|^{-2-\kappa})$, then

$$\tilde{h}'(x) \lesssim x^\kappa \;\Rightarrow\; \tilde{h}(x) = o(x^{\kappa+1}),\; x \to +\infty.$$

**Proof:** If $\kappa \in (-1,0)$, then the weight $|x|^{-2-\kappa}$ satisfies the Muckenhoupt $(A_1)$ condition at infinity, and therefore we have

$$h \in L^1(|x|^{-2-\kappa}) \Rightarrow \tilde{h} \in L^{o(1,\infty)}(|x|^{-2-\kappa}),$$

see [62]. (One can also give an elementary proof for this particular weight.) We then argue as in the proof of Lemma (2.1.3). For example, if $x_* \gg 1$ and $\tilde{h}(x_*) \ge c x_*^{\kappa+1}$, then $\tilde{h} \gtrsim x_*^{\kappa+1}$ on some interval $[x_{**}, x_*]$ of length $\asymp x_*$. The weighted length of this interval is $\asymp x_*^{-1-\kappa}$, which contradicts the weak $L^1$-estimate.

If $\kappa = -1$, then we consider the function

$$h_1(x) = x h(x) \in L^1_\Pi.$$

Since $\tilde{h}_1(x) = x\tilde{h}(x)$, we have

$$\tilde{h}_1'(x) = \tilde{h}(x) + x\tilde{h}'(x) \le x^{-1}\tilde{h}_1(x) + O(1),\; x \to +\infty.$$

By Lemma (2.1.4), we get $\tilde{h}_1(x) = o(x)$ and therefore $\tilde{h}(x) = o(1)$.

**Lemma (2.1.6)[49]:** Let $f \in L^1_\Pi, 0 \notin \operatorname{supp} f$, and let

$$0 \le \alpha \le \beta, \;\text{ or }\; 0 \le \beta < \alpha < 2.$$

Denote

$$g(x) = |x|^{-\alpha} f(x).$$

Then

$$\tilde{f}'(x) \leq \big(1 + o(1)\big)|x|^{\beta} \Rightarrow \tilde{g}'(x) \leq \big(1 + o(1)\big)|x|^{\beta-\alpha}, \quad (11)$$

and

$$x^{-1}\tilde{f}'(x) \leq (1 + o(1))|x|^{\beta-1} \Rightarrow x^{-1}\tilde{g}'(x) \leq (1 + o(1))|x|^{\beta-\alpha-1}. \quad (12)$$

**Proof:** We will prove (11) for $x \to +\infty$. The proof of the other cases is similar. Since the statement is trivial for $\alpha = 0$, we will assume $\alpha > 0$.

It is clear that we can modify $f$ on any finite interval, so we will assume that $f(x) = x^N$ near the origin for some $N \gg 1$. If we specify $\tilde{f}(0) = 0$, then

$$|x|^{-\alpha}\tilde{f}(x) \in L^1_{\Pi}. \quad (13)$$

Indeed, by Lemma (2.1.3) we have $\tilde{f} = O(|x|^N)$, and therefore

$$|x|^{-\alpha}|\tilde{f}| = |x|^{-\alpha}|\tilde{f}|^{\alpha/N}|\tilde{f}|^{1-(\alpha/N)} \lesssim |\tilde{f}|^{1-(\alpha/N)} \in L^1_{\Pi}$$

by Kolmogorov's estimate (6).

Consider the analytic function

$$u(z) + i\tilde{u}(z) := z^{-\alpha}(f + i\tilde{f})(z), \; z \in \mathbf{C}^+,$$

where $z^{-\alpha}$ denotes the branch positive on $\mathbb{R}_+$. Note that

$$u(x) = g(x), \; \tilde{u}(x) = |x|^{-\alpha}\tilde{f}(x) \;\text{ for }\; x \in \mathbb{R}_+,$$

and

$$u(x) = |x|^{-\alpha}[f(x)\cos\alpha\pi + \tilde{f}(x)\sin\alpha\pi] \;\text{ for }\; x \in \mathbb{R}_-.$$

By (13), so if we define

$$g - u \in L^1_{\Pi}, \; g - u = 0 \;\text{ on }\; \mathbb{R}_+,$$

so if we define

$$\delta(x) = \tilde{g}(x) - |x|^{-\alpha}\tilde{f}(x),$$

then $\delta = \bar{g} - \bar{u}$ on $\mathbb{R}_+$, and we have the following representation for the derivative:

$$\delta'(x) = (\widetilde{g - u})'(x) = \int_{-\infty}^{0} \frac{c_1 f(t) + c_2\tilde{f}(t)}{(t - x)^2}\frac{dt}{|t|^{\alpha}}, \; (x > 0).$$

By the dominated convergence theorem

$$\delta'(x) = o(1), \; x \to +\infty,$$

in particular

$$\delta'(x) = o\big(x^{\beta-\alpha}\big) \;\text{ if }\; \beta \geq \alpha.$$

In the case $0 \leq \beta < \alpha < 2$, we consider the integrals involving $f$ and $\tilde{f}$ separately. We have

$$\int_{-\infty}^{-1} \frac{|f(t)|}{(t - x)^2}\frac{dt}{|t|^{\alpha}} \leq \int_{-\infty}^{-x} \frac{1}{|t|^{\alpha}}\frac{|f(t)|dt}{|t|^2} + \frac{1}{x^{\alpha}}\int_{-x}^{-1} \frac{|t|^{2-\alpha}}{x^{2-\alpha}}\frac{|f(t)|dt}{|t|^2}$$

$$\leq \frac{1}{x^{\alpha}}\int_{-\infty}^{-x} \frac{|f(t)|dt}{|t|^2} + \frac{1}{x^{\alpha}}\int_{-x}^{-1} \frac{|f(t)|dt}{|t|^2} = o(x^{-\alpha}) = o\big(x^{\beta-\alpha}\big).$$

Since $\beta \geq 0$, by Lemma (2.1.3) we have

$$\tilde{f}(t) = o\big(|t|^{1+\beta}\big),$$

and since

$$\int_{-\infty}^{-1} \frac{|t|^{1+\beta-\alpha}}{(t - x)^2}dt \leq \int_{-\infty}^{-x} |t|^{\beta-\alpha-1}dt + \frac{1}{x^2}\int_{-x}^{-1} |t|^{1+\beta-\alpha}dt \asymp x^{\beta-\alpha},$$

we have

$$\int_{-\infty}^{-1} \frac{|\tilde{f}(t)|}{(t-x)^2} \frac{dt}{|t|^{\alpha}} = o(x^{\beta-\alpha}).$$

It follows that in all cases we have

$$\delta'(x) = o(x^{\beta-\alpha}),\ x \to +\infty,$$

and therefore

$$\tilde{g}'(x) = x^{-\alpha}\tilde{f}'(x) - \alpha x^{-\alpha-1}\tilde{f}(x) + \delta'(x) \leq x^{\beta-\alpha} + o(x^{\beta-\alpha}).$$

We will also need a converse of (11). We state it only for the range of parameters that will be used later.

**Lemma (2.1.7)[49]:** Let $w \in L^1_\Pi, 0 \notin \operatorname{supp} w$, and

$$0 < \alpha \leq \beta,\ \text{ or }\ 1 \leq \alpha \leq \min\ (2, \beta + 1).$$

Denote

$$h(x) = |x|^{-\alpha} w(x).$$

Then

$$\tilde{h}'(x) \leq (1 + o(1))|x|^{\beta-\alpha}\ \Rightarrow\ \tilde{w}'(x) \leq (1 + o(1))|x|^{\beta}.$$

**Proof:** (a) The case $\beta \geq \alpha$. Let $n$ be an even integer such that

$$\alpha_1 := n - \alpha \in [0,2).$$

Define

$$g(x) := x^{-n} w(x) = |x|^{-\alpha_1} h(x).$$

Since $\alpha_1 < 2$ and $\beta_1 := \beta - \alpha \geq 0$, we can apply Lemma (2.1.6) to $f = h$ and $g$ and obtain the estimate

$$\tilde{g}'(x) \leq |x|^{\beta_1-\alpha_1} + \cdots = |x|^{\beta-n} + \cdots.$$

Since $\tilde{w}(x) = x^n \tilde{g}(x)$, we have

$$\tilde{w}'(x) = nx^{n-1}\tilde{g}(x) + x^n \tilde{g}'(x) \leq nx^{-1}\tilde{w}(x) + |x|^{\beta} + \cdots,$$

and by Lemma (2.1.4) ,

$$\tilde{w}'(x) \leq (1 + o(1))|x|^{\beta}.$$

(b) The case $\alpha \in [1,2]$ and $\beta - \alpha \in [-1,0]$. Note that this implies $\beta \geq 0$. Define the functions

$$g(x) = x^{-1} w(x),\ f(x) = xh(x),$$

so

$$g(x) = |x|^{-\alpha_1} f(x),\ \alpha_1 = 2 - \alpha \in [0,1].$$

Let us show that

$$x^{-1}\tilde{f}' \leq |x|^{\beta_1-1} + \cdots,\ \beta_1 := \beta - \alpha + 1. \tag{14}$$

Since

$$h \in L^1\left(\frac{1}{|x|^{2-\alpha}}\right) \subset L^1\left(\frac{1}{|x|^{2+\kappa}}\right),\ \kappa := \beta - \alpha,$$

by Lemma (2.1.5) we have

$$\tilde{h}(x) = o(|x|^{\kappa+1}),$$

and since $\tilde{f}(x) = x\tilde{h}(x)$, we obtain (14):

$$x^{-1}\tilde{f}'(x) = \tilde{h}'(x) + x^{-1}\tilde{h}(x) \leq |x|^{\kappa} + o(|x|^{\kappa}).$$

We can now apply Lemma (2.1.6) with parameters $\alpha_1$ and $\beta_1$. (Note that $f \in L^1_\Pi$ and the parameters are admissible.) By (12) we get the estimate

$$x^{-1}\tilde{g}'(x) \leq |x|^{\beta_1-\alpha_1-1} + \cdots = |x|^{\beta-2} + \cdots,$$

and from $\tilde{w}(x) = x\tilde{g}(x)$ we derive

$$\tilde{w}'(x) = \tilde{g}(x) + x\tilde{g}'(x) \leq x^{-1}\tilde{w}(x) + |x|^{\beta} + \cdots.$$

Applying Lemma (2.1.4) we conclude the proof.

We prove the first part of Theorem (2.1.22), which gives a sufficient condition for the triviality of a Toeplitz kernel. Let us fix $\kappa \geq 0$ and consider two unimodular functions $U = e^{i\gamma}$ and $S = e^{i\sigma}$ on $\mathbb{R}$ satisfying

$$\gamma'(x) \geq -|x|^{\kappa},\ \sigma'(x) \geq |x|^{\kappa},\ (x \to \infty). \tag{15}$$

**Proposition (2.1.8)[49]:** If $N^{+}[US^{\epsilon}] \neq 0$ for some $\epsilon > 0$, then $\gamma(\mp\infty) = \pm\infty$.

**Proof:** If $N^{+}[US^{\epsilon}] \neq 0$, then by the basic criterion (5) we have

$$\gamma + \epsilon\sigma + \alpha = \tilde{h},\ \alpha' \geq 0,\ h \in L^{1}_{\Pi}.$$

Therefore,

$$\tilde{h}'(x) \gtrsim \gamma'(x) \gtrsim -|x|^{\kappa},$$

and $\tilde{h}(x) = o(|x|^{\kappa+1})$ by Lemma (2.1.3). It follows that

$$\frac{\gamma(x)}{x} \lesssim \frac{\gamma(x)}{x} + \frac{\alpha(x)}{x} = -\epsilon\frac{\sigma(x)}{x} + o(|x|^{\kappa}) \lesssim -|x|^{\kappa},$$

which implies $\gamma(\mp\infty) = \pm\infty$.

Let $c > 0$ be a fixed constant. For an interval $l \subset \mathbb{R}$ we denote by $l'$ and $l''$ the intervals of length $c|l|$ adjacent to $l$ from the left and from the right respectively, and we define

$$\Delta^{*}_{l}[\gamma] = \inf_{l''}\gamma - \sup_{l'}\gamma.$$

**Lemma (2.1.9)[49]:** Let $\epsilon > 0$ and suppose

$$\gamma(\mp) = \pm\infty,\ \sum_{l\in\mathcal{BM}(\gamma)} d^{\kappa-2}l^{2} = \infty. \tag{16}$$

Then there is a constant $c > 0$ and there is a collection of disjoint intervals $\{l_n\}$ in $[1, +\infty)$ or in $(-\infty, -1]$ such that

$$\sum d_n^{\kappa-2}l_n^{2} = \infty,\ 10l_n \leq d_n,\ \text{mult}\{5l_n\} < \infty, \tag{17}$$

and

$$\Delta^{*}_{l_n}[\arg(US^{\epsilon})] \geq c d_n^{\kappa} l_n. \tag{18}$$

Here $5l$ is the notation for the interval of length $5|l|$ concentric with $l$, and mult $\{\cdot\}$ is the multiplicity of the covering.

**Proof:** Suppose the sum (16) over BM intervals in $\mathbb{R}_{+}$is infinite. If there are infinitely many BM intervals $l = (\tilde{a}_n, b_n)$ in $\mathbb{R}_{+}$satisfying $10|l| > d$, then we set

$$l_n = (a_n, b_n),\ a_n := \frac{10}{11}b_n;$$

otherwise we simply enumerate BM intervals such that $10|l| \leq d$. In any case, we get a collection of intervals $l_n = (a_n, b_n)$ satisfying the first two conditions in (17) and also the inequality

$$\gamma(b_n) \geq \gamma(a_n).$$

By (15), the latter implies that the intervals also satisfy (18) for some $c > 0$. Finally, we take a subfamily $\{l_{n_k}\}$ such that $\{5l_{n_k}\}$ is a subcover of $\cup 5l_n$ of finite multiplicity and observe that we still have the divergence of the series $\sum d^{\kappa-2}l^{2}$. Indeed, if $\cup l_j \subset 5l$, then

$$d_j \asymp d,\ \sum l_j^{2} \lesssim l^{2},$$

and so

$$\sum_j d_j^{\kappa-2}l_j^{2} \lesssim d^{\kappa-2}l^{2}.$$

The following proposition completes the proof of the first part of Theorem (2.1.22).

**Proposition (2.1.10)[49]:** Suppose $\gamma'(x) \gtrsim -|x|^{\kappa}$ and suppose there is a collection $\{l\}$ of disjoint intervals in $[1, +\infty)$ such that

$$\forall l,\ \Delta_l^*[\gamma] \geq cld^{\kappa},$$

and

$$\sum d^{\kappa-2} l^2 = \infty,\ 10l \leq d,\ \text{mult}\,\{51\} < \infty,$$

then $N^+\left[e^{i\gamma}\right] = 0$.

**Proof.** The statement corresponds to the so-called Beurling's lemma in the classical BM theory. There are several versions of the proof of Beurling's lemma, e.g. Koosis [66] applies the Beurling-Tsuji estimate of harmonic measure, Nazarov [74] uses the Bellman function, and Kargaev's proof [68] is based on PDE techniques. We suggest yet another approach. According to the criterion (5), we have to exclude the possibility

$$\gamma + \alpha = \tilde{h},\ \alpha \uparrow,\ h \in L^1_{\Pi}.$$

Denote by $h_l$ the restriction of $h$ to the interval $5l$. We say that $l$ is of type I if

$$d^{\kappa-2} l^2 \leq C\|h_l\|_{\Pi}, \tag{19}$$

where $C$ is a sufficiently large constant; otherwise we call $l$ an interval of type II. Clearly, we have

$$\sum_{l \in I} d^{\kappa-2} l^2 < \infty,$$

and to get a contradiction we need to show

$$\sum_{l \in II} d^{\kappa-2} l^2 < \infty \tag{20}$$

Consider the 2D Hilbert transform

$$H(z) = \int_{\mathbb{R}} \frac{h^-(t)dt}{(t-z)^2},\ (z \in \mathbf{C}_+).$$

where $h^- = \max\{0, -h\}$.

**Lemma (2.1.11)[49]:** If $l$ is of type II, then

$$|H(z)| \geq d^{\kappa},\ \forall z \in Q_l := \{x + iy : x \in l, l < y < 2l\}.$$

We prove this lemma, and we now explain how the lemma implies (20). Denote

$$\psi = \sum_{l \in \mathrm{II}} d^{\kappa} l \cdot 1_l.$$

We have

$$\sum_{l \in II} d^{\kappa-2} l^2 \asymp \int_1^{\infty} \frac{\psi(t)dt}{t^2} = \frac{8}{3}\int_1^{\infty} \frac{dA}{A^3}\int_{\frac{A}{2}}^{A} \psi(t)dt. \tag{21}$$

For every $A > 1$ let

$$H_A(z) = \int_{-CA}^{CA} \frac{h^-(t)dt}{(t-z)^2},$$

where $C > 0$ is a large constant, and let II $(A)$ denote the set of all intervals $l \in \mathrm{II}$ intersecting $(A/2, A)$. If $l \in \mathrm{II}(A)$ and $z \in Q_l$, then

$$|H(z) - H_A(z)| \leq \int_{|t|>CA} \frac{h^-(t)dt}{|t-z|^2} \asymp \int_{|t|>CA} \frac{h^-(t)dt}{1+t^2} \ll 1,$$

so by the lemma we have

$$|H_A| \geq A^\kappa \text{ on } \bigcup_{l \in II(A)} Q_l.$$

Applying the weak- $L^1$ estimate for the 2D Hilbert transform, see [53], we get

$$\sum_{l \in \mathrm{II}(A)} l^2 \leq \mathrm{Area}(|H_A| \geqslant A^\kappa) \lesssim A^{-\kappa} \int_{-CA}^{CA} h^-(t)dt,$$

and therefore

$$\int_{A/2}^{A} \psi(t)dt \lesssim A^\kappa \sum_{l \in \mathrm{II}(A)} l^2 \lesssim \int_{-CA}^{CA} h^-(t)dt,$$

Combining this with (21), we conclude

$$\sum_{l \in \mathrm{II}} d^{\kappa-2} l^2 \lesssim \int_1^\infty \frac{dA}{A^3} \int_{-CA}^{CA} h^-(t)dt \lesssim \|h^-\|_\Pi,$$

which proves (20)

**Proof.** Since $\alpha$ in the representation $\tilde{h} = \gamma + \alpha$ is increasing, we have

$$\Delta_l^*[\tilde{h}] \gtrsim l d^\kappa. \tag{22}$$

On the other hand, for intervals of type II we have

$$\Delta_l^*[\tilde{h}_l] \ll l d^\kappa. \tag{23}$$

Indeed, if $\Delta_l^*[\tilde{h}_l] \gtrsim l d^\kappa$, then $|\tilde{h}_l| \gtrsim l d^\kappa$ on either $l'$ or $l''$. Applying the weak type inequality with $A \asymp l d^\kappa$, we get

$$d^{-2} l \lesssim \Pi\{|\tilde{h}_l| > A\} \lesssim A^{-1} \|h_l\|_\Pi,$$

which contradicts the definition of type II.

Denote $f \equiv f_l = 1_{\mathbb{R}\setminus 5l} \cdot h$, so $\tilde{h} = \tilde{h}_l + \tilde{f}_l$. From (22)-(23) we conclude that there are points $a \in l'$ and $b \in l''$ such that

$$\tilde{f}(b) - \tilde{f}(a) \geq \frac{c}{2} l d^\kappa. \tag{24}$$

Represent $f = f^+ - f^-$with $f^+ = \max\{f, 0\}$, and note that the functions $\tilde{f}_\pm := (f^\pm)^2$ are decreasing on $[a, b]$ :

$$\tilde{f}_\pm'(x) = -\frac{1}{\pi} \int_{\mathbb{R}\setminus(5l)} \frac{f^\pm(t)dt}{(t-x)^2} < 0, \ (x \in 5l).$$

From (24) it then follows that

$$\tilde{f}_-(a) - \tilde{f}_-(b) \gtrsim l d^\kappa,$$

so there is a point $x_* \in (a, b)$ such that

$$\frac{1}{\pi} \int \frac{f^-(t)dt}{(t-x_*)^2} = -\tilde{f}_-'(x_*) = \frac{\tilde{f}_-(a) - \tilde{f}_-(b)}{b-a} \gtrsim d^\kappa.$$

Observe that if $z \in Q_l$ and $t \in \mathbb{R} \setminus 5l$, then

$$\Re\left[\frac{1}{(t-z)^2}\right] \asymp \frac{1}{(t-x_*)^2},$$

and we have

$$\left|\int \frac{f^-(t)dt}{(t-z)^2}\right| \geq \Re \int \frac{f^-(t)dt}{(t-z)^2} \asymp \int \frac{f^-(t)dt}{(t-x_*)^2} \gtrsim d^\kappa.$$

It follows that

$$|H(z)| \geq \left|\int \frac{f_l^-(t)dt}{(t-z)^2}\right| - \left|\int \frac{h_l^-(t)dt}{(t-z)^2}\right| \gtrsim d^\kappa,$$

because

$$\left|\int \frac{h_l^-(t)dt}{(t-z)^2}\right| \lesssim \frac{d^2}{l^2}\|h_l\|_\Pi \ll d^\kappa$$

provided that the constant $C$ in (19) is large enough.

We prove the following statement. Let $\kappa \geq 0$ and suppose that $U = e^{i\gamma}, S = e^{i\sigma}$ satisfy conditions (35) of Theorem (2.1.22).

**Proposition (2.1.12)[49]:** If $\gamma$ is almost $(\kappa)$ decreasing, then $N^+[U\bar{S}^\epsilon] \neq 0$ for all $\epsilon > 0$.

**Proof:** By assumption we have

$$\sum_{l\in\mathcal{BM}(\gamma)} d^{\kappa-2}l^2 < \infty. \tag{25}$$

Recall that the BM intervals of $\gamma$ are the components of the open set $\{\gamma^* \neq \gamma\}$, where $\gamma^*(x) = \max\gamma[x,+\infty)$. Denote

$$f = \gamma^* - \gamma,$$

so $f = 0$ outside the union of BM intervals, and $f'(x) \lesssim |x|^\kappa$ on BM intervals. By (25) we have $l \lesssim d$, and therefore

$$0 \leq f \lesssim ld^k \qquad on \quad l, \tag{26}$$

Together with (25) this implies

$$f \in L^1_\Pi.$$

The estimate (26) also shows that we can assume $l \geq d^{-\kappa}$ for all BM intervals; otherwise we can eliminate short intervals by adding a bounded function to $\gamma$ (this will not affect the $N^+$-kernel). In particular, we will assume that BM intervals don't cluster to a finite point. The non-triviality of $N^+[U\bar{S}^\epsilon]$ is a consequence of the following statement which will be verified.

**Lemma (2.1.13)[49]:** For any $\epsilon > 0$, there is a function $\beta$ such that

$$f + \beta \in \tilde{L}^1_\Pi,\ \beta'(x) \leq \epsilon|x|^\kappa \ \text{ for } |x| \gg 1.$$

Indeed, if for instance $\sigma'(x) \geq |x|^\kappa$ near $\pm\infty$, then we can write

$$\gamma - \epsilon\sigma = -(f+\beta) + (\beta - \epsilon\sigma) + \gamma^*.$$

The first term in the RHS is in $\tilde{L}^1_\Pi$, and the last two terms are decreasing near infinities, so we can apply the basic criterion (5).

**Proof.** We will construct disjoint intervals $l_n$ such that they cover all BM intervals and satisfy the following two conditions:

$$\sum_n d_n^{\kappa-2}l_n^2 < \infty, \tag{27}$$

and

$$\forall n\, \exists \epsilon_n \in [0,\epsilon],\ \int_{l_n} \frac{f(x) - \epsilon_n|x|^\kappa T_n(x)}{1+x^2}dx = 0, \tag{28}$$

where $T_n$ is the "tent" function of the interval $l_n$,

$$T_n(x) = \operatorname{dist}(x, \mathbb{R}\setminus l_n).$$

Let us show that the existence of such intervals $l_n$ implies the statement of the lemma. Define

$$\beta(x) = -\sum_n \epsilon_n|x|^\kappa T_n(x),$$

and

$$g(x) = f(x) - \sum \epsilon_n|x|^\kappa T_n(x).$$

Clearly, we have

$$|\beta'(x)| \lesssim \epsilon |x|^{\kappa},$$

and all we need is to check $g \in \tilde{L}^1_\Pi$.

Let us show that $g$ belongs to the real Hardy space $\mathcal{H}^1_\Pi(\mathbb{R})$. We can represent $g$ as follows:

$$g = \sum g_n = \sum \lambda_n \frac{g_n}{\lambda_n} := \sum \lambda_n A_n,$$

where

$$g_n = g \cdot 1_{l_n}, \ \lambda_n = \Pi(l_n)\|g_n\|_\infty.$$

The functions $A_n = \lambda_n^{-1} g_n$ are "atoms":

$$\int A_n d\Pi = \frac{1}{\lambda_n}\int_{l_n} g d\Pi = 0 \quad \text{by } (28),$$

and

$$\|A_n\|_\infty = \frac{\|g_n\|_\infty}{\lambda_n} = \frac{1}{\Pi(l_n)}.$$

Since

$$\|g_n\|_\infty \lesssim d_n^\kappa l_n,$$

(use $l_n \lesssim d_n$ and (26) for the BM intervals covered by $l_n$ ), we have

$$\sum \lambda_n \lesssim \sum \frac{l_n}{d_n^2} d_n^\kappa l_n < \infty \text{ by } \ (27).$$

It follows that $\sum \lambda_n A_n \in \mathcal{H}^1_\Pi(\mathbb{R})$, see [54].

Let us assume that all BM intervals $l$ lie in $[1, +\infty)$. In the general case we will need to apply the procedure described below to BM intervals in $(-\infty, -1]$ and in $[1, +\infty)$ separately.

We construct our intervals $l_1, l_2, \ldots$ by induction. The left endpoint $a_1$ of $l_1$ will be the left endpoint of the leftmost BM interval. Suppose the left endpoint $a_n$ of $l_n$ has been constructed so that $a_n$ is also the left endpoint of some BM interval $l = \left(a_{(l)}, b_{(l)}\right)$, i.e. $a_n = a_{(l)}$. Consider the function

$$F(b) = \int_{a_n}^{b} \left[f - \epsilon |x|^\kappa T_{(a_n, b)}\right] d\Pi,$$

where $T_{(a_n,b)}(\cdot) = \mathrm{dist}(\cdot, \{a_n, b\})$ is the tent function. We define $b_n$, the right endpoint of $l_n$, as the nearest point in the complement of BM intervals at which $F$ is non-positive,

$$b_n = \min \ \left\{b \ge b_{(l)} \colon f(b) = 0, F(b) \le 0\right\}.$$

Since $f \in L^1_\Pi$, we have $F(+\infty) = -\infty$ and so $b_n < \infty$. Finally, we define $a_{n+1}$ as the leftmost endpoint of BM intervals not covered by $l_1 \cup \cdots \cup l_n$. (Recall that we assumed that there are no finite cluster points.)

It is clear from the construction that the intervals $l_n$ cover all BM intervals. We also get (28) by defining $\epsilon_n$ from the equation

$$\int_{a_n}^{b_n} \left[f - \epsilon_n |x|^\kappa T_{(a_n, b_n)}\right] d\Pi = 0;$$

clearly we have $0 < \epsilon_n \le \epsilon$. In remains to verify (27). We have three types of intervals $l_n$ :

(a) $F(b_n) < 0$ but there is a BM interval $l \subset l_n$ such that $|l| \asymp |l_n|$,

(b) $F(b_n) = 0$,

(c) other intervals.

Property (27) is obvious for the collection of intervals of type (a): we have $l \ll d$ (except for finitely many $l'$ s) and therefore $d \asymp d_n$ and $d^{\kappa-2}l^2 \asymp d_n^{\kappa-2}l_n^2$.
To prove (27) for the collection of intervals of type (b), we note that $\epsilon_n - \epsilon$ if $l_n \in$ (b), and since

$$\int_a^b x^\kappa T_{(a,b)}(x)\frac{dx}{x^2} \geq \frac{a^\kappa}{b^2}\int_a^b T_{(a,b)} \geq \frac{a^\kappa}{b^2}(b-a)^2,$$

we have

$$\sum_{(b)} \frac{d_n^\kappa l_n^2}{(d_n+l_n)^2} \lesssim \sum_{(b)} \int_{a_n}^{b_n} |x|^\kappa T_{(a_n,b_n)} d\Pi = \frac{1}{\epsilon}\int_{\cup_{(b)} l_n} f d\Pi < \infty.$$

Since $d_n \to \infty$, it follows that there are only finitely many intervals $l_n \in (b)$ satisfying $d_n \leq l_n$, so the last estimate implies

$$\sum_{(b)} d_n^{\kappa-2} l_n^2 < \infty.$$

The argument for intervals of type (c) is the same if we can show that if $l_n \in (c)$, then $\epsilon_n > \epsilon/2$, i.e.

$$\int_{a_n}^{b_n} \left[f - \frac{\epsilon}{2}|x|^\kappa T_{(a_n,b_n)}\right] d\Pi > 0. \tag{29}$$

Since $l_n$ is not of type (b), we have $F(b_n) < 0$ and by construction, $b_n$ is the right endpoint of some BM interval $l = (c, b_n)$. Note that $|l| \ll |l_n|$ because $l_n$ is not of type (a). Since $f > 0$ on $l$, we have

$$\int_{a_n}^{b_n} \left(f - \frac{\epsilon}{2}x^\kappa T_{(a_n,b_n)}\right) d\Pi > \int_{a_n}^{c} \left(f - \epsilon x^\kappa T_{(a_n,c)}\right) d\Pi +$$
$$+\left[\epsilon \int_{a_n}^{c} x^\kappa T_{(a_n,c)} d\Pi - \frac{\epsilon}{2}\int_{a_n}^{b_n} x^\kappa T_{(a_n,b_n)} d\Pi\right].$$

The first term in the RHS is equal to $F(c)$ and therefore positive by construction. Since $|l| \ll |l_n|$, the second term in the RHS is also positive, and we get (29)

Let $S$ be a unimodular function and let $0 < p \leq \infty$. If $w \in L^1_\Pi$ is a real function, then we write

$$w \in \mathcal{M}_p(S)$$

if the outer function

$$W = e^{w+i\tilde{w}}$$

satisfies the following condition:

$$\forall \epsilon > 0,\ \exists G \in N^+[\bar{S}^\epsilon],\ WG \in H^p.$$

In other words, $w \in \mathcal{M}_p(S)$ if the corresponding outer function belongs to $H^p$ up to an arbitrarily small (compared to $S$) factor.
We can restate this property in terms of Toeplitz kernels.
**Lemma (2.1.14)[49]:** $w \in \mathcal{M}_p(S)$ iff

$$\forall \epsilon > 0,\ N^p\left[\bar{S}^\epsilon \frac{\bar{W}}{W}\right] \neq 0.$$

**Proof:** ⇒ Let $G \in N^+[\bar{S}^\epsilon]$ be such that $F := GW \in H^p$. Then

$$F \in N^p\left[\bar{S}^\epsilon \frac{\bar{W}}{W}\right],$$

and the Toeplitz kernel is non-trivial. Indeed,

$$\bar{S}^{\epsilon}\frac{\bar{W}}{W}F=(\bar{S}^{\epsilon}G)\bar{W}\in\mathcal{N}^{-}\cap L^{p}=\bar{H}^{p}.$$

$\Leftarrow$ If $F$ is in the Toeplitz kernel, i.e. $F\in H^{p}$ and $F\bar{S}^{\epsilon}\bar{W}/W\in\bar{H}^{p}$, then we define $G=F/W\in\mathcal{N}^{+}$. Since

$$\bar{S}^{\epsilon}G=\bar{S}^{\epsilon}\frac{\bar{W}}{W}F\frac{1}{\bar{W}}\in\mathcal{N}^{-},$$

we have $G\in N^{+}[\bar{S}^{\epsilon}]$ and $WG\in H^{p}$.

**Corollary (2.1.15)[49]:** Suppose $(\arg S)'\gtrsim|x|^{\kappa}$. If a real function $w_0\in L^1_{\Pi}$ satisfies the following condition:

$$\forall\epsilon>0,\ \exists w\in L^1_{\Pi},\ w\geq w_0,\ \tilde{w}'>-\epsilon|x|^{\kappa}+o(|x|^{\kappa}),$$

then $w_0\in\mathcal{M}_p(S)$ for all $p<1$.

**Proof:** Without loss of generality, $(\arg S)'\geq 2|x|^{\kappa}$ for $|x|\gg 1$. We have

$$-\arg\left[\bar{S}^{2\epsilon}\frac{\bar{W}_0}{W_0}\right]=2(\epsilon\arg S+\tilde{w})+2(\tilde{w}_0-\tilde{w}):=\alpha+\tilde{g},$$

where

$$\alpha'\geq\epsilon|x|^{\kappa},\ (|x|\gg 1),$$

and

$$g\in L^1_{\Pi},\ g\leq 0.$$

For sufficiently large $N$, the function $\alpha+\arg b^N$, where $b$ is the Blaschke factor (8), is monotone increasing on $\mathbb{R}$, and therefore there is an inner function $\Phi$, not a finite Blaschke product, such that

$$\alpha+\arg b^N=\arg\Phi+\delta,\ \ \|\delta\|_{\infty}\leq\pi.$$

Clearly, $N^{\infty}\left[e^{-i\tilde{g}}\right]\neq 0$, i.e.

$$N^{\infty}\left[e^{i\delta}\bar{b}^N\Phi\bar{S}^{2\epsilon}\frac{\bar{W}_0}{W_0}\right]\neq 0.$$

By (7), we also have

$$N^{p}\left[e^{-i\delta}\bar{b}^N\right]\neq 0,$$

provided that $p<1$ and $N$ is sufficiently large, and of course

$$N^{\infty}[b^{2N}\bar{\Phi}]\neq 0.$$

It follows that

$$N^{p}\left[\bar{S}^{2\epsilon}\frac{\bar{W}_0}{W_0}\right]\neq 0.$$

The main result is the following version of the Beurling-Malliavin multiplier theorem.

**Theorem (2.1.16)[49]:** Suppose $(\arg S)'\gtrsim|x|^{\kappa}$, and let $w_0\in L^1_{\Pi}$ be a real function. Then

$$|x|^{-\frac{2+\kappa}{2}}w_0(x)\in\mathcal{D}(\mathbb{R},\infty)\ \Rightarrow\ w_0\in\mathcal{M}_p(S),\ (\forall p<1).$$

Here the notation $f\in\mathcal{D}(\mathbb{R},\infty)$ means that there is a neighborhood of infinity where $f$ coincides with some function from the Dirichlet space $\mathcal{D}(\mathbb{R})$. Recall that the Hilbert space $\mathcal{D}(\mathbb{R})$ consists of functions $h\in L^1_{\Pi}$ such that the harmonic extension $u=u(z)$ of $h$ to $\mathbf{C}_+$has a finite gradient norm,

$$\|h\|^2_{\mathcal{D}}\equiv\|u\|^2_{\nabla}=\int_{\mathbf{C}_+}|\nabla u|^2dA<\infty,$$

( is the area measure). If $h\in\mathcal{D}(\mathbb{R})$ is a smooth function, then we also have

$$\| h \|_{\mathcal{D}}^2 = \int_{\mathbb{R}} \bar{h}\tilde{h}' dx.$$

We use some ideas from the proof of Theorem 64 in [55].

**Proof.** It is clear that we can assume that the function $w_0 + i\tilde{w}_0$ is analytic and has a zero of sufficiently large multiplicity at the origin; in particular

$$h_0(x) := |x|^{-\frac{2+\kappa}{2}} w_0(x) \in \mathcal{D}(\mathbb{R}).$$

Let us fix $\epsilon > 0$. According to the last corollary we need to construct $w$ such that

(i) $w \in L^1_\Pi$,

(ii) $w \geq w_0$,

(iii) $\tilde{w}' > -\epsilon|x|^\kappa + o(|x|^\kappa)$.

We define

$$w(x) = |x|^{\frac{2+\kappa}{2}} h(x),$$

where $h$ is a solution of the following extremal problem:

$$\mathrm{m}\ \{I(h) : h \geq h_0\},\ I(h) := \| h \|_{\mathcal{D}}^2 + \epsilon \int |x|^{\frac{2+\kappa}{2}} |h(x)| d\Pi(x).$$

The existence of a solution follows from the usual argument: the set

$$\mathcal{A} = \left\{h : \| h \|_{\mathcal{D}} \leq I(h_0),\ h \geq h_0 \text{ a.e.} \right\} \subset \mathcal{D}(\mathbb{R})$$

is bounded, closed, and convex in $\mathcal{D}(\mathbb{R})$, therefore it is weakly compact. Let $I_0$ denote the minimum of $I(h)$ over $\mathcal{A}$. Then there is a sequence of functions $h_n \in \mathcal{A}$ such that $I(h_n) \to I_0$ and $h_n$ weakly converge to some function $g \in \mathcal{A}$. It is then routine to see that

$$I_0 \leq I(g) \leq \liminf I(h_n) = I_0,$$

so $g$ is a solution of the extremal problem.

By construction, $w$ satisfies (i) and (ii). To prove (iii) we first note that

$$\tilde{h}'(x) \geq -\epsilon|x|^{\frac{\kappa-2}{2}}. \tag{30}$$

Indeed, by the extremality of $h$ we have

$$\| \phi \|_{\mathcal{D}}^2 + 2\int \phi\tilde{h}' + \epsilon \int (|h+\phi| - |h|)(x) \frac{|x|^{\frac{2+\kappa}{2}}}{1+x^2} dx = I(h+\phi) - I(h) \geq 0$$

for all smooth test functions $\phi = \phi(x) \geq 0$. (The integral $\int \phi h'$ has to be interpreted in the sense of the theory of distributions.) Since

$$\frac{\phi(x)}{x^2} \geq \frac{|h(x)+\phi(x)| - |h(x)|}{1+x^2},$$

we conclude

$$\| \phi \|_{\mathcal{D}}^2 + 2\int \phi(x)\left[\tilde{h}'(x) + \epsilon|x|^{\frac{\kappa-2}{2}}\right] dx \geq 0$$

Replacing $\phi(x)$ with $\delta\phi(x)$ and letting $\delta \to 0$, we get

$$\int \phi(x)\left[\tilde{h}'(x) + \epsilon|x|^{\frac{\kappa-2}{2}}\right] dx \geq 0$$

for all $\phi \geq 0$, which proves (30)

To derive (iii) from (30) we apply Lemma (2.1.7) with

$$\alpha = 1 + \frac{\kappa}{2},\ \beta = \kappa,\ \beta - \alpha = \frac{\kappa}{2} - 1.$$

The parameters $\alpha$ and $\beta$ are admissible because for $\kappa \geq 2$ we have $\alpha \leq \beta$, and if $0 \leq \kappa \leq 2$ then $1 \leq \alpha \leq 2$ and $\alpha \leq \beta + 1$.

**Proposition (2.1.17)[49]:** If $w \in L^1_\Pi, w \geq 0$, and $\tilde{w}' \lesssim |x|^\kappa$, then

$$|x|^{-\frac{2+\kappa}{2}}w(x) \in \mathcal{D}(\mathbb{R}, \infty).$$

**Proof:** We will assume that the function $w_0 + i\tilde{w}_0$ is analytic and has a zero of sufficiently large multiplicity at the origin. Let $u = u(z)$ be the harmonic extention of $|x|^{-\frac{2+\kappa}{2}}w(x)$ to the upper half plane $\mathbf{C}_+$, and let $v = \tilde{u}$. We need to show that the gradient norm of $u + iv$ in $\mathbf{C}_+$is finite,

$$\| u + iv \|_\nabla^2 = \lim_{r\to\infty} \int_{\partial D(r)} u dv < \infty,$$

where $D(r)$ is the semidisc $\{|z| < r\} \cap \mathbf{C}_+$.

We first prove that the integrals over $\partial D(r) \cap \mathbb{R}$ are uniformly bounded from above. Applying Lemma (2.1.6) with (admissible) parameters

$$\alpha = 1 + \frac{\kappa}{2},\ \beta = \kappa,\ \beta - \alpha = \frac{\kappa}{2} - 1$$

to the functions $f = w$ and $g = u$, we see that

$$v'(x) \lesssim |x|^{\frac{\kappa-2}{2}},\ x \in \mathbb{R}.$$

Since $u > 0$ we have

$$\int_{\partial D(r)\cap\mathbb{R}} u dv = \int_{-r}^{r} v'u \lesssim \int_{-r}^{r} |x|^{\frac{\kappa-2}{2}} |x|^{-\frac{2+\kappa}{2}} w(x) dx \lesssim \| w \|_\Pi < \infty.$$

To finish the proof of the proposition it remains to show that the integrals

$$\int_{\partial D(r)\setminus\mathbb{R}} u dv = rI'(r),\ I(r) := \frac{1}{2}\int_0^\pi u^2(re^{i\theta}) d\theta,$$

don't tend to $+\infty$ as $r \to \infty$. In fact, it is enough to show

$$I(r) \nrightarrow \infty,$$

because if $rI'(r) \to +\infty$, then $I'(r) \geq 1/r$ for all $r \gg 1$, and we have $I(r) \to \infty$.

Since $\kappa \geq 0$, we can apply the following lemma.

**Lemma (2.1.18)[49]:** If $u \in L^1(1 + |x|^{-1})$, then $I(r) \nrightarrow \infty$.

**Proof:** We will prove an equivalent statement for functions in the unit disc $\mathbf{D}$. Let $f = u + i\tilde{u}$ be an analytic function in $\mathbf{D}$ such that

$$\frac{u(\zeta)}{1 - |\zeta|} \in L^1(\partial\mathbf{D}).$$

Define

$$h(z) = \frac{1+z}{1-z} u(z),\ z \in \mathbf{D},$$

and denote by $h^*(\zeta), \zeta \in \partial\mathbf{D}$, the angular maximal function. By Hardy-Littlewood maximal theorem,

$$h^* \in L^1_{\text{weak}}(\partial\mathbf{D}). \tag{31}$$

Let us show that as $\epsilon \to 0$,

$$\frac{1}{\epsilon}\int_{C_\epsilon} |f(z)|^2 |dz| \nrightarrow \infty,\ C_\epsilon := \{|1 - z| = \epsilon\} \cap \mathbf{D}.$$

We have

$$\frac{1}{\epsilon}\int_{C_c} |f|^2 = \epsilon \int_{C_c} |h|^2 \lesssim [\epsilon h^*(\zeta)]^2 + [\epsilon h^*(\bar{\zeta})]^2,$$

where $\zeta \in \partial\mathbf{D}, |1 - \zeta| = \epsilon$. The RHS can not tend to infinity because otherwise for all small $\epsilon$, we would have

$$h^*(\zeta) + h^*(\bar{\zeta}) \gg \frac{1}{\epsilon}$$
on an interval of length $\epsilon$, which would contradict (31).

**Proposition (2.1.19)[49]:** Suppose $(\arg S)' \gtrsim |x|^\kappa$ and let $\Theta$ be a meromorphic inner function satisfying $|\Theta'| \lesssim |x|^\kappa$. Then
$$W \in N^+[\bar{\Theta}] \Rightarrow \log|W| \in \mathcal{M}_p(S), (\forall p < 1).$$
**Proof:** We have $W\bar{\Theta} = \bar{H}$ for some $H \in \mathcal{N}^+$. Define
$$W_1 = WH + \Theta,$$
and let $W_1^e$ be the outer part of $W_1$. From the identity
$$\bar{\Theta}^2 W_1 = \bar{\Theta} W \bar{\Theta} H + \bar{\Theta} = \bar{H}\bar{W} + \bar{\Theta} = \bar{W}_1,$$
we deduce
$$|W_1| = |W\bar{W}\Theta + \Theta| = 1 + |W|^2 \geq 1,$$
and
$$|W_1^e| \geq 1, |W| \leq |W_1^e|, (\arg W_1^e)' \lesssim |x|^\kappa.$$
By Proposition (2.1.17) and the multiplier theorem, we have $\log|W_1| \in \mathcal{M}_p(S)$ and therefore
$$\log|W| \in \mathcal{M}_p(S).$$
**Corollary (2.1.20)[49]:** Let $S$ and $\Theta$ be as above. Then for any meromorphic inner function $J$ and any $p < 1$, we have
$$N^+[\bar{\Theta}J] \neq 0 \Rightarrow \forall\epsilon, N^p[\bar{S}^\epsilon \bar{\Theta} J] \neq 0.$$
**Proof:** Take an outer function $W \in N^+[\bar{\Theta}J]$. Then $W \in N^+[\bar{\Theta}]$, and by the last proposition,
$$\exists G \in N^+[\bar{S}^\epsilon]. \ WG \in H^p.$$
It then follows that
$$WG \in N^+[\bar{S}^\epsilon \bar{\Theta} J] \cap H^p = N^p[\bar{S}^\epsilon \bar{\Theta} J].$$
We finish the proof of Theorems (2.1.22) and (2.1.23).

It is well known that given any two intertwining discrete sets $A = \{a_n\}$ and $B = \{b_n\}$ of real numbers, $\ldots a_n < b_n < a_{n+1} \ldots$, there exists a meromorphic inner function $\Theta$ such that
$$\{\Theta = 1\} = A, \{\Theta = -1\} = B. \tag{32}$$
Indeed, the sequences $A, B$ determine the set
$$E = \{\Im\Theta > 0\} \cap \mathbb{R} = \cup (a_n, b_n),$$
and we can define $\Theta$ in $\mathbf{C}_+$ by the (Krein's shift) formula
$$\frac{1}{\pi i}\log\frac{\Theta + 1}{\Theta - 1} = \mathcal{S}u + ic, \ u := 1_E - \frac{1}{2}, \ c \in \mathbb{R}, \tag{33}$$
where $\mathcal{S}u$ is the Schwarz integral (4), so $\Re[\mathcal{S}u]$ is the Poisson extension of $u$ to the halfplane. (Note that $u$ is the boundary function of the expression in the LHS of (33), provided that $\Theta$ is an inner function with level sets $A$ and $B$, and in fact Krein's shift formula parametrizes all such inner functions.)

An immediate consequence of this construction is the following statement:
for any increasing, continuous function $\sigma: \mathbb{R} \to \mathbb{R}$, there exists a meromorphic inner function $\Theta = e^{i\theta}$ such that
$$\| \theta - \sigma \|_\infty \leq \pi.$$
We will need the following version of this statement.

**Lemma (2.1.21)[49]:** If $\sigma'(x) \asymp |x|^\kappa$, then there is a meromorphic inner function $\Theta = e^{i\theta}$ such that
$$\theta - \sigma \in L^\infty(\mathbb{R}), \ \theta'(x) \asymp |x|^\kappa.$$

**Proof:** We can assume that $\sigma$ is strictly increasing on $\mathbb{R}$. Define the intertwining sequences $A=\{a_n\}$ and $B=\{b_n\}$ by the equations

$$\sigma(a_n)=2\pi n,\ b_n=\frac{a_n+a_{n+1}}{2},\ (n\in\mathbf{Z}),$$

so we have

$$a_n\asymp(\operatorname{sign}n)|n|^{\frac{1}{1+\kappa}},$$

and

$$\delta_n:=b_n-a_n\asymp|a_n|^{-\kappa}.$$

Let $\Theta$ be an inner function satisfying (32),

$$\|\ \theta-\sigma\ \|_\infty\leq 2\pi,$$

and let $\mu_1,\mu_{-1}$ be the corresponding (Aleksandrov-Clark's) measures defined by the Herglotz representation

$$\frac{1+\Theta}{1-\Theta}=\mathcal{S}\mu_1+\ \text{const},\ \frac{1-\Theta}{1+\Theta}=\mathcal{S}\mu_{-1}+\ \text{const}.$$

The measures $\mu_1,\mu_{-1}$ have the following form:

$$\mu_1=\sum\ \alpha_n\delta_{a_n},\ \mu_{-1}=\sum\ \beta_n\delta_{b_n}$$

for some positive numbers $\alpha_n,\beta_n$. (It is easy to see that $\mu_{\pm1}\{\infty\}=0$ though we don't actually need this fact.) We claim that

$$\alpha_n\asymp\delta_n,\ \beta_n\asymp\delta_n. \tag{34}$$

The estimate $\theta'(x)\asymp|x|^\kappa$ easily follows from (34). Since

$$|\Theta'|\asymp|1-\Theta^2||(\mathcal{S}\mu_1)'|,\ |\Theta'|\asymp|1+\Theta^2||(\mathcal{S}\mu_{-1})'|,$$

we have

$$\theta'(x)\asymp\min\left\{\sum\ \frac{\alpha_n}{(x-a_n)^2},\sum\ \frac{\beta_n}{(x-b_n)^2}\right\},\ (x\in\mathbb{R}).$$

It follows that if $x\in(a_m,a_{m+1})$, then by (34)

$$\theta'(x)\asymp\int_{|t-x|\gtrsim\delta_m}\frac{dt}{(x-t)^2}\asymp\delta_m^{-1}\asymp|a_m|^\kappa\asymp|x|^\kappa.$$

**Proof of (34).** We will explain the estimate for $\alpha_n$ 's; the proof for $\beta_n$ 's is similar. According to (33), we have

$$\frac{1-\Theta}{1+\Theta}=\ \text{const}\ e^{Ku},$$

where $u=1_E-1/2$,

$$E=\bigcup_{k=-\infty}^{\infty}\ (a_k,b_k),$$

and $Ku$ is the improper integral

$$Ku(z)=\int\ \frac{u(t)dt}{t-z},\ (z\in\mathbf{C}_+).$$

By construction,

$$\alpha_n=\ \text{const}\ \operatorname{Res}_{a_n}e^{Ku}.$$

Denote

$$g_n(z)=\exp\left\{\int_{b_{n-1}}^{b_n}\frac{u(t)dt}{t-z}\right\}=\frac{\sqrt{(b_n-z)(b_{n-1}-z)}}{a_n-z},$$

and

$$A_n = \exp\left\{\int_{\mathbb{R}\setminus(b_{n-1},b_n)} \frac{u(t)dt}{t-a_n}\right\},$$

So

$$Res_{a_n} e^{Ku} = A_n\, Res_{a_n} g_n, \qquad \left|\text{Res}_{a_n}\, g_n\right| \asymp \delta_n.$$

It remains to show that $A_n = e^{O(1)}$. This can be done as follows.

For $j > n$ we have

$$\begin{aligned}\int_{a_j}^{a_{j+1}} \frac{u(t)dt}{t-a_n} &= \log\frac{b_j - a_n}{a_j - a_n} - \log\frac{a_{j+1}-a_n}{b_j - a_n}\\ &= \log\left(1+\frac{\delta_j}{a_j-a_n}\right) - \log\left(1+\frac{\delta_j}{b_j-a_n}\right)\\ &= \frac{\delta_j}{a_j-a_n} - \frac{\delta_j}{b_j-a_n} + O\left(\frac{\delta_j^2}{\left(a_j-a_n\right)^2}\right) = O\left(\frac{\delta_j^2}{\left(a_j-a_n\right)^2}\right),\end{aligned}$$

where we used the relation $\log(1+x) = x + O(x^2)$ for $0 < x \lesssim 1$. Since

$$\begin{aligned}\sum_{j=n+1}^{\infty} \frac{\delta_j^2}{\left(a_j-a_n\right)^2} &\asymp \sum_{j=n+1}^{\infty} \frac{\delta_j}{a_j^{\kappa}\left(a_j-a_n\right)^2}\\ &\asymp \int_{a_n+\delta_n}^{\infty} \frac{dt}{t^{\kappa}(t-a_n)^2} = \int_{a_n+\delta_n}^{2a_n} + \int_{2a_n}^{\infty}\\ &\lesssim \frac{1}{a_n^{\kappa}}\int_{a_n+\delta_n}^{\infty} \frac{dt}{(t-a_n)^2} + \int_{a_n}^{\infty} \frac{dt}{t^{2+\kappa}}\\ &\asymp \frac{1}{a_n^{\kappa}}\frac{1}{\delta_n} + \frac{1}{a_n^{1+\kappa}} = O(1),\end{aligned}$$

we get

$$\int_{b_n}^{\infty} \frac{u(t)dt}{t-a_n} = O(1).$$

A similar estimate holds for the integral over $(-\infty, b_{n-1})$, and we have $A_n = e^{O(1)}$.

**Theorem (2.1.22)[49]:** Let $\kappa \geq 0$, and let $U = e^{i\gamma}$ and $S = e^{i\sigma}$ be smooth unimodular functions on $\mathbb{R}$ such that

$$\gamma'(x) \geq -|x|^{\kappa},\ \sigma'(x) \gtrsim |x|^{\kappa},\ (x \to \infty). \tag{35}$$

(i) If $\gamma$ is not $(\kappa)$-almost decreasing, then $N^{+}[US^{\epsilon}] = 0$ for all $\epsilon > 0$.

(ii) If $\gamma$ is $(\kappa)$-almost decreasing, then $N^{p}[U\bar{S}^{\epsilon}] \neq 0$ for all $\epsilon > 0$ and all $p < \frac{1}{3}$.

Here and throughout the notation $f(x) \geq g(x)$ means that $f(x) \geq \text{c}g(x)$ for some $c > 0$ and all $x$ such that $|x| \geqslant 1$.

Given two unimodular functions $U$ and $S$ as in Theorem (2.1.22), we can consider the family of symbols

$$U\bar{S}^{a} = e^{i\gamma_a},\ \gamma_a = \gamma - a\sigma,\ (a \in \mathbb{R}).$$

If $a_1 > a$ and if $\gamma_a$ is decreasing near $\infty$, then $\gamma_{a_1}$ is also decreasing. It is not difficult to see that the same is true for almost decreasing functions, so we can define the transition parameter

$$c \equiv c(U,S;\kappa) = \inf\ \left\{a\colon \gamma_a \text{ is } (\kappa)\text{-almost descreasing}\right\} \in (-\infty,+\infty].$$

**Proof.** The first part of the theorem was established. The second part states that if $\gamma$ is almost decreasing and $\epsilon > 0$, then

$$N^p[U\bar{S}^{2\epsilon}] \neq 0,\ (p < 1/3). \tag{36}$$

By Lemma (2.1.21) there exists an inner function $\Theta$ satisfying

$$(\arg\Theta)' \asymp |x|^\kappa.$$

We will assume that $U^2\Theta$ has an increasing argument (otherwise we can replace $\Theta$ with $\Theta^n$ for a large integer $n$). We will also assume that $S\bar{\Theta}$ has an increasing, unbounded argument (otherwise we replace $S$ with a large power). By Proposition (2.1.12) we have

$$N^+[U\Theta^{1-\epsilon}\bar{\Theta}] \neq 0. \tag{37}$$

Since the argument of $U\Theta^{1-\epsilon}$ is increasing, there is an inner function $J$ such that

$$U\Theta^{1-\epsilon} = XJ,\ \ \| \arg X \|_\infty \leq \pi.$$

From (37) we have $N^+[J\bar{\Theta}] \neq 0$, and so by Corollary (2.1.20)

$$N^p[J\bar{\Theta}\bar{S}^\epsilon] \neq 0,\ (p < 1). \tag{38}$$

Note that

$$U\bar{S}^{2\epsilon} = (U\Theta^{1-\epsilon}\bar{\Theta}\bar{S}^\epsilon)(\Theta^\epsilon\bar{S}^\epsilon) = X(J\bar{\Theta}\bar{S}^\epsilon)(\Theta^\epsilon\bar{S}^\epsilon)$$

. Since the argument of $S^\epsilon\bar{\Theta}^\epsilon$ is increasing and unbounded, we can find an infi Blaschke product $\Psi$ such that

$$\Theta^\epsilon\bar{S}^\epsilon = Y\bar{\Psi},\ \ \| \arg Y \|_\infty \leq \pi.$$

Thus the symbol $U\bar{S}^{2\epsilon}$ has the following representation:

$$U\bar{S}^{2\epsilon} = (J\bar{\Theta}\bar{S}^\epsilon)(XY\bar{\Psi}),\ \ \| \arg XY \|_\infty \leq 2\pi,$$

and by (7) we have

$$N^p[XY\bar{\Psi}] \neq 0,\ (p < 1/2). \tag{39}$$

Combining (38) and (39), we get (36) by Hölder's inequality.

**Theorem (2.1.23)[49]:** Let $J$ be a meromorphic inner function, and suppose that a unimodular function $S$ satisfies

$$(\arg S)'(x) \asymp |x|^\kappa,\ x \to \infty.$$

Denote $c = c(J, S; \kappa)$. Then for all $p \leq \infty$ we have

$$N^p[J\bar{S}^a] = 0\ (a < c),\ N^p[J\bar{S}^a] \neq 0\ (a > c).$$

**Proof.** Recall that $J$ is a meromorphic inner function, $S = e^{i\sigma}$ with $\sigma'(x) \asymp |x|^\kappa$, and $c = c(J, S; \kappa)$. Applying Theorem (2.1.22) (or rather its corollary) to $U = J$ we conclude that if $a < c$ then $N^+[J\bar{S}^a] = 0$ and therefore $N^p[J\bar{S}^a] = 0$ for all $p > 0$. On the other hand, if $a > c$, then $N^p[J\bar{S}^a] \neq 0$ for some $p > 0$, and in fact the kernel is infinite dimensional, as we just mentioned. The following proposition completes the proof.

A unimodular function $S$ is called tempered if $\exists n, S'(x) = O(|x|^n)$ as $x \to \infty$.

**Proposition (2.1.24)[49]:** If $S$ is a tempered unimodular function, then for any meromorphic inner function $J$ and any $p > 0$,

$$\dim N^p[J\bar{S}] = \infty\ \Rightarrow\ \dim N^\infty[J\bar{S}] = \infty.$$

**Proof:** First of all we observe that the statement is true if $S$ is a tempered inner function, $S = \Theta$. By Carleson's type embedding theorem [83], all elements in $N^p[J\bar{\Theta}]$ have at most polynomial growth at infinity, see details in [72]. Since the kernel is infinite dimensional, it contains functions with many zeros in $\mathbf{C}_+$. Dividing such functions by appropriate polynomials we obtain functions in $N^\infty[J\bar{\Theta}]$.

Let now $S$ be an arbitrary tempered unimodular function. By Lemma (2.1.21) we can find a tempered inner function $\Theta$ and a bounded real-valued function $\chi$ such that

$$S = \Theta\bar{X},\ X = e^{2i\chi}.$$

By the previous observation, we have

$$\dim N^{\infty}[J\bar{\Theta}] = \infty, \tag{40}$$

and it remains to show that

$$\exists n,\ N^{\infty}[X\bar{b}^{n}] \neq 0, \tag{41}$$

($b$ is the Blaschke factor (8)). Indeed, combining (40) and (41) we conclude that the kernel

$$N^{\infty}[J\bar{S}\bar{b}^{n}] = N^{\infty}[J\bar{\Theta}X\bar{b}^{n}]$$

is infinite dimensional, which allows us to get rid of $b^{n}$.
To prove (41), consider the outer function

$$H = e^{\tilde{\chi} - i\chi},\ \text{ so }\ X = \frac{\bar{H}}{H}.$$

We will have

$$(z+i)^{-n}H(z) \in N^{\infty}[X\bar{b}^{n}],\ (n \gg 1)$$

if we can show that $h := |H| = e^{\tilde{x}}$ has at most polynomial growth at infinity. Without loss of generality, we can assume that the $L^{\infty}$-norm of $\chi$ is so small that $h \in L^{2}_{\Pi}$. We have

$$h(x) - h(0) \leq \int_{0}^{x} |h'| = \int_{0}^{x} h|\tilde{\chi}'| \lesssim \| h \|_{L^{2}_{\Pi}} (1+x^{2})^{\frac{1}{2}} \left( \int_{0}^{x} |\tilde{\chi}'|^{2} \right)^{\frac{1}{2}}. \tag{42}$$

Since $|\chi'(t)| \lesssim |t|^{n}$ by construction, for each $x > 0$ we can represent $\chi$ as the sum of two smooth functions,

$$\chi = \chi_1 + \chi_2,$$

such that

$$\|\chi_1'\|_{L^2} \lesssim |x|^{n},\ \|\chi_2\|_{L^{\infty}} \asymp 1,\ \chi_2 = 0 \ \text{ on }\ (-2x, 2x).$$

(For example, take $\chi_1 = \phi\chi$, where $\phi$ is a smooth "bump" function such that $\phi$ is equal to 1 on $(-2x, 2x)$ and 0 on $\mathbb{R} \setminus (-3x, 3x)$.) Then we have

$$\|\tilde{\chi}_1'\|_{L^2} \lesssim |x|^{n},\ |\tilde{\chi}_2'| \lesssim 1 \ \text{ on }\ (0, x),$$

and so (42) shows that $h$ has at most polynomial growth.

**Section (2.2): Sets and Measures**

For $\mu$ be a non-zero finite complex measure on the real line. By $\hat{\mu}$ we denote its Fourier transform

$$\hat{\mu}(z) = \int \exp(-izt)\, d\mu(t).$$

Various properties of the Fourier transform of a measure have been studied by harmonic analysts for more than a century. One of the reasons for such a prolonged interest is the natural physical sense of the quantity $\hat{\mu}(t)$. In quantum mechanics, if $\mu$ is a spectral measure of a Hamiltonian then $|\hat{\mu}(t)|^2$ represents the so-called survival probability of the particle, i.e. the probability to find the particle in its initial state at the moment $t$. The problems considered belong to the area of the Uncertainty Principle in Harmonic Analysis, whose name itself suggests relations and similarities with physics.
The Uncertainty Principle in Harmonic Analysis, as formulated in [58], says that a measure (function, distribution) and its Fourier transform cannot be simultaneously small. This broad statement gives rise to a multitude of exciting mathematical problems, each corresponding to a particular sense of "smallness."
One of such problems is the well-known Gap Problem. Here the smallness of $\mu$ and $\hat{\mu}$ is understood in the sense of porosity of their supports. The statement that one hopes to obtain is that if the support of $\hat{\mu}$ has a large gap then the support of $\mu$ cannot be too "rare." As usual,

the ultimate challenge is to obtain quantative estimates relating the two supports, something that we will attempt to do.
Beurling's Gap Theorem says that if the sequence of gaps in the support of $\mu$ is long, in the sense given by (50), then the support of $\hat{\mu}$ cannot have any gaps, unless $\mu$ is trivial, see [87]. The proof used some of the methods of an earlier gap theorem by Levinson [70]. In [55] de Branges proved that existence of a measure with a given spectral gap is equivalent to existence of a certain entire function of exponential type. We discuss a version of this result. For further results and for gap problem see [87], [86], [66], [70].
We find an "if and only if" condition for a closed set $X$ on the real line to support a non-trivial complex measure with a given size of the spectral gap. We introduce a new metric characteristic of a closed set, $C_X$, for the definition. The main result is Theorem (2.2.9) that says that the supremum of the lengths of the spectral gaps, taken over all non-trivial measures supported on $X$, is equal to $2\pi C_X$.
The definition of $C_X$ contains two conditions that, for the purposes, we call the density condition and the energy condition. As discussed, the density condition is similar to some of the definitions of densities used in this area. The physical flavor of the energy condition seems to suggest new connections for the gap problem that are yet to be fully understood.

The gap problem can be equivalently reformulated as follows. Let $\mu$ be a finite complex measure on $\mathbb{R}$. Find the supremum, over all non-trivial $f \in L^1(|\mu|)$, of the size of the spectral gap of the measure $f\mu$. If one replaces $L^1$ with $L^2$ in this statement, one obtains (via simple duality) another famous problem, the problem of Wiener and Kolmogorov on completness of families of exponentials in $L^2(\mu)$, see [94] or [90]. In [94] the problem is formulated in the language of Krein strings. Since for finite measures $L^2 \subset L^1$, our result gives an upper estimate for the Wiener-Kolmogorov problem. The $L^\infty$-version was considered by Koosis, see [66].
Via duality the gap problem admits a reformulation in terms of Bernstein's weighted approximation. From that point of view, $2\pi C_X$ is the minimal size of the interval such that continuous functions on $X$ admit weighted approximation by trigonometric polynomials with frequencies from that interval. The "approximative" relatives of the gap problem and connections with other classical areas, such as stationary Gaussian processes, are discussed in [91], [93], [66], [95].
Our methods are based on the approach developed by N. Makarov [72] and [49]. We utilize close connections between most problems from this area of harmonic analysis and the problem of injectivity of Toeplitz operators. In the case of the gap problem, this connection is expressed by Theorem (2.2.6) below. The Toeplitz approach for similar problems was first suggested by Nikolski in [75], see also [98]. Our main proof utilizes several important ideas of the Beurling-Malliavin theory [50], [51], [49], including its famous multiplier theorem.
One of the advantages of the Toeplitz approach is that it reveals hidden connections between various problems of analysis and mathematical physics, see [72]. The relations between the gap problem and the Beurling-Malliavin theory on completeness of exponentials in $L^2$ on an interval have been known to experts, at a rather intuitive level, for several decades. Now we can see this connection formulated in precise mathematical terms. Namely, the Beurling-Malliavin problem is equivalent to the problem of triviality of the kernel of a Toeplitz operator with the symbol

$$\phi = \exp(-iax)\,\theta,$$

for a suitable meromorphic inner function $\theta$, whereas the gap problem is equivalent to the triviality of the kernel of the Toeplitz operator with the symbol

$$\bar{\phi} = \exp(iax)\bar{\theta},$$

see [72].

We organized as follows:

(i) We discuss an alternative formulation of the gap problem and show that the maximal size of the gap for a fixed measure taken over all possible densities is a property of the support of the measure.

(ii) We look at the gap problem from the point of view of Bernstein's weighted approximation of continuous functions by trigonometric polynomials.

(iii) We restate the gap problem in terms of kernels of Toeplitz operators and introduce the approach that will be used in the main proof.

(iv) We contain the main definition and its discussion. For a closed real set $X$ we define a metric characteristic $C_X$ that determines the maximal size of the gap over all non-zero complex measures supported on $X$.

(v) We contain the main result and its proof.

(vi) We prove several technical lemmas and corollaries used.

(vii) We can be viewed as an appendix. It contains a Toeplitz version of the statement and proof of theorem 66 from [55].

Let $M$ be a set of all finite Borel complex measures on the real line. If $X$ is a closed subset of the real line denote

$$G_X = \sup\{a \mid \exists \mu \in M, \mu \not\equiv 0, \operatorname{supp}\mu \subset X, \text{ such that } \hat{\mu} = 0 \text{ on } [0,a]\}. \quad (43)$$

Now let $\mu \in M$. Denote

$$G_\mu = \text{s } \{a \mid \exists f \in L^1(|\mu|) \text{ such that } \widehat{f\mu} = 0 \text{ on } [0,a]\}. \quad (44)$$

**Proposition (2.2.1)[84]:**

$$G_\mu = G_{\operatorname{supp}\mu}. \quad (45)$$

**Proof.** Obviously, $G_{\operatorname{supp}\mu} \geq G_\mu$. To prove the opposite inequality, notice that by Lemma (2.2.20) there exists a finite discrete measure

$$\nu = \sum \alpha_n \delta_{x_n}, \{x_n\} \subset \operatorname{supp}\mu,$$

such that $\hat{\mu}$ has a gap of the size greater than $G_{\operatorname{supp}\mu} - \varepsilon$. Around each $x_n$ choose a small neighborhood $V_n = (a_n, b_n)$ so that for any sequence of points

$$Y = \{y_n\}, y_n \in V_n$$

there exists a non-trivial measure $\eta_Y = \sum \beta_n \delta_{y_n}$ such that $\hat{\eta}$ has a gap of the size greater than $G_{\operatorname{supp}\eta} - \varepsilon$. The existence of such a collection of neighborhoods follows from the results of [86] (for some sequences), from [88] as well as from Theorem (2.2.9) below.

Now one can choose a family of finite measures $\eta_\tau, \tau \in [0,1]$ with the following properties:

- for each $\tau$,

$$\eta_\tau = \sum \beta_n^\tau \delta_{y_n^\tau}$$

such that $y_n^\tau \in V_n$ and $\hat{\eta}_\tau$ has a gap of the size greater than $G_{\operatorname{supp}\eta} - \varepsilon$ centered at 0 ;

- the measure

$$\gamma = \int_0^1 \eta_\tau d\tau$$

is non-trivial and absolutely continuous with respect to $\mu$.

It remains to notice that then $\hat{\gamma}$ has a gap of the size greater than $G_{\text{supp}}\,\eta - \varepsilon$.

**Definition (2.2.2)[84]:** Let $X \subset \mathbb{R}$ be a closed set. A weight is any lower semicontinuous function $W: \mathbb{R} \to [1, \infty)$ that tends to $\infty$ as $x \to \pm\infty$. For any given weight $W$ we define $C_0(W, X)$ to be the space of all continuous functions on $X$ satisfying

$$\lim_{x\in X, x\to\pm\infty} \frac{f(x)}{W(x)} = 0.$$

(If $X$ is bounded from below or from above, the corresponding limit is dropped from the definition. In particular, for bounded $X, C_0(W, X)$ is just $C(X)$.)

We define the norm in $C_0(W, X)$ as

$$\| f \| = \|fW^{-1}\|_\infty.$$

As usual, we say that a system of functions is complete in a space if finite linear combinations of functions from that system are dense in the space.

If $a > 0$ we denote by $\mathcal{E}_a$ the set of complex exponentials with frequencies between 0 and $a$:

$$\mathcal{E}_a = \{\exp(i\lambda t) \mid \lambda \in [0, a]\}.$$

**Definition (2.2.3)[84]:** If $X \subset \mathbb{R}$ is a closed set, define the approximative capacity of $X$, $A_X$, as

$$A_X = \inf\{a \mid \mathcal{E}_a \text{ is complete in } C_0(W, X) \text{ for any weight } W\}$$

or $\infty$ if the set is empty.

The same quantity can be defined in a different way. Denote by $C_0(X)$ the space of all continuous functions on $X$ tending to 0 at infinity, with the usual supnorm. If one now wants to discuss approximation by trigonometric polynomials in $C_0(X)$, he encounters a small problem: exponential functions are no longer inside the space. The solution is to consider "linear combinations" of exponentials, e.g. the Payley-Wiener space

$$PW_a = \{\hat{f} \mid f \in L^2([0, a]\},$$

and define

$$A_X = \inf\{a \mid PW_a \text{ is dense in } C_0(X)\}$$

or $\infty$ if the set is empty.

It is not difficult to show that the above definitions of $A_X$ are equivalent. The following statement is a product of the standard duality argument.

**Proposition (2.2.4)[84]:**

$$A_X = G_X.$$

Together with Theorem (2.2.9), this statement gives a formula for $A_X$.

By $H^2$ we denote the Hardy space in the upper half-plane $\mathbb{C}_+$. We say that an inner function $\theta(z)$ in $\mathbb{C}_+$is meromorphic if it allows a meromorphic extension to the whole complex plane. The meromorphic extension to the lower half-plane $\mathbb{C}_-$ is given by

$$\theta(z) = \frac{1}{\theta^\#(z)}.$$

Each inner function $\theta(z)$ determines a model subspace

$$K_\theta = H^2 \ominus \theta H^2$$

of the Hardy space $H^2(\mathbb{C}_+)$. These subspaces play an important role in complex and harmonic analysis, as well as in operator theory, see [98].

Each inner function $\theta(z)$ determines a positive harmonic function

$$\Re \frac{1 + \theta(z)}{1 - \theta(z)}$$

and, by the Herglotz representation, a positive measure $\sigma$ such that

$$\Re\frac{1+\theta(z)}{1-\theta(z)} = py + \frac{1}{\pi}\int \frac{y d\sigma(t)}{(x-t)^2+y^2}, \; z = x+iy, \tag{46}$$

for some $p \geq 0$. The number $p$ can be viewed as a point mass at infinity. The measure $\sigma$ is singular, supported on the set where non-tangential limits of $\theta$ are equal to 1 and satisfies

$$\int \frac{d\sigma(t)}{1+t^2} < \infty. \tag{47}$$

The measure $\sigma + p\delta_\infty$ on $\hat{\mathbb{R}}$ is called the Clark measure for $\theta(z)$. (Following standard notations, we will often denote the Clark measure defined in (46) by $\sigma_1$.)
Conversely, for every positive singular measure $\sigma$ satisfying (47) and a number $p \geq 0$, there exists an inner function $\theta(z)$ determined by the formula (46).

Every function $f \in K_\theta$ can be represented by the formula

$$f(z) = \frac{p}{2\pi i}(1-\theta(z))\int f(t)\overline{(1-\theta(t))}dt + \frac{1-\theta(z)}{2\pi i}\int \frac{f(t)}{t-z}d\sigma(t). \tag{48}$$

If the Clark measure does not have a point mass at infinity, the formula is simplified to

$$f(z) = \frac{1}{2\pi i}(1-\theta(z))Kf\sigma$$

where $Kf\sigma$ stands for the Cauchy integral

$$Kf\sigma(z) = \int \frac{f(t)}{t-z}d\sigma(t).$$

This gives an isometry of $L^2(\sigma)$ onto $K_\theta$. In the case of meromorphic $\theta(z)$, every function $f \in K_\theta$ also has a meromorphic extension in $\mathbb{C}$, and it is given by the formula (48). The corresponding Clark measure is discrete with atoms at the points of $\{\theta = 1\}$ given by

$$\sigma(\{x\}) = \frac{2\pi}{|\theta'(x)|}.$$

For more details on Clark measures We may consult [100].
Each meromorphic inner function $\theta(z)$ can be written as $\theta(t) = e^{i\phi(t)}$ on $\mathbb{R}$, where $\phi(t)$ is a real analytic and strictly increasing function. The function $\phi(t) = \arg\theta(t)$ is the continuous argument of $\theta(z)$.
Recall that the Toeplitz operator $T_U$ with a symbol $U \in L^\infty(\mathbb{R})$ is the map

$$T_U: H^2 \to H^2, \; F \mapsto P_+(UF),$$

where $P_+$is the orthogonal projection in $L^2(\mathbb{R})$ onto the Hardy space $H^2 = H^2(\mathbb{C}_+)$.
We will use the following notation for kernels of Toeplitz operators (or Toeplitz kernels in $H^2$ ):

$$N[U] = \ker T_U.$$

For example, $N[\bar{\theta}] = K_\theta$ if $\theta$ is an inner function. Along with $H^2$-kernels, one may consider Toeplitz kernels $N^p[U]$ in other Hardy classes $H^p$, the kernel $N^{1,\infty}[U]$ in the "weak" space $H^{1,\infty} = H^p \cap L^{1,\infty}, 0 < p < 1$, or the kernel in the Smirnov class $N^+(\mathbb{C}_+)$:

$$N^+[U] = \left\{f \in N^+ \cap L^1_{\text{loc}}(\mathbb{R}) : \bar{U}\bar{f} \in N^+\right\}.$$

For more on such kernels see [72], [49].
For any inner function $\theta$ in the upper half-plane we denote by $\text{spec}_\theta$ the set $\{\theta = 1\}$, the set of points on the line where the non-tangential limit of $\theta$ is equal to 1 , plus the infinite point if the corresponding Clark measure has a point mass at infinity, i.e. if $p$ in (46) is positive. If $\text{spec}_\theta \subset \mathbb{R}$ like in the next definition, then $p$ in (46) is 0. Throughout, $S$ stands for the exponential inner function $S(z) = \exp(iz)$.

**Definition (2.2.5)[84]:** If $X \subset \mathbb{R}$ is a closed set, denote

$$T_X = \sup\{a \mid N[\bar{\theta}S^a] \neq 0 \text{ for some meromorphic inner } \theta, \operatorname{spec}_\theta \subset X\}.$$

The following theorem shows the connection between the gap problem and the problem of triviality of Toeplitz kernels. This connection will be used throughout.

**Theorem (2.2.6)[84]:**

$$T_X = G_X.$$

We call a sequence of real points discrete if it has no finite accumulation points. Note that $\operatorname{spec}_\theta$ is discrete if and only if $\theta$ is meromorphic.

**Proof.** Let $T_X = d$. Then for any $\varepsilon > 0$ there exists a discrete sequence $\Lambda \subset X$ such that the kernel $N[\bar{\theta}S^a]$ is non-trivial for some/any meromorphic inner $\theta$, spec $_\theta = \Lambda$ and $a = d - \varepsilon$ (if the kernel is non-trivial for some $\theta_1$ with $a = d - \varepsilon/2$ then it is non-trivial for any $\theta_2, \operatorname{spec}_{\theta_1} = \operatorname{spec}_{\theta_2}$, with $a = d - \varepsilon$, see [72] ). Let $f \in N[\bar{\theta}S^a]$. Then $h = S^a f \in K_\theta$ and by the Clark formula

$$h = \frac{1}{2\pi i}(1 - \theta)Kh\sigma_1,$$

where $\sigma_1$ is the Clark measure corresponding to $\theta$. Notice that the function $1 - \theta$ decays at most as $y^{-1}$ along the positive $y$-axis. Hence the Cauchy integral $Kh\sigma_1$ decays as $\exp(-ay)$ as $y \to \infty$. Hence the measure $\nu = h\sigma_1$ satisfies $\operatorname{supp} \nu = \operatorname{supp} \sigma_1 = \operatorname{spec}_\theta \subset X$.

The function $\theta$ can be chosen so that $\sigma_1$ is finite: one can start by choosing any finite positive $\sigma_1$ supported by $\Lambda$ and then simply take $\theta$ corresponding to that measure. Then $\nu$ is finite as well. It remains to observe that $\nu$ has a spectral gap of the size at least $d - \varepsilon$ and $G_X \geq d$ (see, for instance, Lemma (2.2.19)).

In the opposite direction, if $G = d$ then, by Lemma (2.2.15), for any $\varepsilon > 0$ there exists a finite measure $\nu$ with a spectral gap at least $d - \varepsilon$ concentrated on a discrete subset of $X$. Let $\theta$ be the inner function corresponding to $|\nu|$. Then the function $h = (1 - \theta)K\nu$ belongs to $K_\theta$ and can be represented as $h = S^{a-\varepsilon}f$ for some $f \in H^2$. Hence $f \in N[\bar{\theta}S^{a-\varepsilon}]$ and $T_X \geq d$.

Let $\Lambda = \{\lambda_1, \dots, \lambda_n\}$ be a finite set of points on $\mathbb{R}$. Consider the quantity

$$E(\Lambda) = \sum_{\lambda_k, \lambda_j \in \Lambda} \log|\lambda_k - \lambda_l|. \tag{49}$$

According to the 2-dimensional Coulomb's law, $E(\Lambda)$ is the energy of a system of "flat" electrons placed at the points of $\Lambda$. The 2D Coulomb-gas formalism corresponds to the planar potential theory with logarithmic potential and assumes the potential energy at infinity to be equal to $-\infty$, see for instance [89], [97], [101].

Physically, the 2D Coulomb's law can be derived from the standard 3D law via a method of "reduction." According to this method, one replaces each electron in the plane with a uniformly charged string orthogonal to the plane. After that one applies the 3D law and a renormalization procedure.

Let $I \subset \mathbb{R}$ be an interval, $\Lambda = I \cap \mathbb{Z} = \{n + 1, \dots, n + k\}$. Then

$$E(\Lambda) = k^2 \log|I| + O(|I|^2)$$

as follows from Stirling's formula. Here $|I|$ stands for the length of $I$ and the notation $O(|I|^2)$ corresponds to the direction $|I| \to \infty$ (see [92]).

We call a sequence of disjoint intervals $\{I_n\}$ on the real line long (in the sense of Beurling and Malliavin) if

$$\sum_n \frac{|I_n|^2}{1+\text{dist}^2(0,I_n)} = \infty. \tag{50}$$

If the sum is finite we call $\{I_n\}$ short.

Let

$$\dots < a_{-2} < a_{-1} < a_0 = 0 < a_1 < a_2 < \cdots$$

be a two-sided sequence of real points. We say that the intervals $I_n = (a_n, a_{n+1}]$ form a short partition of $\mathbb{R}$ if $|I_n| \to \infty$ as $n \to \pm\infty$ and the sequence $\{I_n\}$ is short.

Let $\Lambda = \{\lambda_n\}$ be a sequence of real points. We write $C_\Lambda \geq a$ if there exists a short partition $\{I_n\}$ such that

$$\Delta_n \geq a|I_n| \text{ for all } n \text{ (density condition)} \tag{51}$$

and

$$\sum_n [\Delta_n^2 \log|I_n| - E_n]/(1+\text{dist}^2(0,I_n)) < \infty \ \text{ (energy condition)} \tag{52}$$

where

$$\Delta_n = \#(\Lambda \cap I_n) \ \text{ and } \ E_n = E(\Lambda \cap I_n) = \sum_{\lambda_k,\lambda_l \in I_n, \lambda_k \neq \lambda_l} \log|\lambda_k - \lambda_l|.$$

If $X$ is a closed subset of $\mathbb{R}$ we put

$$C_X = \sup\{a \mid \text{there exits a sequence } \Lambda \subset X \text{ such that } C_\Lambda \geq a\} \tag{53}$$

**Example (2.2.7)[84]:** As discussed above, if the points of the sequence are spread uniformly over the interval then $E_n = \sum_{\lambda_i,\lambda_j \in I_n} \log|\lambda_i - \lambda_j|$ is roughly (up to $O(|I_n|^2)$ which is small for short sequences of $I_n$ ) equal to $\Delta_n^2 \log|I_n|$ as follows from Stirling's formula. This happens for instance when the sequence $\Lambda$ is separated, i.e. satisfies $|\lambda_n - \lambda_{n+1}| > \delta > 0$ for all $n$. Thus for separated sequences $\Lambda$ the energy condition disappears and

$$G_\Lambda = d_i(\Lambda)$$

where $d_i, i = 1,2,3,4$ is any of the equivalent densities defined in the previous remark. This is one of the results of [96].

**Example (2.2.8)[84]:** Let $\Lambda$ be a real sequence such that the density condition (51) holds for some $a > 0$ and some partition $\{I_n\}$ that satisfies a stronger shortness condition:

$$\sum_n \frac{|I_n|^2 \log|I_n|}{1+\text{dist}^2(0,I_n)} < \infty.$$

Then we will automatically have that

$$\sum_n \frac{\Delta_n^2 \log|I_n| - \sum_{\lambda_i,\lambda_j \in I_n} \log_+|\lambda_i - \lambda_j|}{1+\text{dist}^2(0,I_n)} < \infty.$$

Hence condition (52) will be significantly simplified and one will only need to check that

$$\sum_{\lambda_i,\lambda_j \in \Lambda, \lambda_i \neq \lambda_j} \frac{\log_-|\lambda_i - \lambda_j|}{1+\lambda_j^2} < \infty$$

to conclude that $C_\Lambda \geq a$.

**Theorem (2.2.9)[84]:**

$$G_X = 2\pi C_X.$$

If $f$ is a function on $\mathbb{R}$ and $I \subset \mathbb{R}$ we denote by $f|_I$ the function that is equal to $f$ on $I$ and to 0 on $\mathbb{R} \setminus I$.

In our estimates we write $a(n) \lesssim b(n)$ if $a(n) < Cb(n)$ for some positive constant $C$, not depending on $n$, and large enough $|n|$. Similarly, we write $a(n) \asymp b(n)$ if $ca(n) < b(n) < Ca(n)$ for some $C \geq c > 0$. Some formulas will have other parameters in place of $n$ or no parameters at all.
By II we denote the Poisson measure $dx/(1+x^2)$ on the real line. In particular, $L^p_\Pi = L^p(\mathbb{R}, dx/(1+x^2))$.
We will denote by $\mathcal{D}(\mathbb{R})$ the standard Dirichlet space on $\mathbb{R}$ (in $\mathbb{C}_+$). Recall that the Hilbert space $\mathcal{D}(\mathbb{R})$ consists of functions $h \in L^1_\Pi$ such that the harmonic extension $u = u(z)$ of $h$ to $\mathbb{C}_+$has a finite gradient norm,

$$\| h \|^2_{\mathcal{D}} \equiv \| u \|^2_\nabla \overset{\text{def}}{=} \int_{\mathbb{C}_+} |\nabla u|^2 dA < \infty,$$

where $dA$ is the area measure. If $h \in \mathcal{D}(\mathbb{R})$ is a smooth function, then we also have

$$\| h \|^2_{\mathcal{D}} = \int_{\mathbb{R}} \bar{h}\tilde{h}' dx,$$

where $\tilde{h}$ denotes a harmonic conjugate function.
**Proof:** I) First suppose that $C_X > \frac{1}{2\pi}$. We will show that $G_X \geq 1$.
Choose $\varepsilon > 0$. If $C_X > \frac{1}{2\pi}$, there exists a sequence $\Lambda = \{\lambda_n\} \subset X, C_\Lambda > \frac{1}{2\pi}$. Let

$$I_n = (a_n, a_{n+1})$$

be the corresponding short monotone partition, see remark 4 . WLOG

$$\frac{1}{2\pi}|I_n| < \#(\Lambda \cap I_n) \leq \frac{1}{2\pi}|I_n| + 1$$

(otherwise just delete some of the points from $\Lambda$). We will assume that $|I_n| >> 1/\varepsilon >> 1$ for all $n$.
By Lemma (2.2.10) and Corollary (2.2.14) we can assume that the lengths of the intervals $(\lambda_n, \lambda_{n+1})$ are bounded from above. It will be convenient for us to assume that the endpoints of $I_n$ belong to $\Lambda$, i.e. that $I_n = (\lambda_{k_n}, \lambda_{k_{n+1}}]$ for some $\lambda_{k_n}, \lambda_{k_{n+1}} \in \Lambda$. We will also include the endpoints of the intervals into the energy condition by defining $E_n$ as

$$E_n = \sum_{\lambda_{k_n} \leq \lambda_k, \lambda_l \leq \lambda_{k_{n+1}}, \lambda_k \neq \lambda_l} \log|\lambda_k - \lambda_l| \tag{54}$$

and assuming that (52) is satisfied with these $E_n$. Such an assumption can be made because if the sum in (52) becomes infinite with $E_n$ defined by (54) one can, for instance, delete the first point, $\lambda_{n_k+1}$, from $\Lambda$ on all $I_n$ for large $n$. After the addition of $\lambda_{k_n}$ and deletion of $\lambda_{n_k+1}$ in the sum defining $E_n$, each term in (52) will become smaller and the sum will remain finite. At the same time, since

$$|I_n| \asymp \#\{\Lambda \cap I_n\} \to \infty,$$

the subsequence will still have more than $|I_n|$ points on each $I_n$ and will satisfy the density condition.
We show that $G_\Lambda \geq 1$ by producing a measure on $\Lambda$ with spectral gap of the size arbitrarily close to 1. Due to connections discussed, existence of such a measure will follow from non-triviality of a certain Toeplitz kernel.
Since the lengths of $(\lambda_n, \lambda_{n+1})$ are bounded from above, we can apply Lemma (2.2.17). Denote by $\theta$ the corresponding meromorphic inner function with $\text{spec}_\theta = \Lambda$.

Let $u = \arg(\theta\bar{S}) = \arg\theta - x$. First, we choose a larger partition $J_n = (b_n, b_{n+1})$ and a small"correction" function $v$ so that $u - v$ becomes an atom on each $J_n$ :

**Claim 1.** There exists a subsequence $\{b_n\}$ of the sequence $\{a_n\}$ and smooth functions $v_1, v_2$ such that:

1 $|v_1'| < \varepsilon/2$ and $u - v_1 = 0$ at all $a_n$;
2 $J_n = (b_n, b_{n+1})$ is a short monotone partition;
3 $|v_2'| < \varepsilon/2$ and $u - v = u - (v_1 + v_2) = 0$ at all $b_n$;
4 $\int_{J_n} (u - v)dx = 0$ for all $n$;
5 $\tilde{u} - \tilde{v} \in L^1_\Pi$.

**Proof of claim.** First, choose a smooth function $v_1$ satisfying 1 . Such a function exists because

$$|2\pi\Delta_n - |I_n|| \leq 2\pi << \frac{\varepsilon}{2}|I_n|.$$

Notice that because the sequence $I_n$ is short and

$$(u - v_1)' > -1 - \frac{\varepsilon}{2},$$

1 implies

$$u - v_1 \in L^1_\Pi. \tag{55}$$

Choose $b_0 = a_0 = 0$. Choose $b_1 = a_{n_1} > b_0$ to be the smallest element of $\{a_k\}$ satisfying

$$\left|\int_{b_0}^{a_{n_1}} (u - v_1)dx\right| < \frac{\varepsilon}{8}\left(a_{n_1} - b_0\right)^2.$$

Notice that because of (55) such an $a_{n_1}$ will always exist. After that proceed choosing $b_2, b_3, \ldots$ in the following way: If $b_i$ is chosen, choose $b_{i+1} = a_{n_{i+1}}$ to be the smallest element of $\{a_k\}$ satisfying $a_{n_{i+1}} > b_i$,

$$\left|\int_{b_i}^{a_{n_{i+1}}} (u - v_1)dx\right| < \frac{\varepsilon}{8}\left(a_{n_{i+1}} - b_i\right)^2 \tag{56}$$

and

$$a_{n_{i+1}} - b_i \geq b_i - b_{i-1}.$$

Choose $b_k, k < 0$ in the same way.

We claim that the resulting sequence $J_k = (b_{k-1}, b_k)$ forms a short monotone partition.

Let $k$ be positive. By our construction, $I_{n_k}$ is the last (rightmost) among the intervals $I_n$ contained in $J_k$. Notice that because of monotonicity $I_{n_k}$ is the largest interval among the intervals $I_n$ contained in $J_k$. We will show that for each $k$

$$|J_k| < \left(\left[\frac{10}{\varepsilon}\right] + 1\right)|I_{n_k}| \tag{57}$$

where [.] stands for the entire part of a real number.

This can be proved by induction. The basic step: By our construction $b_1 = a_{n_1}$ and

$$\left|\int_{b_0}^{a_{n_1-1}} (u - v_1)dx\right| \geq \frac{\varepsilon}{8}\left(a_{n_1-1} - b_0\right)^2.$$

Since $(u - v_1)' > -1 - \varepsilon$ and $u - v_1 = 0$ at all $a_n, |u - v_1| \leq (1 + \varepsilon)\left|I_{n_1-1}\right|$ on $\left(b_0, a_{n_1-1}\right)$. Hence

$$(1 + \varepsilon)\left|I_{n_1-1}\right|\left(a_{n_1-1} - b_0\right) \geq \left|\int_{b_0}^{a_{n_1-1}} (u - v_1)dx\right| \geq \frac{\varepsilon}{8}\left(a_{n_1-1} - b_0\right)^2$$

and

$$\left(a_{n_1-1}-b_0\right)\leq 8\frac{1+\varepsilon}{\varepsilon}\left|I_{m_k}\right|.$$

It follows that

$$|J_1|=\left(a_{n_1-1}-b_0\right)+\left|I_{n_1}\right|\leq 9\varepsilon^{-1}\left|I_{n_1-1}\right|+\left|I_{n_1-1}\right|\leq\frac{10}{\varepsilon}\left|I_{n_1-1}\right| \quad (58)$$

(if $\varepsilon$ is small enough). For the inductional step, assume that (57) holds for $k=l-1$. For $J_l=(b_{l-1},b_l), b_l=a_{n_l}$ there are two possibilities:

$$\left|\int_{b_{l-1}}^{a_{n_l-1}}(u-v_1)dx\right|\geq\frac{\varepsilon}{8}\left(a_{n_l-1}-b_{l-1}\right)^2$$

or

$$a_{n_l-1}-b_{l-1}<b_{l-1}-b_{l-2}.$$

In the first case we prove (58) in the same way as in the basic step. In the second case we notice that by monotonicity of $I_n$ the number of intervals $I_n$ inside $\left(b_{l-1},a_{n_l-1}\right)$ is at most $\left(a_{n_l-1}-b_{l-1}\right)/\left|I_{n_{l-1}}\right|$ which is strictly less than $|J_{l-1}|/\left|I_{n_{l-1}}\right|\leq[10/\varepsilon]+1$. Hence the number of intervals in $\left(b_{l-1},a_{n_l-1}\right)$ is at most $[10/\varepsilon]$. Therefore the number of intervals in $J_l=(b_{l-1},b_l)$ is at most $[10/\varepsilon]+1$. Now, since $I_{n_l}$ is the largest interval in $J_l$ we again get (57), which implies shortness of $J_n$. The monotonicity follows from our construction.
Now define the function $v_2$ on each $J_k$ in the following way. First consider the tent function $T_k$ defined on $\mathbb{R}$ as

$$T_k(x)=\frac{\varepsilon}{4}\operatorname{dist}(x,\mathbb{R}\setminus J_k).$$

Notice that because of (56), for each $k$ there exists a constant $C_k, |C_k|\leq 1$ such that

$$\int_{J_k}[(u-v_1)-C_kT_k]dx=0.$$

Now define $v_2$ as a smoothed-out sum $\sum C_kT_k$ that satisfies $|v_2'|<\varepsilon/2$ and still has the properties that $v_2(b_k)=0$ and

$$\int_{J_k}[(u-v_1)-v_2]dx=0$$

for each $k$. Finally, let $v=v_1+v_2$. The last condition of the claim will be satisfied because the restrictions $(u-v)|_{J_k}$ form a collection of atoms with a finite sum of $L^1_\Pi$-norms:

$$\left\|(u-v)|_{J_k}\right\|_{L^1_\Pi}\lesssim\frac{|J_k|^2}{1+\operatorname{dist}^2(0,J_k)}$$

(for more on atomic decompositions see [54]).
The function $v$ from the last claim is a smooth function satisfying $|v'|\leq\varepsilon$. Therefore it can be represented as $v=v_+-v_-$where $v_\pm$are smooth growing functions, $0\leq v_\pm'\leq\varepsilon$. Hence one can choose two meromorphic inner functions $I_\pm$ satisfying

$$\{\arg I_\pm=k\pi\}=\{\arg v_\pm=k\pi\}$$

and

$$\left|I_\pm'\right|\lesssim\varepsilon$$

(the existence of such $I_\pm$follows, for instance, from Lemma (2.2.17) or from the lemma in section 6.1 of [49] ).
Note that then, automatically, $|\arg(\bar{I}_+I_-)-v|<2\pi$. The harmonic conjugate of $\arg(\theta\bar{S}I_+\bar{I}_-)$still belongs to $L^1_\Pi$-.

WLOG$\arg(\theta\bar{S}I_+\bar{I}_-)=0$ at 0.
**Claim 2.** The function arg $(\theta(x)\bar{S}I_+(x)\bar{I}_-(x))/x$ belongs to the Dirichlet class $\mathcal{D}(\mathbb{R})$.
**Proof of claim.** We will actually prove that $w/x, w=\arg\theta-x-v\in\mathcal{D}$ instead (again, WLOG $w(0)=0$ with large multiplicity). The difference between $-v$ and $\arg(I_+\bar{I}_-)$is a bounded function with bounded derivative that obviously belongs to $\mathcal{D}$.
Let $q(z)$ be the harmonic extension of $w/x$ to the upper half plane. We need to show that the gradient norm of $q+i\tilde{q}$ in $\mathbb{C}_+$is finite, i.e. that

$$\| q+i\tilde{q} \|_\nabla^2=\lim_{r\to\infty}\int_{\partial D(r)} qd\tilde{q}<\infty,$$

where $D(r)$ is the semidisc $\{|z|<r\}\cap\mathbb{C}_+$.
We first prove that the integrals over $\partial D(r)\cap\mathbb{R}$ are uniformly bounded from above, i.e. that

$$\int_{\mathbb{R}} qd\tilde{q}<\infty.$$

First, notice that the harmonic conjugate of $\frac{w}{x}$ is $\frac{\tilde{w}}{x}$ and $\left(\frac{w}{x}\right)'=\frac{w'}{x}-\frac{w}{x^2}$, where $\frac{w}{x^2}$ is a bounded function since $w'$ is bounded and has a zero at zero. Hence

$$\int_{\mathbb{R}} qd\tilde{q}\asymp\int_{\mathbb{R}} w'\tilde{w}\frac{dx}{x^2}$$

and we can estimate the last integral instead.
If $I$ is an interval then $2I$ denotes the interval with the same center as $I$ satisfying $|2I|=2|I|$.
Put $w_n=w|_{J_n}$. Then

$$\int_{\mathbb{R}} w'\tilde{w}\frac{dx}{x^2}=\sum_n\sum_k\int_{J_n} w'\tilde{w}_k\frac{dx}{x^2}.$$

To estimate the last integral, first let us consider the case when the intervals $J_n$ and $J_k$ are far from each other:

$$\max\ (|J_n|,|J_k|)\le\operatorname{dist}(J_n,J_k).$$

In this case

$$\int_{J_n} w'\tilde{w}_k\frac{dx}{x^2}\lesssim\int_{J_n}|w'|\frac{|J_k|^3}{\operatorname{dist}^2(J_k,x)}\frac{dx}{x^2}\lesssim \frac{|J_k|^3}{1+\operatorname{dist}^2(J_n,0)}\int_{J_n}\frac{dx}{\operatorname{dist}^2(J_k,x)}. \tag{59}$$

Here we used the property that each $w_k$ is an atom supported on $J_k$ whose $L^1$-norm is $\lesssim |J_k|^2$ and employed the standard estimates from the theory of atomic decompositions, see [54]. In the last inequality we used the property

$$\int_{J_n}|w'(x)|dx\lesssim|J_n|. \tag{60}$$

Now let us consider the "mid-range" case when

$$\min\ (|J_n|,|J_k|)\le\operatorname{dist}(J_n,J_k)<\max\ (|J_n|,|J_k|).$$

Assume that $0<k<n$. Then by monotonicity $|J_k|\le|J_n|$. In this case

$$\int_{J_n} w'\tilde{w}_k \frac{dx}{x^2} \lesssim \frac{|J_k|^2}{1+\mathrm{dist}^2(J_k,0)} \int_{J_n} |w'| \frac{|J_k|}{\mathrm{dist}^2(J_k,x)} dx$$

$$\le \frac{|J_k|^2}{1+\mathrm{dist}^2(J_k,0)} \frac{|\cdot J_k|}{\mathrm{dist}^2(J_k,J_n)} \int_{J_n} |w'| \lesssim \frac{|J_k||J_n|}{1+\mathrm{dist}^2(J_k,0)}. \quad (61)$$

Finally, the last case is

$$dist(J_n,J_k) < \min\ (|J_n|,|J_k|).$$

Again we assume that $n,k > 0$. Then by monotonicity either $n = k$ or $|n-k| = 1$, i.e. the intervals are either the same or adjacent. The estimates in this case are more complicated and will be done differently. First, integrating by parts we get

$$\int_{J_n} w'\tilde{w}_k \frac{dx}{x^2} = \int_{J_n} w' \left[\int_{J_k} \frac{w(t)dt}{t-x}\right] \frac{dx}{x^2} =$$

$$-\int_{J_n} w' \left[\int_{J_k} \log|t-x|\, w'(t)dt\right] \frac{dx}{x^2}.$$

Next we would like to conclude that, for $k \le n$,

$$-\int_{J_n} w' \left[\int_{J_k} \log|t-x| w'(t)dt\right] \frac{dx}{x^2} \lesssim$$

$$-\frac{1}{1+\mathrm{dist}^2(J_n,0)} \left[\iint_{J_n\times J_k} \log|t-x|\, w'(x)w'(t)dxdt + |J_n|^2\right] (62)$$

and work with the latter integral instead of the former. Since $\mathrm{dist}(0,J_n) < |x|$ for $x \in |J_n|$, this estimate would be obvious if the function under the integral were negative. In our case, however, it will require some work.

To prove the last relation denote

$$J_n^+ = \{w' > 0\},\ J_n^- = \{w' \le 0\}.$$

Since $-\int_{J_k} \log|t-x|w'(t)dt = \tilde{w}_k(x)$ and $w_k$ is an atom,

$$\left\| \int_{J_k} \log|t-x|w'(t)dt \right\|_{L^1_\Pi} \lesssim \frac{|J_k|^2}{1+\mathrm{dist}^2(J_k,0)}.$$

Since $\{J_n\}$ is a short sequence, $|J_n| \le \mathrm{dist}(J_n,0)$ for large enough $n$. For the rest of the proof we will assume that this holds for all $n \ne 0,-1$. Since $w'$ is bounded from below and

$$\mathrm{dist}^2(J_n,0) \le x^2 \le 2\mathrm{dist}^2(J_n,0)$$

on $J_n$,

$$-\int_{J_n^-} w' \left[\int_{J_k} \log|t-x|w'(t)dt\right] \frac{dx}{x^2} \lesssim$$

$$-\int_{J_n^-} w' \left[\int_{J_k} \log|t-x|\, w'(t)dt\right] \frac{dx}{1+\mathrm{dist}^2(J_n,0)} + \frac{|J_k|^2}{1+\mathrm{dist}^2(J_k,0)}. (63)$$

To deal with the integral over $J_n^+$notice, that $\tilde{w}_k$ is "almost" positive on $J_n$, i.e.

$$-\int_{J_k} \log|t-x|w'(t)dt \gtrsim -|J_n|$$

for any $x \in J_n$. Indeed

$$-\int_{J_k} \log|t-x|\, w'(t)dt = \int_{J_k} \log_-|t-x|\, w'(t)dt - \int_{J_k} \log_+|t-x|\, w'(t)dt,$$

where

$$\int_{J_k} \log_- |t-x| w'(t)dt = \int_{J_k} \log_- |t-x| (\arg\theta)'(t)dt - \int_{J_k} \log_- |t-x|(x+v)'(t)dt \geq$$
$$\int_{J_k} (\arg\theta)'(t)dt - \int_{J_k} (1+\varepsilon)dt \geq -\varepsilon|J_k| + \text{ const.}$$
Here we used the property that $\int_{J_k} (\arg\theta)'(t)dt = |J_k| + \text{const}$. Also
$$-\int_{J_k} \log_+ |t-x| w'(t)dt$$
$$= \int_{J_k} \log_+ |t-x|(x+v)'(t)dt - \int_{J_k} \log_+ |t-x| (\arg\theta)'(t)\, dt.$$
Recall that $J_k = (b_k, b_{k+1}]$ and $x \in J_n, n \geq k$. If $n > k$, sing Lemma (2.2.18) part 6 we obtain
$$-\int_{J_k} \log_+ |t-x| w'(t)dt$$
$$\geq (\log_+ |b_k - x| - C) \int_{J_k} (1+v')dx - \log_+|b_k - x| \int_{J_k} (\arg\theta)' dx$$
$$\geq -|J_k|,$$
which establishes (64). For $n = k$ the same relation can be obtained using Lemma (2.2.18) part 5 .

To finish the proof of (62) notice that
$$-\int_{J_n} w' \left[ \int_{J_k} \log|t-x| w'(t)dt \right] \frac{dx}{x^2} =$$
$$\int_{J_n^+} w' \left[ -\int_{J_k} \log|t-x|w'(t)dt \right] \frac{dx}{x^2} + \int_{J_n^-} w' \left[ -\int_{J_k} \log|t-x|w'(t)dt \right] \frac{dx}{x^2} =$$
$$\int_{J_n^+} w'(x)\,\mathrm{m} \left( \left[ -\int_{J_k} \log|t-x|w'(t)dt \right], 0 \right) \frac{dx}{x^2} +$$
$$\int_{J_n^+} w'(x)\,\mathrm{m} \left( \left[ -\int_{J_k} \log|t-x|w'(t)dt \right], 0 \right) \frac{dx}{x^2} +$$
$$\int_{J_n^-} w' \left[ -\int_{J_k} \log|t-x|w'(t)dt \right] \frac{dx}{x^2}.$$
Regarding the last three integrals, if one replaces $\frac{dx}{x^2}$ with $\frac{dx}{1+\text{dist}^2(J_n,0)}$ in the first integral, it will get $\gtrsim$ than before because the function under the integral is positive and $x^2 \gtrsim 1 + \text{dist}^2(J_n, 0)$. Under the same operation, the second integral will decrease at most by
$$C|J_k| \int_{J_n} |w'| \frac{dx}{x^2} \lesssim \frac{|J_n||J_k|}{1+\text{dist}^2(J_n,0)}$$
because of (64). Finally, the last integral is estimated in (63). Since $\text{dist}(0, J_k) \asymp \text{dist}(0, J_n)$ and $|J_k| \leq |J_n|$, this finishes (62).

To estimate the integral in the right-hand side of (62), denote $p = \arg\theta - x - v_1 = w + v_2$ where the functions $v_1, v_2$ are from claim 1. Also denote $p_n = p|_{J_n}$ and $v_2^n = v_2|_{J_n}$. The key properties of $v_1$ that we will use are that $\arg\,\theta - x - v_1 = 0$ at the endpoints of all $I_n, v_2 = 0$ at the endpoints of $J_n$ and that $|v_1'|, |v_2'| < \varepsilon$. Then

$$-\iint_{J_n\times J_k} \log|t-x|w'(x)w'(t)dxdt = -\iint_{J_n\times J_k} \log|t-x|p'(x)p'(t)dxdt -$$

$$+\int_{J_n} (\tilde{p}_n v_2' + \tilde{v}_2^n p' + \tilde{v}_2^n v_2')dx.$$

Notice that

$$\left|\int_{J_n} \tilde{p}_n v_2' dx\right| \le \varepsilon \|\tilde{p}_n\|_1 \le \varepsilon \parallel p_n||_2\sqrt{|J_n|} \lesssim \varepsilon |J_n|^2$$

because $|p_n| \lesssim |J_n|$. Also,

$$\left|\int_{J_n} p'\tilde{v}_2^n dx\right| = |< p, v_2^n >_{\mathcal{D}}| = \left|\int_{J_n} \tilde{p}_n v_2' dx\right| \lesssim |J_n|^2.$$

Similarly to the first integral,

$$\int_{J_n} \tilde{v}_2^n v_2' dx \lesssim \varepsilon^2 |J_n|^2.$$

Hence

$$-\iint_{J_n\times J_k} \log|t-x|\, w'(x)w'(t)dxdt$$

$$= -\iint_{J_n\times J_k} \log|t-x|\, p'(x)p'(t)dxdt + O(|J_n|^2). \tag{65}$$

For the last integral we have

$$-\iint_{J_n\times J_k} \log|t-x|\, p'(x)p'(t)dxdt$$

$$= -\sum_{I_i\subset J_k}\sum_{I_j\subset J_n}\iint_{I_i\times I_j} \log|t-x|\, p'(x)p'(t)dxdt. \tag{66}$$

To estimate

$$-\iint_{I_i\times I_j} \log|t-x|p'(x)p'(t) =$$

$$\iint_{I_i\times I_j} \log_-|t-x|p'(x)p'(t) - \iint_{I_i\times I_j} \log_+|t-x|p'(x)p'(t) \tag{67}$$

we consider 3 cases. First, to estimate the integral in the case when $i = j$, notice that, since $1 + v_1'$ is bounded,

$$\int_{I_j} \log_-|x-t|(1+v_1'(x))dx < \text{ const}$$

for any $t \in I_j$. Once again, the positive functions $\arg'\theta$ and $v_1' + 1$ satisfy

$$\int_{I_l} \arg'\theta = \int_{I_l} (v_1'+1) = 2\pi\Delta_l + O(1) = |I_l| + O(1). \tag{68}$$

Hence

$$\iint_{I_j\times I_j} \log_- |t-x| p'(x)p'(t) =$$

$$\iint_{I_j\times I_j} \log_- |t-x| (\arg\theta)'(x)(\arg\theta)'(t) - 2\iint_{I_j\times I_j} \log_- |t-x|(1+v_1'(x))(\arg\theta)'(t) +$$

$$\iint_{I_j\times I_j} \log_- |t-x|(1+v_1'(x))(1+v_1'(t)) = \iint_{I_j\times I_j} \log_- |t-x|(\arg\theta)'(x)(\arg\theta)'(t) + O(|I$$

For the last integral we have

$$\iint_{I_j\times I_j} \log_- |t-x|(\arg\theta)'(x)(\arg\theta)'(t) =$$

$$\sum_{\lambda_l,\lambda_{l+1}\subset I_j} \sum_{\lambda_m,\lambda_{m+1}\subset I_j} \int_{\lambda_l}^{\lambda_{l+1}} \int_{\lambda_m}^{\lambda_{m+1}} (\arg\theta)'(x)(\arg\theta)'(t).$$

Using the properties that

$$\int_{\lambda_s}^{\lambda_{s+1}} \arg\theta' = 2\pi$$

and

$$\arg\theta' \lesssim [\min\ (|I_{s-1}|,|I_s|,|I_{s+1}|)]^{-2}\ \text{ on }\ (\lambda_s,\lambda_{s+1}),$$

for all $s$, we can apply Lemma (2.2.18) , parts 1-3. Assuming that $\lambda_l \le \lambda_m$ we conclude that

$$\int_{\lambda_l}^{\lambda_{l+1}} \int_{\lambda_m}^{\lambda_{m+1}} \log_- |t-x|(\arg\theta)'(x)(\arg\theta)'(t) \lesssim$$

$$\begin{cases} \log_-(\lambda_m-\lambda_{l+1}) & \text{if } \lambda_m > \lambda_{l+1} \\ \max\ \left(\log_-(\lambda_{l-1}-\lambda_l),\log_-(\lambda_l-\lambda_{l+1}),\log_-(\lambda_{l+1}-\lambda_{l+2})\right)+1 & \text{if } \lambda_m = \lambda_{l+1} \\ \max\ \left(\log_-(\lambda_{l-1}-\lambda_l),\log_-(\lambda_l-\lambda_{l+1}),\log_-(\lambda_{l+1}-\lambda_{l+2})\right)+1 & \text{if } \lambda_m = \lambda_l, \end{cases}$$

which implies

$$\iint_{I_j\times I_j} \log_- |t-x| p'(x)p'(t)dxdt \lesssim$$

$$\sum_{\lambda_{k_n}\le\lambda_k,\lambda_l\le\lambda_{k_{n+1}},\lambda_k\ne\lambda_l} \log_-|\lambda_k-\lambda_l| + |I_j|. \tag{69}$$

To estimate the integral of $\log_+$, first notice that by Lemma (2.2.18) , part 5 , and (68),

$$\int_{I_j} \log_+ |x-t|(1+v_1'(x))dx = |I_j|\log_+|I_j| + O(|I_j|)$$

for any $t\in I_j$.

Together with part 4 of Lemma (2.2.18) and (68) we get:

$$\iint_{I_j\times I_j} \log_+ |t-x|p'(x)p'(t) = \iint_{I_j\times I_j} \log_+ |t-x|\arg'\theta(x)\arg'\theta(t) -$$

$$2\iint_{I_j\times I_j} \log_+ |t-x|(v_1'(x)+1)\arg'\theta(t) + \iint_{I_j\times I_j} \log_+ |t-x|(v_1'(x)+1)(v_1'(t)+1) =$$

$$\sum_{\lambda_l,\lambda_m\in I_j}\int_{\lambda_l}^{\lambda_{l+1}}\int_{\lambda_m}^{\lambda_{m+1}}\log_+|t-x|\arg'\theta(x)\arg'\theta(t)-\left|I_j\right|^2\log\left|I_j\right|+$$
$$4\pi^2\sum_{\lambda_l,\lambda_m\subset I_j}\log_+|\lambda_l-\lambda_m|-\left|I_j\right|^2\log\left|I_j\right|+O\left(\left|I_j\right|^2\right) \qquad (70)$$

Next, let us consider the case when $i\neq j$ and the intervals $I_i, I_j$ are not adjacent. This estimate is similar to (61), but we will do it using a different technique. Assume for instance that $j>i+1$. For $\log_-$, recalling that $|I_k|>1$ for all $k$, we get

$$-\iint_{I_i\times I_j}\log_-|t-x|p'(x)p'(t)dxdt=0. \qquad (71)$$

For $\log_+$ we have

$$-\iint_{I_i\times I_j}\log_+|t-x|p'(x)p'(t)=-\int_{a_i}^{a_{i+1}}\int_{a_j}^{a_{j+1}}\log_+|t-x|p'(x)p'(t)=$$
$$-\int_{a_i}^{a_{i+1}}\int_{a_j}^{a_{j+1}}\log_+|t-x|(\arg\theta-x-v_1)'(x)(\arg\theta-x-v_1)'(t)\leq$$
$$-\int_{I_j}\left(\log|a_{i+1}-t|\int_{I_i}\arg'\theta(x)dx-\log|a_i-t|\int_{I_i}(v_1'+1)(x)dx\right)\arg'\theta(t)dt+$$
$$\int_{I_j}\left(\log|a_i-t|\int_{I_i}\arg'\theta(x)dx-\log|a_{i+1}-t|\int_{I_i}(v_1'+1)(x)dx\right)(v_1'+1)(t)dt\leq$$
$$2\left|I_i\parallel I_j\right| \qquad (72)$$

Here we used the properties that $\operatorname{dist}(I_i,I_j)\geq|I_i|$ by monotonicity and (68).
In the case when $I_i$ and $I_j$ are adjacent, i.e. $j=i+1$, the estimate can be done differently. Note that $p_n=p_n|_{I_n}$ is a compactly supported function with bounded derivative (the bound depends on $n$). Therefore it belongs to the Dirichlet space $\mathcal{D}$. The estimates (69) and (70) yield

$$\|p_n\|_{\mathcal{D}}^2\lesssim\frac{1}{4\pi^2}|I_n|^2\log|I_n|-E_n+|I_n|^2.$$

Hence

$$\iint_{I_i\times I_{i+1}}\log|t-x|p'(x)p'(t)=<p_i,p_{i+1}>_{\mathcal{D}}\leq||p_i||^2+||p_{i+1}||^2\lesssim$$
$$\left(\frac{1}{4\pi^2}|I_i|^2\log|I_i|-E_i\right)+\left(\frac{1}{4\pi^2}|I_{i+1}|^2\log|I_{i+1}|-E_{i+1}\right)+$$
$$|I_i|^2+|I_{i+1}|^2 \qquad (73)$$

Now we can return to estimating

$$\int_{J_n}w'\tilde{w}_k\frac{dx}{x^2}$$

in the case when $|k-n|\leq 1$. Using the estimates (62), (65) and (66) we obtain

$$\int_{J_n}w'\tilde{w}_k\frac{dx}{x^2}=\frac{-\sum_{I_i\subset J_k}\sum_{I_j\subset J_n}\iint_{I_i\times I_j}\log|t-x|p'(x)p'(t)dxdt+O(|J_n|^2)}{1+\operatorname{dist}^2(0,J_n)}.$$

The estimates (69)-(73) yield

$$-\sum_{I_i\subset J_n}\sum_{I_j\subset J_k}\iint_{I_i\times I_j}\log|t-x|p'(x)p'(t)dxdt\lesssim$$
$$\sum_{I_i\subset J_k\cup J_n}\left(\frac{1}{4\pi^2}|I_i|^2\log|I_i|-E_i+|I_i|^2\right)+\sum_{I_i,I_j\subset J_n}|I_i||I_j|\le$$
$$\sum_{I_i\subset J_k\cup J_n}\left(\frac{1}{4\pi^2}|I_i|^2\log|I_i|-E_i\right)+|J_n|^2+|J_k|^2.$$
All in all, in the case $|n-k|\le 1$, we have
$$\int_{J_n}w'\tilde{w}_k\frac{dx}{x^2}\lesssim$$
$$\frac{1}{1+\mathrm{dist}^2(0,J_n)}\left[\sum_{I_i\subset J_k\cup J_n}\left(\frac{1}{4\pi^2}|I_i|^2\log|I_i|-E_i\right)+|J_n|^2+|J_k|^2\right].\quad(6.210$$
Combining the estimates (69),(70),(71),(72),(73) with (66),(65) and (62) we get
$$\sum_n\sum_k\int_{J_n}w'\tilde{w}_k\frac{dx}{x^2}=$$
$$\sum_n\left[\sum_{k:\max\,(|J_k|,|J_n|\le\mathrm{dist}(J_k,J_n))}\right]+$$
$$\sum_n\left[\sum_{k:\min\,(|J_k|,|J_n|)\le\mathrm{dist}(J_k,J_n)<\max\,(|J_k|,|J_n|)}\right]+$$
$$\sum_n\left[\sum_{k:\mathrm{dist}(J_k,J_n)<\min\,(|J_k|,|J_n|)}\right]$$
$$I+II+III.$$
For the first sum, by (59), we get
$$I\quad\le\sum_k\left[\sum_{n:n>k,\mathrm{dist}(J_k,J_n)>|J_k|}\frac{|J_k|^3}{1+\mathrm{dist}^2(J_n,0)}\int_{J_n}\frac{dx}{\mathrm{dist}^2(J_k,x)}\right]\le$$
$$\sum_k\frac{|J_k|^3}{1+\mathrm{dist}^2(J_k,0)}\frac{1}{|J_k|}=\sum_k\frac{|J_k|^2}{1+\mathrm{dist}^2(J_k,0)}<\infty.\quad(75)$$
For the second sum, by (61),
$$II\lesssim\sum_n\left[\sum_{k:k\ne n,\mathrm{dist}(J_k,J_n)<\max\,(|J_k|,|J_n|)}\frac{|J_k||J_n|}{1+\mathrm{dist}^2(J_n,0)}\right].$$
Recall that by our assumption $|J_n|\le\mathrm{dist}(J_n,0)$ for all $n\ne 0,-1$. We can also assume that $|J_{-1}|=|J_0|$. Then in each term in the last sum $k$ and $n$ have the same sign. Let us estimate the part of the sum with non-negative $k,n$.

$$\sum_{n\geq 0}\left[\sum_{k\geq 0}\left[\sum_{k\neq n,\operatorname{dist}(J_k,J_n)<\max(|J_k|,|J_n|)}\frac{|J_k||J_n|}{1+\operatorname{dist}^2(J_n,0)}\right]=\right.$$

$$2\sum_{n\geq 0}\left[\sum_{k\geq 0:k<n,\operatorname{dist}(J_k,J_n)<|J_n|}\frac{|J_k||J_n|}{1+\operatorname{dist}^2(J_n,0)}\right]\leq 4\sum_{n\geq 0}\frac{|J_n|^2}{1+\operatorname{dist}^2(J_n,0)}<\infty.$$

The negative terms can be estimated similarly to conclude that

$$II\lesssim\sum_n\frac{|J_n|^2}{1+\operatorname{dist}^2(J_n,0)}<\infty.$$

Finally, for the third sum by (74),

$$III\lesssim\sum_n\left[\sum_{k:|k-n|\leq 1}\frac{1}{1+\operatorname{dist}^2(0,J_n)}\left[\sum_{I_i\subset J_k\cup J_n}\left(\frac{1}{4\pi^2}|I_i|^2\log|I_i|-E_i\right)+|J_n|^2+|J_k|^2\right]\right.$$

$$\sum_n\frac{1}{1+\operatorname{dist}^2(0,J_n)}\left[\sum_{I_i\subset J_n}\left(\frac{1}{4\pi^2}|I_i|^2\log|I_i|-E_i\right)+|J_n|^2\right]<\infty$$

Because $\Lambda$ satisfies the energy condition on $I_n$. Altogether these estimates give us

$$\int_{\mathbb{R}}\tilde{w}w'\frac{dx}{x^2}<\infty.$$

The integrals over the circular part of $\partial D(r)$ can be estimated like in [49]. We need to show that the integrals

$$\int_{\partial D(r)\setminus\mathbb{R}}q\,d\tilde{q}=rI'(r),\ I(r):=\frac{1}{2}\int_0^\pi q^2\left(re^{i\phi}\right)d\phi,$$

do not tend to $+\infty$ as $r\to\infty$. In fact, it is enough to show

$$I(r)\nrightarrow\infty,$$

because if $rI'(r)\to+\infty$, then $I'(r)\geq 1/r$ for all $r\gg 1$, and we have $I(r)\to\infty$.
As we will see shortly, $I(r)\nrightarrow\infty$ for any $q\in L^1(1+|x|^{-1})$. It will be more convenient for us to prove an equivalent statement in the unit disk $\mathbb{D}$.
Let $q+i\tilde{q}$ be an analytic function in $\mathbb{D}$ such that

$$\frac{q(\zeta)}{1-|\zeta|}\in L^1(\partial\mathbb{D}).$$

Define

$$h(z)=\frac{1+z}{1-z}\left(q(z)+i\tilde{q}(z)\right),\ z\in\mathbb{D},$$

and denote by $h^M(\zeta),\zeta\in\partial\mathbb{D}$, the angular maximal function. Then $\Im h\in L^1(\partial\mathbb{D})$ and by the Hardy-Littlewood maximal theorem,

$$h^M\in L^{1,\infty}(\partial\mathbb{D}). \tag{76}$$

Let us show that as $\epsilon\to 0$,

$$\frac{1}{\epsilon}\int_{C_\epsilon}|q+i\tilde{q}|^2|dz|\to\infty,\ C_\epsilon:=\{|1-z|=\epsilon\}\cap\mathbb{D}.$$

We have

$$\frac{1}{\epsilon}\int_{C_\epsilon}\left|q+i\tilde{q}\right|^2\leq\epsilon\int_{C_\epsilon}\left|h\right|^2\lesssim[\epsilon h^M(\zeta)]^2+[\epsilon h^M(\bar{\zeta})]^2,$$

where $\zeta \in \partial\mathbb{D}, |1-\zeta| = \epsilon$. The right-hand side cannot tend to infinity because otherwise, for all small $\epsilon$, we would have

$$h^M(\zeta) + h^M(\bar{\zeta}) \gg \frac{1}{\epsilon}$$

on an interval of length $\epsilon$, which would contradict (76).

Let

$$\phi = \arg(\theta \bar{S} I_+ \bar{I}_-)/2.$$

Recall that $\tilde{\phi} \in L^1_\Pi$. By the last claim $\phi/x$ belongs to the Dirichlet class. Hence, by the Beurling-Malliavin multiplier theorem, see for instance [49], there exists a smooth function $m$ on $\mathbb{R}$ satisfying

$$m' < \varepsilon, \tilde{m} \in L^1_\Pi \text{ and } \tilde{m} \geq \max\ (0, -\tilde{\phi}).$$

In other words, if $\Phi$ and $M$ are outer functions,

$$\Phi = \exp(i\phi - \tilde{\phi}), M = \exp(im - \tilde{m}),$$

then $\Phi M$ is bounded in $\mathbb{C}_+$.

Since $m' < \varepsilon, \varepsilon x - m$ is an increasing function. There exists a meromorphic inner function $J$ such that

$$\{J = \pm 1\} = \{2(\varepsilon x - m) = k\pi\}.$$

Denote

$$d_1 = 2(\varepsilon x - m) \text{ and } d_2 = \arg J.$$

Then the difference

$$d = 2(\varepsilon x - m) - \arg J = d_1 - d_2$$

satisfies $|d| < \pi$.

Put

$$l(x) = \frac{2\varepsilon x - \arg J}{2}.$$

Notice that $\tilde{l} \in L^1_\Pi$ because $2l = d + 2m$ where $d$ is bounded and $\tilde{m} \in L^1_\Pi$.

Consider an outer function $\Psi = \exp(il - \tilde{l})$. Then

$$\bar{S}^{2\varepsilon}\Psi = \bar{J}\bar{\Psi}$$

or equivalently

$$\bar{S}^{2\varepsilon} J \Psi = \bar{\Psi}$$

on $\mathbb{R}$. Thus $\Psi \in N^+[\bar{S}^{2\varepsilon} J]$.

Moreover, the ratio $\Psi/M$ is equal to $\exp\left(i\frac{d}{2} - \frac{\tilde{d}}{2}\right)$. Since $|d| < \pi, \Psi/M$ belongs to any $L^p_\Pi, p < 1$. Our next goal is to construct another "small" outer multiplier function $k$ so that $k\Psi/M \in L^2_{\Pi^*}$.

Consider the step function

$$\alpha(x) = \frac{\pi}{5}\left[\frac{5}{\pi} d_1\right] - \frac{\pi}{5}\left[\frac{5}{\pi} d_2\right],$$

where $[\cdot]$ again denotes the entire part of a real number. Then

$$|d - \alpha| < \frac{2\pi}{5}. \tag{77}$$

Since $d_1 = d_2 = \pi n$ at the points $\{c_n\} = \{J = \pm 1\}$, the function $\alpha$ only takes values $k\frac{\pi}{5}, k = -4, -3, \ldots, 4$. Therefore $\alpha$ can be represented as

$$\alpha = \frac{\pi}{5}\left(\sum_{n=1}^{4} \beta_n - \sum_{n=5}^{8} \beta_n\right),$$

where $\beta_n$ are elementary step functions, each taking only two values, 0 and 1 , and making at most one positive and one negative jump on each interval $[c_n, c_{n+1}]$. For each $n = 1,2, \dots ,8$ one can choose an inner function $Q_n$ so that

$$\frac{1-Q_n}{1+Q_n} = \text{ const } e^{\pi K \beta_n}.$$

Notice that then

$$\exp(-i\pi\alpha + \pi\bar{\alpha}) = \text{const} \sqrt[n]{\prod_{n=1}^{4} \frac{1+Q_n}{1-Q_n} \prod_{n=5}^{8} \frac{1-Q_n}{1+Q_n}}$$

Because of (77) we have

$$\left| \prod_{n=1}^{4} (1+Q_n) \prod_{n=5}^{s} (1-Q_n)\Psi/M \right| \lesssim \left| \sqrt[10]{\prod_{n=1}^{4} \frac{1+Q_n}{1-Q_n} \prod_{n=5}^{8} \frac{1-Q_n}{1+Q_n} \Psi/M} \right|$$

$$= \text{ const } \exp\left[\frac{d}{2} - \frac{\bar{\alpha}}{2}\right] \in L^2_\Pi(\mathbb{R})$$

and since the function $M\Phi$ is bounded,

$$\prod_{n=1}^{4} (1+Q_n) \prod_{n=5}^{8} (1-Q_n)\Psi\Phi = \prod_{n=1}^{4} (1+Q_n) \prod_{n=5}^{8} (1-Q_n) \frac{\Psi}{M} M\Phi \in L^2_\Pi(\mathbb{R}).$$

Now notice that since $N^+[\bar{S}^{2\varepsilon} J] \neq 0$, the set $\{c_n\} = \{J = \pm 1\}$ has BeurlingMalliavin density at most $2\varepsilon$, see section 7 or [72]. By our construction the BeurlingMalliavin density of each of the sets $\{Q_n = \pm 1\}$ is the same as that of $\{c_n\}$, i.e. at most $2\varepsilon$. Hence the kernel $N^\infty[\bar{S}^{17\varepsilon} \prod_n Q_n]$ contains a non-zero function $\tau$, see section 7 or [72].
Similarly, since the Beurling-Malliavin density of $\{I_+ = 1\}$ is less than $\varepsilon$, the kernel $N^\infty[\bar{S}^\varepsilon I_+]$is infinite-dimensional. Hence it contains a non-trivial function $\eta$ with at least one zero $a$ in $\mathbb{C}_+$. Then the function $\kappa = \eta/(z-a)$ also belongs to $N^\infty[\bar{S}^\varepsilon I_+]$and satisfies $|\kappa| \lesssim (1+|x|)^{-1}$ on $\mathbb{R}$.
Therefore

$$\bar{\theta} S^{1-20\varepsilon} \kappa\tau \prod_{n=1}^{4} (1+Q_n) \prod_{n=5}^{8} (1-Q_n)\Psi\Phi =$$

$$(\bar{S}^\varepsilon I_+ \kappa) \left( \bar{S}^{17\varepsilon} \prod_{n=1}^{4} (1+Q_n) \prod_{n=5}^{8} (1-Q_n)\tau \right) (\bar{S}^{2\varepsilon}\Psi)(\bar{\theta} S^1 \bar{I}_+ \Phi) \in \bar{H}^2.$$

Hence the space $K_\theta$ contains the function

$$f = S^{1-20\varepsilon} \kappa\tau \prod_{n=1}^{4} (1+Q_n) \prod_{n=5}^{8} (1-Q_n)\Psi\Phi.$$

Now we could simply refer to Theorem (2.2.6) to conclude this part of the proof.
By the Clark representation formula

$$f = (1+\theta) K f \sigma_1$$

where $\sigma_1$ is the Clark measure corresponding to $\theta$ concentrated on $\{\theta = 1\} = 1$. Since $1 + \theta$ is bounded in the upper half-plane and $f$ decreases faster than $\exp[-(1-21\varepsilon)y]$ along the positive $y$-axis, so does $Kf\mu$. Hence $f\mu$ is the measure concentrated on $\Lambda$ with the spectral gap at least $(1-21\varepsilon)$.

II) Now suppose that $G_X > 1$ but $C_X < \frac{1}{2\pi}$.
By Corollary (2.2.21) there exists a discrete increasing sequence $\Lambda = \{\lambda_n\}_{n\in\mathbb{Z}} \subset X$ and a measure $\nu$, $\operatorname{supp}\nu = \Lambda$ such that $\nu$ has a spectral gap of the size 1 and $K\nu$ does not have any zeros in $\mathbb{C}$.
Similarly to the previous part, we assume that $\sup_n(\lambda_n - \lambda_{n-1}) < \infty$. The general case is discussed at the end of the proof. If $\sup_n(\lambda_n - \lambda_{n-1}) < \infty$, we can apply Lemma (2.2.17) and consider the inner function $\theta$ corresponding to $\Lambda$. A function $f \in N[\phi]$ is called purely outer if $f$ is outer in the upper half-plane and $\phi f = \bar{g}$ is outer in the lower half plane. Since $K\nu$ is divisible by $S$, the function

$$f = S^{-1}(1-\theta)K\nu \in K_\theta^{1,\infty}$$

is a purely outer element of $N^{1,\infty}[\bar{\theta}S]$. Note that $f = \exp(i\phi - \tilde{\phi})$ in $\mathbb{C}_+$, where $2\phi = \arg\theta - x$.
Denote by $\Gamma_n$ the middle one-third of the interval $(\lambda_n, \lambda_{n+1})$. Our plan is to calculate the integral

$$\int_{\cup\Gamma_n} \phi'\tilde{\phi}\frac{dx}{x^2} \tag{78}$$

in two different ways and arrive at a contradiction by obtaining two different answers.
First let us choose a short monotone partition $\{I_n\}$ of $\mathbb{R}$ such that $\Lambda$ satisfies the density condition (51) with $a = \frac{1}{2\pi}$ on that partition:
Put $a_0 = 0$. Choose $a_1 > a_0$ to be the smallest point such that $\#\ \Lambda \cap (a_0, a_1] \geq \frac{1}{2\pi}(a_1 - a_0)$. Note that such a point always exists because $\Lambda$ supports a measure with a spectral gap greater than 1: otherwise we would be able to choose a long sequence of intervals satisfying (86) in Lemma (2.2.11) with $a = \frac{1}{2\pi}$ and arrive at a contradiction. After $a_i, i \geq 1$ is chosen, choose $a_{i+1} > a_i$ as the smallest point such that

$$\#\Lambda \cap (a_i, a_{i+1}] \geq \frac{1}{2\pi}(a_{i+1} - a_i) \text{ and } (a_{i+1} - a_i) \geq (a_i - a_{i-1}).$$

Choose $a_i, i < 0$ in a similar way. Put $I_n = (a_n, a_{n+1}]$. Again by Lemma (2.2.11), $\{I_n\}$ has to be short.
In what follows we will assume, WLOG, that $\frac{1}{2\pi}|I_n| = \#\Lambda \cap I_n$.
Note that since $C_X < 1$, the sum in the energy condition (52) has to be infinite. At the same time, a part of that sum has to be finite:
**Claim 3**.

$$\sum_n \frac{\log_-(\lambda_{n+1} - \lambda_n)}{\lambda_n^2} < \infty.$$

**Proof of claim.** Suppose that the sum is infinite. Put $\mu = |\nu|$ and let $\Phi$ be the inner function corresponding to $\mu$. Let $\psi = \arg\Phi - x$.
Define the intervals $J_n$ and the function $v$ like in part I) of the proof, with $\Phi$ replacing $\theta$. Put $w = \psi - v = \arg\Phi - x - v$. Let again $w_n = \quad |_{J_n}$, $w_n^* = w - w_n$. Then $\tilde{w} \in L^1_\Pi$ because $w_n$ are atoms with summable $L^1_\Pi$-norms.
Like in part I) we can use "atomic" estimates to show that if $\operatorname{dist}(J_k, J_n) > \max(|J_k|, |J_n|)$ and $x \in J_n$ then

$$|\tilde{w}_k(x)| \lesssim \frac{|J_k|^3}{\operatorname{dist}^2(x,J_k)}.$$

By monotonicity and shortness of $J_k$ we conclude that

$$\sum_{\lambda_i \in J_n} \frac{|\tilde{w}_k(\lambda_i)|}{\lambda_i^2} \lesssim \sum_{\lambda_i \in J_n} \frac{|\tilde{w}_k(\lambda_i)|}{1+\operatorname{dist}^2(0,J_n)} \lesssim \frac{1}{1+\operatorname{dist}^2(0,J_n)} \int_{J_n} \frac{|J_k|^3}{\operatorname{dist}^2(x,J_k)} dx.$$

Hence, similarly to (75),

$$\sum_n \left[ \sum_{k:\operatorname{dist}(J_k,J_n)>\max\ (|J_k|,|J_n|)} \left[ \sum_{\lambda_i \in J_n} \frac{|\tilde{w}_k(\lambda_i)|}{\lambda_i^2} \right]\right] =$$

$$\sum_n \left[ \sum_{k:k<n,\operatorname{dist}(J_k,J_n)>\max\ (|J_k|,|J_n|)} \left[ \sum_{\lambda_i \in J_n} \frac{|\tilde{w}_k(\lambda_i)|}{\lambda_i^2} \right]\right] \lesssim$$

$$\sum_n \left[ \sum_{k:k<n,\operatorname{dist}(J_k,J_n)>\max\ (|J_k|,|J_n|)} \frac{1}{1+\operatorname{dist}^2(0,J_n)} \int_{J_n} \frac{|J_k|^3}{\operatorname{dist}^2(x,J_k)} dx \right] < \infty.$$

In other words, on each $J_n$

$$\sum_{k:\operatorname{dist}(J_k,J_n)>\max\ (|J_k|,|J_n|)} |\tilde{w}_k^*| \leq g_1$$

where $g_1$ is a positive function satisfying

$$\sum_n \frac{g_1(\lambda_n)}{1+\lambda_n^2} < \infty.$$

Also for any $x \in J_n$

$$\tilde{w}_k(x) = \int \frac{dt}{t-x} w_k(t) = -\int_{J_k} \log|t-x| w'(t)dt.$$

If $k < n$ then

$$-\int_{J_k} \log|t-x| w'(t)dt \geq$$

$$-\int_{J_k} \log_+|t-x|(\arg\Phi)'(t)dt + \int_{J_k} \log_+|t-x|(1+v')(t)dt - \text{const} \gtrsim -|J_k|. \quad (79)$$

Here we applied Lemma (2.2.18) , part 6 , to the second integral in the second line and used the estimate

$$-\int_{J_k} \log_+|t-x|(\arg\Phi)'(t)dt \geq -\log(x-b) \int_{J_k} (\arg\Phi)',$$

where $b$ is the left endpoint of $J_k$, for the first integral.

Thus for $x \in J_n$

$$\sum_{k:k\neq n,\operatorname{dist}(J_k,J_n)<\max\ (|J_k|,|J_n|)} \tilde{w}_k(x) \geq g_2(x),$$

where again

$$\sum_n \frac{|g_2(\lambda_n)|}{1+\lambda_n^2} < \infty.$$

Also, for $x \in J_n$

$$-\int_{J_n} \log_+ |x-t| w'(t)dt =$$

$$-\int_{J_n} \log_+ |x-t| \arg' \Phi(t)dt + \int_{J_n} \log_+ |x-t|(1-v')(t)dt - \text{ const } \gtrsim -|J_n|. \quad (80)$$

Here we again used the property that $\int_{J_n} \arg' \Phi = \int_{J_n} (v'-1) = |J_n| + O(1)$ and applied Lemma (2.2.18), part 5, to the second integral in the second line.
Hence for any $x \in \mathbb{R}$

$$\tilde{w}(x) \geq \int \log_- |x-t| w'(t)dt + g(x)$$

for some function $g$ satisfying

$$\sum_n \frac{|g(\lambda_n)|}{1+\lambda_n^2} < \infty.$$

Therefore,

$$\sum_n \frac{\tilde{w}(\lambda_n)}{1+\lambda_n^2} \geq \text{ const } + \sum_n \frac{\int_{\lambda_n-1}^{\lambda_n+1} \log_- |\lambda_n - x| w' dx}{1+\lambda_n^2} \geq$$

$$\text{const } + \sum_n \frac{\int_{\lambda_{n-1}}^{\lambda_n} \log_- |\lambda_n - x| (\arg \Phi)' dx}{1+\lambda_n^2} \geq \text{ const } + \sum_n \frac{\log_- |\lambda_n - \lambda_{n-1}|}{1+\lambda_n^2}.$$

Let $f = (1+\Phi)K\nu$. Then $f$ is an outer function in $\mathbb{C}_+$that belongs to $H^2$ and satisfies

$$f = \exp\left(i\frac{\psi}{2} - \frac{\tilde{\psi}}{2}\right).$$

Since $|f(\lambda_n)| = |\nu(\{\lambda_n\})|/\mu(\{\lambda_n\}) = 1$, we have that $\log|f(\lambda_n)| = 2\tilde{\psi}(\lambda_n) = 0$ for all $n$.
Recall that $\tilde{w}(\lambda_n) = \tilde{\psi}(\lambda_n) + \tilde{v}(\lambda_n) = \tilde{v}(\lambda_n)$. It is left to show that

$$\sum_n \frac{\tilde{v}(\lambda_n)}{1+\lambda_n^2} < \infty.$$

Recall that $v \in L^1_\Pi, \tilde{v} = \tilde{w} - \tilde{\psi} = \tilde{w} - \log|f|/2 \in L^1_\Pi$ and $v'$ is bounded on $\mathbb{R}$. Therefore the harmonic extension of $v$ into $\mathbb{C}_+$has a bounded $x$-derivative in $\mathbb{C}_+$. Hence $\tilde{v}_y$ is bounded in $\mathbb{C}_+$as well.
On each interval $J_n$ choose $\lambda_{k_n}$ so that $v(\lambda_{k_n}) = \max_{\lambda_i \in J_n} v(\lambda_i)$. If the last sum is positive infinite then so is

$$\sum_n |J_n| \frac{\tilde{v}(\lambda_{k_n})}{1+\text{dist}^2(0, J_n)}.$$

Because of the boundedness of $\tilde{v}_y, \tilde{v}(\lambda_{k_n} + i|J_n|) \geq \tilde{v}(\lambda_{k_n}) - C|J_n|$ and therefore

$$\sum_n |J_n| \frac{\tilde{v}(\lambda_{k_n} + i|J_n|)}{1+\text{dist}^2(0, J_n)} = \infty.$$

Denote by $(\tilde{v})^M$ the maximal non-tangential function of $\tilde{v}$ in $\mathbb{C}_+$. The last equation implies that $(\tilde{v})^M \notin L^1_\Pi$. But that contradicts the property that both $\tilde{v}$ and $v$ belong to $L^1_\Pi$.
Now notice that if $x \in \Gamma_n$ then

$$|f(x)| = |(1-\theta(x))K\nu(x)| \leq 2\left|\int \frac{1}{t-x} d\nu(t)\right| \leq 6||\nu \parallel |\lambda_{n+1} - \lambda_n|^{-1}. \quad (81)$$

Hence

$$\int_{\cup\Gamma_n} \phi' \tilde{\phi} \frac{dx}{x^2} \lesssim \sum_n \frac{1}{\lambda_n^2} \int_{\Gamma_n} \arg' \theta \log_+ |f| \, dx + \text{ const } \lesssim$$
$$\sum_n \frac{1}{\lambda_n^2} \int_{\Gamma_n} \arg' \theta dx \log_- |\lambda_{n+1} - \lambda_n| + \text{ const } \lesssim \qquad (82)$$
$$\sum_n \frac{1}{\lambda_n^2} \log_- |\lambda_{n+1} - \lambda_n| + \text{ const } < \infty.$$

It follows that

$$\int_{\cup\Gamma_n} w' \tilde{w} \frac{dx}{x^2} = \int_{\cup\Gamma_n} \phi' \dot{\phi} \frac{dx}{x^2} - \int_{\cup\Gamma_n} \phi' \tilde{v} \frac{dx}{x^2} -$$
$$\int_{\cup\Gamma_n} v' \tilde{\phi} \frac{dx}{x^2} + \int_{\cup\Gamma_n} v' \tilde{v} \frac{dx}{x^2} < \infty. \qquad (83)$$

Indeed, arguing like at the end of the proof of the last claim, from the property that $(\tilde{v})^M \in L^1_\Pi$ we deduce that

$$\sum_n |J_n| \frac{\sup_{x \in J_n} |\tilde{v}(x)|}{1 + \text{dist}^2(0, J_n)} < \infty.$$

Therefore

$$\left| \int_{\cup\Gamma_k \cap J_n} \phi' \tilde{v} \frac{dx}{x^2} \right| \leq \int_{\cup\Gamma_k \cap J_n} |\phi'| dx \frac{\sup_{x \in J_n} |\tilde{v}(x)|}{1 + \text{dist}^2(0, J_n)} \asymp |J_n| \frac{\sup_{x \in J_n} |\tilde{v}(x)|}{1 + \text{dist}^2(0, J_n)}$$

and summing over all $n$ we get

$$\left| \int_{\cup\Gamma_k} \phi' \tilde{v} \frac{dx}{x^2} \right| < \infty.$$

The second integral on the right-hand side of (83) is finite because $v'$ is bounded and $\tilde{\phi} = \log |f|$ is in $L^1_\Pi$. The last integral is finite because $v'$ is bounded and $\tilde{v} = \tilde{\phi} - \tilde{w} \in L^1_\Pi$.

Denote

$$L_n = \cup_{\text{dist}(J_k, J_n) < \max\ (|J_k|, |J_n|)} J_k, \; q_n = w|_{L_n} \text{ and } q*_n = w - q_n.$$

Then

$$\int_{\cup\Gamma_k} w' \tilde{w} \frac{dx}{x^2} = \sum_n \left[ \int_{\cup\Gamma_k \cap J_n} w' \tilde{q}_n^* \frac{dx}{x^2} + \int_{\cup\Gamma_k \cap J_n} w' \tilde{q}_n \frac{dx}{x^2} \right].$$

The first integral can be, once again, estimated like in (59), i.e. using the property that each $w_i$ is an atom, and the sum of such integrals shown to be finite. For the second integral, applying similar arguments that were used in the first part to prove (62) we obtain

$$\int_{\cup\Gamma_k \cap J_n} w' \tilde{q}_n \frac{dx}{x^2} \geq$$
$$-\frac{1}{1 + \text{dist}^2(J_n, 0)} \int_{\cup\Gamma_k \cap J_n} w'(x) \left[ \sum_{J_l \subset L_n} \int_{J_l} \log |x - t| w'(t) dt - |J_n||J_l| \right] dx. \; (84)$$

Furthermore, because of (79),

$$-\int_{\cup\Gamma_k\cap J_n} w'(x)\left[\sum_{J_l\subset L_n}\int_{J_l}\log|x-t|w'(t)dt\right]dx \geqslant$$

$$-\int_{\cup\Gamma_k\cap J_n} w'(x)\left[\int_{J_n}\log|x-t|w'(t)dt\right]dx - \sum_{J_l\subset L_n}|J_l||J_n|.$$

Let us remark right away that

$$\sum_n \frac{1}{1+\mathrm{dist}^2(J_n,0)}\sum_{J_l\subset L_n}|J_l||J_n|\sum$$

$$\sum_{kn,\mathrm{dist}(J_l,J_n)<\max(|J_l|,|J_n|)}\frac{|J_l||J_n|}{1+\mathrm{dist}^2(J_n,0)} \lesssim \sum_n \frac{|J_n|}{1+\mathrm{dist}^2(J_n,0)} < \infty.$$

To continue the estimates let us split the last integral:

$$-\int_{\cup\Gamma_k\cap J_n} w'(x)\left[\int_{J_n}\log|x-t|w'(t)dt\right]dx =$$

$$\int_{\cup\Gamma_k\cap J_n}(\arg\theta)'(x)\left[\int_{J_n}\log|x-t|(v'(t)+1)(t)dt\right]dx$$

$$-\int_{\cup\Gamma_k\cap J_n}(\arg\theta)'(x)\left[\int_{J_n}\log|x-t|(\arg\theta)'(t)(t)dt\right]dx$$

$$-\int_{\cup\Gamma_k\cap J_n}(v'(x)+1)\left[\int_{J_n}\log|x-t|(v'(t)+1)dt\right]dx$$

$$+\int_{\cup\Gamma_k\cap J_n}(v'(x)+1)\left[\int_{J_n}\log|x-t|(\arg\theta)'(t)(t)dt\right]dx =$$

$$I+II+III+IV.$$

To estimate $III$ and $IV$ denote by $C$ the constant satisfying

$$\int_{\cup\Gamma_k\cap J_n}(v'(x)+1)dx = C|J_n|.$$

Notice that because $1-2\varepsilon < v'+1 < 1+2\varepsilon$ and $\int_{J_n} v'+1 = \int_{J_n}(\arg\theta)' = |J_n|$, for any $y\in J_n$,

$$\int_{J_n}\log|y-t|(v'(t)+1)dt = |J_n|\log|J_n| + O(|J_n|) \qquad (85)$$

and

$$III = -\int_{\cup\Gamma_k\cap J_n}(v'(x)+1)\left[\int_{J_n}\log|x-t|\,(v'(t)+1)dt\right]dx$$

$$= -C|J_n|^2\log|J_n| + O(|J_n|^2).$$

To estimate $IV$ observe that for any $t\in J_n$, if $\mathrm{dist}\big(t,(\lambda_k,\lambda_{k+1})\big)\geq 1$ then

$$\int_{\Gamma_k}(v'(x)+1)\log_+|x-t|dx \geq$$

$$\int_{\Gamma_k}(v'(x)+1)dx\frac{\int_{\lambda_k}^{\lambda_{k+1}}\log_+|x-t|dx}{\lambda_{k+1}-\lambda_k} - (\lambda_{k+1}-\lambda_k)\log 3$$

(recall that $\Gamma_k$ is the middle one-third of $(\lambda_k, \lambda_{k+1})$ ). Consider a positive step function $\alpha(x)$ defined on each $(\lambda_k, \lambda_{k+1})$ as

$$\frac{\int_{\Gamma_k} (v'(x)+1)dx}{\lambda_{k+1}-\lambda_k}.$$

Then $|\alpha - 1| \leq \varepsilon$ on $J_n$. Hence one can apply Lemma (2.2.18) part 5 to conclude that, for any $t \in J_n$,

$$\int_{\cup\Gamma_k \cap J_n} (v'(x)+1)\log_+|x-t|dx \geq$$

$$\int_{J_n} \alpha(x)\log_+|x-t|dx - \text{ const } |J_n| \geq \left(\int_{J_n} \alpha(x)\right)\log|J_n| - \text{ const } |J_n| =$$

$$C|J_n|\log|J_n| \text{ - const } |J_n|.$$

Therefore

$$IV = \int_{\cup\Gamma_k \cap J_n} (v'(x)+1)\left[\int_{J_n} \log|x-t|(\arg\theta)'(t)(t)dt\right]dx \geq$$

$$\left(\int_{J_n} (\arg\theta)'\right)(C|J_n|\log|J_n| - \text{const}|J_n|) = C|J_n|^2\log|J_n| - \text{const}|J_n|^2.$$

Combining the estimates we get

$$III + IV \geq -|J_n|^2.$$

To estimate $II$ notice that

$$II = -\sum_{\Gamma_k \subset J_n} \int_{\Gamma_k} (\arg\theta)'(x)dx \sum_{\lambda_j, \lambda_{j+1} \in J_n} \int_{\lambda_j}^{\lambda_{j+1}} \log|x-t|(\arg\theta)'(t)dt.$$

If $t \in (\lambda_j, \lambda_{j+1})$ and $x \in \Gamma_k$ then

$$\log\ |x-t| \geq \begin{cases} \log|\lambda_j - \lambda_{k+1}| \text{ if } j < k \\ \log|\lambda_k - \lambda_{j+1}| \text{ if } j > k. \\ \log|\lambda_{j+1} - \lambda_j| \text{ if } j = k \end{cases}$$

Put $\alpha_k = \int_{\Gamma_k} (\arg\theta)'$. Then

$$II \geq -\sum_{\Gamma_k \subset J_n} \alpha_k \sum_{\lambda_j \in J_n, j \neq k} 2\pi\log|\lambda_k - \lambda_j| + A_n$$

where the constants $A_n$ satisfy

$$\sum_n \frac{|A_n|}{1+\text{dist}^2(0, J_n)} < \infty.$$

Using (85), I can be rewritten as

$$I = \sum_{\Gamma_k \subset J_n} \int_{\Gamma_k} (\arg\theta)'(x)\left[\int_{J_n} \log|x-t|(v'(t)+1)(t)dt\right]dx =$$

$$\left(\sum_{\Gamma_k \subset J_n} \alpha_k\right)|J_n|\log|J_n| + B_n$$

where again

$$\sum_n \frac{|B_n|}{1+\operatorname{dist}^2(0,J_n)} < \infty.$$

By the left-hand side of the inequality in part 2) of Lemma (2.2.17),

$$\alpha_k = \int_{\Gamma_k} (\arg\theta)' > c > 0$$

for all $k$. Therefore, since there are $\frac{1}{2\pi}|J_n|$ intervals $\Gamma_k$ in $J_n$,

$$I + II = \left(\sum_{\Gamma_k\subset J_n} \alpha_k\right)|J_n|\log|J_n| -$$

$$\sum_{\Gamma_k\subset J_n} \alpha_k \sum_{\lambda_j\in J_n, j\neq k} 2\pi\log|\lambda_k-\lambda_j| + A_n + B_n \gtrsim$$

$$\frac{1}{2\pi}|J_n|^2\log|J_n| - \sum_{\lambda_j,\lambda_k\in J_n} 2\pi\log|\lambda_k-\lambda_j| + A_n + B_n.$$

Now, going back to (84), we obtain

$$\sum_n \int_{\cup\Gamma_k\cap J_n} w'\tilde{w}_n \frac{dx}{x^2} \gtrsim$$

$$\sum_n \frac{\frac{1}{4\pi^2}|J_n|^2\log|J_n| - \sum_{\lambda_j,\lambda_k\in J_n}\log|\lambda_k-\lambda_j| - |J_n|^2 + A_n + B_n}{1+\operatorname{dist}^2(J_n,0)} + \text{ const.}$$

It remains to notice that the sum on right-hand side is positive infinite because otherwise $\Lambda$ would satisfy the energy condition (52) and $C_X$ would be at least 1 . This contradicts (83).
It remains to discuss the case when $\sup_n(\lambda_n-\lambda_{n-1}) = \infty$. If $\Lambda$ is such a sequence, choose a large constant $C$ and consider the set of all gaps $R_k$ of $\Lambda$ of the size larger than :

$$R_k = (\lambda_{n_k}, \lambda_{n_k+1}), \lambda_{n_k+1} - \lambda_{n_k} > C.$$

After that one can add a separated set of points in every $R_k$ and consider a slightly larger sequence $\Lambda' = \{\lambda'_n\} \supset \Lambda$ that satisfies $\sup_n(\lambda'_n - \lambda'_{n-1}) \le C$ and

$$\inf_{\lambda'_n,\lambda'_{n-1}\in\Lambda'\setminus\Lambda}(\lambda'_n-\lambda'_{n-1}) \ge C/2.$$

Since $C_\Lambda < 1$, for large enough $C$ the sequence $\Lambda'$ will still satisfy $C_{\Lambda'} < 1$.
The inner function $\theta$ should then be chosen for the sequence $\Lambda'$ instead of $\Lambda$. Consider the outer function

$$h = (1-\theta)K\nu \in K_\theta^{1,\infty}.$$

Then $h$ is divisible by $S$ and has zeros at $\Upsilon = \Lambda' \setminus \Lambda$. Since $\Upsilon$ is a separated sequence, there exists an inner function $I, \operatorname{spec}_I = \Upsilon$ such that $(\arg I)'$ is bounded, see for instance Lemma (2.2.10)6 in [55]. If $C$ is large enough, $|(\arg I)'| << \varepsilon$. The function

$$g = \frac{Ih}{1-I}$$

is divisible by $S$ and satisfies

$$\bar{\theta}g = \bar{\theta}\frac{Ih}{1-I} = \bar{\theta}h\frac{1}{1-\bar{I}}$$

on $\mathbb{R}$. Therefore $g \in K_\theta^+$. At the same time, $g$ no longer has zeros on $\mathbb{R}$. Denote $f = g/IS$. Then $f \in N^+[\bar{\theta}S]$ is an outer function whose argument on $\mathbb{R}$ is equal to $(\arg\theta - x -$

$\arg I)/2$. Now we can apply claim 1 to $u = \arg\theta - x - \arg I$ to obtain functions $v = v_1 + v_2$ satisfying the properties 1-5.
If one denotes by $\Gamma_n$ the middle one-third of $(\lambda'_n, \lambda'_{n+1})$, then similarly to (81),
$$|S^{-1}(x)h(x)| = |(1-\theta(x))Kv(x)| \le 6 \parallel v \parallel |\lambda'_{n+1} - \lambda'_n|^{-1}.$$
The argument of the function $h/S$ is $\arg\theta - x$. Note that claim 3 still holds with $\Lambda'$ in place of $\Lambda$, because $\Upsilon$ is separated. Hence (82) still holds for $\phi = \arg\theta - x$. After that, using the property that $|(\arg I)'| << \varepsilon$, one can "absorb" $\arg I$ into $v_1$ and replace $v_1$ with $y = v_1 + \arg I$. The rest of the estimates of the integral in (78) can be done in the same way as before, with $v = y + v_2$ in place of $v = v_1 + v_2$.
If $\Lambda$ is a real sequence we define its Beurling-Malliavin density as
$$d_{BM}(\Lambda) = \sup \ \{d \mid \exists \text{ long } \{I_n\} \text{ such that } \#\Lambda \cap I_n \ge d|I_n| \forall n\}$$
if $\Lambda$ is discrete and as $\infty$ otherwise.
An equivalent definition is given in [72]:
$$d_{BM}(\Lambda) = \sup \ \{a: N[\bar{S}^{2\pi a}\theta] = 0\},$$
where $\theta(z)$ denotes some/any meromorphic inner function with $\mathrm{spec}_\theta = \Lambda$.
Note that the Beurling-Malliavin multiplier theorem implies that $N[\bar{S}^{2\pi a}\theta]$ in the above definition can be replaced with any $N^p[\bar{S}^{2\pi a}\theta], 0 < p \le \infty$, the kernel in the Hardy space $H^p$, or with $N^+[\bar{S}^{2\pi a}\theta]$, the kernel in the Smirnov class.
**Lemma (2.2.10)[84]:** Let $X \subset \mathbb{R}$ be a closed set and let $\Lambda$ be a discrete sequence. Then
$$G_{X\cup\Lambda} \le G_X + 2\pi d_{BM}(\Lambda).$$
**Proof.** Let $d_{BM}(\Lambda) = d_1, G_X = d_2$ and $G_{X\cup\Lambda} = d_3$. Let $\varepsilon > 0$ be a small number. By Theorem (2.2.6), $N[\bar{\theta}S^{d_3-\varepsilon}] \ne 0$ for some meromorphic inner $\theta, \mathrm{spec}_\theta \subset X \cup \Lambda$. Let $f \in N[\bar{\theta}S^{d_3-\varepsilon}]$. Let $I$ be an inner function such that $\mathrm{spec}\ _I = \Lambda$.
By the above definition of the Beurling-Malliavin density, there exists a function $g \in N^\infty[\bar{S}^{2\pi d_1+\varepsilon}I]$. Then the function $h = (1-I)g$ belongs to $N^\infty[\bar{S}^{2\pi d_1+\varepsilon}]$ and is equal to 0 on $\Lambda$. The function $fh$ belongs to $N[\bar{\theta}S^{d_3-2\pi d_1-2\varepsilon}]$ and is zero on $\Lambda$ (obviously, we assume that $d_3 - 2\pi d_1 - 2\varepsilon > 0$). Finally, the function $l = S^{d_3-2\pi d_1-2\varepsilon}fh$ belongs to $N[\bar{\theta}] = K_\theta$ and is still zero on $\Lambda$. By the Clark representation
$$l = \frac{1}{2\pi i}(1-\theta)Kl\sigma$$
where $\sigma$ is the Clark measure for $\theta$, $\mathrm{supp}\,\sigma = \mathrm{spec}_\theta \subset \Lambda \cup X$. Since $l$ is divisible by $S^{d_3-2\pi d_1-2\varepsilon}$ in $\mathbb{C}_+$and $(1-\theta)$ is an outer function in $\mathbb{C}_+, Kl\sigma$ is divisible by $S^{d_3-2\pi d_1-2\varepsilon}$ in $\mathbb{C}_+$. Equivalently, the measure $l\sigma$ has a spectral gap of the size $d_3 - 2\pi d_1 - 2\varepsilon$. Since $l$ is zero on $\Lambda$, the measure $l\sigma$ is supported on $X$. Hence
$$G_X \ge d_3 - 2\pi d_1 - 2\varepsilon = G_{X\cup\Lambda} - 2\pi d_{BM}(\Lambda) - 2\varepsilon.$$
The following statement can be viewed as a version of the first Beurling-Malliavin theorem, see [72], [49].
**Lemma (2.2.11)[84]:** Let $\Lambda$ be a real sequence. Suppose that there exists a long sequence of intervals $I_n$ such that
$$\#(\Lambda \cap I_n) \le a|I_n| \qquad (86)$$
for all $n$, for some $a \ge 0$. Then $G_\Lambda \le 2\pi a$.
**Proof.** Suppose that $G_\Lambda = 2\pi a + 3\varepsilon$ for some $\varepsilon > 0$. Then by Theorem (2.2.6), $N[\bar{\theta}S^{2\pi a+2\varepsilon}] \ne 0$ for some inner function $\theta, \mathrm{spec}_\theta \subset \Lambda$. But (86) implies that the argument of the symbol increases greatly on $I_n$, which leads to a contradiction. More precisely, denote
$$\gamma = \arg(\bar{\theta}S^{2\pi a+2\varepsilon}) = (2\pi a + 2\varepsilon)x - \arg\theta.$$

For each $I_n = (a_n, a_{n+1}]$ denote

$$\delta_n = \inf_{I_n''}\gamma - \sup_{I_n'}\gamma,$$

where

$$I_n' = \left(a_n, a_n + \frac{\varepsilon|I_n|}{6(\pi a + \varepsilon)}\right) \text{ and } I_n'' = \left(a_{n+1} - \frac{\varepsilon|I_n|}{6(\pi a + \varepsilon)}, a_{n+1}\right).$$

Then (86) implies that $\delta_n \geq \frac{\varepsilon}{3}|I_n|$. Hence by a theorem in [72], $N[\bar{\theta}S^{a+2\varepsilon}]$ has to be trivial.

**Lemma (2.2.12)[84]:** Let $I = [a, b]$ be an interval on $\mathbb{R}$ and let $\Lambda = \{\lambda_1, \dots, \lambda_N\}, a \leq \lambda_1 < \dots < \lambda_N \leq b$ be a set of points on I. Let $C > 1$ be a constant and suppose that for some subinterval $J = [c, d] \subset I$,

$$\#\Lambda \cap J \leq \frac{|J|}{C} - 1.$$

Then one can spread the points of $\Lambda$ on $J$ without a large decrease in the energy $E(\Lambda)$. More precisely, if

$$\Gamma = \{\gamma_1, \dots, \gamma_N\},\ a \leq \gamma_1 < \cdots < \gamma_N \leq b$$

is another set of points on $I$ with the properties that

(i) $\gamma_k = \lambda_k$ for all $k$ such that $\lambda_k \notin J$ and

(ii) $|\gamma_k - \gamma_j| \geq C$ for all $\gamma_k, \gamma_j \in J, \gamma_k \neq \gamma_j$

then

$$E(\Gamma) \geq E(\Lambda) - \frac{\log C}{C}|J|N.$$

where $E$ is defined by (49)

**Proof.** Notice that $\sum_{\gamma_k,\gamma_j \in J} \log_-|\gamma_k - \gamma_j| = 0$ and that

$$\log_+|\gamma_k - \gamma_j| \geq \log_+|\lambda_k - \lambda_j| - \log_+ C$$

for all $k, j$.

**Corollary (2.2.13)[84]:** Let $\Lambda$ be a sequence of real points that satisfies the density (51) and energy (52) conditions for some partition $I_n$ and $d > 0$. Let $C > 1$. Let $J_k$ be $a$ sequence of disjoint intervals such that for every $k, J_k \subset I_n$ for some $n$ and

$$\#\Lambda \cap J_k \leq \frac{|J_k|}{C} - 1$$

for all $k$. Let $\Gamma$ be a sequence of points obtained from $\Lambda$ by spreading the points on each interval $J_k$ like in the last lemma. Then $\Gamma$ satisfies the density and energy conditions with the same partition $I_n$ and $d$.

**Corollary (2.2.14)[84]:** Let $\Lambda = \{\lambda_n\}$ be a monotone sequence of real points such that $C_\Lambda \geq d > 0$. Then for any $\varepsilon > 0$ there exists a monotone sequence $\Gamma = \{\gamma_n\}$ such that

(i) $C_\Gamma \geq d$,

(ii) $d_{BM}(\Gamma \setminus \Lambda) < \varepsilon$ and

(iii) $\sup_n(\gamma_{n+1} - \gamma_n) < \infty$.

**Proof.** Choose $C > 0$ so that $1/C << d$ and $1/C << \varepsilon$. Let $[\lambda_{n_k}, \lambda_{n_k+1}]$ be a sequence of all "gaps" of $\Lambda$ satisfying $\lambda_{n_k+1} - \lambda_{n_k} > C$.

Since $C_\Lambda \geq d$, there exists a partition $I_n$ such that $\Lambda$ satisfies (51) and (52) for $I_n$ and $d$. One can choose a sequence of disjoint intervals $J_k$ such that for every $k$, $J_k \subset I_n$ for some $n$,

$$\cup [\lambda_{n_k}, \lambda_{n_k+1}] \subset \cup J_k \text{ and } \frac{|J_k|}{2C} \leq \#\Lambda \cap J_k \leq \frac{|J_k|}{C} - 1 \text{ for all } k$$

(the choice of $J_k$ can be made by a version of the "shading" algorithm, see for instance [66]), volume 2, pp $507-508$. Let $\Gamma$ be a sequence of points obtained from $\Lambda$ by spreading the points on each interval $J_k$ like in Lemma (2.2.12) . Then (i) is satisfied by the previous corollary and the supremum in (iii) is at most $2C$. Since the distances between the points of $\Gamma$ on $\cup J_k$ are at least $C$,

$$d_{BM}(\Gamma \setminus \Lambda) \leq C^{-1} < \varepsilon.$$

**Lemma (2.2.15)[84]:** Let $\Lambda$ be a sequence of real points and let $\{I_n\}$ be a short partition such that $\Lambda$ satisfies

$$a|I_n| < \#\Lambda \cap I_n$$

for all $n$ with some $a > 0$ and the energy condition (52) on $\{I_n\}$. Then for any short partition $\{J_n\}$, there exists a subsequence $\Gamma \subset \Lambda$ that satisfies

$$\#(\Lambda \setminus \Gamma) \cap J_n = o(|J_n|)$$

as $n \to \pm\infty$, and the energy condition (52) on $\{J_n\}$.

**Proof.** To simplify the estimates we will assume that the endpoints of $I_n$ belong to $\Lambda$, i.e. that $I_n = \left(\lambda_{k_n}, \lambda_{k_{n+1}}\right]$ for each $n$, and that the energy condition (52) is satisfied on $I_n$ with $E_n$ defined by (54), see the explanation there.

(To include the endpoints in $E_n$ one may need to compensate by deleting a point on each $I_n$, as explained in the beginning of the proof of Theorem (2.2.9). This is where one may need to pass from $\Lambda$ to a subsequence $\Gamma$. Since $|I_n| \to \infty, \Gamma$ will satisfy $\#(\Lambda \setminus \Gamma) \cap J_n = o(|J_n|).$)

We will also assume that $\#\Lambda \cap I_n = |I_n|$ for all $n$. In this case one can choose $\Gamma = \Lambda$.

Fix $n$ and suppose that the intervals $I_l, \dots, I_{l+N}$ cover $J_n$. To estimate the energy expression for $J_n$ let us first consider the case when and $\cup_l^{l+N} I_j = J_n$. Denote by $u$ a piecewise linear continuous function on $\mathbb{R}$ that is zero outside of $J_n$ and grows linearly by 1 between $\lambda_n$ and $\lambda_{n+1}$ for each $\lambda_n, \lambda_{n+1} \in J_n$. Denote

$$p(x) = \begin{cases} u(x) - x + \lambda_{k_l} & \text{on } J_n = \left(\lambda_{k_l}, \lambda_{k_{l+N+1}}\right] \\ 0 & \text{on } \mathbb{R} \setminus J_n \end{cases}.$$

Then $p\left(\lambda_{k_n}\right) = 0$ for all $l \leq n \leq l+N+1$. Denote by $p_n$ the restriction $p_n|_{I_n}$.

On each $(\lambda_i, \lambda_{i+1})$ the function $u'$ satisfies the same estimates as $|\theta'|$ from the statement of Lemma (2.2.17) , parts 2 and 3 . Therefore for the function $p$ one can apply the same argument as in the first part of the proof of Theorem (2.2.9), where $p$ was defined as arg $\theta - x - v_2$ (we will simply assume that $v_2 \equiv 0$ ).

First, one can show that

$$-\iint_{J_n \times J_n} \log|t-x| p'(t) p'(x) dt dx =$$

$$|J_n|^2 \log|J_n| - \sum_{\lambda_{k_l} \leq \lambda_i, \lambda_j \leq \lambda_{k_{l+N+1}}} \sum_{\lambda_i \neq \lambda_j} \log\left|\lambda_i - \lambda_j\right| + \text{ const } |J_n|^2.$$

To estimate the last integral rewrite it as

$$-\iint_{J_n \times J_n} \log|t-x| p'(t) p'(x) dt dx = \sum_{I_i \subset J_n} \sum_{I_j \subset J_n} -\iint_{I_i \times I_j} \log|t-x| p'(t) p'(x) dt dx.$$

For the last integral, when $i = j$ by (69) and (70) we have

$$-\iint_{I_i \times I_i} \log|t-x| p'(t) p'(x) dt dx \lesssim |I_i|^2 \log|I_i| - E_i + \text{ const } |I_i|^2.$$

If $I_i$ does not intersect $2I_j$ we can apply (71) and (72) (where we used that $\operatorname{dist}(I_j, I_k) \geq |I_j|$) to obtain

$$-\iint_{I_i \times I_j} \log_+|t-x|\, p'(t)p'(x)dtdx \lesssim |I_i||I_j|.$$

For the case when $I_i$ intersects $2I_j$ but not contained in $2I_j$, or when $I_i$ is adjacent to $I_j$, (note that there are at most four of such $I_i$ for each $I_j$ ) we can estimate the integral like in (73) to conclude that

$$-\iint_{I_i \times I_j} \log_+|t-x| p'(t)p'(x)dtdx \lesssim$$

$$(|I_i|^2 \log|I_i| - E_i) + \left(|I_j|^2 \log|I_j| - E_j\right) + |I_i|^2 + |I_j|^2.$$

Finally, in the case when $I_i \subset 2I_j, j > i+1$, notice that for any $x \in I_i$ and any $s, t \in I_j, t > s, \log_+|s-t| < \log_+|x-t|$. Hence (we assume that $\operatorname{dist}(x, I_j) \geq |I_{i+1}| > 1$ to skip the estimates of $\log_-$)

$$-\iint_{I_j \times I_j} \log_+|s-t|p'(t)p'(s)dtds = -2\int_{\lambda_{k_j}}^{\lambda_{k_{j+1}}} \int_s^{\lambda_{k_{j+1}}} \log_+|s-t|p'(t)p'(s)dtds =$$

$$2\int_{\lambda_{k_j}}^{\lambda_{k_{j+1}}} \int_s^{\lambda_{k_{j+1}}} \log_+|s-t|dtds - 2\int_{\lambda_{k_j}}^{\lambda_{k_{j+1}}} \int_s^{\lambda_{k_{j+1}}} \log_+|s-t|u'(t)u'(s)dtds \geq$$

$$2\left(|I_j|^2 \log|I_j| - |I_j| \int_{I_j} \log_+|x-t|u'(t)dt\right) + \text{ const } |I_j|^2.$$

Also for any $x \in I_i$

$$\int_{I_j} \log_+|x-t|dt - \int_{I_j} \log_+|x-t|u'(t)dt \gtrsim -|I_j|.$$

Therefore

$$-\iint_{I_i \times I_j} \log_+|t-x|p'(t)p'(x)dtdx \leq$$

$$\int_{I_i} |p'(x)| \left(\int_{I_j} \log_+|x-t|dt - \int_{I_j} \log_+|x-t|u'(t)dt + \text{ const } |I_j|\right) dx \leq$$

$$2|I_i|\left(|I_j|\log|I_j| - \int_{I_j} \log_+|x-t|u'(t)dt\right) + \text{ const } |I_i||I_j| \leq$$

$$-\frac{|I_i|}{|I_j|}\iint_{I_j \times I_j} \log_+|s-t|p'(t)p'(s)dtds + \text{ const } |I_i||I_j| =$$

$$\frac{|I_i|}{|I_j|}\|p_j\|_{\mathcal{D}} + \text{ const } |I_i||I_j| \lesssim$$

$$\frac{|I_i|}{|I_j|}\left(|I_j|^2 \log|I_j| - E_j\right) + |I_i||I_j|.$$

Combining the estimates and using the shortness of $\{J_n\}$, we obtain that $\Lambda$ satisfies the energy condition on $\{J_n\}$.

In the case when the intervals $I_l, \dots, I_{l+N}$ cover $J_n$ but $\cup_l^{l+N} I_j \neq J_n$, i.e. when $I_l, I_{l+N} \cap J_n \neq \emptyset$ but at least one of $I_l, I_{l+N}$ is not a subset of $J_n$, denote $I_l^* = I_l \cap J_n$ and $I_{l+N}^* = I_{l+N} \cap J_n$. Notice that by remark 3 and the fact that $\log|I_l^*| < \log|I_l|$,

$$|I_l^*|^2 \log|I_l^*| - E_l^* \leq |I_l|^2 \log|I_l| - E_l.$$

Similarly,

$$|I_{l+N}^*|^2 \log|I_{l+N}^*| - E_{l+N}^* \leq |I_{l+N}|^2 \log|I_{l+N}| - E_{l+N}.$$

Now we can use the previous case with $I_l^*, I_{l+N}^*$ in place of $I_l, I_{l+N}$.

**Corollary (2.2.16)[84]:** Let $\Lambda$ be a sequence of real points and let $\{I_n\}$ be a short partition such that $\Lambda$ satisfies the density condition (51) with some $a > 0$ and the energy condition (52). Then for any $\varepsilon > 0$ there exists a subsequence $\Gamma \subset \Lambda$ and a short monotone partition $J_n$ such that $\Gamma$ satisfies (51), with $d - \varepsilon$ in place of $d$, and (52) on $J_n$.

**Proof.** One can choose a short monotone partition $\{J_n\}$ satisfying

$$(a - \varepsilon)|J_n| \leq \#\Lambda \cap J_n$$

for all $n$. Such a partition can be constructed in the same way as in the second part of the proof of Theorem (2.2.9) , see the second paragraph before claim 3 . Then $\Gamma$ can be found by Lemma (2.2.15).

**Lemma (2.2.17)[84]:** Let $A = \{a_n\}_{n\in\mathbb{Z}}$ be a real sequence satisfying

$$a_n < a_{n+1}, a_{n+1} - a_n < C < \infty$$

and $a_n \to \pm\infty$ as $n \to \pm\infty$. Denote $I_n = (a_n, a_{n+1})$. Then there exists an inner function $\theta$ satisfying

(i) $\mathrm{spec}_\theta = A$,

(ii) $|I_n|^{-1} \lesssim |\theta'| \lesssim |I_n|^{-2}$ on the middle one-third of $I_n$, for all $n$;

(iii) $|\theta'| \lesssim [\min(|I_{n-1}|, |I_n|, |I_{n+1}|)]^{-1}$ on the rest of $I_n$, for all $n$.

**Proof.** Define a second sequence $B$ as the sequence of midpoints of complimentary intervals of $A$ in $\mathbb{R}$: $b_n = (a_n + a_{n+1})/2$.

Define the inner function $\theta$ to satisfy

$$\frac{1-\theta}{1+\theta} = \text{ const } e^{Ku}, \tag{87}$$

where $u = 1_E - 1/2$,

$$E = \bigcup_{k=-\infty}^{\infty} (a_k, b_k),$$

and $Ku$ is the improper integral

$$Ku(z) = \int \frac{u(t)dt}{t-z}, \ (z \in \mathbb{C}_+).$$

The integral converges since $u$ is a convergent sum of atoms $u|_{[a_n, a_{n+1}]}$.

(Formulas similar to (87) are often used in perturbation theory. In those settings, $u$ is the Krein-Lifshits shift function and $\theta$ is the characteristic function of the perturbed operator, see for instance [99], [102])

Let $\mu_1, \mu_{-1}$ be the Clark measures for $\theta$ defined by the Herglotz representation

$$\frac{1+\theta}{1-\theta} = \frac{1}{\pi i}\int_{\mathbb{R}} \left[\frac{1}{t-z} - \frac{t}{1+t^2}\right] dt\mu_1(t) + \text{ const},$$

$$\frac{1-\theta}{1+\theta} = \frac{1}{\pi i}\int_{\mathbb{R}} \left[\frac{1}{t-z} - \frac{t}{1+t^2}\right] d\mu_{-1}(t) + \text{ const}.$$

The measures $\mu_1, \mu_{-1}$ have the following form:

$$\mu_1 = \sum \alpha_n \delta_{a_n},\ \mu_{-1} = \sum \beta_n \delta_{b_n}$$

for some positive numbers $\alpha_n, \beta_n$. (It is easy to see that $\mu_{\pm 1}\{\infty\} = 0$ although we don't actually need this fact.)

Put $\delta_n = a_{n+1} - a_n$. We claim that

$$\delta_n^2 \lesssim \beta_n \lesssim \delta_n. \tag{88}$$

Assuming that this estimate holds, we could finish as follows. Since

$$|\theta'| \asymp |1-\theta|^2 |(\mathcal{S}\mu_1)'|,\ |\theta'| \asymp |1+\theta|^2 |(\mathcal{S}\mu_{-1})'|,$$

we have

$$|\theta'(x)| \asymp \min\left\{\sum \frac{\alpha_n}{(x-a_n)^2}, \sum \frac{\beta_n}{(x-b_n)^2}\right\},\ (x \in \mathbb{R}).$$

Now if $x$ belongs to the middle one-third of one of the intervals $(a_m, a_{m+1})$, then $|\theta'(x)|$ can be estimated as

$$|\theta'(x)| \asymp \sum \frac{\alpha_n}{(x-a_n)^2} \asymp \sum \frac{\alpha_n}{(b_m - a_n)^2} = \beta_m^{-1}$$

and the estimate follows from (88). On the rest of the interval $|\theta'(x)|$ can be estimated by $\sum \frac{\beta_n}{(x-b_n)^2}$ which together with the right half of (88) gives the desired estimate.

It remains to prove (88). As follows from (87),

$$\beta_n = \text{ const } \mathrm{Res}_{b_n} e^{-Ku}.$$

Denote

$$g_n(z) = \exp\left\{-\int_{a_n}^{a_{n+1}} \frac{u(t)dt}{t-z}\right\} = \frac{\sqrt{(a_n - z)(a_{n+1} - z)}}{b_n - z},$$

and

$$A_n = \exp\left\{-\int_{\mathbb{R}\setminus(a_n, a_{n+1})} \frac{u(t)dt}{t-b_n}\right\},$$

so

$$Res_{b_n} e^{-Ku} = A_n Res_{b_n} g_n, \quad \left|\mathrm{Res}_{b_n} g_n\right| = \frac{1}{2}\delta_n.$$

To prove the right half of (88) notice that $A_n \lesssim 1$. Indeed, to the right from $a_{n+1}$, on each $(a_j, a_{j+1})$ the function $u$ is positive on the half of the interval that is closer to $b_n$ and negative on the half that is further from it. Thus

$$-\int_{(a_{n+1},\infty)} \frac{u(t)dt}{t-b_n} < 0.$$

Similarly

$$\int_{(-\infty,a_n)} \frac{u(t)dt}{t-b_n} < 0.$$

To prove the left half of (88) one needs to show that $\delta_n \lesssim A_n$. Notice that, since $\delta_n < C$,

$$-\sum_{\mathrm{dist}(b_n,(a_j,a_{j+1}))\geq 1} \int_{(a_j,a_{j+1})} \frac{u(t)dt}{t-b_n} > \text{ const } > -\infty.$$

As for the remaining part,

$$-\sum_{0<\mathrm{dist}(b_n,(a_j,a_{j+1}))\leq 1} \int_{(a_j,a_{j+1})} \frac{u(t)dt}{t-b_n} > -\int_{\delta_n/2}^{1+C} \frac{dx}{x} = \log\delta_n + \text{ const}.$$

**Lemma (2.2.18)[84]:** Let $a_1 < a_2$ and $b_1 < b_2$ be points on the real line. Let $\alpha$ and $\beta$ be nonnegative functions on the intervals $(a_1, a_2)$ and $(b_1, b_2)$ correspondingly satisfying

$$\int_{a_1}^{a_2} \alpha = \int_{b_1}^{b_2} \beta = 1, \alpha < A \text{ and } \beta < B$$

where $A, B > 1$. Then

(i) $\log_-(a_2 - a_1) \leq \int_{a_1}^{a_2}\int_{a_1}^{a_2} \log_-(x-y)\alpha(x)\alpha(y)dxdy \leq \log_-\frac{1}{A} + 1$.

(ii)If $a_2 < b_1$ then

$$\log_-(b_2 - a_1) \leq \int_{a_1}^{a_2}\int_{b_1}^{b_2} \log_-(x-y)\,\alpha(x)\beta(y)dxdy \leq \log_-(b_1 - a_2).$$

(iii) If $a_2 = b_1$ then

$$\int_{a_1}^{a_2}\int_{b_1}^{b_2} \log_-(x-y)\alpha(x)\beta(y)dxdy \leq \mathrm{m}\ \left(\log_-\frac{1}{A}, \log_-\frac{1}{B}\right) + 1.$$

(iv) If $a_2 \leq b_1$ then

$$\log_+(b_1 - a_2) \leq \int_{a_1}^{a_2}\int_{b_1}^{b_2} \log_+(x-y)\alpha(x)\beta(y)dxdy \leq \log_+(b_2 - a_1).$$

(v)If $A/2 \leq \alpha(x) \leq A$ on $(a_1, a_2)$ then for any $y \in (a_1, a_2)$

$$\log_+|a_2 - a_1| - C \leq \int_{a_1}^{a_2} \log_+(x-y)\alpha(x)dx \leq \log_+|a_2 - a_1| + C$$

for some absolute constant $C$.

(vi) If $A/2 \leq \alpha(x) \leq A$ on $(a_1, a_2)$ then for any $y > a_2$

$$\log_+|y - a_1| - C \leq \int_{\mathrm{a}_1}^{\mathrm{a}_2} \log_+(x-y)\alpha(x)dx \leq \log_+|y - a_1|$$

for some absolute constant $C$.

We discuss a Toeplitz version of theorem 66 from [55].

We say that a finite measure $\mu$ on $\mathbb{R}$ annihilates $K_\theta$ if $\int f d\mu = 0$ for a dense set of $f \in K_\theta$. Note that the integral always exists for a dense set of functions since, for instance, the disk algebra is dense in every $K_\theta$.

We say that the Cauchy integral $K\mu$ is divisible by an inner function $\theta$ if $K\mu/\theta = K\eta$ in $\mathbb{C} \setminus \mathbb{R}$ for some finite complex measure $\eta$ on $\mathbb{R}$.

**Lemma (2.2.19)[84]:** [85] Let $\mu$ be a finite complex measure on $\mathbb{R}$ and let $\theta$ be an inner function in $\mathbb{C}_+$. Then the following statements are equivalent:

(i) $\mu$ annihilates $K_\theta$;

(ii) The Cauchy integral of the conjugate measure $\bar{\mu}, K\bar{\mu}$, is divisible by $\theta$.

**Proof.** (i) $\Rightarrow$ (ii). We will assume that the reproducing kernels of $K_\theta$ belong to the dense set annihilated by $\mu$ (otherwise one needs to use a standard limiting procedure). If $\lambda \in \mathbb{C}_+$then

$$0 = \overline{\frac{1 - \bar{\theta}(\lambda)\theta(z)}{\bar{\lambda} - z}}\, d\mu(z) = \theta(\lambda)K\bar{\theta}\bar{\mu}(\lambda) - K\bar{\mu}(\lambda)$$

which implies the statement.

(ii) $\Rightarrow$ (i). Let $J$ be the inner function whose Clark measure is $|\mu|$. Then the function $f = (1 - J)K\bar{\mu}$ belongs to $K_J$ and is divisible by $\theta$. Let $f = \theta g$. We will assume that $\theta(i) = 0$. The general case can be treated similarly.

Let

$$\theta = \theta^* \frac{z-i}{z+i}.$$

If $I$ is an inner divisor of $\theta^*$ then $Ig/(z+i)$ also belongs to $K_j$ and is summable on the real line. Recall that the integral over $\mathbb{R}$ of a function from $H^1(\mathbb{C}_+)$is 0 . At the same time, by the Clark representation, for each function $h$ from $K_J$,

$$\int_{\mathbb{R}} h(x)dx = \int h(x)d|\mu|(x).$$

Hence

$$0 = \int \frac{(Ig)(x)}{x+i} dx = \int \frac{(Ig)(x)}{x+i} d|\mu|(x) = \int \frac{I(x)\bar{\theta}(x)f(x)}{x+i} d|\mu(x)| = \int \overline{\frac{\theta \bar{I}}{x-\imath}} d\mu.$$

Since functions $\frac{\theta}{I(z-i)}$ are complete in $K_\theta$, we obtain the statement.

Like in de Branges' proof of theorem 66 , we will use the Krein-Milman theorem on the existence of extreme points in a convex set to obtain the following lemma:

**Lemma (2.2.20)[84]:** Let $\theta$ be an inner function in $\mathbb{C}_+$. Let $\mu$ be a finite complex measure whose Cauchy integral $K\mu$ is divisible by $\theta$ (or, equivalently, $\bar{\mu}$ annihilates $K_\theta$ ). Then there exists a finite complex measure $\nu$ such that

(i) $\operatorname{supp}\nu \subset \operatorname{supp}\mu$;

(ii)$K\nu$ is divisible by $\theta$ ( $\bar{\nu}$ annihilates $K_\theta$ );

(iii) K\nu has no zeros outside of $\operatorname{supp}\nu$, except the zeros of $\theta$ in $\mathbb{C}_+$;

(iv) if $\theta$ is a meromorphic inner function then $\nu$ is concentrated on a discrete set.

**Proof.** First, let us symmetrize $\mu$. Since together with any $f \in K_\theta, \theta\bar{f} \in K_\theta$, the measure $\bar{\theta}\mu$, just like $\bar{\mu}$, annihilates $K_\theta$ and $K\theta\bar{\mu}$ is divisible by $\theta$. Consider $\eta = \mu + \theta\bar{\mu}$. WLOG $\| \eta \| \leq 1$.

Denote $\Sigma = \operatorname{supp}\mu$. Let $A_\Sigma^\theta$ be the set of all finite complex measures $\sigma$ such that $\| \sigma \| \leq 1, \operatorname{supp}\sigma \subset \Sigma$, the Cauchy integral of $\sigma$ is divisible by $\theta$ and

$$\theta\bar{\sigma} = \sigma. \quad (89)$$

Since $\eta \in A_\Sigma^\theta$, this set is not empty. It is also convex. By the Krein-Milman theorem it contains a non-zero extremal point $\nu$. We claim that this is the desired measure.

First, let us show that the set of real $L^\infty(|\nu|)$-functions $h$ such that $Kh\nu$ is divisible by $\theta$ is one-dimensional, and therefore $h = c \in \mathbb{R}$. (This is equivalent to the statement that the closure of $K_\theta$ in $L^1(|\nu|)$ has deficiency 1, i.e. the space of its annihilators is one dimensional)

Let there be a bounded real $h$ such that $Kh\nu$ is divisible by $\theta$. WLOG $h \geq 0$, since one can add constants, and $\| h\nu \| = 1$. Choose $0 < \alpha < 1$ so that $|\alpha h| < 1$. Consider probability measures $\nu_1 = h\nu$ and $\nu_2 = (1-\alpha)^{-1}(\nu - \nu_1)$. Then both of them belong to $A_\Sigma^\theta$ and $\nu = \alpha\nu_1 + (1-\alpha)\nu_2$ which contradicts the extremality of $\nu$.

Now let us show that $\nu$ is a singular measure. Let $g$ be a continuous compactly supported real function such that $\int g d\nu = 0$. By the previous part, there exists a sequence $f_n \in K_\theta, f_n \to g$ in $L^1(|\nu|)$ (otherwise the defect is larger than 1). Since $\bar{\nu}$ annihilates $K_\theta$ and $(f_n(z) - f_n(w))/(z-w) \in K_\theta$ for every fixed $w \in \mathbb{C} \setminus \mathbb{R}$,

$$0 = \int \frac{f_n(z) - f_n(w)}{z-w} d\bar{\nu}(z) = Kf_n\bar{\nu}(w) - f_n(w)K\bar{\nu}(w)$$

and therefore

$$f_n = \frac{Kf_n\bar{\nu}}{K\bar{\nu}}.$$

Taking the limit,

$$f = \lim\ f_n = \frac{Kg\bar{\nu}}{K\bar{\nu}}.$$

Since all of $f_n$ have pseudocontinuations, one can show that the limit function $f$ must have one as well. Since the numerator is analytic outside the compact support of $h$, the measure in the denominator must be singular (Cauchy integrals of nonsingular measures have jumps at the real line on the support of the a.c. part).

Moreover, $f$ must be analytically continuable through the real line outside of clos spec $_\theta$, like all of $f_n$. In particular, if $\theta$ is meromorphic, the zero set of $f$ has to be discrete. Since $\nu$ is singular, $K\nu$ tends to $\infty$ at $\nu$-a.e. point and $f = 0$ at $\nu$-a.e. point outside of the support of $g$. Choosing two different $g$ with disjoint supports we prove that if $\theta$ is meromorphic, then $\nu$ is concentrated on a discrete set.

We show that $K\nu$ does not have any zeros in $\mathbb{C} \setminus \operatorname{supp} \nu$ other than the zeros of $\theta$. Let $J$ be the inner function corresponding to $|\nu|(|\nu|$ is the Clark measure for $J)$. Denote $G = (1 - J)K\nu \in K_J$. Since $K\nu$ is divisible by $\theta$ and $K|\nu|$ is outer, $G$ is divisible by $\theta$. Let us first show that $G/\theta$ does not have an inner component in the upper half-plane. Suppose that $G = \theta U H$ for some inner $U$. Then $\theta(1 + U)^2 H$ also belongs to $K_J$. Denote

$$\gamma = \theta(1 + U)^2 H|\nu|.$$

Then $\gamma = h\nu$ for a bounded non-constant function $h = (1 + U)^2/U$. The Cauchy integral of $\gamma$ is divisible by $\theta$ because $(1 - J)K\gamma = \theta(1 + U)^2 H$. We obtain a contradiction with the property that the space of annihilators is one dimensional.

Thus $G/\theta \in K_J$ is outer in $C_+$. By (89),

$$G\bar{\theta}|\nu| = \bar{\theta}\nu = \bar{\nu}$$

and the Clark representation formula implies

$$\bar{J}G = \overline{(1 - J)K\bar{\nu}} = \overline{(1 - J)KG\bar{\theta}|\nu|} = \overline{G/\theta},$$

so the pseudocontinuation of $G$ does not have zeros in $\mathbb{C}_-$.

If $G$ has a zero at $x = a \in \mathbb{R}$ outside of $\operatorname{spec}_J$ then

$$\frac{G}{x - a} \in K_J$$

and the measure

$$\gamma = \frac{G}{x - a}|\nu|$$

leads to a similar contradiction with the property that the space of annihilators is one-dimensional, since $(x - a)^{-1}$ is bounded on the support of $\nu$. Since

$$G = (1 - J)K\nu \in K_J,$$

$K\nu$ does not have any extra zeros.

Recall that a function $f \in N[\phi]$ is called purely outer if $f$ is outer in the upper half-plane and $\phi f = \bar{g}$ is outer in the lower half plane.

**Corollary (2.2.21)[84]:** Let $I, \theta$ be inner functions in $\mathbb{C}_+$. Suppose that the kernel $N[\bar{I}\theta]$ is non-trivial.

Then there exists an inner function $J$ in $\mathbb{C}_+$such that $\operatorname{spec}_J \subset \operatorname{spec}_I$ and the kernel $N[\bar{J}\theta]$ contains a purely outer function $f$ that does not have any zeros on $\mathbb{R} \setminus \operatorname{spec}_J$. If $\sigma_1$ is the Clark measure of $J$ then $f$ is also non-zero $\sigma_1$-a.e. on $\operatorname{spec}_J$. If $\theta$ is a meromorphic function, then $J$ can be chosen as a meromorphic function.

**Section (2.3): The Hilbert Transform: A Real Variable Characterization**

Define a truncated Hilbert transform of a locally bounded signed measure $v$ by

$$\mathrm{H}_{\epsilon,\delta}v(\mathrm{x}) := \int_{\epsilon<|y-\mathrm{x}|<\delta} \frac{\mathrm{d}v(\mathrm{y})}{\mathrm{y}-\mathrm{x}},\ 0<\epsilon<\delta.$$

Given weights (i.e. locally bounded positive Borel measures) $\sigma$ and $w$ on the real line $\mathbb{R}$, we consider the following two weight norm inequality for the Hilbert transform,

$$\sup_{0<\epsilon<\delta} \int_{\mathbb{R}} \left|\mathrm{H}_{\epsilon,\delta}(f\sigma)\right|^2 \mathrm{d}w \le \mathcal{N}^2 \int_{\mathbb{R}} |f|^2 d\sigma,\ f\in \mathrm{L}^2(\sigma), \quad (90)$$

where $\mathcal{N}$ is the best constant in the inequality, uniform over all truncations of the Hilbert transform kernel. Below, we will write the inequality above as $\| \mathrm{H}(f\sigma) \|_{L^2(w)} \le \mathcal{N} \| \mathrm{f} \|_{\mathrm{L}^2(w)}$, that is the uniformity over the truncation parameters is suppressed.

The primary question is to find a real variable characterization of this inequality, and the theorem below is an answer to the beautiful conjecture of Nazarov-Treil-Volberg, see [130]. Set

$$\mathrm{P}(\sigma,\mathrm{I}) := \int_{\mathbb{R}} \frac{|\mathrm{I}|}{\mathrm{I}|^2 + \mathrm{dist}(\mathrm{x},\mathrm{I})^2}\sigma(\mathrm{d}x),$$

which is approximately the Poisson extension of $\sigma$ to the upper half plane, evaluated at $(\mathrm{x}_1, |\mathrm{I}|)$, where $\chi_I$ is the center of $I$.

**Theorem (2.3.1)[103]:** Let $\sigma$ and $w$ be locally finite positive Borel measures on the real line $\mathbb{R}$ with no common point masses. Then, the two weight inequality (90) holds if and only if these three conditions hold uniformly over all intervals I,

$$\mathrm{P}(\sigma,\mathrm{I})\mathrm{P}(w,\mathrm{I}) \le \mathcal{A}_2, \quad (91)$$

$$\int_I |\mathrm{H}(\mathbf{1}_1\sigma)|^2\,\mathrm{d}w \le \mathcal{T}^2\sigma(\mathrm{I}),\ \int_I |\mathrm{H}(\mathbf{1}_1 w)|^2\,\mathrm{d}\sigma \le \mathcal{T}^2 w(\mathrm{I}). \quad (92)$$

There holds

$$\mathcal{N} \approx \mathcal{A}_2^{\frac{1}{2}} + \mathcal{J} =: \mathcal{H}, \quad (93)$$

where $\mathcal{A}_2$ and $\mathcal{J}$ are the best constants in the inequalities above.

It is well known [130] that the $A_2$ condition is necessary for the norm inequality, and the inequalities (92) are obviously necessary, thus the content of the Theorem is the sufficiency of the $A_2$ and testing inequalities. We will carry out a 'global to local' reduction in the proof of sufficiency, with the analysis of the 'local' problem being carried out in part II of this series [108].

The Nazarov-Treil-Volberg conjecture has only been verified before under additional hypotheses on the pair of weights, hypotheses which are not necessary for the two weight inequality. The so-called pivotal condition of [130] is not necessary, as was proved in [111]. The pivotal condition is still an interesting condition: It is all that is needed to characterize the boundedness of the Hilbert transform, together with the maximal function in both directions. But, the boundedness of this triple of operators is decoupled in the two weight setting [125].

Our argument has these attributes. Certain degeneracies of the pair of weights must be addressed, the contribution of the innovative 2004 of Nazarov-Treil-Volberg [119], also see [130], which was further sharpened with the property of energy in [111], a crucial property of the Hilbert transform. This theme is further developed herein, with notion of functional energy.

The proof should proceed through the analysis of the bilinear form $\langle \mathrm{H}(\sigma f), gw\rangle$, as one expects certain paraproducts to appear. Still, the paraproducts have no canonical form, suggesting that the proof be highly non-linear in $f$ and $g$. The non-linear point of view was initiated in [112]. A particular feature of our arguments is a repeated appeal to certain quasiorthogonality arguments, providing (many) simplifications over prior arguments. For instance, we never find ourselves constructing auxiliary measures, and verifying that they are Carleson, a frequent step in many related arguments.
One can phrase a two weight inequality question for any operator T, a question that became apparent with the foundational of Muckenhoupt [113] on $A_p$ weights for the maximal function. Indeed, the case of Hardy's inequality was quickly resolved by Muckenhoupt [114]. The maximal function was resolved by one of us [127], as well as the fractional integrals, and, essential, Poisson integrals [128]. The latter established a result which closely paralleled the contemporaneous $T1$ theorem of David and Journé [104]. This connection, fundamental in nature, was not fully appreciated until the innovative work of Nazarov-Treil-Volberg [116], [117], [118] in developing a non-homogeneous theory of singular integrals. The two weight problem for dyadic singular integrals was only resolved recently [120]. Partial information about the two weight problem for singular integrals [123] was basic to the resolution of the $A_2$ conjecture [106], and several related results [107], [109], [123], [124]. The result is the first real variable characterization of a two weight inequality for a continuous singular integral.
Interest in the two weight problem for the Hilbert transform arises from its natural occurrence in questions related to operator theory [122], [126], spectral theory [122], and model spaces [100], and analytic function spaces [71]. In operator theory Sarason posed the conjecture (See [105].) that the Hilbert transform would be bounded if the pair of weights satisfied the (full) Poisson $A_2$ condition. This was disproved by Nazarov [115]. Advances on these questions have been linked to finer understanding of the two weight question, see for instance [121], [122], which build upon Nazarov's counterexample.

We have stated the main theorem with 'hard' cut-offs in the truncation of the Hilbert transform. There are many possible variants in the choice of truncation, moreover the proof of sufficiency requires a different choice of truncation.
Consider a truncation given by

$$\tilde{H}_{\alpha,\beta}(\sigma f)(x) := \int f(y) K_{\alpha,\beta}(y-x)\sigma(dy)$$

where $\mathrm{K}_{\alpha,\beta}(\mathrm{y})$ is chosen to minimize the technicalities associated with off-diagonal considerations. Specifically, set $\mathrm{K}_{\alpha,\beta}(0)=0$, and otherwise $\mathrm{K}_{\alpha,\beta}(\mathrm{y})$ is odd and for $\mathrm{y}>0$

$$K_{\alpha,\beta}(y) := \begin{cases} -\dfrac{y}{\alpha^2}+\dfrac{2}{\alpha} & 0<y<\alpha, \\ \dfrac{1}{y} & \alpha \le y \le \beta, \\ -\dfrac{y}{\beta^2}+\dfrac{2}{\beta} & \beta<y<2\beta, \\ 0 & 2\beta \le y. \end{cases}$$

This is a $\mathrm{C}^1$ function on $(0,2\beta)$, and is Lipschitz, convex and monotone on $(0,\infty)$. We now argue that we can use these truncations in the proof of the sufficiency bound of our main theorem.

**Proposition (2.3.2)[103]:** If the pair of weights $\sigma, w$ satisfy the $A_2$ bound (91), then, one has the uniform norm estimate with the 'hard' truncations (90) if and only if one has uniform norm estimate for the 'smooth' truncations,

$$\sup_{0<\alpha<\beta} \|\tilde{H}_{\alpha,\beta}(\sigma f)\|_w \leq \mathcal{N} \parallel f \parallel_\sigma.$$

Indeed, $\left|\mathrm{H}_{\alpha,\beta}(\sigma f) - \tilde{\mathrm{H}}_{\alpha,\beta}(\sigma f)\right| \lesssim A_\alpha(\sigma|f|) + A_\beta(\sigma|f|)$, where these last two operators are 'single-scale' averages, namely

$$A_\alpha(\sigma\phi)(x) = \alpha^{-1}\int_{(x-3\alpha,x+3\alpha)} \phi(y)\sigma(dy).$$

But, the (simple) $A_2$ bound is all that is needed to provide a uniform bound on the operators $A_\alpha(\sigma\phi)$. So the proposition follows.

We use the truncations $\bar{H}_{\alpha,\beta}$, and we suppress the tilde in the notation. The particular choice of truncation is motivated by this off-diagonal estimate on the kernels.

**Proposition (2.3.3)[103]:** Suppose that $2|x-x'| < |x-y|$, then

$$\mathrm{K}_{\alpha,\beta}(y-x') - \mathrm{K}_{\alpha,\beta}(y-x) = C_{x,x',y}\frac{x'-x}{(y-x)(y-x')},$$

$$\text{where } C_{x,x',y} = 1\ 2\alpha < |x-y| < \frac{1}{2}\beta, \tag{94}$$

and is otherwise positive and never more than 4.

**Proof.** The assumptions imply that $y-x'$ and $y-x$ have the same sign. Assume, without loss of generality that $0 < y-x' < y-x$. If $2\alpha < |x-y| < \frac{1}{2}\beta$, it follows that $\alpha < |x'-y| < \beta$, and so by the definition

$$\mathrm{K}_{\alpha,\beta}(\mathrm{y}-\mathrm{x}') - \mathrm{K}_{\alpha,\beta}(\mathrm{y}-\mathrm{x}) = \frac{1}{\mathrm{y}-x'} - \frac{1}{\mathrm{y}-\mathrm{x}} = \frac{\mathrm{x}'-\mathrm{x}}{(\mathrm{y}-\mathrm{x})(\mathrm{y}-\mathrm{x}')}.$$

And, in the general case, there holds $\left|\frac{\mathrm{d}}{\mathrm{dt}}\mathrm{K}_{\alpha,\beta}(\mathrm{t})\right| \leq 4\mathrm{t}^{-2}$, so that

$$0 \leq K_{\alpha,\beta}(y-x') - K_{\alpha,\beta}(y-x) \leq \int_{y-x'}^{y-x} \frac{4}{t^2}dt = 4\frac{x'-x}{(y-x)(y-x')}.$$

A collection of intervals $\mathcal{G}$ is a grid if for all $\mathrm{G}, \mathrm{G}' \in \mathcal{G}$, we have $\mathrm{G} \cap \mathrm{G}' \in \{\emptyset, \mathrm{G}, \mathrm{G}'\}$. By a dyadic grid we mean a grid $\mathcal{D}$ of intervals of $\mathbb{R}$ such that for each interval $\mathrm{I} \in \mathcal{D}$, the subcollection $\{\mathrm{I}' \in \mathcal{D}: |\mathrm{I}'| = |\mathrm{I}|\}$ partitions $\mathbb{R}$, aside from endpoints of the intervals. In addition, the left and right halves of $I$, denoted by $I_\pm$, are also in $\mathcal{D}$.

For $I \in \mathcal{D}$, the left and right halves $I_\pm$are referred to as the children of $I$. We denote by $\pi_{\mathcal{D}}I$ the unique interval in $\mathcal{D}$ having I as a child, and we refer to $\pi_{\mathcal{D}}I$ as the $\mathcal{D}$-parent of $I$.

We will work with subsets $\mathcal{F} \subset \mathcal{D}$. We say that I has $\mathcal{F}$-parent $\pi_{\mathcal{F}}\mathrm{I} = \mathrm{F}$ if $\mathrm{F} \in \mathcal{F}$ is the minimal element of $\mathcal{F}$ that contains I.

Let $\sigma$ be a weight on $\mathbb{R}$, one that does not assign positive mass to any endpoint of a dyadic grid $\mathcal{D}$. If $\mathrm{I} \in \mathcal{D}$ is such that $\sigma$ assigns non-zero weight to both children of I, the associated Haar function is

$$\mathrm{h}_1^\sigma := \sqrt{\frac{\sigma(\mathrm{I}_-)\sigma(\mathrm{I}_+)}{\sigma(\mathrm{I})}}\left(-\frac{\mathrm{I}_-}{\sigma(\mathrm{I}_-)} + \frac{\mathrm{I}_+}{\sigma(\mathrm{I}_+)}\right).$$

In this definition, we are identifying an interval with its indicator function, and we will do so throughout the remainder. This is an $L^2(\sigma)$-normalized function, and has $\sigma$-integral zero.

For any dyadic interval $I_0$, it holds that $\{\sigma(I_0)^{-1/2}I_0\} \cup \{h_1^\sigma : I \in \mathcal{D}, I \subset I_0\}$ is an orthogonal basis for $L^2(I_0,\sigma)$.
We will use the notations $\hat{f}(\mathrm{I}) = \langle \mathrm{f}, \mathrm{h}_1^\sigma \rangle_\sigma$, as well as

$$\Delta_1^\sigma f = \langle f, h_1^\sigma \rangle_\sigma h_1^\sigma = I_+ \mathbb{E}_{1_+}^\sigma f + I_- \mathbb{E}_{1\_}^\sigma f - I \mathbb{E}_1^\sigma f.$$

The second equality is the familiar martingale difference equality, and so we will refer to $\Delta_1^\sigma f$ as a martingale difference. It implies the familiar telescoping identity $\mathbb{E}_\mathrm{J}^\sigma \mathrm{f} = \sum_1 : \mathrm{I}_2\ \mathrm{J}^\sigma \mathbb{E}_\mathrm{J}^\sigma \mathrm{f}$.
For any function the Haar support of $f$ is the collection $\{I \in \mathcal{D}: \hat{f}(I) \neq 0\}$.

With a choice of dyadic grid $\mathcal{D}$ understood, we say that $J \in \mathcal{D}$ is $(\epsilon, r)$-good if and only if for all intervals $I \in \mathcal{D}$ with $|I| \geq 2^{r-1}|J|$, the distance from $J$ to the boundary of either child of $I$ is at least $|J|^\epsilon |I|^{1-\epsilon}$.
For $f \in L^2(\sigma)$ we set $P_{\mathrm{good}}^\sigma f = \sum_1 \sum_{\substack{1\in\mathcal{D} \\ (\epsilon,\mathrm{r})-\mathrm{good}}} \Delta_1^\sigma f$. The projection $P_{\mathrm{good}}^w g$ is defined similarly. To make the two reductions below, one must make a random selection of grids, as is detailed in [111], [130]. The use of random dyadic grids has been a basic tool since the foundational work of [116], [117], [118]. Important elements of the suppressed construction of random grids are that
(i) It suffices to consider a single dyadic grid $\mathcal{D}$.
(ii) For any fixed $0 < \epsilon < \frac{1}{2}$, we can choose integer r sufficiently large so that it suffices to consider $f$ such that $f = P_{\mathrm{good}}^\sigma f$, and likewise for $g \in L^2(w)$. Namely, it suffices to estimate the constant below, for arbitrary dyadic grid $\mathcal{D}$,

$$|\langle \mathrm{H}_\sigma \mathrm{f}, \mathrm{g} \rangle_w| \leq \mathcal{N}_{\mathrm{good}} \parallel \mathrm{f} \parallel_\sigma \parallel \mathrm{g} \parallel_w,$$

where it is required that $f = P_{\mathrm{good}}^\sigma \in L^2(\sigma)$ and $g = P_{\mathrm{good}}^w \in L^2(w)$.
That the functions are good is, at some moments, an essential property.
A reduction, using randomized dyadic grids, allows one the extraordinarily useful reduction in the next Lemma. This is a well-known reduction, due to Nazarov-Treil-Volberg, explained in full detail in the current setting, in [119]. Below, $\mathcal{H}$ is as in (93), the normalized sum of the $A_2$ and testing constants.
**Lemma (2.3.4)[103]:** For all sufficiently small €, and sufficiently large r, this holds. Suppose that for any dyadic grid $\mathcal{D}$, such that no endpoint of an interval I $\in \mathcal{D}$ is a point mass for $\sigma$ or $w$, [1] there holds

$$\left|\left\langle \mathrm{H}_\sigma \mathrm{P}_{\mathrm{good}}^\sigma \mathrm{f}, \mathrm{P}_{\mathrm{good}}^w\ \mathrm{g} \right\rangle_w\right| \lesssim \mathcal{H} \parallel \mathrm{f} \parallel_\sigma \parallel \mathrm{g} \parallel_w. \tag{95}$$

Then, the same inequality holds without the projections $P_{\mathrm{good},}^\sigma$, and $P_{\mathrm{good}}^w$.
Inequality (95) should be understood as an inequality, uniform over the class of smooth truncations of the Hilbert transform. But, we can suppress this in the notation without causing confusion. The bilinear form only needs to be controlled for $(\epsilon, r)$-good functions $f$ and $g$, goodness being defined with respect to a fixed dyadic grid. Suppressing the notation, we write 'good' for ' $(\epsilon, r)$-good,' and it is always assumed that the dyadic grid $\mathcal{D}$ is fixed, and only good intervals are in the Haar support of $f$ and $g$, though is also suppressed in the notation.

We reduce the analysis of the bilinear form in (95) to the local estimate, (96). It is sufficient to assume that $f$ and $g$ are supported on an interval $I^0$; by trivial use of the interval testing condition, we can further assume that $f$ and $g$ are of integral zero in their respective spaces. Thus, $f$ is in the linear span of (good) Haar functions $h_1^{\sigma}$ for $I \subset I^0$, and similarly for $g$, and

$$\langle \mathrm{H}_\sigma \mathrm{f}, \mathrm{g}\rangle_w = \sum_{1,\, \mathrm{J}:1,\, \mathrm{J}\subset 1^{\circ}} \left\langle \mathrm{H}_\sigma \Delta_{\mathrm{I}}^{\sigma} \mathrm{f}, \Delta_{\mathrm{J}}^{w}\, \mathrm{g}\right\rangle_w .$$

The argument is independent of the choice of truncation that implicitly appears in the inner product above.

The double sum is broken into different summands. Many of the resulting cases are elementary, and we summarize these estimates as follows. Define the bilinear form

$$B^{\text{above}}(f,g) := \sum_{I: I\subset I^0} \sum_{J: J e I_J} \mathbb{E}_{I_J}^{\sigma} \Delta_I^{\sigma} f \cdot \left\langle H_\sigma I_J, \Delta_J^{w} g\right\rangle_w$$

where here and throughout, $J \Subset I$ means $J \subset I$ and $2^r|J| \leq |I|$. In addition, the argument of the Hilbert transform, $\mathrm{I_J}$, is the child of I that contains J, so that $\Delta_{\mathrm{I}}^{\sigma}\mathrm{f}$ is constant on $\mathrm{I_J}$. Define $\mathrm{B}^{\text{below}}(f,g)$ in the dual fashion.

**Lemma (2.3.5)[103]:** There holds, with the notation of (93),

$$\left|\langle \mathrm{H}_\sigma \mathrm{f}, \mathrm{g}\rangle_w - \mathrm{B}^{\text{above}}(\mathrm{f},\mathrm{g}) - \mathrm{B}^{\text{below}}(\mathrm{f},\mathrm{g})\right| \leqslant \mathcal{H} \parallel \mathrm{f} \parallel_\sigma \parallel \mathrm{g} \parallel_w .$$

This is a common reduction in a proof of a Tl theorem, and in the current context, it only requires goodness of intervals and the $A_2$ condition. For a proof, one can consult [119], [130]. The Lemma is specifically phrased and proved in this way in [112].

These definitions are needed to phrase the global to local reduction. The following definition depends upon the essential energy inequality (106).

**Definition (2.3.6)[103]:** Given any interval $F_0$, define $\mathcal{F}_{\text{energy}}(F_0)$ to be the maximal subintervals $F \subseteq F_0$ such that

$$\mathrm{P}(\sigma \mathrm{F}_0, \mathrm{F})^2 \mathrm{E}(w, \mathrm{F})^2 w(\mathrm{F}) > 10 \mathrm{C}_0 \mathcal{H}^2 \sigma(\mathrm{F}),$$

where $E(w,F)$ is defined in (105), and $C_0$ is the constant in Proposition (2.3.13). There holds $(\cup \{F : F \in \mathcal{F}(F_0)\}) \leq \frac{1}{10}\sigma(F_0)$.

**Definition (2.3.7)[103]:** Let $\mathrm{I}_0$ be an interval, and let $\mathcal{S}$ be a collection of disjoint intervals contained in $\mathrm{I}_0$. A function $\mathrm{f} \in \mathrm{L}_0^2(\mathrm{I}_0, \sigma)$ is said to be uniform (w.r.t. ) if these conditions are met:

(i) Each energy stopping interval $F \in \mathcal{F}_{\text{energy}}(I_0)$ is contained in some $S \in \mathcal{S}$.

(ii) The function $f$ is constant on each interval $S \in \mathcal{S}$.

(iii) For any interval $I \subset I_0$ which is not contained in any $S \in \mathcal{S}, \mathbb{E}_{\mathrm{I}}^{\sigma}|f| \leq 1$.

We will say that g is weakly adapted to a function f uniform w.r.t. $\mathcal{S}$, if $\mathrm{J} \Subset \mathrm{S}$ for some interval $\mathrm{S} \in \mathcal{S}$ implies that $\left\langle \mathrm{g}, \mathrm{h}_{\mathrm{J}}^{w}\right\rangle_w = 0$. We will also say that g is weakly adapted to $\mathcal{S}$.

The constant $\mathcal{L}$ is defined as the best constant in the local estimate:

$$\left|B^{\text{above}}(f,g)\right| \leq \mathcal{L}\left\{\sigma(I_0)^{\frac{1}{2}} + \parallel f \parallel_\sigma\right\} \parallel g \parallel_w , \qquad (96)$$

where $f, g$ are of mean zero on their respective spaces, supported on an interval $I_0$. Moreover, $f$ is uniform and $g$ is weakly adapted to $f$. The inequality above is homogeneous in $g$, but not $f$, since the term $\sigma(\mathrm{I}_0)^{1/2}$ is motivated by the bounded averages property of f.

A reduction of this type is a familiar aspect of many proofs of a T1 theorem, proved by exploiting standard off-diagonal estimates for Calderón-Zygmund kernels, but in the current setting, it is a much deeper fact, a consequence of the functional energy inequality. We make the following construction for an $f \in L^2(I^0, \sigma)$, of $\sigma$-integral zero. Add $I^0$ to $\mathcal{F}$, and set $\alpha_f(I^0) := \mathbb{E}_{10}^{\sigma}|f|$. In the inductive stage, if $\mathrm{F} \in \mathcal{F}$ is minimal, add to $\mathcal{F}$ those maximal descendants $\mathrm{F}'$ of F such that $\mathrm{F}' \in \mathcal{F}_{\text{energy}}(\mathrm{F})$ or $\mathrm{E}_{\mathrm{F}'}^{\sigma}|\mathrm{f}| \geq 10\alpha_{\mathrm{f}}(\mathrm{F})$. Then define

$$\alpha_f(F') := \begin{cases} \alpha_f(F) & \mathbb{E}_{\mathrm{F}'}^{\sigma}|f| < 2\alpha_f(F) \\ \mathbb{E}_{\mathrm{F}'}^{\sigma}|f| & \text{otherwise} \end{cases}$$

If there are no such intervals $\mathrm{F}'$, the construction stops. We refer to $\mathcal{F}$ and $\alpha_f(\cdot)$ as CalderónZygmund stopping data for f, following the terminology of [112]. Their key properties are collected here.

**Lemma (2.3.8)[103]:** For $\mathcal{F}$ and $\alpha_f(\cdot)$ as defined above, there holds

(i) $I_0$ is the maximal element of $\mathcal{F}$.

(ii) For all $\mathrm{I} \in \mathcal{D}, \mathrm{I} \subset \mathrm{I}^0$, we have $\mathbb{E}_1^{\sigma}|\mathrm{f}| \leq 10\alpha_{\mathrm{f}}(\pi_{\mathcal{F}}\mathrm{I})$.

(iii) $\alpha_{\mathrm{f}}$ is monotonic: If $\mathrm{F}, \mathrm{F}' \in \mathcal{F}$ and $\mathrm{F} \subset \mathrm{F}'$ then $\alpha_{\mathrm{f}}(\mathrm{F}) \geq \alpha_{\mathrm{f}}(\mathrm{F}')$.

(iv) The collection $\mathcal{F}$ is $\sigma$-Carleson in that

$$\sum_{\mathrm{F} \in \mathcal{F}: \mathrm{F} \subset \mathrm{S}} \sigma(\mathrm{F}) \leq 2\sigma(\mathrm{S}),\ \mathrm{S} \in \mathcal{D}. \tag{97}$$

(v) We have the inequality

$$\left\| \sum_{\mathrm{F} \in \mathcal{F}} \alpha_{\mathrm{f}}(\mathrm{F}) \cdot \mathrm{F} \right\|_{\sigma} \lesssim \| \mathrm{f} \|_{\sigma}. \tag{98}$$

**Proof.** The first three properties are immediate from the construction. The fourth, the $\sigma$-Carleson property is seen this way. It suffices to check the property for $S \in \mathcal{F}$. Now, the $\mathcal{F}$-children can be in $\mathcal{F}_{\text{energy}}(S)$, which satisfy

$$\sum_{F' \in \mathcal{F}_{\text{energgy}}(S)} \sigma(F') \leq \frac{1}{10}\sigma(S).$$

Otherwise, note that by choice of $\alpha_f(\cdot)$, we have $\mathbb{E}_S^{\sigma}|f| \leq 2\alpha_f(S)$. These intervals $F'$, satisfy $\mathbb{E}_{\mathrm{F},}^{\sigma}|\mathrm{f}| \geq 10\alpha_{\mathrm{f}}(\mathrm{S}) \geq 5\mathbb{E}_{\mathrm{S}}^{\sigma}|\mathrm{f}|$. These intervals satisfy the display above with $\frac{1}{10}$ replaced by $\frac{1}{5}$. Hence, (97) holds.

For the final property, let $\mathcal{G} \subset \mathcal{F}$ be the subset at which the stopping values change: If $F \in \mathcal{F} - \mathcal{G}$, and G is the $\mathcal{G}$-parent of F, then $\alpha_{\mathrm{f}}(\mathrm{F}) = \alpha_{\mathrm{f}}(\mathrm{G})$. Set

$$\Phi_{\mathrm{G}} := \sum_{\mathrm{F} \in \mathcal{F}: \pi_{\mathcal{G}}\mathrm{F} = \mathrm{G}} \mathrm{F}.$$

Define $G_k := \{\Phi_G \geq 2^k\}$, for $k = 0,1, \ldots$. The $\sigma$-Carleson property implies integrability of all orders in $\sigma$-measure of $\Phi_G$. Using the third moment, we have $\sigma(G_k) \lesssim 2^{-3k}\sigma(G)$. Then, estimate

$$\left\|\sum_{\mathrm{F}\in\mathcal{F}} \alpha_{\mathrm{f}}(\mathrm{F})\cdot \mathrm{F}\right\|_{\sigma}^{2} = \left\|\sum_{\mathrm{G}\in\mathcal{G}} \alpha_{\mathrm{f}}(\mathrm{G})\Phi_{\mathrm{G}}\right\|_{\sigma}^{2}$$

$$\leq \left\|\sum_{\mathrm{k}=0}^{\infty} (\mathrm{k}+1)^{+1-1}\sum_{\mathrm{G}\in\mathcal{G}} \alpha_{\mathrm{f}}(\mathrm{G})2^{\mathrm{k}}1_{\mathrm{G_k}}\right\|_{\sigma}^{2}$$

$$\overset{*}{\lesssim} \sum_{\mathrm{k}=0}^{\infty} (\mathrm{k}+1)^{2}\left\|\sum_{\mathrm{G}\in\mathcal{G}} \alpha_{\mathrm{f}}(\mathrm{G})2^{\mathrm{k}}1_{\mathrm{G_k}}(\mathrm{x})\right\|_{\sigma}^{2}$$

$$\overset{**}{\lesssim} \sum_{\mathrm{k}=0} (\mathrm{k}+1)^{2}\sum_{\mathrm{G}\in\mathcal{G}} \alpha_{\mathrm{f}}(\mathrm{G})^{2}2^{2\mathrm{k}}\sigma(\mathrm{G_k})$$

$$\lesssim \sum_{G\in\mathcal{G}} \alpha_{f}(G)^{2}\sigma(G) \lesssim \| Mf \|_{\sigma}^{2} \lesssim \| f \|_{\sigma}^{2}.$$

Note that we have used Cauchy-Schwarz in k at the step marked by an $*$. In the step marked with $**$, for each point $x$, the non-zero summands are a (super)-geometric sequence of scalars, so the square can be moved inside the sum. Finally, we use the estimate on the $\sigma$-measure of $G_k$, and compare to the maximal function Mf to complete the estimate.

We will use the notation

$$P_F^{\sigma} f := \sum_{I\in\mathcal{D}:\pi_F I=F} \Delta_I^{\sigma} f,\ F\in\mathcal{F}.$$

and similarly for $\mathrm{Q}_{\mathrm{F}}^{w}$, but rather than use $\pi_{\mathcal{F}}\mathrm{J}$, in the definition, we use $\dot{\pi}_{\mathcal{F}}\mathrm{J}$, defined to be the minimal $F\in\mathcal{F}$ with $J \Subset F$. Without this alternate definition, some delicate case analysis would be forced upon us. The inequality (98) allows us to estimate

$$\sum_{\mathrm{F}\in\mathcal{F}} \left\{\alpha_{\mathrm{f}}(\mathrm{F})\sigma(\mathrm{F})^{1/2}+\|\mathrm{P}_{\mathrm{F}}^{\sigma}\mathrm{f}\|_{\sigma}\right\}\|Q_{\mathrm{F}}^{w} g\|_{w}$$

$$\leq \left[\sum_{\mathrm{F}\in\mathcal{F}} \left\{\alpha_{\mathrm{f}}(\mathrm{F})^{2}\sigma(\mathrm{F}) + \|\mathrm{P}_{\mathrm{F}}^{\sigma}\mathrm{f}\|_{\sigma}^{2}\right\}\times\sum_{\mathrm{F}\in\mathcal{F}} \|\mathrm{Q}_{\mathrm{F}}^{w} g\|_{w}^{2}\right]^{\frac{1}{2}} \lesssim \| \mathrm{f} \|_{\sigma} \| \mathrm{g} \|_{w}. \quad (99)$$

We will refer to this as the quasi-orthogonality argument, and we remark that it only requires orthogonality of the projections $Q_F^w g$. It is very useful.

**Lemma (2.3.9)[103]:** There holds

$$\left|B^{\text{above}}(f,g) - B_{\mathcal{F}}^{\text{above}}(f,g)\right| \lesssim \mathcal{H} \| f \|_{\sigma} \| g \|_{w},$$

$$\text{where } B_{\mathcal{F}}^{\text{above}}(f,g) := \sum_{F\in\mathcal{F}} B^{\text{above}}(P_F^{\sigma} f, Q_F^{w} g).$$

**Proof.** We apply functional energy. Observe that $f = \sum_{F\in\mathcal{F}} P_F^{\sigma} f$, and

$$\sum_{\mathrm{J}:\mathrm{J}\Subset\mathrm{I}_0} \Delta_{\mathrm{J}}^{w}\ \mathrm{g} = \sum_{\mathrm{F}\in\mathcal{F}} \mathrm{Q}_{\mathrm{F}}^{w}\ \mathrm{g}.$$

From the definition of $B^{\text{above}}(f,g)$, we can assume that $g$ equals the sum above. Therefore,

$$B^{\text{above}}(f,g) = \sum_{\mathrm{F}'\in\mathcal{F}}\sum_{\mathrm{F}\in\mathcal{F}} B^{\text{above}}\left(P_{F'}^{\sigma}, f, Q_F^{w} g\right).$$

In the sum above, we can also add the restriction that $F' \cap F \neq \emptyset$, for otherwise $B^{\text{above}}(P_{\mathrm{F}}^{\sigma}, f, Q_{\mathrm{F}}^{w} g) = 0$. For a pair of intervals $J \Subset I_J$, note that this implies that $J \Subset \pi_{\mathcal{F}} I$, that is $\pi_{\mathcal{F}} J \subset \pi_{\mathcal{F}} I$. Therefore, we can add the restriction $F \subset F'$. The case of $F' = F$ is the definition of $B_{\mathcal{F}}^{\text{above}}(f, g)$, so that it suffices to estimate

$$\sum_{\substack{\mathrm{F}, \mathrm{F}' \in \mathcal{F} \\ \mathrm{F}' \supsetneq \mathrm{F}}} B^{\text{above}}\left(P_{F'}^{\sigma}, f, Q_F^W g\right). \tag{100}$$

Observe that the functions $g_F := Q_F^w g$ are $\mathcal{F}$ adapted in the sense of Definition (2.3.15), and by construction $\mathcal{F}$ satisfies the Carleson measure condition (97). We take these steps to apply functional energy inequality. The argument of the Hilbert transform is $\mathrm{I_F}$, the child of I that contains $F$. Write $I_F = F + (I_F - F)$, and use linearity of $H_\sigma$. Note that by the standard martingale difference identity and the construction of stopping data,

$$\left|\sum_{I : I_Z F} \mathbb{E}_{I_F}^{\sigma} \Delta_I^{\sigma} f\right| \lesssim \alpha_f(F),\ F \in \mathcal{F}.$$

Hence, invoking interval testing,

$$\begin{aligned}\left|\sum_{\mathrm{F} \in \mathcal{F}} \sum_{\mathrm{I} : \mathrm{I} \supsetneq \mathrm{F}} \mathbb{E}_{\mathrm{I_F}}^{\sigma} \Delta_{\mathrm{I}}^{\sigma} \mathrm{f} \cdot \langle \mathrm{H}_\sigma \mathrm{F}, \mathrm{g_F} \rangle_w\right| &\lesssim \sum_{\mathrm{F} \in \mathcal{F}} \alpha_{\mathrm{f}}(\mathrm{F}) |\langle \mathrm{H}_\sigma \mathrm{F}, \mathrm{g_F} \rangle_w| \\ &\lesssim \mathcal{H} \sum_{\mathrm{F} \in \mathcal{F}} \alpha_{\mathrm{f}}(\mathrm{F}) \sigma(\mathrm{F})^{\frac{1}{2}} \|\mathrm{g_F}\|_w.\end{aligned}$$

Quasi-orthogonality bounds this last expression.
For the second expression, when the argument of the Hilbert transform is $\mathrm{I_F} - \mathrm{F}$, first note that

$$\left|\sum_{\mathrm{I} : \mathrm{I} \geq \mathrm{F}} \mathbb{E}_{\mathrm{I_F}}^{\sigma} \Delta_{\mathrm{I}}^{\sigma} \mathrm{f} \cdot (\mathrm{I_F} - \mathrm{F})\right| \lesssim \Phi := \sum_{\mathrm{F}' \in \mathcal{F}} \alpha_{\mathrm{f}}(\mathrm{F}') \cdot \mathrm{F}',\ \mathrm{F} \in \mathcal{F}.$$

Therefore, by the definition of $\mathcal{F}$-adapted, the monotonicity property (104) applies, and yields

$$\left|\sum_{I : I_Z F} \mathbb{E}_{I_F}^{\sigma} \Delta_I^{\sigma} f \cdot \langle H_\sigma(I_F - F), g_F \rangle_w\right| \lesssim \sum_{J \in \mathcal{J}^*(F)} P(\Phi\sigma, J) \left\langle \frac{x}{|J|}, J \bar{g}_F \right\rangle_w,\ F \in \mathcal{F}.$$

Here, $\mathcal{J}^*(\mathrm{F})$ are the maximal good intervals $\mathrm{J} \Subset \mathrm{F}$, and $\bar{\mathrm{g}}_{\mathrm{F}} := \sum_{\mathrm{J} \in \mathcal{J}(\mathrm{F}) : \mathrm{J} \in \mathrm{F}} |\hat{\mathrm{g}}(\mathrm{J})| \cdot \mathrm{h}_{\mathrm{J}}^{w}$, so that every term has a positive inner product with $x$. The sum over $F \in \mathcal{F}$ of this last expression is controlled by functional energy, and the property that $\|\Phi\|_\sigma \lesssim \|f\|_\sigma$. This completes the bound for (100).

**Theorem (2.3.10)[103]:** [Global to Local Reduction] There holds

$$\left|B^{\text{above}}(f, g)\right| \lesssim \{\mathcal{H} + \mathcal{L}\} \|f\|_\sigma \|g\|_w.$$

The same inequality holds for the dual form $B^{bolow}(f, g)$.

**Proof.** By Lemma (2.3.9), it remains to control $\mathrm{B}_{\mathcal{F}}^{\text{above}}(f, g)$. Keeping the quasiorthogonality argument in mind, we see that appropriate control on the individual summands is enough to control it. For each $\mathrm{F} \in \mathcal{F}$, let $\mathcal{S}_{\mathrm{F}}$ be the $\mathcal{F}$-children of F. Observe that the function

$$\left(C\alpha_f(F)\right)^{-1} P_F^\sigma f \tag{101}$$
is uniform on $F$ w.r.t. $\mathcal{S}_F$, for appropriate absolute constant $C$. Moreover, the function $Q_F^w g$ does not have any interval $J$ in its Haar support strongly contained in an interval $S \in \mathcal{S}_F$. That is, it is weakly adapted to the function in (101). Therefore, by assumption,
$$\left|B^{\text{above}}\left(P_F^\sigma f, Q_F^w g\right)\right| \le \mathcal{L}\left\{\alpha_F(F)\sigma(F)^{1/2} + \|P_F^\sigma f\|_\sigma\right\}\|Q_F^w g\|_w.$$
The sum over $F \in \mathcal{F}$ of the right hand side is bounded by the quasi-orthogonality argument of (99).

Our Theorem is particular to the Hilbert transform, and so depends upon special properties of it. They largely extend from the fact that the derivative of $-1/y$ is positive. The following Monotonicity Property for the Hilbert transform was observed in [112], and is basic to the analysis of the functional energy inequality.

**Lemma (2.3.11)[103]:** Let I and J be two intervals which share an endpoint a, at which neither $\sigma$ nor $w$ have a point mass. Then,
$$\sup_{0<\alpha<\beta}\left|\left\langle H_{\alpha,\beta}\sigma I, J\right\rangle_w\right| \lesssim \mathcal{A}_2^{\frac{1}{2}}\sqrt{\sigma(I)w(J)}. \tag{102}$$
**Proof.** If $|I| \simeq |J|$, this inequality is the weak boundedness principle of [111]. So, let us assume that $10|I| < |J|$. Then, it remains to bound
$$\begin{aligned}\left|\left\langle H_{\alpha,\beta}\sigma I, (J \setminus 10I)\right\rangle_w\right| &\le \sum_{n=11}^{\infty} \frac{\sigma(I)w(J \cap ((n+1)I \setminus nI))}{n|I|} \\ &\le \frac{\sigma(I)}{|I|^{1/2}} P(w, I)^{1/2} w(J)^{1/2} \lesssim \mathcal{A}_2^{1/2}\sqrt{\sigma(I)w(J)}.\end{aligned}$$
This depends upon obvious kernel bounds, and an application of Cauchy-Schwarz to derive the Poisson term above.

**Lemma (2.3.12)[103]:** (Monotonicity Property). Let $K \supsetneq I$ be two intervals, and assume that $\sigma$ does not have point masses at the end point of I. Then, for any function $g \in L^2(I, w)$, with $w$-integral zero, and $\beta > 2|K|$,
$$P(\sigma \cdot (K - I), I)\left\langle \frac{x}{|I|}, \bar{g}\right\rangle_w \lesssim \liminf_{\alpha \downarrow 0}\left\langle H_{\alpha,\beta}\left(\sigma(K - I)\right), \bar{g}\right\rangle_w. \tag{103}$$
Here, $\bar{g} = \sum_J' |\hat{g}(J')| h_J^w$ is a Haar multiplier applied to g. If J is a good interval, $J \Subset I$, then, for function $g \in L^2(J, w)$, with $w$-integral zero, and signed measures $v$ and $\mu$ supported on $K - I$, with $|v| \le \mu$, it holds that
$$\sup_{0<\alpha<\beta}\left|\left\langle H_{\alpha,\beta} v, g\right\rangle_w\right| \lesssim P(\mu, J)\left\langle \frac{x}{|J|}, \bar{g}\right\rangle_w. \tag{104}$$
**Proof.** By linearity, it suffices to prove (103) in the case of $g = h_I^w$. The point is to separate the supports of the functions involved. Since I does not have a point mass at the end point of I, we have $\sigma(\lambda I \setminus I) \downarrow 0$ as $\lambda \downarrow 1$. It follows that we can fix a $\lambda > 1$ sufficiently small so that $P(\sigma(K - I), I) \simeq P(\sigma(K - \lambda I), I)$, and one more condition that we will come back to. Then, for $0 < \alpha < \frac{1}{2}(\lambda - 1)|I|$, we estimate as below, where $x_1$ is the center of $I$,
$$\left\langle H_{\alpha,\beta}(\sigma(K - \lambda I)), h_I^w\right\rangle_w = \int_{K-\lambda I}\int_I \left\{K_{\alpha,\beta}(y - x) - K_{\alpha,\beta}(y - x_I)\right\} h_I^w(x) w(dx)\sigma(dy)$$

$$= \int_{K-\lambda I} \int_I \frac{x - x_I}{(y-x)(y-x_J)} h_I^w(x) w(\,dx) \sigma(dy)$$
$$\gtrsim P(\sigma(K-I), I) \left\langle \frac{x - x_I}{|I|}, h_I^w \right\rangle_w^* .$$
We have subtracted the term, since $h_1^w$ has integral zero, then applied (94) with $C_{x,x_J,y} = 1$, as follows from our choices of $\alpha$ and $\beta$. Then, note that $(x - x_J)h_1^w \geq 0$, so that we can pull out the Poisson term. The last line follows by our selection of $\lambda$ sufficiently close to 1 . Then, the last condition needed, is to select $\lambda$ sufficiently close to one that, in view of (102),
$$\sup_{\alpha,\beta} \left| \langle H_{\alpha,\beta}(\lambda I \setminus I), h_I^w \rangle_w \right| \lesssim \mathcal{A}_2^{1/2} \sqrt{\sigma(\lambda I \setminus I)} < cP(\sigma(K-I), I) \left\langle \frac{x - \chi_I}{|I|}, h_I^w \right\rangle_w .$$
In the last line, $c > 0$ is an absolute constant. This completes the proof of (103).
Turn to (104). The estimate (94) applies.
$$| \langle H_{\alpha,\beta} v, g \rangle_w \| \quad = \left| \int_{K-I} \int_J \{K_{\alpha,\beta}(y-x) - K_{\alpha,\beta}(y-x_J)\} h_J^w(x) w(dx) v(dy) \right|$$
$$= \left| \int_{K-I} \int_J C_{x,x_J,y} \frac{(x-x_J)}{(y-x)(y-x_J)} h_J^w(x) w(dx) v(dy) \right|$$
But recall that $0 \leq C_{x,x_1,y} \leq 4$, and equals one for $\alpha$ sufficiently small. Moreover, $y - x$ and $y - x_J$ have the same sign, and $(x - x_J)h_J^w(x) \geq 0$. So an upper bound is obtained by passing from $v$ to $\mu$.
$$\left| \langle H_{\alpha,\beta} v, g \rangle_w \right| \quad \leq \int_{K-I} \int_J \frac{(x-x_J)}{(y-x)(y-x_J)} h_J^w(x) w(dx) \mu(dy)$$
$$\simeq P(\mu, J) \left\langle \frac{x}{|J|}, h_J^w \right\rangle_w .$$
The concept of energy is fundamental to the subject. For interval I, define
$$E(w, I)^2 := \mathbb{E}_I^{w(dx)} \mathbb{E}_I^{w(dx')} \frac{(x - x')^2}{|I|^2} = \frac{2}{w(I)} \sum_{J \subset I} \left\langle \frac{x}{|I|}, h_J^w \right\rangle_w^2 . \quad (105)$$
Now, consider the energy constant, the smallest constant, the smallest constant $\mathcal{E}$ such that this condition holds, as presented or in its dual formulation. For all dyadic intervals $I_0$, all partitions $\mathcal{P}$ of $I_0$ into dyadic intervals, it holds that
$$\sum_{I \in \mathcal{P}} P(\sigma I_0, I)^2 E(w, I)^2 w(I) \leq \varepsilon^2 \sigma(I_0). \quad (106)$$
This was shown in [111]

**Proposition (2.3.13)[103]:** For a finite constant $C_0, \varepsilon^2 \leq C_0 \left\{ \mathcal{A}_2^{1/2} + \mathcal{T} \right\}^2 = C_0 \mathcal{H}^2$.

We will always estimate $\mathcal{E}$ by $\mathcal{H}$. The proof is recalled here.

**Proof.** It suffices to consider the case of finite partitions $\mathcal{P}$ of I. We first prove a version of the energy inequality with 'holes' in the argument of the Poisson. It follows from (103) that we can fix $0 < \alpha < \beta$ such that
$$P(\sigma(I_0 - I), I)^2 E(w, I)^2 w(I) \lesssim \| H_{\alpha,\beta}(\sigma(I_0 - I)) \|_{L^2(I,\sigma)}^2, \; I \in \mathcal{P}.$$
Then, using linearity and interval testing, we have

$$\sum_{\mathrm{I}\in\mathcal{P}} \|\mathrm{H}_{\alpha,\beta}(\sigma\cdot \mathrm{I}_0)\|^2_{\mathrm{L}^2(\mathrm{I},\sigma)} \lesssim \|\mathrm{H}_{\alpha,\beta}(\sigma\cdot \mathrm{I}_0)\|^2_{\mathrm{L}^2(\mathrm{I},\sigma)} \lesssim \mathcal{H}^2\sigma(\mathrm{I}_0),$$

$$\text{and} \sum_{\mathrm{I}\in\mathcal{P}} \|\mathrm{H}_{\alpha,\beta}(\sigma\cdot \mathrm{I})\|^2_{\mathrm{L}^2(\mathrm{I},\sigma)} \lesssim \mathcal{H}^2 \sum_{\mathrm{I}\in\mathcal{P}} \sigma(\mathrm{I}) \lesssim \mathcal{H}^2\sigma(\mathrm{I}_0).$$

Then, by the $A_2$ bound, we have $P(\sigma\cdot I,I)^2E(w,I)^2w(I)\lesssim\sigma(I)$, which we can sum over the partition. This completes the proof.

One should keep in mind that the concept of energy is related to the tails of the Hilbert transform. The energy inequality, and its multi-scale extension to the functional energy inequality, show that the control of the tails is very subtle in this problem.

We also need the following elementary Poisson estimate from [130]; used occasionally in this argument, it is crucial to the proof of Lemma (2.3.5).

**Lemma (2.3.14)[103]:** Suppose that $\mathrm{J}\Subset\mathrm{I}\subset\mathrm{I}_0$, and that J is good. Then

$$|\mathrm{J}|^{2\epsilon-1}\mathrm{P}(\sigma(\mathrm{I}_0-\mathrm{I}),\mathrm{J}) \lesssim |\mathrm{I}|^{2\epsilon-1}\mathrm{P}(\sigma(\mathrm{I}_0-\mathrm{I}),\mathrm{I}). \tag{107}$$

**Proof.** We have $\mathrm{dist}(\mathrm{J},\mathrm{I}_0-\mathrm{I})\geq|\mathrm{J}|^\epsilon|\mathrm{I}|^{1-\epsilon}$, so that for any $x\in\mathrm{I}_0-\mathrm{I}$, we have

$$\frac{|J|^{2\epsilon}}{(|J|+\mathrm{dist}(x,J))^2} \lesssim \frac{|I|^{2\epsilon}}{(|I|+\mathrm{dist}(x,I))^2}.$$

Integrating this last expression, it follows that

$$\begin{aligned} |J|^{2\epsilon-1}P(\sigma\cdot(I_0-I),J) &= |J|^{2\epsilon-1}\int_{I_0-I}\frac{|J|}{(|J|+\mathrm{dist}(x,J))^2}\,d\sigma \\ &\lesssim |I|^{2\epsilon}\int_{I_0-1}\frac{1}{(|J|+\mathrm{dist}(x,J))^2}\,d\sigma. \end{aligned}$$

And this proves the inequality.

We state an important multi-scale extension of the energy inequality (106).

**Definition (2.3.15)[103]:** Let $\mathcal{F}$ be a collection of dyadic intervals. A collection of (good) functions $\{g_F\}_{F\in\mathcal{F}}$ in $L^2(w)$ is said to be $\mathcal{F}$-adapted if for all $F\in\mathcal{F}$, the Haar support of the function $g_F$ is contained in $\{\mathrm{J}:\dot{\pi}_{\mathcal{F}}\mathrm{J}=\mathrm{F}\}$.

**Definition (2.3.16)[103]:** Let $\mathcal{F}$ be the smallest constant in the inequality below, or its dual form. The inequality holds for all non-negative $h\in L^2(\sigma)$, all $\sigma$-Carleson collections $\mathcal{F}$, and all $\mathcal{F}$-adapted collections $\{g_F\}_{F\in\mathcal{F}}$ :

$$\sum_{\mathrm{F}\in\mathcal{F}}\sum_{\mathrm{J}^*\in\mathcal{J}^*(\mathrm{F})}\mathrm{P}(h\sigma,\mathrm{J}^*)\left|\left\langle\frac{x}{|\mathrm{J}^*|},g_{\mathrm{F}}\mathrm{J}^*\right\rangle_w\right| \leq \mathcal{F}\parallel \mathrm{h}\parallel_\sigma\left[\sum_{\mathrm{F}\in\mathcal{F}}\|g_{\mathrm{F}}\|^2_w\right]^{1/2}.$$

Here $\mathcal{J}^*(F)$ consists of the maximal good intervals $J\Subset F$. Note that the estimate is universal in $h$ and $\mathcal{F}$, separately.

This constant was identified in [112], and is herein shown to be necessary from the $A_2$ and interval testing inequalities. Recall the definition of $\mathcal{H}$ in (93).

**Theorem (2.3.17)[103]:** Assume that $\mathcal{F}$ satisfies (97), then, $\mathcal{F}\lesssim\mathcal{H}$.

The first step in the proof is the domination of the constant $\mathcal{F}$ by the best constant in a certain two weight inequality for the Poisson operator, with the weights being determined by $w$ and $\sigma$ in a particular way. This is the decisive step, since there is a two weight inequality for the Poisson operator proved by It reduces the full norm inequality to simpler testing conditions, which are in turn controlled by the $A_2$ and Hilbert transform testing conditions.

Consider the weight

$$\mu \equiv \sum_{F\in\mathcal{F}} \sum_{J\in\mathcal{J}^*(F)} \left\| P^w_{F,J} \frac{x}{|J|} \right\|_w^2 \cdot \delta_{(x_J,|J|)}.$$

Here, $P^w_{F,J} := \sum_{J':J'\subset J,} \pi_F J = F^w_{J'}$. We can replace $x$ by $x - c$ for any choice of $c$ we wish; the projection is unchanged. And $\delta_q$ denotes a Dirac unit mass at a point $q$ in the upper half plane $\mathbb{R}^2_+$. We prove the two-weight inequality for the Poisson integral:

$$\| \mathbb{P}(h\sigma) \|_{L^2(\mathbb{R}^2_+,\mu)} \lesssim \mathcal{H} \| h \|_\sigma,$$

for all nonnegative $h$. Above, $\mathbb{P}(\cdot)$ denotes the Poisson extension operator to the upper half-plane, so that in particular

$$\| \mathbb{P}(h\sigma) \|^2_{L^2(\mathbb{R}^2_+,\mu)} = \sum_{F\in\mathcal{F}} \sum_{J\in\mathcal{J}^*(F)} \mathbb{P}(h\sigma)(x_J,|J|)^2 \left\| P^w_{F,J} \frac{x}{|J|} \right\|_w^2,$$

where $x_J$ is the center of the interval $J$. The proof of Theorem (2.3.17) follows by duality. Phrasing things in this way brings a significant advantage: The characterization of the twoweight inequality for the Poisson operator, [128], reduces the full norm inequality above to these testing inequalities. For any dyadic interval $I \in \mathcal{D}$

$$\int_{\mathbb{R}} \mathbb{P}(\sigma \cdot I)^2 \, d\mu(x,t) \lesssim \mathcal{H}^2 \sigma(I), \tag{108}$$

$$\int_{\mathbb{R}} \mathbb{P}^*(t\hat{I}\mu)^2 \sigma(dx) \lesssim \mathcal{A}_2 \int_{\hat{I}} t^2 d\mu(x,t), \tag{109}$$

where $\hat{I} = I \times [0, |I|]$ is the box over $I$ in the upper half-plane, and $\mathbb{P}^*$ is the dual Poisson operator

$$\mathbb{P}^*(t\hat{I}\mu) = \int_{\hat{I}} \frac{t^2}{t^2 + |x-y|^2} \mu(dy,dt).$$

One should keep in mind that the intervals I are restricted to be in our fixed dyadic grid, a reduction allowed as the integrations on the left in (108) and (109) are done over the entire space, either $\mathbb{R}^2_+$ or $\mathbb{R}$. (Goodness of the intervals I above is not needed.) This reduction is critical to the analysis below.

We concerned with a part of inequality (108): Restrict the integral on the left to the set $\hat{I} \subset \mathbb{R}^2_+$.

$$\int_{\hat{I}} \mathbb{P}(\sigma \cdot I)^2 d\mu(x,t) \lesssim \mathcal{H}^2 \sigma(I).$$

Since $(x_J, |J|) \in \hat{I}$ if and only if $J \subset I$, we have

$$\int_{\hat{I}} \mathbb{P}(\sigma \cdot I)(x,t)^2 \, d\mu(x,t) = \sum_{F\in\mathcal{F}} \sum_{J\in\mathcal{J}^*(F):J\subset I} \mathbb{P}(\sigma \cdot I)(x_J,|J|)^2 \left\| P^w_{F,J} \frac{x}{|J|} \right\|_w^2$$

For each $J$,

$$\left\| P^w_{F,J} \frac{x}{|J|} \right\|_w^2 \le \int_J \left| \frac{x - \mathbb{E}^w_J x}{|J|} \right|^2 dw(x) = 2E(w,J)^2 w(J) \le 2w(J). \tag{110}$$

Let $\mathcal{F}_0$ be the maximal $F \in \mathcal{F}$ which are strictly contained in $I$, and let $\mathcal{J}^\sharp$ be those dyadic $J$ such that $(x_J, |J|)$ is in the support of $\mu$, but has no parent in $\mathcal{F}_0$. These intervals are necessarily disjoint. Observe that by (110) and the energy inequality,

$$\sum_{J\in\mathcal{J}^\sharp} \mathbb{P}(\sigma F)(x_J,|J|)^2 \mu(x_J,|J|) \lesssim J \in \mathcal{J}^\sharp P(\sigma \cdot F, J)^2 E(w,J)^2 w(J) \lesssim \mathcal{H}^2 \sigma(F). \tag{111}$$

We claim that

$$\sum_{F\in\mathcal{F}_0}\int_{\hat{F}}\mathbb{P}(\sigma(I\setminus F))(x,t)^2\,d\mu(x,t)\lesssim\mathcal{H}\sigma(I). \tag{112}$$

This is sufficient, since

$$\begin{aligned}\int_{\hat{I}}\mathbb{P}(\sigma\cdot I)(x,t)^2\,d\mu(x,t) &\lesssim \mathrm{LHS}(111)+\mathrm{LHS}(112)+\sum_{F\in\mathcal{F}_0}\int_{\hat{F}}\mathbb{P}(\sigma\cdot F)(x,t)^2\,d\mu(x,t)\\ &\lesssim \mathcal{H}^2\sigma(I)++\sum_{F\in\mathcal{F}_0}\int_{\hat{F}}\mathbb{P}(\sigma\cdot F)(x,t)^2\,d\mu(x,t).\end{aligned}$$

The individual terms in the last sum are set up for a recursive application of this inequality. Due to the Carleson condition (97), this recursion will finish the proof.

It remains to prove (112), which is another instance of the energy inequality. For an interval $F_0\in\mathcal{F}_0$, and $F\in\mathcal{F}$ strictly contained in $F_0$, each interval $J\in\mathcal{J}^*(F)$ is contained in some $J_0\in\mathcal{J}^*(F_0)$. Then, the intervals $F\in\mathcal{F}$ are not good, but $J$ and $J_0$ are good, hence

$$\begin{aligned}\mathbb{P}\big(\sigma(I\setminus F_0)\big)\big(x_J,|J|\big)^2\mu\big(x_J,|J|\big) &=\left[\int_{I\setminus F_0}\frac{|J|}{|J|^2+\left|x-x_J\right|^2}\right]^2\left\|P^w_{F,J}\frac{x}{|J|}\right\|_w^2\\ &=\left[\int_{I\setminus F_0}\frac{1}{|J|^2+\left|x-x_J\right|^2}\right]^2\left\|P^w_{F,J}x\right\|_w^2\\ &\lesssim\left[\int_{I\setminus F_0}\frac{|J_0|}{|J_0|^2+\left|x-x_{J_0}\right|^2}\right]^2\left\|P^w_{F,J}\frac{x}{|J_0|}\right\|_w^2.\end{aligned}$$

This follows from goodness: For $x\in I\setminus F_0$,

$$|J|^2+\left|x-x_J\right|^2\geq\left|x-x_J\right|^2\geq\left|x-x_{J_0}\right|^2\geq|J_0|^{\epsilon}|F_0|^{1-\epsilon}.$$

But then, we can add the projections $P^w_{F,J}$, due to orthogonality, and use (110) again to see that

$$\begin{aligned}&\sum_{\substack{F\in\mathcal{F}\\ F\subset F_0}}\sum_{\substack{J\in\mathcal{J}^*(F)\\ J\subset J_0}}\mathbb{P}\big(\sigma(I\setminus F_0)\big)\big(x_J,|J|\big)^2\mu\big(x_J,|J|\big)\\ &\lesssim\mathbb{P}(\sigma\cdot I)\big(x_{J_0},|J_0|\big)^2\sum_{\substack{F\in\mathcal{F}\\ F\subset F_0}}\sum_{\substack{J\in\mathcal{J}^*(F)\\ J\subset J_0}}\left\|P^w_{F,J}\frac{x}{|J_0|}\right\|_w^2\\ &\lesssim\mathbb{P}(\sigma\cdot I)\big(\chi_{J_0},|J_0|\big)^2E(w,J_0)^2w(J_0).\end{aligned}$$

The sum over $F_0\in\mathcal{F}_0$, and $J_0\in\mathcal{J}^*(F_0)$ is controlled by the energy inequality. This completes the proof of (112).

Now we turn to proving the following estimate for the global part of the first testing condition (108):

$$\int_{\mathbb{R}^2_+-\hat{I}}\mathbb{P}(\sigma\cdot I)^2\,d\mu\lesssim\mathcal{A}_2\sigma(I).$$

Decompose the integral on the left into four terms: With $F_J$ the unique $F\in\mathcal{F}$ with $J\in\mathcal{J}^*(F)$, and using (110),

$$\int_{\mathbb{R}^2_+-\hat{\mathrm{I}}} \mathbb{P}(\sigma\cdot \mathrm{I})^2\,\mathrm{d}\mu = \sum_{\mathrm{J}:(\mathrm{x_J},\mathrm{J},\mathrm{I})\in\mathbb{R}^2_+-\hat{\mathrm{I}}} \mathbb{P}(\sigma\cdot \mathrm{I})\big(\mathrm{x_J},|\mathrm{J}|\big)^2 \left\| P^w_{\mathrm{F_J},\mathrm{J}} \frac{x}{|\,\mathrm{J}|}\right\|_w^2$$

$$\leq \left\{ \sum_{\substack{\mathrm{J}:\mathrm{I}\cap 3\mathrm{I}=\emptyset\\ |\mathrm{J}|\leq|\mathrm{I}|}} + \sum_{\mathrm{J}:\mathrm{J}\subset 3\mathrm{I}-\mathrm{I}} + \sum_{\substack{\mathrm{J}:\mathrm{J}\cap\mathrm{I}=\emptyset\\ \mathrm{I}|>|\mathrm{I}|}} + \sum_{\mathrm{J}:\mathrm{J}\geq\mathrm{I}} \right\} \mathbb{P}(\sigma\cdot \mathrm{I})\big(\mathrm{x_J},|\mathrm{J}|\big)^2 w(\,\mathrm{J})$$

$$= A + B + C + D.$$

Decompose term $A$ according to the length of $J$ and its distance from I, to obtain:

$$\begin{aligned} A &\lesssim \sum_{n=0}^{\infty}\sum_{k=1}^{\infty} J_J \sum_{J\subset 3^{k+1}I-3^kI} \left(\frac{2^{-n}|I|}{\mathrm{dist}(J,I)^2}\sigma(I)\right)^2 w(J) \\ &\lesssim \sum_{n=0}^{\infty} 2^{-2n}\sum_{k=1}^{\infty} \frac{|I|^2\sigma(I)w(3^{k+1}I-3^kI)}{|3^kI|^4}\sigma(I) \\ &\lesssim \sum_{n=0}^{\infty} 2^{-2n}\sum_{k=1}^{\infty} 3^{-2k}\left\{\frac{\sigma(3^{k+1}I)w(3^{k+1}I)}{|3^kI|^2}\right\}\sigma(I) \lesssim \mathcal{A}_2\sigma(I). \end{aligned}$$

Decompose term B according to the length of J and then use the Poisson inequality (107), available to use because of goodness of intervals J. We then obtain

$$\begin{aligned} B &\lesssim \sum_{n=0}^{\infty} \sum_{\substack{J:J\subset 3I-I\\ |J|=2^{-n}||}} 2^{-n(2-4\epsilon)}\frac{\sigma(I)^2}{|I|^2} w(J) \\ &\lesssim \sum_{n=0}^{\infty} 2^{-n(2-4\epsilon)}\frac{\sigma(3I)w(3I)}{|3I|^2}\sigma(I) \lesssim \mathcal{A}_2\sigma(I). \end{aligned}$$

For term $C$, for $n = 1,2, ...$, set $\mathcal{J}_n$ to be those good dyadic intervals $J$ with $|J| > |I|, J\cap I = \emptyset$, and

$$(n-1)|J| \leq \mathrm{dist}(I,J) < n|J|.$$

These intervals have bounded overlaps. Indeed, suppose that $J_1 \subsetneq \cdots \subsetneq J_r$ are all members for $\mathcal{J}_1$. Then, by goodness,

$$\begin{aligned} \mathrm{dist}(J_1,I) &\geq \mathrm{dist}(J_r,I) \geq (n-1)2^r|J_1| + \mathrm{dist}(J_1,\partial J_r I \\ &\geq \left\{(n-1)2^r + 2^{r(1-\epsilon)}\right\}|J_1|. \end{aligned}$$

which is a contradiction to membership in $\mathcal{J}_n$. Restricting the sum to intervals in $\mathcal{J}_\mathrm{n}$, there holds

$$\begin{aligned} \sum_{\mathrm{J}\in\mathcal{J}_\mathrm{n}} \mathbb{P}(\sigma\cdot \mathrm{I})\big(\mathrm{x_J},|\mathrm{J}|\big)^2 w(\,\mathrm{J}) &\lesssim \sigma(\mathrm{I})^2 \sum_{\mathrm{J}\in\mathcal{J}_\mathrm{n}} \frac{w(\,\mathrm{J})}{\mathrm{n}^4|\,\mathrm{J}|^2} \\ &\lesssim \frac{\sigma(\mathrm{I})^2}{|\mathrm{I}|}\sum_{\mathrm{J}\in\mathcal{J}_\mathrm{n}} \frac{w(\,\mathrm{J})\cdot|\mathrm{I}|}{\mathrm{n}^4|\,\mathrm{J}|^2} \\ &\lesssim \frac{\sigma(\mathrm{I})}{\mathrm{n}^2}\cdot\frac{\sigma(\mathrm{I})}{|\mathrm{I}|}\mathrm{P}(w,\mathrm{I}) \lesssim \mathcal{A}_2\frac{\sigma(\mathrm{I})}{\mathrm{n}^2}. \end{aligned}$$

And this is summable in $n \in \mathbb{N}$.

In the last term $D$, all the intervals $J$ contain $I$. Note that

$$\sum_{\mathrm{J}:\mathrm{J}\supseteq\mathrm{I}} \mathbb{P}(\sigma\cdot\mathrm{I})\big(\mathrm{x_J},|\mathrm{J}|\big)^2 w(\mathrm{J}) \lesssim \sigma(\mathrm{I})^2 \sum_{\mathrm{J}:\mathrm{J}\supseteq\mathrm{I}} \frac{w(\mathrm{J})}{|\mathrm{J}|^2}$$

$$\lesssim \sigma(\mathrm{I})\cdot\frac{\sigma(\mathrm{I})}{|\mathrm{I}|}\sum_{\mathrm{J}:\mathrm{JI}} \frac{w(\mathrm{I})\cdot|\mathrm{I}|}{|\mathrm{J}|^2}$$

$$\lesssim \sigma(\mathrm{I})\cdot\frac{\sigma(\mathrm{I})}{|\mathrm{I}|}\mathrm{P}(w,\mathrm{I}) \lesssim \mathcal{A}_2\sigma(\mathrm{I}).$$

We are considering (109). Note that there is a power of $t$ on both sides, and that the expressions on the two sides of this inequality are

$$\int_{\hat{I}} t^2\mu(dx,dt) = \sum_{\mathrm{F}\in\mathcal{F}} \sum_{\substack{\mathrm{J}\in\mathcal{J}^*(\mathrm{F})\\ J\subset I}} \left\|P^w_{\mathrm{F,J}}x\right\|_w^2$$

$$\mathbb{P}^*\big(\mathrm{t\hat{I}}\mu\big)(x) = \sum_{\mathrm{F}\in\mathcal{F}} \sum_{\substack{\mathrm{J}\in\mathcal{J}^*(\mathrm{F})\\ J\subset I}} \frac{\left\|P^w_{\mathrm{F,J}}x\right\|_w^2}{|\mathrm{J}|^2+\left|x-x_{\mathrm{J}}\right|^2}.$$

We are to dominate $\|\mathbb{P}^*(t\hat{I}\mu)\|_\sigma^2$ by the first expression above. The squared norm will be the sum over integers $s$ of $T_s$ below, in which the relative lengths of $J$ and $J'$ are fixed by $s$. Suppressing the requirement that $J,J'\subset I$,

$$\mathrm{T_s} := \sum_{\mathrm{F}\in\mathcal{F}} \sum_{\mathrm{J}\in\mathcal{J}^*(\mathrm{F})} \sum_{\mathrm{F}'\in\mathcal{F}} \sum_{\substack{\mathrm{J}'\in\mathcal{J}^*(\mathrm{F})\\ |\mathrm{J}'|=2^{-s}|\mathrm{J}|}} \int \frac{\left\|\mathrm{P}^w_{\mathrm{F,J}}x\right\|_w^2}{|\mathrm{J}|^2+\left|x-x_{\mathrm{J}}\right|^2}\cdot\frac{\left\|\mathrm{P}^w_{\mathrm{F',J'}}x\right\|_w^2}{|\mathrm{J}'|^2+\left|x-x_{\mathrm{J}'}\right|^2}\,\mathrm{d}\sigma$$

$$\leq M_{\mathrm{s}} \sum_{\mathrm{F}\in\mathcal{F}} \sum_{\mathrm{J}\in\mathcal{J}^*(\mathrm{F})} \left\|P^w_{\mathrm{F,J}}x\right\|_w^2$$

where $M_s \equiv \sup_{F\in\mathcal{F}}\sup_{J\in\mathcal{J}^*(F)} \sum_{\mathrm{F}'\in\mathcal{F}} \sum_{\substack{\mathrm{J}'\in\mathcal{J}^*(\mathrm{F})\\ |\mathrm{J}'|=2^{-s}|\mathrm{J}|}} \int \frac{1}{|J|^2+\left|x-x_J\right|^2}\cdot\frac{w(J')\cdot|J'|^2}{|J'|^2+\left|x-x_{J'}\right|^2}\,d\sigma.$

The estimate (110) has been used in the definition of $M_{\mathrm{s}}$. We claim the term $M_{\mathrm{s}}$ is at most a constant times $\mathcal{A}_2 2^{-\mathrm{s}}$, and it is here that the full Poisson $A_2$ condition is used.
Fix $J$, and let $n\in\mathbb{N}$ be the integer chosen so that $(n-1)|J|\leq \mathrm{dist}(J,J')\leq n|J|$. Estimate the integral in the definition of $\mathbf{M}_{\mathrm{s}}$ by

$$\frac{w(\mathrm{J}')}{|\mathrm{J}'|}\int \frac{|\mathrm{J}'|^2}{|\mathrm{J}|^2+\left|x-x_{\mathrm{J}}\right|^2}\cdot\frac{|\mathrm{J}'|}{|\mathrm{J}'|^2+\left|x-x_{\mathrm{J}'}\right|^2}\,d\sigma \lesssim \mathcal{A}_2 2^{-2s}.$$

This estimate is adequate for $n=0,1,2$. Then estimate the sum over $J'$ as follows.

$$\sum_{\mathrm{F}'\in\mathcal{F}} \sum_{\substack{\mathrm{J}'\in\mathcal{J}^*(\mathrm{F}'):|\mathrm{J}'|=2^{-s}|\mathrm{J}|\\ (\mathrm{n}-1)|\mathrm{J}|\leq\mathrm{dist}(\mathrm{J},\mathrm{J}')\leq\mathrm{n}|\mathrm{J}|}} 2^{-2\mathrm{s}} \lesssim 2^{-\mathrm{s}}.$$

because the relative lengths of $J$ and $J'$ are fixed, and each $J'$ is in at most one $\mathcal{J}^*(F)$.
For the case of $n\geq 3$, restrict $J'$ to be to the right of $J$, and let $t_n=\frac{x_1+x_1}{2}$, so that $\left|x_{\mathrm{J}}-\mathrm{t_n}\right|,\left|x_{\mathrm{J}'}-\mathrm{t_n}\right|\simeq \mathrm{n}|\mathrm{J}|$. First, estimate the integral in the definition of $\mathrm{M_s}$ on the interval $[t_n,\infty)$.

$$\frac{w(J')}{|J'|}\int_{t_n}^{\infty}\frac{|J'|^2}{|J|^2+\left|x-x_J\right|^2}\cdot\frac{|J'|}{|J'|^2+\left|x-x_{J'}\right|^2}d\sigma\lesssim\mathcal{A}_2\frac{2^{-2s}}{n^2}$$

Then estimate the sum over $J'$ as follows.

$$\sum_{F'\in\mathcal{F}}\sum_{\substack{J'\in\mathcal{J}^e(F'):|J'|=2^{-s}|J|\\(n-1)|J|\le \mathrm{dist}(J,J')\le n|J|}}\frac{2^{-2s}}{n^2}\lesssim\frac{2^{-s}}{n^2}.$$

This is clearly summable in $n\geq 4$.

Now, estimate on the integral on the interval $(-\infty,t_n)$,

$$\begin{aligned}\frac{w(J')}{|J'|}\int_{-\infty}^{t_n}&\frac{|J'|^2}{|J|^2+\left|x-x_J\right|^2}\cdot\frac{|J'|}{|J'|^2+\left|x-x_{J'}\right|^2}d\sigma\\&=2^{-2s}\frac{w(J')}{|J|}\int_{-\infty}^{t_n}\frac{|J|}{|J|^2+\left|x-x_J\right|^2}\cdot\frac{|J|^2}{|J'|^2+\left|x-x_{J'}\right|^2}d\sigma\\&\lesssim 2^{-2s}\frac{w(J')}{n^2|J|}P(\sigma,J).\end{aligned}$$

Drop the term with the geometric decay in $s$, and sum over $n$ and $J'$ to see that

$$\sum_{n=4}^{\infty}\sum_{F'\in\mathcal{F}}\sum_{\substack{J'\in\mathcal{J}^c(F'):|J'|=2^{-s}|J|\\(n-1)|J|\le \mathrm{dis}(J,J')\le n|J|}}\frac{w(J')}{n^2|J|}P(\sigma,I)\lesssim P(w,J)P(\sigma,J)\lesssim\mathcal{A}_2.$$

Here, we have appealed to the full Poisson $A_2$ condition. This completes the control of the dual Poisson testing condition.

# Chapter 3
# Toeplitz Lemma and Regular Operator Mappings

We show that the mean convergence versions of the Toeplitz lemma, Cesàro mean convergence theorem, and the Kronecker lemma are presented and a general mean convergence theorem for a normed sum of independent random variables is established. Some additional problems are posed. We propose different types of updating conditions that seems natural in many applications and prove that each of these conditions, together with a few other natural axioms, uniquely defines the geometric mean for any number of operator variables. The means defined in this way are given by explicit formulas and are computationally tractable. We introduce two classes of complete moment convergence, which are stronger versions of mean convergence and consider the Toeplitz lemma, the Cesàro mean convergence theorem, and the Kronecker lemma under these two classes of complete moment convergence.

## Section (3.1): Convergence in Probability and Mean Convergence

The Toeplitz lemma is a result in mathematical analysis which is a useful tool for proving a wide variety of probability limit theorems. It is stated as follows and its proof may be found in [138].

**Theorem (3.1.1)[131]:** (Toeplitz Lemma). Let $\{a_{nk}, 1 \leq k \leq k_n, n \geq 1\}$ be a double array of real numbers such that $\lim_{n\to\infty} a_{nk} = 0$ for all $k \geq 1$ and $\sup_{n\geq 1} \sum_{k=1}^{k_n} |a_{nk}| < \infty$. Let $\{x_n, n \geq 1\}$ be a sequence of real numbers.

(i) If $\lim_{n\to\infty} x_n = 0$, then $\lim_{n\to\infty} \sum_{k=1}^{k_n} a_{nk} x_k = 0$.

(ii) If $\lim_{n\to\infty} x_n = x$ finite and $\lim_{n\to\infty} \sum_{k=1}^{k_n} a_{nk} = 1$, then $\lim_{n\to\infty} \sum_{k=1}^{k_n} a_{nk} x_k = x$.

The Toeplitz lemma contains, as corollaries, the following well-known and important results.

**Corollary (3.1.2)[131]:** (Cesarro Mean Convergence Theorem). Let $\{x_n, n \geq 1\}$ be a sequence of real numbers and let $\bar{x}_n = \sum_{k=1}^{n} x_k / n, n \geq 1$. If $\lim_{n\to\infty} x_n = x$ finite, then $\lim_{n\to\infty} \bar{x}_n = x$.

**Corollary (3.1.3)[131]:** (Kronecker Lemma). Let $\{x_n, n \geq 1\}$ and $\{b_n, n \geq 1\}$ be sequences of real numbers with $0 < b_n \uparrow \infty$. If the series $\sum_{k=1}^{n} x_k / b_k$ converges, then $\lim_{n\to\infty} \sum_{k=1}^{n} x_k / b_n = 0$.

The proof of the Cesaro mean convergence theorem follows immediately from the Toeplitz lemma $(ii)$ by taking $k_n = n, n \geq 1$ and $a_{nk} = n^{-1}, 1 \leq k \leq n, n \geq 1$.

See [138] for a proof of the Kronecker lemma, which also follows from the Toeplitz lemma $(ii)$.

It is clear that the Toeplitz lemma and its corollaries are are valid when the numerical sequence $\{x_n, n \geq 1\}$ and real number $x$ are replaced, respectively, by a sequence of random variables $\{X_n, n \geq 1\}$ and random variable $X$ provided the convergence statements involving the random variables are couched in terms of almost sure $(a.s.)$ convergence.

It is natural to inquire as to whether or not the Toeplitz lemma and its corollaries hold when the mode of convergence is changed from a.s. convergence to convergence in probability or to mean convergence of some order. We demonstrate by three examples that both corollaries of the Toeplitz lemma fail when a.s convergence is replaced by convergence in probability, and that a variety of possible limiting behaviors can prevail. We present several "mean convergence" versions of the Toeplitz lemma, Cesàro mean convergence theorem, and the Kronecker lemma. $A$ general mean convergence theorem for a normed sum of independent random variables is established .

Dugué [136] investigated the "convergence in probability" problem for the sequence of Cesàro means $\{\bar{X}_n = \sum_{k=1}^{n} X_k, n \geq 1\}$ where $\{X_n, n \geq 1\}$ is a sequence of independent random variables and proved that if $\bar{X}_n \xrightarrow{P} c$ for some constant $c$, then $\frac{1}{n}\min_{1\leq k\leq n} X_k \xrightarrow{P} 0$ and $\frac{1}{n}\max_{1\leq k\leq n} X_k \xrightarrow{P} 0$. Then, for a sequence of independent random variables $\{X_n, n \geq 1\}$ where $X_n$ has distribution function

$$F_n(x) = \left(1 - \frac{1}{x+n}\right) I_{(0,\infty)}(x), x \in \mathbb{R}, n \geq 1,$$

Dugué showed $X_n \xrightarrow{P} 0, \frac{1}{n}\max_{1\leq k\leq n} X_k \xrightarrow{P} 0$, and, consequently, $\bar{X}_n \xrightarrow{P} 0$. However, for his example, Dugué did not actually characterize the weak limiting behavior of $\bar{X}_n$ We present sequences of independent random variables $\{X_n, n \geq 1\}$ wherein $X_n \xrightarrow{P} 0$ and $\bar{X}_n \xrightarrow{P} 0$ and we characterize the weak limiting behavior of $\bar{X}_n$.

We present three counterexamples. In the first example, $X_n \xrightarrow{P} 0$ yet the corresponding sequence of Cesàro means $\bar{X}_n$ has a nondegenerate limiting distribution. The following two lemmas are used in the verification of Example (3.1.6).

**Lemma (3.1.4)[131]:** For $x > -1, \frac{x}{x+1} \leq \log(1+x) \leq x$.

**Proof.** This is well known; see [133].

**Lemma (3.1.5)[131]:** For all $t \in \mathbb{R}$,

$$\lim_{n\to\infty} \max_{1\leq k\leq n} \frac{\left|e^{kt/n} - 1\right|}{k} = 0. \tag{1}$$

**Proof.** For $t \geq 0$ and $2 \leq m \leq n$,

$$\begin{aligned} 0 \quad &\leq \max_{1\leq k\leq n} \frac{\left|e^{kt/n} - 1\right|}{k} \\ &\leq \max_{1\leq k\leq m-1} \frac{\left|e^{kt/n} - 1\right|}{k} + \max_{m\leq k\leq n} \frac{\left|e^{kt/n} - 1\right|}{k} \\ &\leq e^{(m-1)t/n} - 1 + e^t/m \overset{n\to\infty}{\longrightarrow} e^t/m \overset{m\to\infty}{\longrightarrow} 0. \end{aligned}$$

$A$ similar argument works for $t < 0$ with the right-hand side of the last inequality replaced by $1 - e^{(m-1)t/n} + m^{-1}$ and so (1) holds for all $t \in \mathbb{R}$.

**Example (3.1.6)[131]:** Let $\{X_n, n \geq 1\}$ be a sequence of independent random variables with $P(X_n = 0) = 1 - n^{-1}$ and $P(X_n = n) = n^{-1}, n \geq 1$. It is clear that $X_n \overset{P}{\to} 0$. Set $\bar{X}_n = \sum_{k=1}^{n} X_k/n, n \geq 1$. For $k \geq 1$, the moment generating function of $X_k$ is

$$m_{X_k}(t) = 1 + \frac{e^{kt} - 1}{k},\ t \in \mathbb{R}$$

and so the cumulant generating function of $\bar{X}_n$ is

$$\kappa_n(t) = \sum_{k=1}^{n} \log\left(1 + \frac{e^{\frac{kt}{n}} - 1}{k}\right),\ t \in \mathbb{R}, n \geq 1.$$

Let $t \in \mathbb{R}$ and $n \geq 1$. By Lemma (3.1.4),

$$\kappa_n(t) \leq \sum_{k=1}^{n} \frac{e^{\frac{kt}{n}} - 1}{k} = \frac{1}{n}\sum_{k=1}^{n} \frac{e^{\frac{tk}{n}} - 1}{\frac{k}{n}} \overset{n\to\infty}{\longrightarrow} \int_0^1 \frac{e^{tx} - 1}{x} dx,$$

where we have used the fact that the upper bound for $\kappa_n(t)$ is a Riemann sum. Hence,

$$\limsup_{n\to\infty} \kappa_n(t) \leq \int_0^1 \frac{e^{tx} - 1}{x} dx. \tag{2}$$

Again let $t \in \mathbb{R}$ and fix $0 < \epsilon < 1$. For all large $n$,

$$\kappa_n(t) \geq \sum_{k=1}^{n} \frac{\left(e^{kt/n} - 1\right)/k}{1 + (e^{kt/n} - 1)/k} = \sum_{k=1}^{n} \frac{\frac{\mathrm{e}^{tk/n} - 1}{k/n}\frac{1}{n}}{1 + \frac{e^{kt/n-1}}{k}} \geq \frac{1}{1 \pm \epsilon}\frac{1}{n}\sum_{k=1}^{n} \frac{e^{tk/n} - 1}{k/n},$$

where $\pm$ is taken to be $+$ when $t \geq 0$ and $-$ when $t < 0$. Thus,

$$\liminf_{n\to\infty} \kappa_n(t) \geq \frac{1}{1 \pm \epsilon}\int_0^1 \frac{e^{tx} - 1}{x} dx,$$

and since $0 < \epsilon < 1$ is arbitrary, we have

$$\liminf_{n\to\infty} \kappa_n(t) \geq \int_0^1 \frac{e^{tx} - 1}{x} dx. \tag{3}$$

Combining (2) and (3) gives

$$\lim_{n\to\infty} \kappa_n(t) = \int_0^1 \frac{e^{tx} - 1}{x} dx, t \in \mathbb{R}.$$

Thus the sequence of moment generating functions $\{m_{\bar{X}_n}(\cdot), n \geq 1\}$ corresponding to $\{\bar{X}_n, n \geq 1\}$ satisfies

$$\lim_{n\to\infty} m_{\bar{X}_n}(t) = \lim_{n\to\infty} e^{K_n(t)} = \exp\left(\int_0^1 \frac{e^{tx}-1}{x}\right),\ t \in \mathbb{R}.$$

Then by the continuity theorem for moment generating functions [135], the function

$$m(t) = \exp\left(\int_0^1 \frac{e^{tx}-1}{x} dx\right),\ t \in \mathbb{R}$$

is the moment generating function of a random variable and the sequence of Cesarro means $\bar{X}_n$ has a limiting distribution with moment generating function $m(\cdot)$. It is clear that the limiting distribution of $\bar{X}_n$ is nondegenerate.

In the next example, $X_n \xrightarrow{P} 0$ yet the corresponding sequence of Cesarro means $\bar{X}_n$ approaches $\infty$ in probability.

**Example (3.1.7)[131]:** Let $\alpha > 1$ and let $\{X_n, n \geq 1\}$ be a sequence of independent random variables with $P(X_n = 0) = 1 - n^{-1}$ and $P(X_n = n^2) = n^{-1}, n \geq 1$. It is clear that $X_n \xrightarrow{P} 0$. Set $\bar{X}_n = \sum_{k=1}^n X_k/n, n \geq 1$. Let $M \geq 1$ be arbitrary. Let $k_n$ be the smallest integer greater than or equal to $(Mn)^{1/\alpha}, n \geq 1$. Then, for all large $n$,

$$\begin{aligned} P(\bar{X}_n \geq M) \quad &= P\left(\sum_{k=1}^n X_k \geq Mn\right) \geq P\left(\bigcup_{k=k_n}^n [X_k = k^\alpha]\right) \\ &= 1 - P\left(\bigcap_{k=k_n}^n [X_k \neq k^\alpha]\right) = 1 - \prod_{k=k_n}^n \frac{k-1}{k} \\ &= 1 - \frac{k_n - 1}{n} > 1 - \frac{(Mn)^{1/\alpha}}{n} \overset{n\to\infty}{\longrightarrow} 1. \end{aligned}$$

Thus, $\bar{X}_n \xrightarrow{P} \infty$ since $M \geq 1$ is arbitrary.

The next example demonstrates that a convergence in probability version of the Kronecker lemma also fails; we note, however, that for the Kronecker lemma to fail we must have dependence among the $X_k$ since otherwise convergence in probability of $\sum_{k=1}^n X_k/b_k$ to a random variable implies convergence $a.s.$ (see, e.g., [134]), in which case the traditional Kronecker lemma yields $\sum_{k=1}^n X_k/b_n \to 0\ a.s.$ and hence in probability.

**Example (3.1.8)[131]:** Let $\{Y_n, n \geq 1\}$ be a sequence of independent random variables with $P(Y_n = 0) = 1 - \frac{1}{2n-1}$ and $P(Y_n = 16^{n-1}) = \frac{1}{2^{n-1}}, n \geq 1$. For $n \geq 1$ define $X_{2n-1} = Y_n$ and $X_{2n} = -2Y_n$. Then

$$\frac{X_{2n-1}}{2^{2n-1}} + \frac{X_{2n}}{2^{2n}} = 0,\ n \geq 1$$

and so for $n \geq 1$

$$\sum_{k=1}^{n} \frac{X_k}{2^k} = \frac{X_n}{2^n} I(n \text{ is odd }).$$

Then, since it is clear that $X_n \overset{P}{\to} 0$, we have that $\sum_{k=1}^{n} \frac{X_k}{2^t} \overset{P}{\to} 0$ as well.

To show that $\sum_{k=1}^{n} X_k/2^n \overset{P}{\not\to} 0$, we show that the convergence fails along a subsequence, namely, that

$$\frac{\sum_{k=1}^{4n} X_k}{2^{4n}} \overset{P}{\not\to} 0. \tag{4}$$

Note that for all odd positive integers $k$,

$$X_k + X_{k+1} = X_k - 2X_k = -X_k$$

and consequently for all $n \geq 1$,

$$\frac{\sum_{k=1}^{2n} X_k}{2^{2n}} = \frac{-\sum_{k=1}^{n} X_{2k-1}}{2^{2n}}.$$

Thus, (4) will hold if we can show that

$$\frac{\sum_{k=1}^{2n} X_{2k-1}}{16^n} \overset{P}{\not\to} 0. \tag{5}$$

To this end,

$$\begin{aligned} P\left(\frac{\left|\sum_{k=1}^{2n} X_{2k-1}\right|}{16^n} \geq 1\right) &\geq P\left(\sum_{k=n+1}^{2n} X_{2k-1} \geq 16^n\right) \\ &\geq P\left(\bigcup_{k=n+1}^{2n} [X_{2k-1} \neq 0]\right) = 1 - \prod_{k=n+1}^{2n} P(X_{2k-1} = 0) \\ &\geq 1 - \prod_{k=n+1}^{2n} \left(1 - \frac{1}{2k}\right) \geq 1 - \left(1 - \frac{1}{4n}\right)^n \\ &\to 1 - e^{-1/4} > 0, \end{aligned}$$

proving (5).

We present "mean convergence" versions of the Toeplitz lemma (Theorem (3.1.9)), the Cesaro mean convergence theorem (Corollary (3.1.10)), and the Kronecker lemma (Theorems (3.1.11) and (3.1.13)). In Theorems (3.1.9)−(3.1.13) and Corollary (3.1.10), no independence conditions are imposed on the random variables $\{X_n, n \geq 1\}$.

For $p \geq 1$, the space $\mathcal{L}_p$ of the absolute $pth$ power integrable random variables is a Banach space with norm $\| X \|_p = (E|X|^p)^{1/p}, X \in \mathcal{L}_p$. Now the proof of the Toeplitz lemma in [138] carries over to a Banach space setting, We, thus, immediately obtain the following "mean convergence" version of the Toeplitz lemma.

**Theorem (3.1.9)[131]:** Let $\{a_{nk}, 1 \leq k \leq k_n, n \geq 1\}$ be a double array of real numbers such that $\lim_{n\to\infty} a_{nk} = 0$ for all $k \geq 1$ and $\sup_{n\geq 1} \sum_{k=1}^{k_n} |a_{nk}| < \infty$. Let $\{X_n, n \geq 1\}$ be a sequence of $\mathcal{L}_p$, random variables for some $p \geq 1$.

(i) If $X_n \xrightarrow{\mathrm{L}_p} 0$, then $\sum_{k=1}^{k_n} a_{nk} X_k \xrightarrow{\mathrm{L}_p} 0$.

(ii) If there exists a random variable $X$ such that $X_n \xrightarrow{L_p} X$ and $\lim_{n\to\infty} \sum_{k=1}^{k_n} a_{nk} = 1$, then $\sum_{k=1}^{k_n} a_{nk} X_k \xrightarrow{L_p} X$.

Letting $k_n = n, n \geq 1$ and $a_{nk} = n^{-1}, 1 \leq k \leq k_n, n \geq 1$ in Theorem (3.1.9) $(ii)$ yields the following "mean convergence" version of the Cesàro mean convergence theorem.

**Corollary (3.1.10)[131]:** Let $\{X_n, n \geq 1\}$ be a sequence of $\mathcal{L}_p$ random variables where $p \geq 1$ and let $\bar{X}_n = \sum_{\mathrm{k}=1}^{n} X_k/n, n \geq 1$. If there exists a random variable $X$ such that $X_n \xrightarrow{\mathrm{L}_p} X$, then $\bar{X}_n \xrightarrow{L_p} X$.

By employing the Banach space version of the traditional Kronecker lemma (see, e.g., [141]), the following "mean convergence" version of the Kronecker lemma is immediate.

**Theorem (3.1.11)[131]:** Let $\{X_n, n \geq 1\}$ be a sequence of $\mathcal{L}_p$ random variables for some $p \geq 1$ and let $\{b_n, n \geq 1\}$ be a sequence of real numbers with $0 < b_n \uparrow \infty$. If there exists $a$ random variable $S$ such that

$$\sum_{k=1}^{n} \frac{X_k}{b_k} \xrightarrow{L_p} S, \tag{6}$$

then

$$\frac{\sum_{k=1}^{n} X_k}{b_n} \xrightarrow{\mathrm{L}_p} 0. \tag{7}$$

**Remark (3.1.12)[131]:** For a sequence of independent mean 0 random variables $\{X_n, n \geq 1\}$ and a sequence of real numbers $\{b_n, n \geq 1\}$ with $0 < b_n \uparrow \infty$, a sufficient condition for the existence of a random variable $S$ satisfying (6) with $p \in [1,2]$ is that

$$\sum_{n=1}^{\infty} \frac{E|X_n|^p}{b_n^p} < \infty. \tag{8}$$

This follows readily from [132] and the Cauchy convergence criterion (see, e.g., [134]). However, for $p \geq 2$, a necessary and sufficient condition for the existence of a random variable $S$ satisfying (6) is that

$$\sum_{n=1}^{\infty} \frac{E|X_n|^p}{b_n^p} < \infty \text{ and } \sum_{n=1}^{\infty} \frac{EX_n^2}{b_n^2} < \infty.$$

This follows readily from [140] and the Cauchy convergence criterion.

It is easy to construct an example showing that when $p \in [1,2)$, the condition (8) is not necessary for the existence of a random variable $S$ satisfying (6); that is, (6) can hold when (8) fails.

We now establish a version of Theorem (3.1.11) for a sequence of nonnegative random variables $\{X_n, n \geq 1\}$. In view of Theorem (3.1.11), Theorem (3.1.13) is of interest only when $0 < p < 1$.

**Theorem (3.1.13)[131]:** Let $\{X_n, n \geq 1\}$ be a sequence of nonnegative $L_p$ random variables for some $0 < p < \infty$ and let $\{b_n, n \geq 1\}$ be a sequence with $0 < b_n \uparrow \infty$. If there exists $a$ random variable $S$ such that (6) holds, then (7) holds as well.

**Proof.** The $\mathcal{L}_p$ convergence in (6) implies convergence in probability which, in turn, implies $a.s.$ convergence since $\sum_{k=1}^{n} X_k/b_k$ is nondecreasing. Then by the traditional Kronecker lemma, $\sum_{k=1}^{n} X_k/b_n \to 0\ a.s.$ On the other hand, $\mathcal{L}_p$ convergence of $\sum_{k=1}^{n} \frac{X_k}{b_k}$ implies that the sequence $\{|\sum_{k=1}^{n} X_k/b_k|^p, n \geq 1\}$ is uniformly integrable by the mean convergence criterion (see, $e.g.$, [134]). By nonnegativity,

$$0 \leq \frac{\sum_{k=1}^{n} X_k}{b_n} \leq \sum_{k=1}^{n} \frac{X_k}{b_k}$$

and so the sequence $\{|\sum_{k=1}^{n} X_k/b_n|^p, n \geq 1\}$ is uniformly integrable as well. Combining this with $\sum_{k=1}^{n} X_k/b_n \to 0\ a.s.$ and employing the mean convergence criterion gives (7).

We close by establishing in Theorem (3.1.17) a general mean convergence theorem for a normed sum of independent random variables. Its proof uses three results which will now be stated.

The following result is the famous de La Vallée Poussin criterion for uniform integrability; a proof of it may be found in [139].

**Proposition (3.1.14)[131]:** $A$ sequence of random variables $\{U_n, n \geq 1\}$ is uniformly integrable if and only if there exists a nondecreasing convex function $G$ defined on $[0, \infty)$ with

$$G(0) = 0,\ \lim_{u \to \infty} \frac{G(u)}{u} = \infty, \text{ and } \sup_{n \geq 1} E\big(G(|U_n|)\big) < \infty. \tag{9}$$

The next result is a so-called "contraction principle" and is due to [137].

**Proposition (3.1.15)[131]:** Let $\varphi$ be a nonnegative nondecreasing convex function defined on $[0,\infty)$, let $n \geq 1$, let $\{\lambda_1, \dots, \lambda_n\} \subset \mathbb{R}$, and let $Y_1, \dots, Y_n$ be independent symmetric random variables. Then

$$E\left(\varphi\left(\left|\sum_{k=1}^{n} \lambda_k Y_k\right|\right)\right) \leq E\left(\varphi\left(\left(\max_{1\leq k\leq n} |\lambda_k|\right)\left|\sum_{k=1}^{n} Y_k\right|\right)\right).$$

The next result is referred to as a "symmetrization moment inequality" and its proof may be found in [138].

**Proposition (3.1.16)[131]:** Let $U^*$ be a symmetrized version of a random variable $U$ and let $m$ be any median of $U$. Then for all $p > 0$,

$$E|U-m|^p \leq 2E|U^*|^p.$$

**Theorem (3.1.17)[131]:** Let $\{X_n, n \geq 1\}$ be a sequence of independent $\mathcal{L}_p$ random variables for some $p \geq 1$ and let $\{b_n, n \geq 1\}$ and $\{B_n, n \geq 1\}$ be sequences of real numbers with $0 < B_n \uparrow \infty$. Suppose that the sequence

$$\left\{\left|\sum_{k=1}^{n} \frac{X_k}{b_k}\right|^p, n \geq 1\right\} \text{ is uniformly integrable.} \tag{10}$$

(i) If $b_n = O(B_n)$ and

$$\frac{\sum_{k=1}^{n} X_k}{B_n} \xrightarrow{P} 0, \tag{11}$$

then

$$\frac{\sum_{k=1}^{n} X_k}{B_n} \xrightarrow{Y_p} 0. \tag{12}$$

(ii) If $b_n = o(B_n)$ and $m_n = o(B_n)$ where $m_n$ is any median of $\sum_{k=1}^{n} X_k, n \geq 1$, then

$$\frac{\sum_{k=1}^{n} X_k}{B_n} \xrightarrow{Y_p} 0. \tag{13}$$

**Proof.** Let $\{X'_n, n \geq 1\}$ be an independent copy of $\{X_n, n \geq 1\}$ and consider the sequence of symmetrized random variables $\{X_n^* = X_n - X'_n, n \geq 1\}$. Now (10) holds with $X_k$ replaced by $X'_k, k \geq 1$ and since

$$\left|\sum_{k=1}^{n} \frac{X_k^*}{b_k}\right|^p \leq 2^{p-1}\left(\left|\sum_{k=1}^{n} \frac{X_k}{b_k}\right|^p + \left|\sum_{k=1}^{n} \frac{X'_k}{b_k}\right|^p\right), n \geq 1,$$

the sequence

$$\left\{\left|\sum_{k=1}^{n}\frac{X_k^*}{b_k}\right|^p, n\geq 1\right\} \text{ is uniformly integrable.} \tag{14}$$

By (14) and Proposition (3.1.14), there exists a nondecreasing convex function $G$ defined on $[0,\infty)$ satisfying (9) with $U_n = |\sum_{k=1}^{n} X_k^*/b_k|^p, n \geq 1$. Set $b_n^* = \max_{1\leq k\leq n}|b_k|, n \geq 1$. Now the function $\varphi(u) = G(u^p), u \geq 0$ is a nonnegative nondecreasing convex function and so by Proposition (3.1.15)

$$\begin{aligned}
\sup_{n\geq 1} E\left(G\left(\left|\sum_{k=1}^{n}\frac{X_k^*}{b_n^*}\right|^p\right)\right) &= \sup_{n\geq 1} E\left(\varphi\left(\left|\sum_{k=1}^{n}\frac{b_k}{b_n^*}\frac{X_k^*}{b_k}\right|\right)\right) \\
&\leq \sup_{n\geq 1} E\left(\varphi\left(\left(\max_{1\leq k\leq n}\left|\frac{b_k}{b_n^*}\right|\right)\left|\sum_{k=1}^{n}\frac{X_k^*}{b_k}\right|\right)\right) \\
&= \sup_{n\geq 1} E\left(\varphi\left(\left|\sum_{k=1}^{n}\frac{X_k^*}{b_k}\right|\right)\right) \\
&= \sup_{n\geq 1} E\left(G\left(\left|\sum_{k=1}^{n}\frac{X_k^*}{b_k}\right|\right)\right) < \infty
\end{aligned}$$

recalling (9). Then since $G(0) = 0$ and $\lim_{u\to\infty}\frac{G(u)}{u} = \infty$, again by applying Proposition (3.1.14) with the same function $G$ we get that the sequence

$$\left\{\left|\frac{\sum_{k=1}^{n} X_k^*}{b_n^*}\right|^p, n\geq 1\right\} \text{ is uniformly integrable.} \tag{15}$$

We first prove part $(i)$. It follows from $b_n = O(B_n)$ and $B_n \uparrow \infty$ that $b_n^* = O(B_n)$ and so by (15) the sequence

$$\left\{\left|\frac{\sum_{k=1}^{n} X_k^*}{B_n}\right|^p, n\geq 1\right\} \text{ is uniformly integrable.} \tag{16}$$

Now (11) also holds with $X_k$ replaced by $X_k', k \geq 1$ and so

$$\frac{\sum_{k=1}^{n} X_k^*}{B_n} = \frac{\sum_{k=1}^{n} X_k}{B_n} - \frac{\sum_{k=1}^{n} X_k'}{B_n} \xrightarrow{P} 0. \tag{17}$$

Then by (16), (17), and the mean convergence criterion,

$$\frac{\sum_{k=1}^{n} X_k^*}{B_n} \xrightarrow{\mathcal{L}_p} 0. \tag{18}$$

Let $m_n$ be any median of $\sum_{k=1}^{n} X_k, n \geq 1$. It follows from (11) that

$$m_n = o(B_n). \tag{19}$$

Then, by Proposition (3.1.16), (18), and (19),

$$
\begin{aligned}
E\left|\frac{\sum_{k=1}^{n} X_k}{B_n}\right|^p & \leq 2^{p-1}\left(\frac{E\left|\sum_{k=1}^{n} X_k-m_n\right|^p}{B_n^p}+\frac{\left|m_n\right|^p}{B_n^p}\right) \\
& \leq 2^{p-1}\left(\frac{2E\left|\sum_{k=1}^{n} X_k^*\right|^p}{B_n^p}+\frac{\left|m_n\right|^p}{B_n^p}\right) \\
& \rightarrow 0
\end{aligned}
$$

proving (12) thereby completing the proof of part $(i)$.

Next, we prove part $(ii)$. It follows from $b_n=o(B_n)$ and $B_n \uparrow \infty$ that $b_n^*=o(B_n)$. Then,

$$
E\left|\frac{\sum_{k=1}^{n} X_k^*}{B_n}\right|^p=\frac{b_n^{*p}}{B_n^p} E\left|\frac{\sum_{k=1}^{n} X_k^*}{b_n^*}\right|^p=o(1) O(1)=o(1)
$$

recalling (15). Since $m_n=o(B_n)$ by hypothesis, the conclusion (13) follows by the same argument used to complete the proof of part $(i)$.

**Section (3.2): Multivariate Geometric Means**

The geometric mean of two positive definite operators was introduced by Pusz and Woronowicz [19], and their definition was soon put into the context of the axiomatic approach to operator means developed by Kubo and Ando [149]. Subsequently a number of authors [148], [2], [17], [5], [151], [150] have suggested several ways of defining means of operators for several variables as extensions of the geometric mean of two operators.

There is no satisfactory definition of a geometric mean of several operator variables that is both computationally tractable and satisfies a number of natural conditions put forward by Ando, Li, and Mathias [2]. We put the emphasis on methods to extend a geometric mean of $k$ variables to a mean of $k+1$ variables, and in the process we challenge one of the requirements to a geometric mean put forward by Ando, Li, and Mathias.

The symmetry condition of a geometric mean is mathematically very appealing, but the condition makes no sense in a number of applications. If for example positive definite matrices $A_1, A_2, \ldots, A_k$ correspond to measurements made at times $t_1<t_2<\cdots<t_k$ then there is no way of permuting the matrices since time only goes forward. It makes more sense to impose an updating condition

$$
G_{k+1}(A_1, \ldots, A_k, 1)=G_k(A_1, \ldots, A_k)^{\frac{k}{k+1}} \tag{20}
$$

when moving from a mean $G_k$ of $k$ variables to a mean $G_{k+1}$ of $k+1$ variables. The condition corresponds to taking the geometric mean of $k$ copies of $G_k(A_1, \ldots, A_k)$ and one copy of the unit matrix. $A$ variant condition would be to impose the equality

$$
G_{k+1}(A_1, \ldots, A_k, 1)=G_k\left(A_1^{k/(k+1)}, \ldots, A_k^{k/(k+1)}\right) \tag{21}
$$

when updating from $k$ to $k+1$ variables. It is an easy exercise to realise that if we set $G_1(A)=A$, then either of the conditions (20) or (21) together with homogeneity uniquely defines the geometric mean of $k$ commuting operators.

We furthermore prove that by setting $G_1(A)=A$ and by demanding homogeneity and a few more natural conditions, then either of the updating conditions (20) or (21) leads to unique but different solutions to the problem of defining a geometric mean of $k$ operators.

The means defined in this way are given by explicit formulas, and they are computationally tractable. They are given by explicit formulas, and they are computationally tractable. They means discussed in [2] with the notable exception of symmetry. If one emphasises either of the updating conditions (20) or (21) we are thus forced to abandon symmetry.
Efficient averaging techniques of positive definite matrices are important in many practical applications; for example in radar imaging, medical imaging, and the analysis of financial data.

Let $B(\mathcal{H})$ denote the set of bounded linear operators on a Hilbert space $\mathcal{H}. A$ function $F: \mathcal{D} \to B(\mathcal{H})$ defined in a convex domain $\mathcal{D}$ of self-adjoint operators in $B(\mathcal{H})$ is called a spectral function, if it can be written on the form $F(x) = f(x)$ for some function $f$ defined in a real interval $I$, where $f(x)$ is obtained by applying the functional calculus.
The definition contains some hidden assumptions. The domain $\mathcal{D}$ should be invariant under unitary transformations and

$$F(u^*xu) = u^*F(x)u \; x \in \mathcal{D} \tag{22}$$

for every unitary transformation $u$ on $\mathcal{H}$. Furthermore, to pairs of mutually orthogonal projections $p$ and $q$ acting on $\mathcal{H}$, the element $pxp + qxq$ should be in $\mathcal{D}$ and the equality

$$F(pxp + qxq) = pF(pxp)p + qF(qxq)q \tag{23}$$

should hold for any $x \in B(\mathcal{H})$ such that $pxp$ and $qxq$ are in $\mathcal{D}$. An operator function is a spectral function if and only if (22) and (23) are satisfied, cf. [143], [147].

The notion of spectral function is not immediately extendable to functions of several variables. However, we may consider the two properties of spectral functions noticed by $C$. Davis as a kind of regularity conditions, and they are readily extendable to functions of more than one variable.
The notion of a regular map of two operator variables were studied by Effros and [144], $cf$. also [146].
**Definition (3.2.1)[142]:** Let $F: \mathcal{D} \to B(\mathcal{H})$ be a mapping of $k$ variables defined in a convex domain $\mathcal{D} \subseteq B(\mathcal{H}) \times \cdots \times B(\mathcal{H})$. We say that $F$ is regular if
(i) The domain $\mathcal{D}$ is invariant under unitary transformations of $\mathcal{H}$ and

$$F(u^*x_1u, \dots, u^*x_ku) = u^*F(x_1, \dots, x_k)u$$

for every $x = (x_1, \dots, x_k) \in \mathcal{D}$ and every unitary $u$ on $\mathcal{H}$.
(ii) Let $p$ and $q$ be mutually orthogonal projections acting on $\mathcal{H}$ and take arbitrary $k$-tuples $(x_1, \dots, x_k)$ and $(y_1, \dots, y_k)$ of operators in $B(\mathcal{H})$ such that the compressed tuples

$$(px_1p, \dots, px_kp) \text{ and } (qy_1q, \dots, qy_kq)$$

are in the domain $\mathcal{D}$. Then the $k$-tuple of diagonal block matrices

$$(px_1p + qy_1q, \dots, px_kp + qy_kq)$$

is also in the domain $\mathcal{D}$ and

$$F(px_1p + qy_1q, \dots, px_kp + qy_kq)$$
$$= pF(px_1p, \dots, px_kp)p + qF(qy_1q, \dots, qy_kq)q.$$

By choosing $q$ as the zero projection in the second condition in the above definition we obtain

$$F(px_1p, \dots, px_kp) = pF(px_1p, \dots, px_kp)p,$$

which shows that $F$ for any orthogonal projection $p$ on $\mathcal{H}$ may be considered as a regular operator mapping

$$F: \mathcal{D}_p \to B(p\mathcal{H}),$$

where the compressed domain

$$\mathcal{D}_p = \left\{ (x_1, \dots, x_k) \in \bigoplus_{m=1}^{k} B(p\mathcal{H}) \mid (x_1 \oplus 0(1-p), \dots, x_k \oplus 0(1-p)) \in \mathcal{D} \right\}.$$

With this interpretation we may unambiguously calculate block matrices by the formula

$$F\left( \begin{pmatrix} x_1 & 0 \\ 0 & y_1 \end{pmatrix}, \dots, \begin{pmatrix} x_k & 0 \\ 0 & y_k \end{pmatrix} \right) = \begin{pmatrix} F(x_1, \dots, x_k) & 0 \\ 0 & F(y_1, \dots, y_k) \end{pmatrix}$$

which is well-known from mappings generated by the functional calculus.

We consider throughout the domain

$$\mathcal{D}^k = \{(A_1, \dots, A_k) \mid A_1, \dots, A_k \geq 0\}$$

of $k$-tuples of positive semi-definite operators acting on an infinite dimensional Hilbert space $\mathcal{H}$. It is convenient to consider an infinite dimensional Hilbert space since in this case $\mathcal{H}$ is isomorphic to $\mathcal{H} \oplus \mathcal{H}$ which allows us to use block matrix techniques without imposing dimension conditions.

**Theorem (3.2.2)[142]:** Consider a convex regular mapping

$$F: \mathcal{D}^k \to B(\mathcal{H})_{sa}$$

of $\mathcal{D}^k$ into self-adjoint operators acting on $\mathcal{H}$.

(i) Let $C$ be a contraction on $\mathcal{H}$. If $F(0, \dots, 0) \leq 0$ then the inequality

$$F(C^*A_1C, \dots, C^*A_kC) \leq C^*F(A_1, \dots, A_k)C$$

holds for $k$-tuples $(A_1, \dots, A_k)$ in $\mathcal{D}^k$.

(ii) Let $X$ and $Y$ be operators acting on $\mathcal{H}$ with $X^*X + Y^*Y = 1$. Then the inequality

$$F(X^*A_1X + Y^*B_1Y, \dots, X^*A_kX + Y^*B_kY) \leq X^*F(A_1, \dots, A_k)X + Y^*F(B_1, \dots, B_k)Y$$

holds for $k$-tuples $(A_1, \dots, A_k)$ and $(B_1, \dots, B_k)$ in $\mathcal{D}^k$.

**Proof.** By setting $T = (1 - C^*C)^{1/2}$ and $S = (1 - CC^*)^{1/2}$ we obtain that the block matrices

$$U = \begin{pmatrix} C & S \\ T & -C^* \end{pmatrix} \text{ and } V = \begin{pmatrix} C & -S \\ -T & -C^* \end{pmatrix}$$

are unitary operators on $\mathcal{H} \oplus \mathcal{H}$. Furthermore,

$$\frac{1}{2}U^* \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} U + \frac{1}{2}V^* \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} V = \begin{pmatrix} C^*AC & 0 \\ 0 & SAS \end{pmatrix}$$

for any operator $A \in B(\mathcal{H})$. By using that $F$ is a convex regular map we obtain

$$\begin{aligned}
&\begin{pmatrix} F(C^*A_1C, \dots, C^*A_kC) & 0 \\ 0 & F(SA_1S, \dots, SA_kS) \end{pmatrix} \\
&= F\left( \begin{pmatrix} C^*A_1C & 0 \\ 0 & SA_1S \end{pmatrix}, \dots, \begin{pmatrix} C^*A_kC & 0 \\ 0 & SA_kS \end{pmatrix} \right) \\
&= F\left( \frac{1}{2}U^* \begin{pmatrix} A_1 & 0 \\ 0 & 0 \end{pmatrix} U + \frac{1}{2}V^* \begin{pmatrix} A_1 & 0 \\ 0 & 0 \end{pmatrix} V, \dots, \frac{1}{2}U^* \begin{pmatrix} A_k & 0 \\ 0 & 0 \end{pmatrix} U \right. \\
&\left. + \frac{1}{2}V^* \begin{pmatrix} A_k & 0 \\ 0 & 0 \end{pmatrix} V \right) \\
&\leq \frac{1}{2}F\left( U^* \begin{pmatrix} A_1 & 0 \\ 0 & 0 \end{pmatrix} U, \dots, U^* \begin{pmatrix} A_k & 0 \\ 0 & 0 \end{pmatrix} U \right) \\
&+ \frac{1}{2}F\left( V^* \begin{pmatrix} A_1 & 0 \\ 0 & 0 \end{pmatrix} V, \dots, V^* \begin{pmatrix} A_k & 0 \\ 0 & 0 \end{pmatrix} V \right) \\
&= \frac{1}{2}U^*F\left( \begin{pmatrix} A_1 & 0 \\ 0 & 0 \end{pmatrix}, \dots, \begin{pmatrix} A_k & 0 \\ 0 & 0 \end{pmatrix} \right) U + \frac{1}{2}V^*F\left( \begin{pmatrix} A_1 & 0 \\ 0 & 0 \end{pmatrix}, \dots, \begin{pmatrix} A_k & 0 \\ 0 & 0 \end{pmatrix} \right) V
\end{aligned}$$

$$= \frac{1}{2}U^*\begin{pmatrix} F(A_1,\dots,A_k) & 0 \\ 0 & F(0,\dots,0) \end{pmatrix}U + \frac{1}{2}V^*\begin{pmatrix} F(A_1,\dots,A_k) & 0 \\ 0 & F(0,\dots,0) \end{pmatrix}V$$
$$\le \frac{1}{2}U^*\begin{pmatrix} F(A_1,\dots,A_k) & 0 \\ 0 & 0 \end{pmatrix}U + \frac{1}{2}V^*\begin{pmatrix} F(A_1,\dots,A_k) & 0 \\ 0 & 0 \end{pmatrix}V$$
$$= \begin{pmatrix} C^*F(A_1,\dots,A_k)C & 0 \\ 0 & SF(A_1,\dots,A_k)S \end{pmatrix},$$
where we used convexity in the first inequality, and in the second inequality used $F(0,\dots,0) \le 0$. The first statement now follows.

In order to prove $(ii)$ we define the map
$$G(A_1,\dots,A_k) = F(A_1,\dots,A_k) - F(0,\dots,0)\ (A_1,\dots,A_k) \in \mathcal{D}^k.$$
Unitary invariance of $F$ implies that $F(0,\dots,0)$ is a multiple of the unit operator and thus commutes with all projections. Therefore $G$ is regular and convex with $G(0,\dots,0) = 0$. We then define block matrices
$$C = \begin{pmatrix} X & 0 \\ Y & 0 \end{pmatrix} \text{ and } Z_m = \begin{pmatrix} A_m & 0 \\ 0 & B_m \end{pmatrix},\ m = 1,\dots,k$$
and notice that
$$C^*Z_mC = \begin{pmatrix} X^*A_mX + Y^*B_mY & 0 \\ 0 & 0 \end{pmatrix}$$
for $m = 1,\dots,k$. Finally we use $(i)$ to obtain
$$\begin{pmatrix} G(X^*A_1X + Y^*B_1Y,\dots,X^*A_kX + Y^*B_kY) & 0 \\ 0 & 0 \end{pmatrix}$$
$$= \begin{pmatrix} G(X^*A_1X + Y^*B_1Y,\dots,X^*A_kX + Y^*B_kY) & 0 \\ 0 & G(0,\dots,0) \end{pmatrix}$$
$$= G(C^*Z_1C,\dots,C^*Z_kC)$$
$$\le C^*G(Z_1,\dots,Z_k)C = C^*\begin{pmatrix} G(A_1,\dots,A_k) & 0 \\ 0 & G(B_1,\dots,B_k) \end{pmatrix}C$$
$$= \begin{pmatrix} X^*G(A_1,\dots,A_k)X + Y^*G(B_1,\dots,B_k)Y & 0 \\ 0 & 0 \end{pmatrix}$$
from which we deduce that
$$F(X^*A_1X + Y^*B_1Y,\dots,X^*A_kX + Y^*B_kY)$$
$$= G(X^*A_1X + Y^*B_1Y,\dots,X^*A_kX + Y^*B_kY) + F(0,\dots,0)$$
$$\le X^*G(A_1,\dots,A_k)X + Y^*G(B_1,\dots,B_k)Y + F(0,\dots,0)$$
$$= X^*F(A_1,\dots,A_k)X + Y^*F(B_1,\dots,B_k)Y$$
$$- X^*F(0,\dots,0)X - Y^*F(0,\dots,0)Y + F(0,\dots,0).$$
Since as above $F(0,\dots,0) = c \cdot 1$ for some real constant $c$ we obtain
$$-X^*F(0,\dots,0)X - Y^*F(0,\dots,0)Y + F(0,\dots,0)$$
$$= -c(X^*X + Y^*Y) + c \cdot 1 = 0,$$
and the statement of the theorem follows.

We shall for $k = 1,2,\dots$ consider the convex domain
$$\mathcal{D}_+^k = \{(A_1,\dots,A_k) \mid A_1,\dots,A_k > 0\}$$
of positive definite and invertible operators acting on the Hilbert space $\mathcal{H}$.

**Proposition (3.2.3)[142]:** Let $F$ be a regular map of $\mathcal{D}_+^k$ into self-adjoint operators acting on $\mathcal{H}$. We assume that

(i) $F$ is convex

(ii) $F(tA_1,\dots,tA_k) = tF(A_1,\dots,A_k)\ t > 0,\ (A_1,\dots,A_k) \in \mathcal{D}_+^k$.

Then

$$F(C^*A_1C, \dots, C^*A_kC) = C^*F(A_1, \dots, A_k)C$$
for any invertible operator $C$ on $\mathcal{H}$ and $(A_1, \dots, A_k) \in \mathcal{D}_+^k$.

**Proof.** Assume first that $C$ is an invertible contraction on $\mathcal{H}$. Jensen's subhomogeneous inequality is only available for regular mappings defined in $\mathcal{D}^k$.

To $\varepsilon > 0$ we therefore consider the mapping $F_\varepsilon: \mathcal{D}^k \to B(\mathcal{H})$ by setting
$$F_\varepsilon(A_1, \dots, A_k) = F(\varepsilon + A_1, \dots, \varepsilon + A_k) - F(\varepsilon, \dots, \varepsilon).$$
By unitary invariance of $F$ we realise that $F(\varepsilon, \dots, \varepsilon)$ is a multiple of the unity. Therefore, $F_\varepsilon$ is regular and convex with $F_\varepsilon(0, \dots, 0) = 0$.

We may thus use Jensen's sub-homogeneous inequality for regular mappings and obtain
$$F_\varepsilon(C^*A_1C, \dots, C^*A_kC) \le C^*F_\varepsilon(A_1, \dots, A_k)C,$$
where we now restrict $(A_1, \dots, A_k)$ to the domain $\mathcal{D}_+^k$ and rearrange the inequality to
$$\begin{aligned}&F(\varepsilon + C^*A_1C, \dots, \varepsilon + C^*A_kC)\\ &\le C^*F(A_1 + \varepsilon, \dots, A_k + \varepsilon)C + F(\varepsilon, \dots, \varepsilon) - C^*F(\varepsilon, \dots, \varepsilon)C.\end{aligned}$$
Since $F$ is positively homogeneous the term $F(\varepsilon, \dots, \varepsilon) = \varepsilon F(1, \dots, 1)$ is vanishing for $\varepsilon \to 0$ and we obtain
$$F(C^*A_1C, \dots, C^*A_kC) \le C^*F(A_1, \dots, A_k)C \tag{24}$$
for invertible $C$. Again using homogeneousness we obtain inequality (24) also for arbitrary invertible $C$. Then by repeated application of (24) we obtain
$$F(A_1, \dots, A_k) \le C^{*-1}F(C^*A_1C, \dots, C^*A_kC)C^{-1} \le F(A_1, \dots, A_k)$$
and the statement follows.

**Definition (3.2.4)[142]:** Let $F: \mathcal{D}_+^k \to B(\mathcal{H})$ be a regular mapping. The perspective map $\mathcal{P}_F$ is the mapping defined in the domain $\mathcal{D}_+^{k+1}$ by setting
$$\mathcal{P}_F(A_1, \dots, A_k, B) = B^{1/2}F\left(B^{-1/2}A_1B^{-1/2}, \dots, B^{-1/2}A_kB^{-1/2}\right)B^{1/2}$$
for positive invertible operators $A_1, \dots, A_k$ and $B$ acting on $\mathcal{H}$.

It is a small exercise to prove that the perspective $\mathcal{P}_F$ is a regular mapping which is positively homogeneous in the sense that
$$\mathcal{P}_F(tA_1, \dots, tA_k, tB) = t\mathcal{P}_F(A_1, \dots, A_k, B)$$
for arbitrary $(A_1, \dots, A_k, B) \in \mathcal{D}_+^{k+1}$ and real numbers $t > 0$. The following theorem generalises a result of Effros [145] for functions of one variable.

**Theorem (3.2.5)[142]:** The perspective $\mathcal{P}_F$ of a convex regular map $F: \mathcal{D}_k^+ \to B(\mathcal{H})$ is convex.

**Proof.** Consider tuples $(A_1, \dots, A_{k+1})$ and $(B_1, \dots, B_{k+1})$ in $\mathcal{D}_+^{k+1}$ and take $\lambda \in [0,1]$. We define the operators
$$\begin{aligned}C &= \lambda A_{k+1} + (1-\lambda)B_{k+1}\\ X &= \lambda^{1/2}A_{k+1}^{1/2}C^{-1/2}\\ Y &= (1-\lambda)^{1/2}B_{k+1}^{1/2}C^{-1/2}\end{aligned}$$
and calculate that
$$X^*X + Y^*Y = C^{-1/2}\lambda A_{k+1}C^{-1/2} + C^{-1/2}(1-\lambda)B_{k+1}C^{-1/2} = 1$$
and
$$\begin{aligned}&X^*A_{k+1}^{-1/2}A_iA_{k+1}^{-1/2}X + Y^*B_{k+1}^{-1/2}B_iB_{k+1}^{-1/2}Y\\ &= C^{-1/2}\lambda^{1/2}A_{k+1}^{1/2}A_{k+1}^{-1/2}A_iA_{k+1}^{-1/2}\lambda^{1/2}A_{k+1}^{1/2}C^{-1/2}\\ &+ C^{-1/2}(1-\lambda)^{1/2}B_{k+1}^{1/2}B_{k+1}^{-1/2}B_iB_{k+1}^{-1/2}(1-\lambda)^{1/2}B_{k+1}^{1/2}C^{-1/2}\\ &= C^{-1/2}(\lambda A_i + (1-\lambda)B_i)C^{-1/2}\end{aligned}$$
for $i = 1, \dots, k$. We thus obtain

$$
\begin{aligned}
&\mathcal{P}_F(\lambda A_1+(1-\lambda)B_1,\dots,\lambda A_{k+1}+(1-\lambda)B_{k+1})\\
&=C^{1/2}F\big(C^{-1/2}(\lambda A_1+(1-\lambda)B_1)C^{-1/2},\dots,\\
&C^{-1/2}(\lambda A_k+(1-\lambda)B_k)C^{-1/2}\big)C^{1/2}\\
&=C^{1/2}F\Big(X^*A_{k+1}^{-1/2}A_1A_{k+1}^{-1/2}X+Y^*B_{k+1}^{-1/2}B_1B_{k+1}^{-1/2}Y,\dots,\\
&X^*A_{k+1}^{-1/2}A_kA_{k+1}^{-1/2}X+Y^*B_{k+1}^{-1/2}B_kB_{k+1}^{-1/2}Y\Big)C^{1/2}\\
&\le C^{1/2}\left(X^*F\Big(A_{k+1}^{-1/2}A_1A_{k+1}^{-1/2},\dots,A_{k+1}^{-1/2}A_kA_{k+1}^{-1/2}\Big)X\right.\\
&\left.+Y^*F\Big(B_{k+1}^{-1/2}B_1B_{k+1}^{-1/2},\dots,B_{k+1}^{-1/2}B_kB_{k+1}^{-1/2}\Big)Y\right)C^{1/2}\\
&=\lambda A_{k+1}^{1/2}F\Big(A_{k+1}^{-1/2}A_1A_{k+1}^{-1/2},\dots,A_{k+1}^{-1/2}A_kA_{k+1}^{-1/2}\Big)A_{k+1}^{1/2}\\
&+(1-\lambda)B_{k+1}^{1/2}F\Big(B_{k+1}^{-1/2}B_1B_{k+1}^{-1/2},\dots,B_{k+1}^{-1/2}B_kB_{k+1}^{-1/2}\Big)B_{k+1}^{1/2}\\
&=\lambda\mathcal{P}_F(A_1,\dots,A_{k+1})+(1-\lambda)\mathcal{P}_F(B_1,\dots,B_{k+1}),
\end{aligned}
$$

where we used Jensen's inequality for regular mappings.

**Proposition (3.2.6)[142]:** Let $F:\mathcal{D}_+^{k+1}\to B(\mathcal{H})$ be a convex and positively homogeneous regular mapping. Then $F$ is the perspective of its restriction $G$ to $\mathcal{D}_+^k$ given by

$$G(A_1,\dots,A_k)=F(A_1,\dots,A_k,1)$$

for positive invertible operators $A_1,\dots,A_k$ acting on $\mathcal{H}$.

**Proof.** Since $F$ is a convex and positively homogeneous regular mapping we may apply Proposition (3.2.3). Then by setting $C=A_{k+1}^{-1/2}$ we obtain

$$
\begin{aligned}
&A_{k+1}^{-1/2}F(A_1,\dots,A_k,\quad A_{k+1})A_{k+1}^{-1/2}\\
&\qquad=F\Big(A_{k+1}^{-1/2}A_1A_{k+1}^{-1/2},\dots,A_{k+1}^{-1/2}A_kA_{k+1}^{-1/2},1\Big).
\end{aligned}
$$

By rearranging this equation we obtain

$$F(A_1,\dots,A_k,A_{k+1})=A_{k+1}^{1/2}G\Big(A_{k+1}^{-1/2}A_1A_{k+1}^{-1/2},\dots,A_{k+1}^{-1/2}A_kA_{k+1}^{-1/2}\Big)A_{k+1}^{1/2}$$

which is the statement to be proved.

The result in the above proposition may be reformulated in the following way: The perspective $\mathcal{P}_G$ of a convex regular mapping $G:\mathcal{D}_+^k\to B(\mathcal{H})$ is the unique extension of $G$ to a positively homogeneous convex regular mapping $F:\mathcal{D}_+^{k+1}\to B(\mathcal{H})$.

We construct a sequence of multivariate geometric means $G_1,G_2,\dots$ by the following general procedure.

(i) We begin by setting $G_1(A)=A$ for each positive definite invertible operator $A$.

(ii) To each geometric mean $G_k$ of $k$ variables we associate an auxiliary mapping $F_k:\mathcal{D}_+^k\to B(\mathcal{H})$ such that

(a) $F_k$ is a regular map,

(b) $F_k$ is concave,

(c) $F_k(t_1,\dots,t_k)=(t_1\cdots t_k)^{1/(k+1)}$ for positive numbers $t_1,\dots,t_k$.

(iii) We define the geometric mean $G_{k+1}:\mathcal{D}_+^{k+1}\to B(\mathcal{H})$ of $k+1$ variables as the perspective

$$G_{k+1}(A_1,\dots,A_{k+1})=\mathcal{P}_{F_k}(A_1,\dots,A_{k+1})$$

of the auxiliary map $F_k$.

Geometric means defined by this very general procedure are concave and positively homogeneous regular mappings by Theorem (3.2.5) and the preceding remarks. They also satisfy

$$G_k(A_1,\dots,A_k) = (A_1 \cdots A_k)^{\frac{1}{k}} \tag{25}$$

for commuting operators. Indeed, since $G_k$ is the perspective of $F_{k-1}$ and this map satisfies $(c)$ in condition $(ii)$, we obtain $G_k(t_1,\dots,t_k) = (t_1 \cdots t_k)^{1/k}$ for positive numbers. Equality (25) then follows since $G_k$ is regular. The geometric mean of two variables

$$G_2(A_1,A_2) = A_2^{1/2}\left(A_2^{-1/2}A_1A_2^{-1/2}\right)^{1/2}A_2^{1/2} \tag{26}$$

coincides with the geometric mean of two variables $A_1\#A_2$ introduced by Pusz and Woronowicz. This is so since $G_2$ is the perspective of $F_1$ and $F_1(A) = A^{1/2}$. The last statement is obtained since $F_1$ is a regular mapping and satisfies $F_1(t) = t^{1/2}$ for positive numbers by $(c)$ in condition $(ii)$.

There are many ways to associate the auxiliary map $F_k$ in the above procedure, so we should not in general expect much similarity between the geometric means for different number of variables.

We define the auxiliary mapping $F_k\colon \mathcal{D}_+^k \to B(\mathcal{H})$ by setting

$$F_k(A_1,\dots,A_k) = G_k(A_1,\dots,A_k)^{k/(k+1)}$$

for $k = 1,2,\dots$.

**Theorem (3.2.7)[142]:** The means $G_k$ constructed then have the following properties:
(i) $G_k\colon \mathcal{D}_+^k \to B(\mathcal{H})_+$is a regular map for each $k = 1,2,\dots$
(ii) $G_k(tA_1,\dots,tA_k) = tG_k(A_1,\dots,A_k)$ for $t > 0, (A_1,\dots,A_k) \in \mathcal{D}_+^k$ and $k = 1,2,\dots$.
(iii) $G_k\colon \mathcal{D}_+^k \to B(\mathcal{H})$ is concave for each $k = 1,2,\dots$
(iv) $G_{k+1}(A_1,\dots,A_k,1) = G_k(A_1,\dots,A_k)^{k/(k+1)}$ for $(A_1,\dots,A_k) \in \mathcal{D}_+^k$ and $k = 1,2,\dots$.

Any sequence of mappings $\tilde{G}_k$ beginning with $\tilde{G}_1(A) = A$ and satisfying the above conditions coincide with the means $G_k$ for $k = 1,2,\dots$.

**Proof.** Each map $G_k$ is for $k = 2,3,\dots$ the perspective of a regular map and this implies $(i)$ and $(ii)$. The assertion of concavity for $G_1$ is immediate. Suppose now $G_k$ is concave for some $k$. Since the map $t \to t^p$ is both operator monotone and operator concave for $0 \le p \le 1$, we realise that the auxiliary mapping

$$F_k(A_1,\dots,A_k) = G_k(A_1,\dots,A_k)^{k/(k+1)}$$

is concave, and since $G_{k+1}$ is the perspective of $F_k$ we then obtain by Theorem (3.2.5) that also $G_{k+1}$ is concave. Since the first map $G_1$ is concave we have thus proved by induction that $G_k$ is concave for all $k = 1,2\dots$. The last property $(iv)$ follows since $G_{k+1}$ is the perspective of $G_k^{k/(k+1)}$.

Let finally $\tilde{G}_k$ be a sequence of mappings satisfying $(i)$ to $(iv)$. Since each $\tilde{G}_{k+1}$ is concave and homogeneous it follows by Proposition (3.2.6) that $\tilde{G}_{k+1}$ is the perspective of its restriction $\tilde{G}_{k+1}(A_1,\dots,A_k,1)$. Because of $(iv)$ we then realise that $\tilde{G}_{k+1}$ is the perspective of the map

$$\tilde{F}_k(A_1,\dots,A_k) = \tilde{G}_k(A_1,\dots,A_k)^{k/(k+1)}$$

constructed from $\tilde{G}_k$. The $\tilde{G}_k$ mappings are thus constructed by the same algorithm as the mappings $G_k$ for every $k \ge 2$, and since $\tilde{G}_1 = G_1$ they must all coincide.

In addition to the properties listed in the above theorem the means $G_k$ enjoy a number of other properties that we list below.

**Theorem (3.2.8)[142]:** The means $G_k$ constructed have the following additional properties:

(i) The means $G_k$ are increasing in each variable for $k = 1,2 \ldots$.
(ii) The means $G_k$ are congruence invariant. For any invertible operator $C$ on $\mathcal{H}$ the identity

$$G_k(C^*A_1C, \ldots, C^*A_kC) = C^*G_k(A_1, \ldots, A_k)C$$

holds for $(A_1, \ldots, A_k) \in \mathcal{D}_+^k$ and $k = 1,2, \ldots$
(iii) The means $G_k$ are jointly homogeneous in the sense that

$$G_k(t_1A_1, \ldots, t_kA_k) = (t_1 \cdots t_k)^{1/k} G_k(A_1, \ldots, A_k)$$

for scalars $t_1, \ldots, t_k > 0$, operators $(A_1, \ldots, A_k) \in \mathcal{D}_+^k$ and $k = 1,2, \ldots$
(iv) The means $G_k$ are self-dual in the sense that

$$G_k(A_1^{-1}, \ldots, A_k^{-1}) = G_k(A_1, \ldots, A_k)^{-1}$$

for $(A_1, \ldots, A_k) \in \mathcal{D}_+^k$ and $k = 1,2, \ldots$
(v) When restricted to positive definite matrices the determinant identity

$$\det G_k(A_1, \ldots, A_k) = (\det A_1 \cdots \det A_k)^{\frac{1}{k}}$$

holds for $k = 1,2 \ldots$.
**Proof.** The first property follows by the following standard argument for positive concave mappings. Consider positive definite invertible operators $A_m \leq B_m$ for $m = 1, \ldots, k$. By first assuming that the difference $B_m - A_m$ is invertible we may take $\lambda \in (0,1)$ and write

$$\lambda B_m = \lambda A_m + (1-\lambda) C_m \; m = 1, \ldots, k,$$

where $C_m = \lambda(1-\lambda)^{-1}(B_m - A_m)$ is positive definite and invertible. By using concavity we then obtain

$$\begin{aligned} G_k(\lambda B_1, \ldots, \lambda B_k) \quad &\geq \lambda G(A_1, \ldots, A_k) + (1-\lambda) G_k(C_1, \ldots, C_k) \\ &\geq \lambda G(A_1, \ldots, A_k). \end{aligned}$$

Letting $\lambda \to 1$ we obtain $G_k(B_1, \ldots, B_k) \geq G_k(A_1, \ldots, A_k)$ by continuity. In the general case we choose $0 < \mu < 1$ such that

$$\mu A_m < A_m \leq B_m \; m = 1, \ldots, k$$

and obtain $G_k(\mu A_1, \ldots, \mu A_k) \leq G_k(B_1, \ldots, B_k)$. By letting $\mu \to 1$ we then obtain $G_k(A_1, \ldots, A_k) \leq G_k(B_1, \ldots, B_k)$ which shows $(i)$.
Since $G_k$ is concave and homogeneous we obtain $(ii)$ from Proposition (3.2.3). Property $(iii)$ is immediate for $k = 1$ and $k = 2$. Suppose the property is verified for $k$, then

$$\begin{aligned} &G_{k+1}(t_1A_1, \ldots, t_kA_k, t_{k+1}A_{k+1}) \\ &= t_{k+1} A_{k+1}^{1/2} F_k\Big(t_1 t_{k+1}^{-1} A_{k+1}^{-1/2} A_1 A_{k+1}^{-1/2}, \ldots, t_k t_{k+1}^{-1} A_{k+1}^{-1/2} A_k A_{k+1}^{-1/2}\Big) A_{k+1}^{1/2} \\ &= t_{k+1} A_{k+1}^{1/2} G_k\Big(t_1 t_{k+1}^{-1} A_{k+1}^{-1/2} A_1 A_{k+1}^{-1/2}, \ldots, t_k t_{k+1}^{-1} A_{k+1}^{-1/2} A_k A_{k+1}^{-1/2}\Big)^{k/(k+1)} A_{k+1}^{1/2}. \end{aligned}$$

By using the induction assumption we obtain

$$\begin{aligned} &G_{k+1}(t_1A_1, \ldots, t_kA_k, t_{k+1}A_{k+1}) \\ &= t_{k+1}\Big(t_{k+1}^{-1} t_1^{1/k} \cdots t_k^{1/k}\Big)^{k/(k+1)} G_{k+1}(A_1, \ldots, A_k, A_{k+1}) \\ &= (t_1 \cdots t_k t_{k+1})^{1/(k+1)} G_{k+1}(A_1, \ldots, A_k, A_{k+1}) \end{aligned}$$

which shows $(iii)$.
Property $(iv)$ is immediate for $k = 1$ and $k = 2$. Suppose the property is verified for $k$, then

$$\begin{aligned} &G_{k+1}(A_1^{-1}, \ldots, A_k^{-1}, A_{k+1}^{-1}) \\ &= A_{k+1}^{-1/2} F_k\Big(A_{k+1}^{1/2} A_1^{-1} A_{k+1}^{1/2}, \ldots, A_{k+1}^{1/2} A_k^{-1} A_{k+1}^{1/2}\Big) A_{k+1}^{-1/2} \\ &= A_{k+1}^{-1/2} G_k\Big(A_{k+1}^{1/2} A_1^{-1} A_{k+1}^{1/2}, \ldots, A_{k+1}^{1/2} A_k^{-1} A_{k+1}^{1/2}\Big)^{k/(k+1)} A_{k+1}^{-1/2}. \end{aligned}$$

By using the induction assumption we obtain

$$
\begin{aligned}
&G_{k+1}(A_1^{-1},\dots,A_k^{-1},A_{k+1}^{-1})\\
&= A_{k+1}^{-1/2}G_k\left(A_{k+1}^{-1/2}A_1A_{k+1}^{-1/2},\dots,A_{k+1}^{-1/2}A_kA_{k+1}^{-1/2}\right)^{-k/(k+1)}A_{k+1}^{-1/2}\\
&= \left(A_{k+1}^{1/2}G_k\left(A_{k+1}^{-1/2}A_1A_{k+1}^{-1/2},\dots,A_{k+1}^{-1/2}A_kA_{k+1}^{-1/2}\right)^{k/(k+1)}A_{k+1}^{1/2}\right)^{-1}\\
&= G_{k+1}(A_1,\dots,A_k,A_{k+1})^{-1}
\end{aligned}
$$

which shows $(iv)$.

Notice that since $\det A = \exp(\operatorname{Tr}\log A)$ for positive definite $A$, we have $\det A^p = (\det A)^p$ for all real exponents $p$. Property $(v)$ is easy to calculate for $k = 1$ and $k = 2$. Suppose the property is verified for $k$. Since as above

$$
\begin{aligned}
&G_{k+1}(A_1,\dots,A_k,A_{k+1})\\
&= A_{k+1}^{1/2}G_k\left(A_{k+1}^{-1/2}A_1A_{k+1}^{-1/2},\dots,A_{k+1}^{-1/2}A_kA_{k+1}^{-1/2}\right)^{k/(k+1)}A_{k+1}^{1/2}
\end{aligned}
$$

we obtain

$$
\begin{aligned}
&\det G_{k+1}(A_1,\dots,A_k,A_{k+1})\\
&= \det A_{k+1}(\det A_{k+1}^{-1}\det A_1 \cdots \det A_{k+1}^{-1}\det A_k)^{1/(k+1)}\\
&= (\det A_1 \cdots \det A_k \cdot \det A_{k+1})^{1/k+1}
\end{aligned}
$$

which shows $(v)$.

**Theorem (3.2.9)[142]:** The geometric means $G_k$ are for $k = 1,2,\dots$ bounded between the symmetric harmonic and arithmetic means. That is,

$$
\frac{k}{A_1^{-1}+\cdots+A_k^{-1}} \le G_k(A_1,\dots,A_k) \le \frac{A_1+\cdots+A_k}{k}
$$

for arbitrary $(A_1,\dots,A_k) \in \mathcal{D}_+^k$ and $k = 1,2,\dots$.

**Proof.** The upper bound holds with equality for $k = 1$. Suppose that we have verified the inequality for $k$. Since by classical analysis

$$
X^{k/(k+1)} \le 1 + \frac{k}{k+1}(X-1)
$$

for positive definite $X$, we obtain

$$
\begin{aligned}
F_k(A_1,\dots,A_k) &= G_k(A_1,\dots,A_k)^{k/(k+1)} \le 1 + \frac{k}{k+1}(G_k(A_1,\dots,A_k)-1)\\
&\le 1 + \frac{k}{k+1}\left(\frac{A_1+\cdots+A_k}{k}-1\right) = \frac{A_1+\cdots+A_k+1}{k+1}.
\end{aligned}
$$

By taking perspectives we now obtain

$$
\begin{aligned}
&G_{k+1}(A_1,\dots,A_k,B) = \mathcal{P}_{F_k}(A_1,\dots,A_k,B)\\
&= B^{1/2}F_k\left(B^{-1/2}A_1B^{-1/2},\dots,B^{-1/2}A_kB^{-1/2}\right)B^{1/2}\\
&\le B^{1/2}\frac{B^{-1/2}A_1B^{-1/2}+\cdots+B^{-1/2}A_kB^{-1/2}+1}{k+1}B^{1/2} = \frac{A_1+\cdots A_k+B}{k+1}
\end{aligned}
$$

which proves the upper bound by induction. We next use the upper bound to obtain

$$
G_k(A_1^{-1},\dots,A_k^{-1}) \le \frac{A_1^{-1}+\cdots+A_k^{-1}}{k}.
$$

By inversion we then obtain

$$
\frac{k}{A_1^{-1}+\cdots+A_k^{-1}} \le G_k(A_1^{-1},\dots,A_k^{-1})^{-1} = G_k(A_1,\dots,A_k),
$$

where we in the last equation used self-duality of the geometric mean, $cf$. property $(iv)$ in Theorem (3.2.8).
The means studied are known in the literature as the inductive means of Sagae and Tanabe [152]. By considering the power mean

$$A\#_t B = B^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^t B^{1/2}\ 0 \le t \le 1$$

they established the recursive relation by setting

$$G_{k+1}(A_1, \dots, A_{k+1}) = G_k(A_1, \dots, A_k)\#_{k/(k+1)}A_{k+1}.$$

We established the harmonic-geometric-arithmetic mean inequality of Theorem (3.2.9).
It is possible to prove the crucial concavity property $(iii)$ in Theorem (3.2.7) by induction. It can be done without the general theory of perspectives of regular operator mappings, and it only requires the properties of an operator mean of two variables as studied by Kubo and Ando [149]. However, this is a special situation that only applies to the inductive means.

The inductive geometric means are uniquely specified within the general framework discussed by choosing the updating condition (20), cf. property (iv) in Theorem (3.2.7). We may instead construct geometric means satisfying updating condition (21) by choosing the auxiliary map

$$F_k(A_1, \dots, A_k) = G_k\left(A_1^{k/(k+1)}, \dots, A_k^{k/(k+1)}\right)$$

for $k = 1,2$ ... It is a small exercise to realise that these means satisfy all of the properties listed in Theorem (3.2.7), Theorem (3.2.8), and Theorem (3.2.9) with the only exception that condition $(iv)$ in Theorem (3.2.7) is replaced by updating condition (21). Concavity of these means cannot be reduced to concavity of operator means of two variables but relies on the general theory of regular operator mappings and Theorem (3.2.5).

The Karcher mean $\Lambda_k(A_1, \dots, A_k)$ of $k$ positive definite invertible operator variables is defined as the unique positive definite solution to the equation

$$\sum_{i=1}^{k} \log\left(X^{\frac{1}{2}}A_i X^{\frac{1}{2}}\right) = 0, \tag{27}$$

and it enjoys all of the attractive properties of an operator mean listed by Ando, $Li$, and Mathias, $cf$. [150]. The defining equation (27) immediately implies that the Karcher mean $\Lambda_k: \mathcal{D}_+^k \to B(\mathcal{H})$ is a regular operator mapping, and it may therefore be understood within the general framework discussed by choosing the auxiliary map

$$F_k(A_1, \dots, A_k) = \Lambda_{k+1}(A_1, \dots, A_k, 1).$$

The problem, however, is that we do not have any explicit expression of $F_k$ in terms of $\Lambda_k$.

**Section (3.3): Complete Convergence and Complete Moment Convergence**

The Toeplitz lemma and its two corollaries (the Cesàro mean convergence theorem and the Kronecker lemma) are useful tools in the study of probability limit theorems. We spell out them in the following and their proofs may be found in [138].

**Theorem (3.3.1)[153]:** (Toeplitz lemma) Let $\{a_{nk}, 1 \le k \le k_n, n \ge 1\}$ be a double array of real numbers such that for any $k \ge 1$, $\lim_{n\to\infty} a_{nk} = 0$ and $\sup_{n\ge1}\sum_{k=1}^{\bar{k}_n}|a_{nk}| < \infty$. Let $\{x_n, n \ge 1\}$ ba a sequence of real numbers.

(i) If $\lim_{n\to\infty} x_n = 0$, then $\lim_{n\to\infty}\sum_{k=1}^{k_n} a_{nk}x_k = 0$.

(ii) If $\lim_{n\to\infty} x_n = x \in \mathbf{R}$ and $\lim_{n\to\infty}\sum_{k=1}^{k_n} a_{nk} = 1$, then $\lim_{n\to\infty}\sum_{k=1}^{k_n} a_{nk}x_k = x$.

**Corollary (3.3.2)[153]:** (Cesàro mean convergence theorem) Let $\{x_n, n \geq 1\}$ be a sequence of real numbers and let $\bar{x}_n = \sum_{k=1}^{n} x_k/n, n \geq 1$. If $\lim_{n\to\infty} x_n = x \in \mathbf{R}$, then $\lim_{n\to\infty} \bar{x}_n = x$.

**Corollary (3.3.3)[153]:** (Kronecker lemma) Let $\{x_n, n \geq 1\}$ and $\{b_n, n \geq 1\}$ be sequences of real numbers such that $0 < b_n \uparrow \infty$. If the series $\sum_{k=1}^{\infty} x_k/b_k$ converges, then $\lim_{n\to\infty} \frac{1}{b_n} \sum_{k=1}^{n} x_k = 0$.

By the definition of almost sure $(a.s.)$ convergence, we know that the Toeplitz lemma and its two corollaries (the Cesàro mean convergence theorem and the Kronecker lemma) still hold when the numerical sequence $\{x_n, n \geq 1\}$ and real number $x$ are replaced by a sequence of random variable $\{X_n, n \geq 1\}$ and a random variable $X$, respectively, and the limit is taken to be $a.s.$ convergence.

Recently, [131] showed among other things that "convergence in probability" versions of the Toeplitz lemma, the Cesàro mean convergence theorem and the Kronecker lemma can fail, and their "mean convergence" versions are true.

We will give two examples to show that they can fail in general. Then we give some sufficient conditions for the Cesàro mean convergence theorem under complete convergence.

For $\{X, X_n, n \geq 1\}$ be a sequence of random variables on some probability space $(\Omega, \mathcal{F}, P)$. If $\forall \varepsilon > 0$,

$$\sum_{n=1}^{\infty} P\{|X_n - X| \geq \varepsilon\} < \infty,$$

then $\{X_n, n \geq 1\}$ is said to converge completely to $X$ (write $X_n \overset{\text{c.c.}}{\longrightarrow} X$, or $X_n \to X\ c.c.$ for short).

This concept was introduced by [168]. Let $\{X, X_n, n \geq 1\}$ be a sequence of independent and indentically distributed $(i.i.d.)$ random variables and set $S_n = \sum_{k=1}^{n} X_k, n \geq 1$.

[168] proved that if $E[X] = 0$ and $E[X^2] < \infty$, then $S_n/n \overset{\text{c.c.}}{\longrightarrow} 0$. The converse was proved by [160], [161]. The Hsu-Robbins-Erdös theorem was generalized in various ways, see, [154], [165], [166], [172], [171], [181], [180], [167], and [158].

In view of the relations between convergence in probability and complete convergence, we introduce two classes of complete moment convergences, which are stronger versions of mean convergence. Let $p > 0$.

**Definition (3.3.4)[153]:** $\{X_n, n \geq 1\}$ is said to $s$ - $L^p$ converge to $X$ (denote $X_n \overset{s-L^p}{\longrightarrow} X$ for short), if

$$\sum_{n=1}^{\infty} E[|X_n - X|^p] < \infty.$$

**Definition (3.3.5)[153]:** $\{X_n, n \geq 1\}$ is said to $s^* - L^p$ converge to $X$ (denote $X_n \overset{s^*-L^p}{\longrightarrow} X$ for short), if

$$\sum_{n=1}^{\infty} \|X_n - X\|_p < \infty,$$

where $\|X_n - X\|_p = (E[|X_n - X|^p])^{1/p}$.

**Remark (3.3.6)[153]:** (i) Obviously, if $X_n \stackrel{s-L^p}{\longrightarrow} X$ or $X_n \stackrel{s^*-L^p}{\longrightarrow} X$ for some $p > 0$, then $\|X_n - X\|_p \to 0$.

(ii) By Markov's inequality, we know that if $X_n \stackrel{s-L^p}{\longrightarrow} X$ for some $p > 0$, then $X_n \stackrel{c.c.}{\longrightarrow} X$ and thus $X_n \stackrel{a.s}{\longrightarrow} X$ by the Borel-Cantelli lemma.

(iii) If $p > 1$ and $X_n \stackrel{s^*-L^p}{\longrightarrow} X$, then $X_n \stackrel{s-L^p}{\longrightarrow} X$; if $0 < p < 1$ and $X_n \stackrel{s-L^p}{\longrightarrow} X$, then $X_n \stackrel{s^*-L^p}{\longrightarrow} X$.
[159] first investigated the complete moment convergence, and obtained the following result. Let $\{X, X_n, n \geq 1\}$ be a sequence of $i.i.d.$ random variables with $E[X] = 0$. Let $1 \leq p < 2$ and $\gamma \geq p$. If $E[|X|^\gamma + |X|\log(1 + |X|)] < \infty$, then

$$\sum_{n\geq 1} n^{\frac{2}{p}-2-\frac{1}{p}} E\left[\left(|S_n| - \varepsilon n^{\frac{1}{p}}\right)^+\right] < \infty \text{ for all } \varepsilon > 0, \qquad (28)$$

where $x^+ = \max\{0, x\}$.
Chow's result has been generalized in various directions. [182], [183], [155], [164], [179], [185], and [177] studied complete moment convergence for sums of Banach space valued random elements.
[174], [157], [170], and [187] considered complete moment convergence for moving average processes. [169], [173], [176], [185], [162], [188], and [156] studied precise asymptotics for complete moment convergence. [184], [175], and [163] considered complete moment convergence for negatively associated random variables. $Qiu$ and Chen [177] studied complete moment convergence for $i.i.d.$ random variables, and extended two results in [167] to complete moment convergence.
**Example (3.3.7)[153]:** Let $\{X_n, n \geq 1\}$ be a sequence of random variables with $\sup_{i\geq 1} E[|X_i|] \leq C$ for some positive constant $C$. Then for any $\alpha > 1$, we have $\frac{S_n}{n^2(\ln n)^\alpha} \stackrel{s-L^1}{\longrightarrow} 0$. In fact,

$$\sum_{n=1}^{\infty} E\left[\left|\frac{S_n}{n^2(\ln n)^\alpha}\right|\right] \leq \sum_{n=1}^{\infty} \frac{1}{n^2(\ln n)^\alpha} \sum_{k=1}^{n} E[|X_i|] \leq C \sum_{n=1}^{\infty} \frac{1}{n(\ln n)^\alpha} < \infty.$$

**Example (3.3.8)[153]:** Let $\{X_n, n \geq 1\}$ be a sequence of pairwise uncorrelated random variables with $\sup_{i\geq 1} \operatorname{Var}(X_i) \leq C$ for some positive constant $C$, where $\operatorname{Var}(X_i)$ stands for the variance of $X_i$.

Then for any $\alpha > 1$, we have $\frac{S_n - E[S_n]}{n^{3/2}(\ln n)^\alpha} \stackrel{s^*-L^2}{\longrightarrow} 0$. In fact,

$$\begin{aligned}
\sum_{n=1}^{\infty} \left\|\frac{S_n - E[S_n]}{n^{3/2}(\ln n)^\alpha}\right\|_2 &= \sum_{n=1}^{\infty} \frac{1}{n^{3/2}(\ln n)^\alpha} \left(\operatorname{Var}(S_n)\right)^{1/2} \\
&= \sum_{n=1}^{\infty} \frac{1}{n^{3/2}(\ln n)^\alpha} \left(\sum_{k=1}^{n} \operatorname{Var}(X_i)\right)^{1/2} \\
&\leq \sqrt{C} \sum_{n=1}^{\infty} \frac{1}{n(\ln n)^\alpha} < \infty.
\end{aligned}$$

We consider "$s - L^p$ convergence " versions and " $s^* - L^p$ convergence " versions of the Toeplitz lemma, the Cesàro mean convergence theorem and the Kronecker lemma. Four counterexamples will be given to show that they can fail in general. Some sufficient conditions for the Cesàro mean convergence theorem under these two complete moment convergences will be presented.

We will construct two counterexamples to show that "complete convergence " versions of the Toeplitz lemma, the Cesarro mean convergence theorem and the Kronecker lemma can fail in general.

The next example shows that complete convergence version of the Cesàro mean convergence theorem fails.

**Example (3.3.9)[153]:** Suppose that $\{X_n, n \geq 1\}$ is a sequence of independent random variables such that $P(X_n = n) = \frac{1}{n^2}, P(X_n = 0) = 1 - \frac{1}{n^2}$. For any $\varepsilon > 0$, we have

$$\sum_{n=1}^{\infty} P(|X_n - 0| \geq \varepsilon) = \sum_{n=1}^{\infty} P(X_n = n) = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty,$$

$i.e. X_n \to 0\ c.c.$ Let $\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k, n \geq 1$. In the following, we will show that $\bar{X}_n \nrightarrow 0\ c.c.$

Let $n = 2k, k \geq 2$ and define $k$ sets $A_1, \cdots, A_k$ as follows:

$$\begin{aligned} A_1 &:= \{X_{2k} = 2k\}, \\ A_2 &:= \{X_{2k} = 0, X_{2k-1} = 2k - 1\}, \\ &\cdots \\ A_k &:= \{X_{2k} = 0, \cdots, X_{k+2} = 0, X_{k+1} = k + 1\}. \end{aligned}$$

Then we have $\bigcup_{i=1}^{k} A_i \subset \left\{\bar{X}_n \geq \frac{1}{2}\right\}$, and thus

$$P\left(\bar{X}_n \geq \frac{1}{2}\right) \geq \sum_{i=1}^{k} P(A_i)$$

$$= \frac{1}{(2k)^2} + \left(1 - \frac{1}{(2k)^2}\right)\frac{1}{(2k-1)^2} + \cdots + \prod_{j=2}^{k}\left(1 - \frac{1}{(k+j)^2}\right)\frac{1}{(k+1)^2}$$

$$\geq \prod_{j=2}^{k}\left(1 - \frac{1}{(k+j)^2}\right)\sum_{i=k+1}^{2k}\frac{1}{i^2}.$$

Denote $I_k = \prod_{j=2}^{k}\left(1 - \frac{1}{(k+j)^2}\right)$. Then

$$I_k = \frac{(2k+1)(2k-1)}{(2k)^2}\frac{2k(2k-2)}{(2k-1)^2}\cdots\frac{(k+4)(k+2)}{(k+3)^2}\frac{(k+3)(k+1)}{(k+2)^2}$$

$$= \frac{(2k+1)(k+1)}{2k(k+2)} \to 1 \text{ as } k \to \infty.$$

Thus there exists a large number $K$ such that for any $k \geq K$, we have $I_k \geq \frac{1}{2}$. So, for any $n = 2k \geq 2K$, we have

$$P\left(\bar{X}_n \geq \frac{1}{2}\right) \geq I_k \sum_{i=k+1}^{2k}\frac{1}{i^2} \geq \frac{1}{2}\sum_{i=k+1}^{2k}\frac{1}{(2k)^2} = \frac{1}{8k}.$$

It follows that

$$\sum_{n=1}^{\infty} P\left(\bar{X}_n \geq \frac{1}{2}\right) \geq \sum_{k=K}^{\infty}\frac{1}{8k} = \infty.$$

Hence $\bar{X}_n \nrightarrow 0\ c.c.$

**Example (3.3.10)[153]:** Let $\{Y_n, n \geq 1\}$ be a sequence of independent random variables such that $P(Y_n = 16^{n-1}) = \frac{1}{n^2}, P(Y_n = 0) = 1 - \frac{1}{n^2}$. Denote $X_{2n-1} = Y_n, X_{2n} = -2Y_n, n \geq 1$. Then for any $n \geq 1$, we have

$$\frac{X_{2n-1}}{2^{2n-1}} + \frac{X_{2n}}{2^{2n}} = 0. \tag{29}$$

By the above definitions, for any $\varepsilon > 0$, we have

$$\sum_{n=1}^{\infty} P(|X_n - 0| \geq \varepsilon) \leq 2\sum_{n=1}^{\infty} \frac{1}{n^2} < \infty,$$

and thus $X_n \to 0\ c.c.$ By (29), we know that $\sum_{k=1}^{n} \frac{X_k}{2^k} = \frac{X_n}{2^n} I(n \text{ is odd })$. Hence $\sum_{k=1}^{n} \frac{X_k}{2^k} \to 0\ c.c.$

In the following, we will show that $\frac{1}{2^n}\sum_{k=1}^{n} X_k \to 0\ c.c.$ It's enough to show one of its subsequence

$$\frac{1}{2^{4n}}\sum_{k=1}^{4n} X_k \nrightarrow 0 \text{ c.c.} \tag{30}$$

For any odd integer $k$,

$$X_k + X_{k+1} = X_k - 2X_k = -X_k.$$

Thus, for any $n \geq 1$,

$$\frac{1}{2^{2n}}\sum_{k=1}^{2n} X_k = -\frac{1}{2^{2n}}\sum_{k=1}^{n} X_{2k-1}.$$

And so (30) can be expressed to be

$$\frac{1}{16^n}\sum_{k=1}^{2n} X_{2k-1} \nrightarrow 0 \text{ c.c.} \tag{31}$$

For $k = n+1, \cdots, 2n$, we have

$$P(X_{2k-1} = 16^{k-1}) = P(Y_k = 16^{k-1}) = \frac{1}{k^2}, P(X_{2k-1} = 0) = P(Y_k = 0) = 1 - \frac{1}{k^2}. \tag{32}$$

Define $n$ sets $A_1, \cdots, A_n$ as follows:

$$\begin{aligned} A_1 &:= \{X_{2(2n)-1} = 16^{2n-1}\}, \\ A_2 &:= \{X_{2(2n)-1} = 0, X_{2(2n-1)-1} = 16^{(2n-1)-1}\}, \\ &\ \ \cdots \\ A_n &:= \{X_{2(2n)-1} = 0, \cdots, X_{2(n+2)-1} = 0, X_{2(n+1)-1} = 16^n\}. \end{aligned}$$

Then $\cup_{k=1}^{n} A_k \subset \left\{\left|\frac{1}{16^n}\sum_{k=1}^{2n} X_{2k-1} - 0\right| \geq 1\right\}$, and thus

$$\begin{aligned} &P\left\{\left|\frac{1}{16^n}\sum_{k=1}^{2n} X_{2k-1} - 0\right| \geq 1\right\} \geq \sum_{k=1}^{n} P(A_k) \\ &= \frac{1}{(2n)^2} + \left(1 - \frac{1}{(2n)^2}\right)\frac{1}{(2n-1)^2} + \cdots + \prod_{j=2}^{n}\left(1 - \frac{1}{(n+j)^2}\right) \cdot \frac{1}{(n+1)^2} \end{aligned}$$

$$\geq \prod_{j=2}^{n} \left(1 - \frac{1}{(n+j)^2}\right) \cdot \sum_{k=n+1}^{2n} \frac{1}{k^2}. \tag{33}$$

By (33) and following the deduction in Example (3.3.9), we can obtain that

$$\sum_{n=1}^{\infty} P\left\{\left|\frac{1}{16^n}\sum_{k=1}^{2n} X_{2k-1} - 0\right| \geq 1\right\} = \infty,$$

$i.e.$ (31) holds.

**Proposition (3.3.11)[153]:** Let $\{X_1, X_2, \cdots\}$ be pairwise uncorrelated random variables satisfying

$$\sum_{n=1}^{\infty} \frac{\mathrm{Var}(X_n)}{n^{\alpha}} < \infty, \tag{34}$$

where $\alpha > 0$, then for any $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} n^{1-\alpha} P\left\{\left|\frac{S_n - E(S_n)}{n}\right| \geq \varepsilon\right\} < \infty. \tag{35}$$

If $\{X_1, X_2, \cdots\}$ is a sequence of independent random variables satisfying (34), then for any $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} n^{1-\alpha} P\left\{\max_{1\leq k\leq n} |S_k - E(S_k)| \geq n\varepsilon\right\} < \infty. \tag{36}$$

**Proof.** For any $\varepsilon > 0$, by Chebyshev's inequality and (34), we get

$$\sum_{n=1}^{\infty} n^{1-\alpha} P\left\{\left|\frac{S_n - E(S_n)}{n}\right| \geq \varepsilon\right\} \leq \sum_{n=1}^{\infty} n^{1-\alpha} \frac{\mathrm{Var}(S_n)}{(n\varepsilon)^2}$$

$$= \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{1}{n^{1+\alpha}} \sum_{k=1}^{n} \mathrm{Var}(X_k)$$

$$= \frac{1}{\varepsilon^2} \sum_{k=1}^{\infty} \mathrm{Var}(X_k) \sum_{n=k}^{\infty} \frac{1}{n^{1+\alpha}}$$

$$\leq \frac{M}{\varepsilon^2} \sum_{k=1}^{\infty} \frac{\mathrm{Var}(X_k)}{k^{\alpha}} < \infty, \tag{37}$$

where $M$ is a positive constant. Hence (35) holds. By Kolmogorov's inequality and the deduction of (37), we get (36).

**Corollary (3.3.12)[153]:** Let $\{X_1, X_2, \cdots\}$ be pairwise uncorrelated random variables satisfying $\sum_{n=1}^{\infty} \frac{\mathrm{Var}(X_n)}{n} < \infty$, and $E(X_n) \to 0$. Then $\frac{S_n}{n} \to 0$ $c.c.$

**Proof.** In this case, $\frac{E(S_n)}{n} = \frac{\sum_{k-1}^{n} E(X_k)}{n} \to 0$. Then the result follows from Proposition (3.3.11).

We will construct four counterexamples to show that s $- L^1$ convergence versions and $s^* - L^2$ convergence versions of the Toeplitz lemma, the Cesàro mean convergence theorem and the Kronecker lemma can fail in general.

The next example shows that s- $L^1$ convergence versions of the Cesàro mean convergence theorem and the Toeplitz lemma fail.

**Example (3.3.13)[153]:** Let $\{X_n, n \geq 1\}$ be a sequence of random variables such that $P(X_n = n) = \frac{1}{n^3}, P(X_n = 0) = 1 - \frac{1}{n^3}$. Then we have $E[|X_n|] = \frac{1}{n^2}$ and thus

$$\sum_{n=1}^{\infty} E[|X_n|] = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty,$$

$$i.e. X_n \overset{s-L^1}{\longrightarrow} 0.$$

Let $\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k, n \geq 1$. Then

$$E[|\bar{X}_n|] = \frac{1}{n}\sum_{k=1}^{n} E[|X_k|] = \frac{1}{n}\sum_{k=1}^{n} \frac{1}{k^2}.$$

Since $\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$, there exists a large $N$ such that $\forall n \geq N, \sum_{k=1}^{n} \frac{1}{k^2} \geq \frac{\pi^2}{12}$. Hence

$$\sum_{n=1}^{\infty} E[|\bar{X}_n|] \geq \sum_{n=N}^{\infty} \frac{1}{n} \cdot \frac{\pi^2}{12} = \infty,$$

and so it doesn't hold that $\bar{X}_n \overset{s-L^1}{\longrightarrow} 0$.

The next example shows that $s - L^1$ convergence version of the Kronecker lemma fails. The basic idea comes from Linero and Rosalsky [23].

**Example (3.3.14)[153]:** Let $\{Y_n, n \geq 1\}$ be a sequence of independent random variables such that $P(Y_n = n) = \frac{1}{n^2(\ln n)^{1+\alpha}}, P(Y_n = 0) = 1 - \frac{1}{n^2(\ln n)^{1+\alpha}}, \alpha > 0$. Denote $X_{2n-1} = (2n-1)Y_n, X_{2n} = -2nY_n, n \geq$
1. Then for any $n \geq 1$, we have

$$\frac{X_{2n-1}}{2n-1} + \frac{X_{2n}}{2n} = 0. \tag{38}$$

By (38), we know that $\sum_{k=1}^{n} \frac{X_k}{k} = \frac{X_n}{n} I(n \text{ is odd })$. If $n = 2k-1$, then we have

$$E[|X_n/n|] = E[|Y_k|] = \frac{1}{k(\ln k)^{1+\alpha}},$$

which implies that

$$\sum_{n=1}^{\infty} E\left[\left|\sum_{k=1}^{n} \frac{X_k}{k}\right|\right] = \sum_{k=1}^{\infty} \frac{1}{k(\ln k)^{1+\alpha}} < \infty,$$

$$i.e. \sum_{k=1}^{n} \frac{X_k}{k} \overset{s-L^1}{\longrightarrow} 0.$$

In the following, we will show that $\frac{1}{n}\sum_{k=1}^{n} X_k \overset{s-L^1}{\rightarrow} 0$. It's enough to show one of its subsequence

$$\frac{1}{2n}\sum_{k=1}^{2n} X_k \overset{s-L^1}{\nrightarrow} 0. \tag{39}$$

For any integer $k$, we have $X_{2k-1} + X_{2k} = -Y_k$. Thus (39) can be expressed to be

$$\frac{1}{2n}\sum_{k=1}^{n} Y_k \overset{s-L^1}{\leftrightarrow} 0. \tag{40}$$

By the Fubini theorem, we have

$$\sum_{n=1}^{\infty} E\left[\left\|\frac{1}{2n}\sum_{k=1}^{n} Y_k\right\|\right] = \sum_{n=1}^{\infty} \frac{1}{2n}\sum_{k=1}^{n} E[Y_k] = \sum_{n=1}^{\infty} \frac{1}{2n}\sum_{k=1}^{n} \frac{1}{k(\ln k)^{1+\alpha}}$$
$$= \sum_{k=1}^{\infty} \frac{1}{k(\ln k)^{1+\alpha}}\sum_{n=k}^{\infty} \frac{1}{2n} = \infty.$$

Hence (40) holds.

The next example shows that $s^* - L^2$ convergence versions of the Cesàro mean convergence theorem and the Toeplitz lemma fail.

**Example (3.3.15)[153]:** Let $\{X_n, n \geq 1\}$ be a sequence of random variables such that $P\big(X_n = \sqrt{n}\big) = \frac{1}{n^5}, P(X_n = 0) = 1 - \frac{1}{n^5}$. Then we have $E[|X_n|^2] = \frac{1}{n^4}$ and thus

$$\sum_{n=1}^{n} \|X_n\|_2 = \sum_{n=1}^{\infty} \left(\frac{1}{n^4}\right)^{\frac{1}{2}} = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty,$$
$$i.e. X_n \overset{s^*-L^2}{\longrightarrow} 0.$$

Let $\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k, n \geq 1$. Then

$$E[|\bar{X}_n|^2] = \frac{1}{n^2}\left(\sum_{k=1}^{n} E[|X_k|^2] + 2\sum_{1\leq i<j\leq n} E\big[X_i X_j\big]\right)$$
$$\geq \frac{1}{n^2}\sum_{k=1}^{n} E[|X_k|^2] = \frac{1}{n^2}\sum_{k=1}^{n} \frac{1}{k^4}.$$

Denote $c = \sum_{k=1}^{\infty} \frac{1}{k^4}$. Then $c$ is a positive constant and there exists a large $N$ such that $\forall n \geq N, \sum_{k=1}^{n} \frac{1}{k^4} \geq \frac{c}{2}$. It follows that

$$\sum_{n=1}^{\infty} \|\bar{X}_n\|_2 \geq \sum_{n=N}^{\infty} \left(\frac{1}{n^2}\cdot\frac{c}{2}\right)^{\frac{1}{2}} = \sqrt{\frac{c}{2}}\sum_{n=N}^{\infty} \frac{1}{n} = \infty.$$

Hence it doesn't hold that $\bar{X}_n \overset{s^*-L^2}{\to} 0$.

Following Examples (3.3.14) and (3.3.15), we construct the following example, which shows that $s^* - L^2$ convergence version of the Kronecker lemma fails.

**Example (3.3.16)[153]:** Let $\{Y_n, n \geq 1\}$ be a sequence of independent random variables such that $P(Y_n = \sqrt{n}) = \frac{1}{n^5}, P(Y_n = 0) = 1 - \frac{1}{n^5}$. Denote $X_{2n-1} = (2n-1)Y_n, X_{2n} = -2nY_n, n \geq 1$. Then for any $n \geq 1$, we have

$$\frac{X_{2n-1}}{2n-1} + \frac{X_{2n}}{2n} = 0. \tag{41}$$

By (41), we know that $\sum_{k=1}^{n} \frac{X_k}{k} = \frac{X_n}{n} I(n \text{ is odd })$. If $n = 2k-1$, then we have

$$\|X_n/n\|_2 = \|Y_k\|_2 = \frac{1}{k^2}.$$

Hence

$$\sum_{n=1}^{\infty} \left\|\sum_{k=1}^{n} \frac{X_k}{k}\right\|_2 = \sum_{k=1}^{\infty} \|Y_k\|_2 = \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty,$$

$$i.e. \sum_{k=1}^{n} \frac{X_k}{k} \overset{s^*-L^2}{\longrightarrow} 0.$$

In the following, we will show that $\frac{1}{n}\sum_{k=1}^{n} X_k \overset{s^*-L^2}{\nrightarrow} 0$. It's enough to show one of its subsequence

$$\frac{1}{2n}\sum_{k=1}^{2n} X_k \overset{s^*-L^2}{\nrightarrow} 0. \tag{42}$$

For any integer $k$, we have $X_{2k-1} + X_{2k} = -Y_k$. Thus (42) can be expressed to be

$$\frac{1}{2n}\sum_{k=1}^{n} Y_k \overset{s^*-L^2}{\nrightarrow} 0. \tag{43}$$

Denote $c = \sum_{k=1}^{\infty} \frac{1}{k^4}$. Then $0 < c < \infty$, and there exists $N$ such that for any $n \geq N$, we have $\sum_{k=1}^{n} \frac{1}{k^d} \geq \frac{c}{2}$. Hence we have

$$\begin{aligned}
\sum_{n=1}^{\infty} \left\| \frac{1}{2n}\sum_{k=1}^{n} Y_k \right\|_2 &= \sum_{n=1}^{\infty} \frac{1}{2n}\left(\sum_{k=1}^{n} E[Y_k^2] + 2\sum_{1\leq i<j\leq n} E\left[Y_iY_j\right]\right)^{1/2} \\
&\geq \sum_{n=1}^{\infty} \frac{1}{2n}\left(\sum_{k=1}^{n} E[Y_k^2]\right)^{1/2} = \sum_{n=1}^{\infty} \frac{1}{2n}\left(\sum_{k=1}^{n} \frac{1}{k^4}\right)^{1/2} \\
&\geq \sum_{n=N}^{\infty} \frac{1}{2n}\left(\sum_{k=1}^{n} \frac{1}{k^4}\right)^{1/2} \\
&\geq \sum_{n=N}^{\infty} \frac{1}{2n}\sqrt{\frac{c}{2}} = \infty.
\end{aligned}$$

Hence (43) holds.

By Example (3.3.13), we know that, if $\sum_{n=1}^{\infty} E[|X_n|^p] < \infty$, then we don't have $\sum_{n=1}^{\infty} E[|S_n/n|^p] < \infty$ necessarily. In general, we have the following result.

**Proposition (3.3.17)[153]:** Suppose that $1 \leq p < \infty$ and $\sum_{n=1}^{\infty} E[|X_n|^p] < \infty$, then $\forall \varepsilon > 0$, we have

$$\sum_{n=1}^{\infty} \frac{1}{(\ln n)^{1+\varepsilon}} E[|S_n/n|^p] < \infty.$$

**Proof.** By the convexity of the function $f(x) = |x|^p$, we have

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{(\ln n)^{1+\varepsilon}} E[|S_n/n|^p] &\leq \sum_{n=1}^{\infty} \frac{1}{n(\ln n)^{1+\varepsilon}}\left(\sum_{k=1}^{n} E[|X_k|^p]\right) \\
&\leq \left(\sum_{k=1}^{\infty} E[|X_k|^p]\right)\sum_{n=1}^{\infty} \frac{1}{n(\ln n)^{1+\varepsilon}} < \infty.
\end{aligned}$$

**Proposition (3.3.18)[153]:** Let $\{X_1, X_2, \cdots\}$ be pairwise uncorrelated random variables satisfying $\sum_{n=1}^{\infty} \mathrm{Var}(X_n) < \infty$, then for any $1 < q \leq 2$, we have

$$\sum_{n=1}^{\infty} E\left[\left|\frac{S_n - E(S_n)}{n}\right|^q\right] < \infty,$$

in particular, $\frac{S_n - E(S_n)}{n} \to 0\ c.c.$

**Proof.** By the assumptions, we have

$$\sum_{n=1}^{\infty} E\left[\left|\frac{S_n - E(S_n)}{n}\right|^q\right] = \sum_{n=1}^{\infty}\left(\left\|\frac{S_n - E(S_n)}{n}\right\|_q\right)^q \leq \sum_{n=1}^{\infty}\left(\left\|\frac{S_n - E(S_n)}{n}\right\|_2\right)^q$$
$$= \sum_{n=1}^{\infty}\frac{1}{n^q}\left(\sum_{i=1}^{n}\mathrm{Var}(X_i)\right)^{q/2} \leq \sum_{n=1}^{\infty}\frac{1}{n^q}\left(\sum_{i=1}^{\infty}\mathrm{Var}(X_i)\right)^{\frac{q}{2}} < \infty.$$

By Example (3.3.15), we know that, if $\sum_{n=1}^{\infty}\|X_n\|_p < \infty$, then we don't have $\sum_{n=}^{\infty}\left\|\frac{S_n}{n}\right\|_p < \infty$ necessarily. In general, we have the following two propositions.

**Proposition (3.3.19)[153]:** Suppose that $1 \leq p < \infty$ and $\sum_{n=1}^{\infty}\|X_n\|_p < \infty$, then $\forall \varepsilon > 0$, we have

$$\sum_{n=1}^{\infty}\frac{1}{(\ln n)^{1+e}}\|S_n/n\|_p < \infty. \tag{44}$$

**Proof.** By Minkowski's inequality and the definition of the norm $\|\cdot\|_p$, we have that

$$\|S_n/n\|_p \leq \frac{1}{n}\left(\sum_{k=1}^{n}\|X_k\|_p\right).$$

Then we can prove (44) by following the proof of Proposition (3.3.17).

**Proposition (3.3.20)[153]:** Suppose that $1 < p < \infty$ and $\sum_{n=1}^{\infty}\|X_n\|_p < \infty$, then for any $1 < q \leq p$, we have

$$\sum_{n=1}^{\infty}E[|S_n/n|^q] < \infty,$$

in particular, $S_n/n \to 0\ c.c.$

**Proof.** By the fact that $\|\cdot\|_q \leq \|\cdot\|_p$, Minkowski's inequality and the assumption, we have

$$\sum_{n=1}^{\infty}E[|S_n/n|^q] = \sum_{n=1}^{\infty}\left(\|S_n/n\|_q\right)^q \leq \sum_{n=1}^{\infty}\left(\|S_n/n\|_p\right)^q$$
$$\leq \sum_{n=1}^{\infty}\left(\frac{\sum_{k=1}^{n}\|X_k\|_p}{n}\right)^q$$
$$= \sum_{n=1}^{\infty}\frac{1}{n^q}\left(\sum_{k=1}^{n}\|X_k\|_p\right)^q$$
$$\leq \left(\sum_{k=1}^{\infty}\|X_k\|_p\right)^q\sum_{n=1}^{\infty}\frac{1}{n^q} < \infty.$$

**Proposition (3.3.21)[153]:** Suppose that $\sum_{n=1}^{\infty}\|X_n\|_{\infty} < \infty$. Then

(i) for any $\varepsilon > 0$, we have

$$\sum_{n=1}^{\infty}\frac{1}{(\ln n)^{1+e}}\|S_n/n\|_{\infty} < \infty; \tag{45}$$

(ii) for any $1 < q < \infty$, we have

$$\sum_{n=1}^{\infty} E[|S_n/n|^q] < \infty,$$

in particular, $S_n/n \to 0$ $c.c.$

**Proof.** (i) By the definition of the norm $\|\cdot\|_\infty$, we have that

$$\|S_n/n\|_\infty \leq \frac{1}{n}\left(\sum_{k=1}^{n} \|X_k\|_\infty\right).$$

Then we can prove $(45)$ by following the proof of Proposition (3.3.17).

(ii) It's a direct consequence of Proposition (3.3.20) by noting that for any $1 < p < \infty$ and any random variable $X$, $\| X \|_p \leq \| X \|_\infty$.

# Chapter 4
# Monotone Geometric Mean and Kähler Means

We show that the cost of our approach in term of arithmetic operations for m matrices is of the order $O(mn^2)$. This definition preserves the structure, is simple to calculate, preserves monotonicity and satisfies some other Ando–Li–Mathias properties. A generalization of the mean towards PD (Toeplitz-block) block-Toeplitz matrices is discussed. For PD Toeplitz-block block-Toeplitz matrices, we derive the generalized barycenter, or generalized Kähler mean, and a greedy approximation. This approximation is shown to be close to the generalized mean with a significantly lower computational cost. The proposed definition preserves the structure of the matrices and satisfies some important Ando– Li–Mathias properties, such as monotonicity and continuity. Also, it has low cost and is simple to calculate.

**Section (4.1): A Class of Toeplitz Matrices**

The notion of geometric mean for positive definite matrices naturally appears in several areas, for instance in radar detection [40], [48], image processing [195] and elasticity tensor analysis [10]. $A$ definition of geometric mean of three or more positive definite matrices has been defined by M. Moakher [17] and $R$. Bhatia, $J$. Holbrook [3], [5]. This is usually identified with Karcher mean. $T$. Ando, $C.Li$ and $R$. Mathias in [2] have introduced a definition of geometric mean and have shown some of its properties, called the Ando-$Li$-Mathias $(ALM)$ axioms. These properties should be required for any reasonable notion of geometric mean of the matrices. There is a rich of the geometric means of matrices and methods for computing them, see [190], [191], [36] and the references in [191], [36]. But there are also many unsolved problems in this field yet. The Karcher mean does not preserve structures, for example the Karcher mean of two Toeplitz matrices is not necessarily a Toeplitz matrix. $D.A$. Bini $et\ al.$ in [191] have introduced a definition of geometric mean for structured matrices. This definition satisfies many of the $ALM$ properties except monotonicity. Moreover, this method can not guarantee uniqueness of the structured geometric mean.

We consider only the Toeplitz matrices. $A$ Toeplitz matrix is a matrix in which entries along their diagonals are constant. These matrices have many applications in a wide variety of problems in engineering. For positive definite Toeplitz matrices, there is the interesting notion of mean based on Kähler metric [22], [23] which is not a geometric mean, but satisfies some desirable properties such as permutation invariance and repetition invariance. This mean, called Kähler metric mean, does not coincide with Karcher mean but with this manner the mean of two positive definite Toeplitz matrices will be a Toeplitz matrix again. Unfortunately, the Kähler metric mean is not monotonic.

We introduce a new definition of geometric mean for a class of Toeplitz matrices. This approach is an operator theoretical approach as follows. For every $n \times n$ Toeplitz matrix

$$\mathrm{A} = \begin{bmatrix} a_0 & a_1 & \cdots & & a_{n-1} \\ a_{-1} & a_0 & a_1 & \cdots & \\ \vdots & & \ddots & & \vdots \\ & & \ddots & & a_1 \\ a_{-n+1} & & \cdots & a_{-1} & a_0 \end{bmatrix},$$

we consider the function $a: \mathbb{T} \longrightarrow \mathbb{C}$, where $\mathbb{T} = \{t \in \mathbb{C}: |t| = 1\}$ is the unit circle in the plane by definition $a(t) = \sum_{k=-n+1}^{n-1} a_k t^k$ for all $t \in \mathbb{T}$. Now, let $M(a) \in \mathcal{L}(L^\infty(\mathbb{T}))$ be the multiplication operator associated to the function $a$, *i.e.*, $M(a)f = af$ for all $f \in L^\infty(\mathbb{T})$.

In fact $M(a)$ is a Laurent operator with the so-called 'symbol' function $a$, see [194] and (1) below. We denote the cone of all positive semi-definite $n \times n$ Toeplitz matrices with non-negative symbols by $\mathcal{T}_n^{++}$, see (4) and the Lemma (4.1.6). We introduce a new definition of geometric mean on $\mathcal{T}_n^{++}$which satisfies among other properties, the monotonicity property in the ordering induced by the cone $\mathcal{T}_n^{++}$, see (5) and Theorem (4.1.8). Comparing to the other approaches, our proposed definition admits some important advantages: low cost in terms of arithmetic operations, simple calculations, structure preserving and monotonicity. Moreover, we do not use the non-singularity of matrices, *e.g.*, the zero matrix belongs to $\mathcal{T}_n^{++}$.

We state some basic definitions and theorems and then we define the geometric mean for Laurent operators. We introduce the main idea and explain some properties. We show numerical experiments and some comparisons between our proposed geometric mean, the structured geometric mean [191] and the Kähler metric mean [191], [22], [23].

The Laurent operators form a commutative $*$-algebra of bounded operators act on a Hilbert space as follows. First, We to recall some standard notions and notations. Let $\mathbb{T} = \{e^{i\theta} \mid \theta \in \mathbb{R}\}$ be the unit circle in the complex number plane. For each complex-valued Lebesgue measurable map $f: \mathbb{T} \to \mathbb{C}$ and each $1 \le p < +\infty$, let $\| f \|_p = \left(\frac{1}{2\pi}\int_0^{2\pi} |f(e^{i\theta})|^p d\theta\right)^{\frac{1}{p}}$ and $\| f \|_\infty = \inf\{M > 0 | |f(e^{i\theta})| < M \text{ for } a.e.\, \theta \in \mathbb{R}\}$.

As usual, the spaces $L^p := L^p(\mathbb{T}) = \{f \,|\, \| f \|_p < +\infty\}$ with norm $\|\cdot\|_p$ are Banach spaces for all $1 \le p \le +\infty$. Since T has finite Lebesgue measure we have $L^r \subset L^s$ when $1 \le s \le r \le +\infty$. In the case $p = 2, L^2$ equipped whit the inner product

$$(f, g) = \frac{1}{2\pi}\int_0^{2\pi} f(e^{i\theta})\overline{g(e^{i\theta})}d\theta$$

for $f, g \in L^2$ is a Hilbert space.

Given $f \in L^1$, we define its Fourier coefficients $\{f_n\}$ by $f_n = \frac{1}{2\pi}\int_0^{2\pi} f(e^{i\theta})e^{-in\theta}d\theta$ for all $n \in \mathbb{Z}$. Notice that $f \in L^2$ if and only if $\sum_{n\in\mathbb{Z}} |f_n|^2 < \infty$. If we define the functions $\chi_n$ by $\chi_n(t) = t^n$ for $n \in \mathbb{Z}$ and $t \in \mathbb{T}$, then $\{\chi_n\}_{n\in\mathbb{Z}}$ is an orthogonal basis for $L^2$, see [192]. For each $a \in L^\infty$ and each $1 < p < +\infty$ the operator $M(a): L^p \to L^p$ defined by $f \longmapsto af$, is bounded and $\| M(a) \|_{\mathcal{L}(L^p)} = \| a \|_\infty$ where $\mathcal{L}(L^p)$ is the Banach algebra of all bounded operators on $L^p$. The operator $M(a)$ is called the multiplication operator on $L^p$ generated by the function $a \in L^\infty$ and $a$ is called the symbol function of $M(a)$.

From the definitions, it is obvious that $(M(a)\chi_j, \chi_k) = a_{k-j}$ is the $(k-j)-th$ Fourier coefficients of $a$, see [192]. The following proposition shows that the converse is also true.

**Proposition (4.1.1)[189]:** ([192]) Let $A \in \mathcal{L}(L^p)(1 < p < \infty)$ and suppose there is a sequence $\{a_n\}_{n\in\mathbb{Z}}$ of complex numbers such that $(A\chi_j, \chi_k) = a_{k-j}$. Then there is an $a \in L^\infty$ such that $A = M(a)$ and $\{a_n\}$ is the Fourier coefficient sequence of $a$.

Moreover,

$$\| M(a) \|_{\mathcal{L}(L^p)} = \| a \|_\infty.$$

The proof of the following lemma is straightforward. The definition of Laurent operators are introduced in this lemma.

**Lemma (4.1.2)[189]:** The multiplication operator $M: L^\infty \to \mathcal{L}(L^2)$ is a linear positive isometry from $L^\infty$ to $\mathcal{M} = \{M(a) \mid a \in L^\infty\}$ the space of Laurent operators then the following properties hold:

i) $M(ab) = M(a)M(b)$ for all $a, b \in L^\infty$, hence the space of Laurent operators $\mathcal{M}$ is $a$ commutative algebra.

ii) $M(\bar{a}) = M(a)^*$ for all $a \in L^\infty$.

iii) The operator $M(a)$ is a positive semi-definite operator if and only if $a \in L^\infty$ and $a \geq 0\ a.e.$

iv) The operator $M(a)$ is invertible in $\mathcal{L}(L^2)$ if and only if $a \in L^\infty$ and a is away from zero, $i.e., |a| > \epsilon\ a.e.$ for some $\epsilon > 0$.

v) $M(|a|) = |M(a)|$ for all $a \in L^\infty$.

vi) $M(\sqrt[m]{a}) = \sqrt[m]{M(a)}$ when $m \in \mathbb{N}, a \in L^\infty$ and $a \geq 0\ a.e.$

If we express the multiplication operator $M(a)$ generated by $a \in L^\infty$ via Fourier coefficients, we have the following infinite dimensional matrix representation

$$M(a) = \begin{bmatrix} \ddots & & & & & & \\ \cdots & a_0 & a_{-1} & \cdots & & & \\ \cdots & a_1 & a_0 & a_{-1} & a_{-2} & a_{-3} & \cdots \\ & & \ddots & \ddots & \ddots & & \\ & \cdots & a_2 & a_1 & a_0 & a_{-1} & \cdots \\ \vdots & & & & \ddots & \ddots & \ddots \end{bmatrix}.$$

We denote the class of all positive semi-definite multiplication operators by $\mathcal{M}^+, i.e.$,

$$\mathcal{M}^+ = \{M(a) \mid a \in L^\infty \text{ and } a \geq 0 \text{ a.e. }\}.$$

The following definition naturally motivated by Lemma (4.1.2).

**Definition (4.1.3)[189]:** Let $n \in \mathbb{N}$ and $M(a_1), M(a_2), \cdots, M(a_n) \in \mathcal{M}^+$.The geometric mean of these operators defined by

$$G\big(M(a_1), \cdots, M(a_n)\big) := M\left((a_1 \cdots a_n)\frac{1}{n}\right).$$

We will show that this definition satisfies many of the Ando-$Li$-Mathias ($ALM$) axioms in [2] for operators in $\mathcal{M}^+$. In the following properties, we assume that all operators are in $\mathcal{M}^+$and all scalars are non-negative reals.

■ If $a_1 = a_2 = \cdots = a_n =: a$ then by definition

$$G\big(M(a_1), \cdots, M(a_n)\big) = M(a).$$

■ Joint homogeneity. Since multiplication operators are linear, we have

$$\begin{aligned} G\big(\alpha_1 M(a_1), \cdots, \alpha_n M(a_n)\big) &= G\big(M(\alpha_1 a_1), \cdots, M(\alpha_n a_n)\big) \\ &= M\left((\alpha_1 \ldots \alpha_n)^{\frac{1}{n}}(a_1 \cdots a_n)\frac{1}{n}\right) \\ &= (\alpha_1 \ldots \alpha_n)\frac{\frac{1}{n}}{M}\left((a_1 \cdots a_n)\frac{1}{n}\right) \\ &= (\alpha_1 \ldots \alpha_n)\frac{1}{n}G\big(M(a_1), \cdots, M(a_n)\big). \end{aligned}$$

■ Permutation invariance. For any permutation $\pi$ of $\{1,2, \ldots, n\}$, obviously we have

$$G(M(a_1),\cdots,M(a_n)) = G\left(M(a_{\pi(1)}),\cdots,M(a_{\pi(n)})\right).$$

■ Monotonicity. Since the operator $M$ is positive, see Lemma (4.1.2) $(iii)$, if $M(a_j) \geq M(a'_j)$ for $j = 1,\cdots,n$ then

$$G(M(a_1),\cdots,M(a_n)) \geq G(M(a'_1),\cdots,M(a'_n)).$$

■ Continuity from above. If the sequences $\{M(a_{k,n})\}_{n\in\mathbb{N}}$ converge to the operators $M(a_k)$ for each $k = 1,\cdots,m$, respectively, $i.e.$, $\|M(a_{k,n}) - M(a_k)\|_{\mathcal{L}(L^2)} \to 0$ as $n \to +\infty$, then

$$\left\|G\left(M(a_{1,n}),\cdots,M(a_{m,n})\right) - G(M(a_1),\cdots,M(a_m))\right\|_{\mathcal{L}(L^2)} \to 0,$$

as $n \to +\infty$.

■ Joint concavity. Again by linearity of multiplication operators, for each $0 < \lambda < 1$ we have

$$\begin{aligned} G\quad &(\lambda M(a_1) + (1-\lambda)M(b_1), \lambda M(a_2) + (1-\lambda)M(b_2),\cdots,\lambda M(a_n) + (1-\lambda)M(b_n)) \\ &\geq \lambda G(M(a_1),\cdots,M(a_n)) + (1-\lambda)G(M(b_1),\cdots,M(b_n)). \end{aligned}$$

■ Self-duality. By Lemma (4.1.2)(i) and $(iv)$, for invertible operators, we have

$$\begin{aligned} G(M(a_1),\cdots,M(a_n))^{-1} \quad &= M\left((a_1\cdots a_n)\frac{1}{n}\right)^{-1} \\ &= M\left((a_1^{-1}\cdots a_n^{-1})\frac{1}{n}\right) \\ &= G(M(a_1)^{-1},\cdots,M(a_n)^{-1}). \end{aligned}$$

■The arithmetic-geometric-harmonic mean inequality. For $n \geq 1$ and $a_1,\cdots,a_n \geq 0$, we have

$$(a_1\cdots a_n)^{\frac{1}{n}} \leq \frac{a_1 + \cdots + a_n}{n}.$$

So by Definition (4.1.3) and linearity of multiplication operators, we have

$$G(M(a_1),\cdots,M(a_n)) \leq \frac{M(a_1) + \cdots + M(a_n)}{n}.$$

Combine this by self-duality property and the fact that for each two positive invertible operators $T, S$ we have $T \leq S$ if and only if $S^{-1} \leq T^{-1}$, we will have

$$\left(\frac{M(a_1)^{-1} + \cdots + M(a_n)^{-1}}{n}\right)^{-1} \leq G(M(a_1),\cdots,M(a_n)) \leq \frac{M(a_1) + \cdots + M(a_n)}{n},$$

where $a_1,\cdots,a_n$ are assumed to be away from zero.

We introduce a concept of geometric mean for a cone of positive semi-definite Toeplitz matrices. Let

$$A = \begin{bmatrix} a_0 & a_1 & \cdots & a_{n-1} \\ \bar{a}_1 & a_0 & \cdots & \vdots \\ \vdots & & \ddots & \\ \bar{a}_{n-1} & & \cdots & a_0 \end{bmatrix},$$

be a Hermitian Toeplitz matrix with complex entries. The symbol of $A$ is a function $a \in L^\infty$ defined by

$$a(e^{i\theta}) = a_0 + \sum_{k=1}^{n-1}\left(a_k e^{ik\theta} + \bar{a}_k e^{-ik\theta}\right),\ (\theta \in \mathbb{R}). \tag{1}$$

In fact, by letting $a_{-k} := \bar{a}_k$ for $k = 1,2,\cdots,n-1$, and $a_k = 0$ for $|k| \geq n$ we can write

$$a(t)=\sum_{k=-\infty}^{+\infty} a_k\chi_k(t),\ (t\in\mathbb{T}),$$

where $\{\chi_k\}_{k\in\mathbb{Z}}$ by definition $\chi_k(t)=t^k, (t\in\mathbb{T})$, is the standard basis of $L^2$. Moreover the function $a$ is the symbol of the following Toeplitz operator [192], [194],

$$T_A=\begin{bmatrix} a_0 & a_1 & a_2 & \cdots \\ a_{-1} & a_0 & a_1 & \cdots \\ a_{-2} & \cdots & & \\ \vdots & & \ddots & \end{bmatrix}.$$

Even if the Toeplitz matrix $A$ is positive semi-definite the associated Toeplitz operator $T_A$ may not be (positive) semi-definite operator. For example the matrix $A=\begin{bmatrix}3 & 2\\ 2 & 3\end{bmatrix}$ is a $2\times 2$ positive definite Toeplitz matrix but its associated Toeplitz operator

$$T_A=\begin{bmatrix} 3 & 2 & 0 & 0 & 0 & \cdots \\ 2 & 3 & 2 & 0 & 0 & \cdots \\ 0 & 2 & 3 & 2 & 0 & \cdots \\ \vdots & & \ddots & & & \vdots \end{bmatrix}$$

is not a positive semi-definite operator, to see this let $\xi=\sum_{k=1}^{+\infty}(-1)^k\alpha^{k-1}\chi_k$ for any $\alpha\in\left(\frac{3}{4},1\right)$, then we have

$$\begin{aligned}\langle T_A\xi,\xi\rangle &= 3\sum_{k=1}^{+\infty}\alpha^{2k-2}-4\sum_{k=1}^{+\infty}\alpha^{2k-1}\\ &=\frac{3-4\alpha}{1-\alpha^2}<0.\end{aligned}$$

In fact, the symbol function of this matrix $A$ is $a\left(e^{i\theta}\right)=3+4\cos\theta, (\theta\in\mathbb{R})$, which is not a positive function. In order to use the theory of positive semi-definite Laurent operators to define the geometric mean which satisfies the monotonicity property, we consider only those positive semi-definite Toeplitz matrices whose symbols are non-negative functions.

We are going to specify a class of positive semi-definite Toeplitz matrices with nonnegative symbols. For each $n\in\mathbb{N}$, let $V_n$ be the n-dimensional vector subspace of $L^2$ generated by the set $\{\chi_0,\chi_1,\cdots,\chi_{n-1}\}$. The inclusion map $\iota_n: V_n\to L^2$ and the orthogonal projection map $\pi_n: L^2\to V_n$ are well-defined linear operators that $\|\iota_n\|=\|\pi_n\|=1$ (see [193], Theorem 2.7), note that $V_n$ is a closed subspace of the Hilbert space $L^2$. Now for each $a\in L^\infty$ the associated multiplication operator $M(a): L^2\to L^2$ is a Laurent operator, so one can define the linear transformation $\pi_n\circ M(a)\circ \iota_n: V_n\to V_n$. The matrix representation of $\pi_n\circ M(a)\circ L_n$ in the ordered basis $\{\chi_0,\chi_1,\cdots,\chi_{n-1}\}$ is denoted by $T_n(a)$. By a simple calculation, we have

$$T_n(a)=\begin{bmatrix} a_0 & a_1 & \cdots & a_{n-1} \\ a_{-1} & a_0 & & \vdots \\ \vdots & & \ddots & a_1 \\ a_{-n+1} & \cdots & a_{-1} & a_0 \end{bmatrix}$$

where

$$a_k=<a,\chi_k>=\frac{1}{2\pi}\int_0^{2\pi} a\left(e^{i\theta}\right)e^{-ik\theta}d\theta,\ (k\in\mathbb{Z}),$$

are the Fourier coefficients of $a$.

**Lemma (4.1.4)[189]:** If $a \in L^\infty, a \geq 0$ $a.e.$ and $n \in \mathbb{N}$ then the matrix $T_n(a)$ is a positive semi-definite $n \times n$ Toeplitz matrix but the converse is not true.

**Proof.** Using the notation $\chi_k(t) = t^k$ where $k \in \mathbb{Z}$ and $t = e^{i\theta} \in \mathbb{T}, \theta \in \mathbb{R}$, we can write

$$f = \sum_{k=0}^{n-1} f_k t^k \in V_n$$

and

$$a = \sum_{k=-\infty}^{+\infty} a_k t^k \in L^\infty \qquad (2)$$

as two polynomials on variable $t \in \mathbb{T}$ where $f_k$ $'s$ and $a_k$ $'s$ are complex numbers. Now, we have

$$af = \sum_{k=-\infty}^{+\infty} \left( \sum_{j=0}^{n-1} f_j a_{k-j} \right) t^k$$

and

$$\begin{aligned} \pi_n \circ M(a) \circ \iota_n(f) &= \pi_n(af) \\ &= \sum_{k=0}^{n-1} \left( \sum_{j=0}^{n-1} f_j a_{k-j} \right) t^k. \end{aligned}$$

Recall that $\bar{t} = t^{-1}$ when $t \in \mathbb{T}$, so $\bar{f} = \sum_{m=0}^{n-1} \bar{f}_m t^{-m}$. Notice that if $p(t) = \sum_{k=-\infty}^{+\infty} \alpha_k t^k, (t \in \mathbb{T})$, and $p \in L^1$ then

$$\begin{aligned} \frac{1}{2\pi} \int_0^{2\pi} p(e^{i\theta}) d\theta &= \alpha_0 \\ &= \text{ the constant term of } p. \end{aligned}$$

Therefore

$$\begin{aligned} < \pi_n \circ M(a) \circ \iota(f), f > &= \frac{1}{2\pi} \int_0^{2\pi} \pi_n(af)(e^{i\theta}) \overline{f(e^{\iota\theta})} d\theta \\ &= \text{ the constant term of } \pi_n(af)\bar{f} \qquad (3) \\ &= \sum_{k=0}^{n-1} \sum_{j=0}^{n-1} \bar{f}_k f_j a_{k-j}. \end{aligned}$$

In the other hand

$$|f(t)|^2 = f(t)\overline{f(t)} = \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} \bar{f}_k f_j t^{j-k},$$

using (2), we have

$$\begin{aligned} \frac{1}{2\pi} \int_0^{2\pi} a(e^{i\theta}) \left| f(e^{i\theta}) \right|^2 d\theta &= \text{ the constant term of } a|f|^2 \\ &= \sum_{k=0}^{n-1} \sum_{j=0}^{n-1} \bar{f}_k f_j a_{k-j}, \end{aligned}$$

which equals (3), so

$$< \pi_n \circ M(a) \circ \iota(f), f >= \frac{1}{2\pi} \int_0^{2\pi} a(e^{i\theta}) \left| f(e^{i\theta}) \right|^2 d\theta \geq 0$$

whenever $a \geq 0$ $a.e.$ Hence, we just prove that the linear transformation $\pi_n \circ M(a) \circ \iota(f)$ and its matrix, $T_n(a)$, are positive semi-definite.

The converse does not valid, $e.g.$, the Toeplitz matrix $A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$, which is mentioned before, is positive definite and $A = T_n(a)$ where $a(e^{i\theta}) = 3 + 4\cos\theta$, but $a \ngeq 0$. Now, we introduce the following sets

$\mathcal{T}_n =$ The set of all $n \times n$ Toeplitz matrices,

$$\mathcal{T}_n^+ = \{A \in \mathcal{T}_n \mid A \text{ is positive semi-definite } \}, \quad (4)$$

$\mathcal{T}_n^{++} = \{A \in \mathcal{T}_n \mid \text{the symbol of } A \text{ is a non-negative function } \}$.

Notice that by the Lemma (4.1.4), we have $\mathcal{T}_n^{++} \subsetneqq \mathcal{T}_n^+$.

Now, we summarize the above discussion in the following corollary ($cf.$ Theorem 9.3 in [194]).

**Corollary (4.1.5)[189]:** For any natural number $n \in \mathbb{N}$, the map $T_n: L^\infty \to \mathcal{T}_n$ is a surjective linear operator and

$$\mathcal{T}_n^{++} = \left\{T_n(a) \mid a \in L^\infty, a \geq 0 \text{ a.e. } \right\}.$$

The following lemma shows that there are many matrices in $\mathcal{T}_n^{++}$.

**Lemma (4.1.6)[189]:** Let $n \in \mathbb{N}, a_0 \in \mathbb{R}$ and $a_1, \cdots, a_{n-1} \in \mathbb{C}$ be such that

$$a_0 \geq 2 \sum_{k=0}^{n-1} |a_k|,$$

then

$$A = \begin{bmatrix} a_0 & a_1 & \cdots & a_{n-1} \\ \bar{a}_1 & a_0 & & \\ \vdots & & \ddots & \vdots \\ \bar{a}_{n-1} & \cdots & & a_0 \end{bmatrix} \in \mathcal{T}_n^{++}.$$

**Proof.** We have $A = T_n(a)$, where

$$a(e^{i\theta}) = a_0 + \sum_{k=1}^{n-1} \left(a_k e^{ik\theta} + \bar{a}_k e^{-ik\theta}\right),\ \theta \in \mathbb{R}.$$

So we have to prove that $a \geq 0, a.e.$ by the Corollary (4.1.5). But we have

$$\begin{aligned} a(e^{i\theta}) \quad &= a_0 + 2 \sum_{k=1}^{n-1} \operatorname{Re}\left(a_k e^{ik\theta}\right) \\ &\geq a_0 - 2 \sum_{k=1}^{n-1} \left|a_k e^{ik\theta}\right| \\ &= a_0 - 2 \sum_{k=1}^{n-1} |a_k| \geq 0, \end{aligned}$$

by our assumption.

The converse of the Lemma (4.1.6) is not valid. For example, let

$$A = \begin{bmatrix} 19 & 6 & 9 \\ 6 & 19 & 6 \\ 9 & 6 & 19 \end{bmatrix}$$

then $A$ is a positive definite Toeplitz matrix with symbol $a(e^{i\theta}) = 19 + 12\cos(\theta) + 18\cos(2\theta) \geq 0$, but $19 \ngeq 2(6+9)$.

Notice that the set $\mathcal{T}_n^{++}$is a cone of positive semi-definite Toeplitz matrices, $i.e.$, $A + B \in \mathcal{T}_n^{++}$when $A, B \in \mathcal{T}_n^{++}$and $\alpha A \in \mathcal{T}_n^{++}$when $A \in \mathcal{T}_n^{++}, \alpha \geq 0$. Hence, the cone $\mathcal{T}_n^{++}$naturally induces an order, we denote it by $\prec$, as follows: for any $A, B \in \mathcal{T}_n^{++}$we define

$$A \prec B \Leftrightarrow B - A \in \mathcal{T}_n^{++}. \tag{5}$$

Note that since $\mathcal{T}_n^{++} \subset \mathcal{T}_n^{+}$, $A \prec B$ implies $A \leq B$. But the converse is not true, $e.g.$, let $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}$, then $A, B \in \mathcal{T}_n^{++}$but $B - A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix} \in \mathcal{T}_n^{+} \setminus \mathcal{T}_n^{++}$, $i.e.$, $A \leq B$ but $A \nprec B$. Now we can introduce the main definition.

**Definition (4.1.7)[189]:** Given $m, n \geq 1$ and $A_1, A_2, \cdots, A_m \in \mathcal{T}_n^{++}$and let $a_1, \cdots, a_m \in L^\infty$ be their symbols, respectively. The geometric mean of $A_1, \cdots, A_m$ is defined by the Toeplitz matrix

$$G(A_1, \cdots, A_m) = T_n\left((a_1 \cdots a_m)^{\frac{1}{m}}\right).$$

In the following theorem we collect some properties which are satisfied by this definition.

**Theorem (4.1.8)[189]:** Let $m \geq 1, A_1, \cdots, A_m, B_1, \cdots B_m \in \mathcal{T}_n^{++}$and $\alpha_1, \cdots, \alpha_m \in \mathbb{R}^+$.The following properties hold:

i) If $A_1 = A_2 = \cdots = A_m =: A$ then $G(A_1, \cdots, A_m) = A$.

ii) Joint homogeneity. $G(\alpha_1 A_1, \cdots, \alpha_m A_m) = (\alpha_1 \cdots \alpha_m)\frac{1}{m} G(A_1, \cdots, A_m)$.

iii) Permutation invariance. For any permutation $\pi$ of $\{1, \cdots, m\}$ we have

$$G(A_1, \cdots, A_m) = G\left(A_{\pi(1)}, \cdots, A_{\pi(m)}\right).$$

iv) Monotonicity. $G(A_1, \cdots, A_m) \prec G(A_1 + B_1, \cdots, A_m + B_m)$.

v) Continuity. If $\{A_{k,1}\}_{k\geq 1}, \cdots, \{A_{k,m}\}_{k\geq 1}$ are monotonic decreasing sequences in ordering (5), converging to $A_1, \cdots, A_m$ respectively then the sequence $\{G(A_{k,1},, \cdots, A_{k,m})\}_{k\geq 1}$ converges to $G(A_1, \cdots, A_m)$.

vi) Joint concavity. For each $0 < \lambda < 1$

$$\begin{aligned} G \quad &(\lambda A_1 + (1-\lambda)B_1, \cdots, \lambda A_m + (1-\lambda)B_m) \\ &\prec \lambda G(A_1, \cdots, A_m) + (1-\lambda)G(B_1, \cdots, B_m). \end{aligned}$$

vii) Arithmetic-geometric inequality. $G(A_1, \cdots, A_m) \prec \frac{A_1 + \cdots + A_m}{m}$.

**Proof.** Properties $(i)$ and $(iii)$ trivially hold. We state proof of the others. ii) This follows from the Definition (4.1.7) and linearity of $T_n$, Corollary (4.1.5). iv) Let $a_1, \cdots, a_m$ and $b_1, \cdots, b_m$ be the symbols of $A_1, \cdots, A_m$ and $B_1, \cdots, B_m$, respectively. By assumption $a_1, \cdots, a_m, b_1, \cdots, b_m \geq 0$, since

$$\left((a_1 + b_1) \cdots (a_m + b_m)\right)^{\frac{1}{m}} - (a_1 \cdots a_m)^{\frac{1}{m}} \geq 0,$$

by the Corollary (4.1.5) and (5) the conclusion follows.

v) We have

$$A_{k,1} = T_n(a_{k,1}), \cdots, A_{k,m} = T_n(a_{k,m}),$$

and

$$A_1 = T_n(a_1), \cdots, A_m = T_n(a_m).$$

Since, all norms in finite dimensional vector spaces are equivalent, for $0 \leq i \leq m, A_{k,i} \to A_i$ as $k \to \infty, i.e., \|A_{k,i} - A_i\|_F \to 0$ as $k \to \infty$, where $\|.\|_F$ denotes the Frobenius norm. In the other hand $\|a_{k,i} - a_i\|_2 = \|A_{k,i} - A_i\|_F$, $0 \leq i \leq m$, and

$$\left\|\left(a_{k,1}\cdots a_{k,m}\right)^{\frac{1}{m}}-(a_1\cdots a_m)^{\frac{1}{m}}\right\|_2 \to 0 \text{ as } k\to\infty.$$

Again, by $\| a \|_2 = \|T_n(a)\|_F$, for each $a \in L^\infty$, we have

$$\|G(A_{k,1},\cdots,A_{k,m})-G(A_1,\cdots,A_m)\|_F \longrightarrow 0 \text{ as } k\longrightarrow\infty.$$

vi) Since $A_k = T_n(a_k), B_k = T_n(b_k)$ for $0 \le k \le m$, and

$$\left((\lambda a_1+(1-\lambda)b_1)\cdots(\lambda a_m+(1-\lambda)b_m)\right)^{\frac{1}{m}}$$
$$\le \lambda(a_1\cdots a_m)^{\frac{1}{m}}+(1-\lambda)(b_1\cdots b_m)^{\frac{1}{m}},$$

by concavity property of geometric mean of non-negative reals, the statement is deduced from linearity and positiveness of the operator $T_n$, see Corollary (4.1.5).

vii) Because of

$$(a_1\cdots a_m)\frac{1}{m}\le\frac{a_1+\cdots a_m}{m},$$

and $A_k = T_n(a_k), 0 \le k \le m$, the conclusion follows.

As we mentioned before, some Ando-$Li$-Mathias axioms are not satisfied yet. Now, we illustrate this by the following counterexamples. Let

$$A=\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix},\ B=\begin{bmatrix}2 & -1\\ -1 & 2\end{bmatrix}. \qquad (6)$$

■ Consistency with scalars. We have $AB = BA$ and

$$(AB)^{\frac{1}{2}}=\begin{bmatrix}\sqrt{3} & 0\\ 0 & \sqrt{3}\end{bmatrix},$$

but

$$G(A,B)=\begin{bmatrix}\frac{4}{\pi} & 0\\ 0 & \frac{4}{\pi}\end{bmatrix}.$$

■ Congruence invariance. Let $S=\begin{bmatrix}4 & -1\\ -1 & 4\end{bmatrix}$ and $C=\begin{bmatrix}3 & 1\\ 1 & 3\end{bmatrix}$ we have

$$S^*AS=\begin{bmatrix}26 & 1\\ 1 & 26\end{bmatrix},\ S^*CS=\begin{bmatrix}43 & -7\\ -7 & 43\end{bmatrix},$$

and

$$G(S^*AS,S^*CS)\simeq\begin{bmatrix}33.0912 & -2.1114\\ -2.1114 & 33.0912\end{bmatrix},$$

but

$$S^*G(A,C)S\simeq\begin{bmatrix}32.3134 & -1.1941\\ -1.1941 & 32.3134\end{bmatrix}.$$

■ Self-duality. We have

$$G(A,B)^{-1}=\begin{bmatrix}\frac{\pi}{4} & 0\\ 0 & \frac{\pi}{4}\end{bmatrix},$$

but

$$G(A^{-1},B^{-1})=\begin{bmatrix}\frac{4}{3\pi} & 0\\ 0 & \frac{4}{3\pi}\end{bmatrix}.$$

■ Determinant identity. $\det(G(A,B)) = \frac{16}{\pi^2}$ but $(\det A \cdot \det B)^{\frac{1}{2}} = 3$.
■ The geometric-harmonic mean inequality.

$$G(A,B) - \left(\frac{A^{-1}+B^{-1}}{2}\right)^{-1} = \begin{bmatrix} \frac{4}{\pi} - \frac{3}{2} & 0 \\ 0 & \frac{4}{\pi} - \frac{3}{2} \end{bmatrix}.$$

which is not positive semi-definite, i.e., $\left(\frac{A^{-1}+B^{-1}}{2}\right)^{-1} \nleq G(A,B)$.

Now, we want to glance at the comparison between our definition of the geometric mean and two other well-known concepts of geometric means: the Karcher mean and the Kähler metric mean. First let us to recall their definitions briefly. For a set of $m$ positive definite $n \times n$ matrices $A_1, \cdots, A_m$, the Karcher mean is defined as the minimizer

$$\arg\min_{X \in \mathbb{S}_+^n} \sum_{k=1}^{m} \delta^2(A_k, X)$$

(see [36],p.8 ), where $\mathbb{S}_+^n$ represents the set of symmetric positive definite matrices and $\delta(A,B) = \|\log(A^{-1/2}BA^{-1/2})\|_F$ with $\|.\|_F$ the Frobenius norm is the intrinsic distance on the manifold $\mathbb{S}_+^n$. The definition of the Kähler metric mean for Toeplitz matrices is somewhat more complicated. Let $A_1, \cdots, A_m$ be a set of $m$ positive definite $n \times n$ Toeplitz matrices. Their Kähler mean is defined as minimizer

$$\arg\min_{X} \sum_{k=1}^{m} d^2(A_k, X)$$

where $X$ varies on the manifold of all the positive definite $n \times n$ Toeplitz matrices. This manifold is a Cartan-Hadamard manifold with intrinsic metric $d$, see e.g., [191].

**Example (4.1.9)[189]:** Consider the following matrices of Example 5.1 in [191],

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, B = \begin{bmatrix} 4 & -1 \\ -1 & 4 \end{bmatrix} \text{ and } C = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

As mentioned in [191], we have $A \leq B$ but $\mathcal{B}(A,C) \nleq \mathcal{B}(B,C)$ where $\mathcal{B}$ stands for the Kähler barycenter mean. Also it has been shown that the Kähler barycenter mean $\mathcal{B}(A,C) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ does not coincide with the Karcher mean $(AC)^{\frac{1}{2}} = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & \sqrt{3} \end{bmatrix}$.

Now, applying our definition on the matrices $A, B$ and $C$ we have

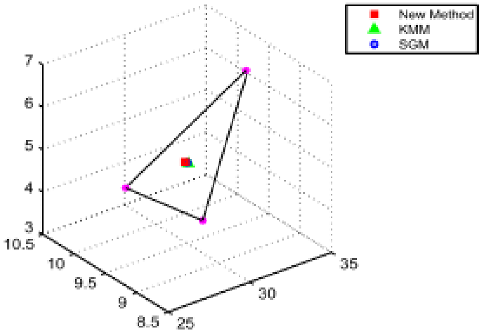
$$G(A,C) = \begin{bmatrix} \frac{4}{\pi} & 0 \\ 0 & \frac{4}{\pi} \end{bmatrix}, \; G(B,C) \simeq \begin{bmatrix} 2.725 & -1.1314 \\ -1.1314 & 2.725 \end{bmatrix}.$$

Since the eigenvalues of the matrix $G(B,C) - G(A,C)$ are 2.580 and 0.317, so

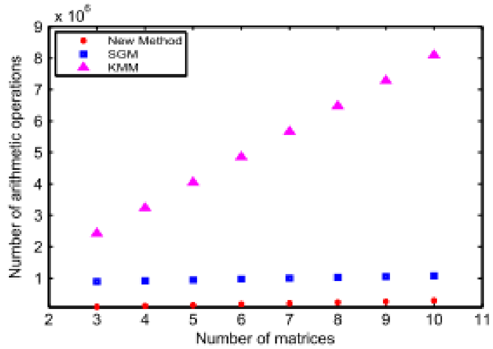
$$G(A,C) \leq G(B,C),$$

therefore, our definition is monotone in this example. Moreover in this method $G(A,C) \neq (AC)\frac{1}{2}$ thus our approach does not coincide with the Karcher mean and the Kähler barycenter mean.

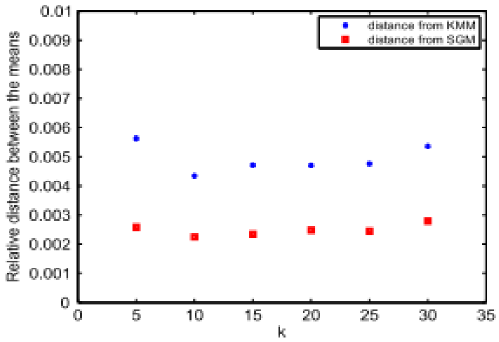
There are many well-known methods to compute these integrals like that trapezoidal rule, Simpson's rule, $etc$. We will use Simpson's rule for numerical integration.




(a) Representation of the three means for three randomly matrices




(b) Required arithmetic operations for different amounts of matrices




(c) Distances between different means, using Frobenius norm

**Fig. 1.** Comparison of the different methods.

For illustrative purposes only, we construct randomly three matrices $3 \times 3$ belong to $\mathcal{T}_3^{++}$such that $|a_0| > 2(|a_1| + |a_2|)$. $Fig. 1(a)$ shows these matrices and their means with our method, structured geometric mean $(SGM)$, and the Kähler metric mean $(KMM)$ in $a$ 3D diagram, note that each $3 \times 3$ Toeplitz matrix characterized by its first row, so corresponds to a point of $\mathbb{R}^3$. As we see, our proposed geometric mean is close other ones. For our method the number of operations equals $O(\mathrm{kmn}^2)$ where $n$ is the size of the matrices, $m$ is the number of matrices and $k$ is the number of subdivision of the interval $[0,2\pi]$ in order to use Simpson's rule for numerical integration, . The cost of structured geometric mean in the case of Toeplitz matrices with approach 4.4(1) in [191] is equal to $O(pn^4 + pmn^3)$, where $m, n$ are same as before and $p$ is the number of iterations of this algorithm. The Kähler metric mean in the case real Toeplitz matrices as described in the Section 5 of [191], requires $O(\mathrm{mn}^4)$ arithmetic operations where $m, n$ are same as before, again. In order to compare the cost of different methods we let $n = 30, m = 3$ to 10, for $SGM$ $p = 1$ and for our method $k = 32$. The results are given in $Fig. 1(b)$. As this figure shows our purposed method has less cost than others. Unfortunately in the generic case there is no reference solution. In order to compare our method to the $SGM$ and the $KMM$ methods, for each $q = 5,10,15,20,25,30$ we construct randomly $q$ triples $(A_1, B_1, C_1), \cdots, (A_q, B_q, C_q)$ of $10 \times 10$ Toeplitz matrices

$$\begin{bmatrix} a_0 & a_1 & \cdots & a_9 \\ a_1 & a_0 & \cdots & \vdots \\ \vdots & & \ddots & \\ a_9 & & \cdots & a_0 \end{bmatrix}$$

such that $|a_0| > 2\sum_{j=1}^{9} |a_k|$, assuring to belong to $\mathcal{T}_{10}^{++}$. Then we compute the average value of these geometric means, $i.e. \frac{1}{q}\sum_{j=1}^{q} G(A_j, B_j, C_j)$ by our method, the $SGM$ method and the

$KMM$ method and denote them by $G, G_{SGM}$ and $G_{KMM}$ respectively. Finally, we show the relative distances $\|G_{SGM} - G\|_F / \| G \|_F$ and $\|G_{KMM} - G\|_F / \| G \|_F$ by using the Frobenius norm in $Fig. 1(c)$. As one can see, the results of these three approaches are not far from each other, nevertheless our method is closer to $SGM$ method.

We have introduced a new definition of geometric mean for the set of all positive semidefinite Toeplitz matrices with nonnegative symbol functions. This definition satisfies in some $ALM$ axioms. It preserves monotonicity and structure of Toeplitz matrices.

Moreover, computing this mean is simple and requires only $O(\mathrm{mn}^2)$ arithmetic operations (for $m$ matrices of size $n$ ). Finally, we compared our approach with the structured geometric mean for Toeplitz matrices [191] and the Kähler metric mean [23], [191] in term of number of arithmetic operations and the relative distances. It is still unclear whether our approach can handle the set of all positive semi-definite Toeplitz matrices such that it preserves monotonicity and the structure of Toeplitz matrices as well. Another interesting question is to characterize the set of $n \times n$ positive semi-definite Toeplitz matrices with non-negative symbol $\mathcal{T}_n^{++}$. This can be a topic for future investigation.

**Section (4.2): Block-Toeplitz Matrices with Toeplitz Structured Blocks**

In radar theory and other signal processing applications [197], [200], [201], [202], [240], autocorrelation matrices are very popular for representing windows of discrete or continuous signals.

For a signal $x(k)$, the element at position $(t_1, t_2)$ in such an autocorrelation matrix is obtained from an averaging operation $E[x(k + t_1)x(k + t_2)^*] = E[x(k + t)x(k)^*]$, with $t = t_1 - t_2$ referred to as the lag. Note that $E[x(k - t)x(k)^*] = E[x(k)x(k + t)^*] = (E[x(k + t)x(k)^*])^*$. Theoretically, this averaging operation is taken over the entire signal, resulting in an infinite sum (for a discrete signal) or integral (for a continuous signal). In practice, the sum/integral is taken over the finite window of interest, where as many entries in the sum/integral as possible are taken considering the lag and size of the window.

For a finite window (or a stationary signal in general), the resulting autocorrelation matrix will be a positive definite $(PD)$ Toeplitz matrix [234]. A popular detection technique in radar theory consists in comparing a certain window in a signal with an average of the signal in the neighboring windows. Translated to the autocorrelation matrices, this means that a $PD$ Toeplitz matrix is compared with an average of its neighboring $PD$ Toeplitz matrices.

One approach to the averaging of $PD$ Toeplitz matrices was proposed by Bini $et\ al.$ [20], and is referred to as the structured geometric mean. The mean emphasizes the natural geometry of $PD$ matrices in a restricted search for the center of mass or barycenter $w.r.t.$ this natural geometry. An alternative could be to focus on the natural geometry of the Toeplitz matrices. But, as a vector space, the set of Toeplitz matrices is naturally endowed with Euclidean geometry, having the arithmetic mean as its corresponding barycenter.

On the other hand, from the applications mentioned above, a transformation of the autocorrelation matrices and a geodesic distance measure based on information geometry theory can be found [202], [203]. This distance measure is derived from a natural geometry in the transformed space and the corresponding averaging operation shows appealing results in applications. We analyze the associated barycenter and discuss how it is derived from the signal processing application.

When the basic signal $x(k)$ is replaced with a multichannel signal $X(k)$, the corresponding autocorrelation matrix can be constructed as a block matrix. Specifically, we obtain a $PD$ block-Toeplitz ($BT$) matrix, which is a $PD$ block matrix with identical blocks along the block diagonals. In some applications, the blocks themselves will also have the Toeplitz structure, resulting in autocorrelation matrices which are $PD$ Toeplitz-block $BT$ ($TBBT$). We derive first order optimization techniques for the computation of these generalized Kähler means and analyze their properties.

We organized in the following way. The transformation of $PD$ Toeplitz matrices and its underlying interpretation are discussed. Afterwards, the natural geometry of the resulting transformed space is presented, and the corresponding barycenter is referred to as the Kähler mean. Two possible generalizations for the transformation of $PD$ Toeplitz matrices towards $PD$ $BT$ matrices are investigated. Moreover, we also discuss two different distance measures for the second generalized transformation. The generalized Kähler means for $PD$ $BT$ matrices and $PDTBBT$ matrices are presented, respectively. Finally, we compare the resulting algorithms in numerical experiments.

The set of $PD$ matrices, denoted by $\mathcal{P}_n$, is defined as the set

$$\mathcal{P}_n = \{A \in \mathbb{C}^{n\times n} \mid x^H A x > 0, \forall x \in \mathbb{C}^n/\{0\}\}.$$

This characterization of $PD$ matrices is equivalent to the condition that $A$ is Hermitian and has positive eigenvalues [3], and is also denoted as $A > 0$. $\mathcal{P}_n$ is naturally endowed with the following distance measure and inner product,

$$d(A,B) = \left\|\log\left(A^{-\frac{1}{2}}BA^{-\frac{1}{2}}\right)\right\|_F, \quad (7)$$

$$\langle E,F\rangle_A = \operatorname{trace}(A^{-1}EA^{-1}F), \quad (8)$$

where $A, B \in \mathcal{P}_n, E, F \in \mathcal{H}_n$, the set of $n\times n$ Hermitian matrices, and $\|\cdot\|_F$ denotes the Frobenius norm.

The vector space of Toeplitz matrices consists of all matrices having identical elements along the diagonals,

$$\mathcal{T}_n = \left\{\begin{bmatrix} t_0 & t_1 & \cdots & t_{n-1} \\ t_{-1} & t_0 & \ddots & t_{n-2} \\ \vdots & \ddots & \ddots & \vdots \\ t_{-n+1} & t_{-n+2} & \cdots & t_0 \end{bmatrix} \mid t_{-n+1},\dots,t_{n-1} \in \mathbb{C}\right\}. \quad (9)$$

The intersection of this set of Toeplitz matrices with the Hermitian matrices $\mathcal{H}_n$ is given by the elements in (9) for which $t_{-i} = t_i^*, i = 0,\dots,n-1$. The set of $PD$ Toeplitz matrices will be denoted as $\mathcal{T}_n^+ := \mathcal{T}_n \cap \mathcal{P}_n$.

We denote by $\mathcal{B}_{n,N}$ the vector space of $BT$ matrices, where the indices $n$ and $N$ indicate that the matrices consist of $n$ by $n$ blocks and each block is an $N\times N$ matrix. As for the Toeplitz matrices, the set containing all $PD$ elements in $\mathcal{B}_{n,N}$ will be denoted by $\mathcal{B}_{n,N}^+$. The subspace of $\mathcal{B}_{n,N}$ where the matrix blocks themselves are also Toeplitz matrices is the vector space of $TBBT$ matrices, which we denote by $\mathcal{T}_{n,N}$. The intersection with the manifold of $PD$ matrices is denoted by $\mathcal{T}_{n,N}^+$.

Several instances of (un)structured matrices can be combined in a least squares approach, and the result is, in general, referred to as the barycenter. For a number of elements $A_1,\dots,A_k$ in a set $\mathcal{S}$ with given distance measure $d_\mathcal{S}$, the barycenter is defined as the minimizer of the sum of squared distances to these given elements:

$$\mathcal{B}_\mathcal{S}(A_1,\dots,A_k) = \arg\min_{X\in\mathcal{S}} \frac{1}{2}\sum_{i=1}^{k} d_\mathcal{S}^2(X,A_i). \quad (10)$$

This concept is known to be a natural method for combining elements, $e.g.$, the barycenter corresponding to the classical Euclidean geometry is the arithmetic mean.
Furthermore, by considering the set $\mathcal{P}_n$ of $PD$ matrices with its natural distance measure (7), this barycenter is identical to the Karcher mean, the main instance of the geometric mean of $PD$ matrices [3], [20], [26], [216], [36], [37], [225], [227], [46]. The structured geometric mean, proposed by Bini $et\ al.$ [20], is obtained by minimizing the cost function of the Karcher mean, where the search space is restricted to the $PD$ matrices of a specified matrix structure.
Another averaging operation, known as the median barycenter or median for short, is defined as the minimizer of the sum of the distances to the given elements (instead of the squared distances). While this median is more robust to outliers, we will focus on barycenters because they behave more smoothly and are better at combining the inherent information of all given elements. In matrix information geometry, the median approach has been applied to $PD$ Toeplitz matrices [197], [198] and $PD\ BT$ matrices [202].
The computation of these barycenters is often performed using Riemannian optimization, a generalized version of optimization which takes the geometry of the search space into account [21]. When the search space is lacking a differentiable structure and associated geometry, other approaches can be used, such as random and deterministic walks in the general setting of Hadamard spaces [207], [225], [231].
Throughout, expressions will be presented containing a multitude of variables. We aim to clearly indicate the difference between main and auxiliary variables by using the following notation. We denote a function $f$, defined as $f(X) = g(A,B,C)$, with auxiliary variables $A = g_1(X), B = g_2(X)$, and $C = g_3(X)$, as

$$f(X) = g(A,B,C),$$
$$\begin{bmatrix} A = g_1(X), \\ B = g_2(X), \\ C = g_3(X), \end{bmatrix}$$

indicating that $f$ is the main variable of interest.
In what follows, the matrix $I_n$ will represent the $n \times n$ identity matrix, and $J_n$ the so-called counteridentity, the $n \times n$ matrix with ones on the antidiagonal and zeros everywhere else. For both matrices, the index might be omitted if the size is clear from the context. The transpose of a matrix $A$ will be denoted by $A^T$, its conjugate transpose by $A^H$, and its elementwise conjugate by $A^*$. Finally, we write $\bar{A}$ to represent the form $JA^*J$. Note that this operation corresponds to taking the conjugate transpose of $A$ and reflecting the result over the antidiagonal.

The set of Toeplitz matrices $\mathcal{T}_n$ is a linear space of matrices and is therefore traditionally associated with Euclidean] geometry. However, we are interested in the intersection of $\mathcal{T}_n$ with the set of positive matrices $\mathcal{P}_n$. Applying the geometry of the latter to the intersected set results in the structured geometric mean which has been discussed by Bini $et\ al.$ [20]. Here, we will discuss a different geometry on $\mathcal{T}_n^+$, along with its underlying interpretation and its properties.

The interpretation of the Kähler mean heavily depends on the linear autoregressive model from signal processing theory:

$$x(k) + \sum_{j=1}^{n} a_j^n x(k-j) = w(k),$$

where $x$ is the signal of interest and $w$ represents its prediction error. Our interest now goes to the so-called prediction coefficients $a_j^n$, and the intermediate factors that arise in their computation.
By applying autocorrelation to the stationary signal $x(k)$, its autocorrelation coefficients $r_t$, defined as $E[x(k+t)x(k)^*]$, can be obtained for different lags $t$. If this autocorrelation is performed on the above autoregressive model, the following system is found:

$$\begin{aligned} R_n\tilde{a}_n &= -\tilde{r}_n, \\ \tilde{a}_n &= [a_1^n, \dots, a_n^n]^T, \\ \tilde{r}_n &= [r_1, \dots, r_n]^T, \end{aligned} \tag{11}$$

where $R_n$ is the $PD$ Toeplitz matrix of size $n$ with elements $[R]_{i,j} = [R]_{j,i}^* = r_{i-j}, i,j = 0, \pm 1, \dots, \pm(n-1)$. Note that the prediction error $w(k)$ is assumed to be uncorrelated to the signal $x(k)$. $A$ recursive method known as the Levinson algorithm [222], [238] can be used to find the solution to system (11) by solving the system for $n = 1$, and sequentially obtaining the prediction coefficients $\tilde{a}_n$ for increasing $n$. The Levinson recurrence relation for the prediction coefficients is given by

$$\begin{aligned} \tilde{a}_1 &= a_1^1 = -\frac{r_1}{r_0}, \\ a_\ell^\ell &= -\frac{r_\ell + \sum_{j=1}^{\ell-1} r_{\ell-j} a_j^{\ell-1}}{r_0 + \sum_{j=1}^{\ell-1} r_j a_j^{\ell-1*}}, \\ \tilde{a}_\ell &= \begin{bmatrix} a_1^\ell \\ \vdots \\ a_{\ell-1}^\ell \\ a_\ell^\ell \end{bmatrix} = \begin{bmatrix} a_1^{\ell-1} \\ \vdots \\ a_{\ell-1}^{\ell-1} \\ 0 \end{bmatrix} + a_\ell^\ell \begin{bmatrix} a_{\ell-1}^{\ell-1*} \\ \vdots \\ a_1^{\ell-1*} \\ 1 \end{bmatrix}, \end{aligned} \tag{12}$$

with $\ell = 2, \dots, n$. It can be shown that the factors $a_\ell^\ell$ all lie within the complex unit disk $\mathbb{D}$, $|a_\ell^\ell| < 1, \forall \ell = 1, \dots, n$.
Our main interest in the above is the one-to-one relation between the $PD$ Toeplitz matrix $R_n$ and the scalars $(r_0, a_1^1, \dots, a_{n-1}^{n-1})$. Note that indices of the prediction coefficients only reach $n-1$, since the computation of $a_n^n$ requires the autocorrelation coefficient $r_n$, which is only given as an element of the right-hand side of (11), but not of $R_n$.
The transformation of the matrix $R_n$ is the following:

$$\begin{aligned} \mathcal{T}_n^+ &\to \mathbb{R}^{++} \times \mathbb{D}^{n-1}, \\ R_n &\mapsto (p_0, \mu_1, \dots, \mu_{n-1}), \end{aligned} \tag{13}$$

where we use the notation $p_0 := r_0, \mu_\ell := a_\ell^\ell$, and $R^{++}$represents the set of strictly positive numbers. This transformation creates a one-to-one mapping between the $PD$ Toeplitz matrices and the parameter space $\mathbb{R}^{++} \times \mathbb{D}^{n-1}$. Note that increasing the size of $R_n$ by 1 (increasing $n$ by 1 ) only requires the computation of 1 additional parameter $\mu_n := a_n^n$, while all other parameters remain fixed. This corresponds to the recursive construction of the Levinson algorithm.
We note that other parametrizations of the matrices are possible [199], [209]. However, these algorithms have mostly been designed for more robust estimation of the correlation coefficients $r_t = E[x(k+t)x(k)^*]$ from a finite number of measurements. Since our goal is the averaging of the correlation matrices, we assume that the estimation of the correlation coefficients has been performed prior to the application of the mean.

Moreover, changing the parametrization will not affect the following theory, since the same parameter space $\mathbb{R}^{++} \times \mathbb{D}^{n-1}$ is obtained for the different algorithms.

In order to define the Kähler metric, the set of $PD$ Toeplitz matrices is considered to be a Kähler manifold [22],[23],[205]. Such a manifold is associated with the concept of a Kähler potential, of which the Hessian form defines the inner product, and hence the geometry, imposed on the manifold. In the field of signal processing (and specifically in the context of Koszul information geometry [205]), the Kähler potential is chosen to be the process entropy $\Phi(R_n)$ [203], [205], defined as follows:

$$\Phi(R_n) = \log(\det R_n^{-1}) - \log(\pi \mathrm{e}), \tag{14}$$

where $\pi$ and e are the well-known mathematical constants. Applying some decomposition rules on the determinant of $R_n$ and by recognizing the components of the transformation (13) of $R_n$, the process entropy $\Phi(R_n)$ can be rewritten as a function of the parameter space $\mathbb{R}^{++} \times \mathbb{D}^{n-1}$ :

$$\Phi(R_n) = -n \log(p_0) - \sum_{\ell=1}^{n-1} (n-\ell) \log(1 - |\mu_\ell|^2) - \log(\pi e),$$

where $R_n$ is identified with its transformation $(p_0, \mu_1, \dots, \mu_{n-1})$. This decomposition of the determinant of $R_n$ is discussed.

The Kähler metric can now be obtained by determining the Hessian of the Kähler potential where complex differentiation should be used for the components $\mu_\ell \in \mathbb{D}$. If we denote $\xi^{(n)} = [p_0, \mu_1, \dots, \mu_{n-1}]^T$, then

$$[H]_{i,j} = \frac{\partial^2 \Phi}{\partial \xi_i^{(n)} \, \bar{\partial} \xi_j^{(n)}}.$$

The desired metric can be found as

$$\begin{aligned} \mathrm{d}s^2 &= \mathrm{d}\xi^{(n)^H} H \, \mathrm{d}\xi^{(n)} \\ &= n \frac{\mathrm{d}p_0^2}{p_0^2} + \sum_{\ell=1}^{n-1} (n-\ell) \frac{|\mathrm{d}\mu_\ell|^2}{(1-|\mu_\ell|^2)^2}. \end{aligned} \tag{15}$$

By examining this differential metric, a natural geometry and distance measure can be found for (each of the components of) the parameter space $\mathbb{R}^{++} \times \mathbb{D}^{n-1}$. The geometry on $\mathbb{R}^{++}$is that of the positive numbers, which is given by the scalar analog of (7) and (8) (up to a scaling with factor $\sqrt{n}$ and $n$, respectively). For the complex unit disk $\mathbb{D}$, the hyperbolic metric of the Poincaré disk can be recognized (up to a scaling of a factor $(n-\ell)/4$). We summarize as follows:

$$\begin{aligned} \forall a, b \in \mathbb{R}^{++}, \forall e, f \in \mathbb{R}: \quad & \langle e, f \rangle_a = n \frac{ef}{a^2}, \\ & d_{\mathbb{R}^{++}}(a,b) = \sqrt{n} \left| \log \frac{b}{a} \right|; \\ & \forall \mu, \nu \in \mathbb{D}, \forall \varepsilon, \varsigma \in \mathbb{C}: \langle \varepsilon, \varsigma \rangle_\mu = \frac{n-\ell}{2} \frac{\varepsilon \varsigma^* + \varsigma \varepsilon^*}{(1-|\mu|^2)^2}, \\ & d_{\mathbb{D}}(\mu, \nu) = \frac{\sqrt{n-\ell}}{2} \log \left( \frac{1 + \left| \frac{\mu - \nu}{1 - \mu \nu^*} \right|}{1 - \left| \frac{\mu - \nu}{1 - \mu \nu^*} \right|} \right), \end{aligned} \tag{16}$$

where $\ell$ is chosen corresponding to the coordinate ($\mu_\ell, \ell = 1, \ldots, n-1$, from (13)) to which it relates.
Combined, we define the Kähler distance $d_{\mathcal{T}_n^+}$ between two $PD$ Toeplitz matrices $T_1$ and $T_2$ as

$$\begin{aligned} d_{\mathcal{T}_n^+}^2(T_1, T_2) \quad &= d_{\mathcal{T}_n^+}^2\left(\left(p_{0,1}, \mu_{1,1}, \ldots, \mu_{n-1,1}\right), \left(p_{0,2}, \mu_{1,2}, \ldots, \mu_{n-1,2}\right)\right) \\ &= n\log^2\left(\frac{p_{0,2}}{p_{0,1}}\right) + \sum_{\ell=1}^{n-1} \frac{n-\ell}{4} \log^2\left(\frac{1 + \left|\frac{\mu_{\ell,1} - \mu_{\ell,2}}{1 - \mu_{\ell,1}\mu_{\ell,2}^*}\right|}{1 - \left|\frac{\mu_{\ell,1} - \mu_{\ell,2}}{1 - \mu_{\ell,1}\mu_{\ell,2}^2}\right|}\right). \end{aligned} \tag{17}$$

By entering this distance measure into definition (10), the Kähler mean is obtained as the barycenter $\mathcal{B}_{\mathcal{T}_n^+}$. Endowing the manifold $\mathcal{T}_n^+$ with the Kähler metric (15) results in a complete, simply connected manifold with nonpositive sectional curvature everywhere, or a Cartan-Hadamard manifold. Hence, existence and uniqueness are guaranteed for the barycenter with respect to this metric [30], [38].

Regarding the properties of the barycenter $\mathcal{B}_{\mathcal{T}_n^+}$, it can easily be seen that it is permutation invariant, repetition invariant, and idempotent (these hold for any barycenter). Moreover, if we denote the transformation (13) of a matrix $T_i \in \mathcal{T}_n^+$ by $\left(p_{0,i}, \mu_{1,i}, \ldots, \mu_{n-1,i}\right)$, then, for any $\alpha_i > 0$, the transformation of $\alpha_i T_i$ is $\left(\alpha_i p_{0,i}, \mu_{1,i}, \ldots, \mu_{n-1,i}\right)$. Hence, from the explicit expression of the first coordinate $p_{0,\mathcal{B}} = \left(p_{0,1} \cdots p_{0,k}\right)^{1/k}$ of the barycenter $\mathcal{B}_{\mathcal{T}_n^+}(T_1, \ldots, T_k)$, we get

$$\mathcal{B}_{\mathcal{T}_n^+}(\alpha_1 T_1, \alpha_2 T_2, \ldots, \alpha_k T_k) = (\alpha_1 \cdots \alpha_k)^{1/k} \mathcal{B}_{\mathcal{T}_n^+}(T_1, \ldots, T_k),$$

that is, joint homogeneity holds.
Properties related to the partial ordering of $PD$ matrices do not hold in general, $e.g.$, monotonicity: suppose $\tilde{T}_1, T_1, T_2 \in \mathcal{T}_n^+$ with $\tilde{T}_1 \geq T_1$, then, in general, $\mathcal{B}_{\mathcal{T}_n^+}\left(\tilde{T}_1, T_2\right) \ngeq \mathcal{B}_{\mathcal{T}_n^+}(T_1, T_2)$.
When experimenting with the Kähler mean, results have shown that its averaging properties cooperate very well with the application from which it was derived [197], [23], [202], [240]. This makes sense since at every step of the derivation, the most natural geometries and concepts, related to this particular model, were chosen from information theory.
Furthermore, the mean also has a computational advantage through its separation of optimization. The separate coordinates of the matrices can be grouped and averaged independently:

$$\begin{array}{ccccccc} T_1 & \mapsto & ( & \boxed{p_{0,1}, \ \vdots \ p_{0,k},} & \boxed{\mu_{1,1}, \ \vdots \ \mu_{1,k},} & \ldots, & \boxed{\mu_{n-1,1} \ \vdots \ \mu_{n-1,k}} \ ) \\ \vdots & & & & & & \\ T_k & \mapsto & ( & & & & ) \\ & & & \downarrow & \downarrow & & \downarrow \\ T & \leftarrow & ( & p_0, & \mu_1, & \ldots, & \mu_{n-1} \ ) \end{array}$$

This results in two main advantages. First, each coordinate group can be averaged in parallel since they have no influence on any of the other coordinate groups, and second, the means we end up computing contain elements of much smaller sizes than the original data (from matrices of size $n$ to scalars), and additional computational time is saved. The computation itself is discussed by Bini $et\ al.$ [20].

Our real interest goes out to the linear autoregressive model for multichannel signals [226], given by

$$X(k) + \sum_{j=1}^{n} A_j^n X(k-j) = W(k),$$

with $X$ and $W$ vectors of signals and the factors $A_j^n$ square matrices. Taking the normal equations of the multichannel model, the so-called Yule-Walker equations are obtained:

$$\begin{aligned} \tilde{A}_n \tilde{R}_n &= -U_n \\ \tilde{A}_n &= [A_1^n, \dots, A_n^n], \\ U_n &= [R_1, \dots, R_n], \\ \tilde{R}_n &= \begin{bmatrix} R_0 & R_1 & \cdots & R_{n-1} \\ R_1^H & R_0 & \ddots & R_{n-2} \\ \vdots & \ddots & \ddots & \vdots \\ R_{n-1}^H & R_{n-2}^H & \cdots & R_0 \end{bmatrix}, \end{aligned} \tag{18}$$

where $\tilde{R}_n \in \mathcal{B}_{n,N}^{+}$is a $PD\ BT$ matrix of $n$ by $n$ blocks. The size of the blocks $(N)$ is equal to the length of the multichannel signal vectors $X$ and $W$.

Some interesting cases of the multichannel model (such as a two dimensional signal, when interpreted as a multichannel signal) result in a matrix $\tilde{R}_n$ which is not only $PD\ BT$, but also has the Toeplitz structure in the individual blocks [221], [222], [233].

Hence it will become a $PD\ TBBT$ matrix. In practice, these Toeplitz blocks will often be Hermitian themselves, $R_\ell = R_\ell^H, \ell = 0, \dots, n-1$, but we will develop our theory for the more general case in which only the entire matrix $\tilde{R}_n$ is Hermitian. The results remain valid in the more specified setting.

The transformation. With $R_n$ now defined as a $PD\ TBBT$ matrix, we would like to generalize the transformation (13) to $\mathcal{T}_{n,N}^{+}$. Similarly to the link between the recursion (12) and the transformation (13), this generalization is obtained using a recursive computation of the prediction matrices in $\tilde{A}_n$. This recursive computation goes as follows [222], [226], [237], [239]:

$$A_1^1 = -R_1 R_0^{-1}, \tag{19}$$

$$A_\ell^\ell = -\Delta_\ell P_{\ell-1}^{-1}, \tag{20}$$

$$\begin{cases} \Delta_\ell = R_\ell + \sum_{j=1}^{\ell-1} A_j^{\ell-1} R_{\ell-j}, \\ P_{\ell-1} = R_0 + \sum_{j=1}^{\ell-1} J A_j^{\ell-1*} J R_j = R_0 + \sum_{j=1}^{\ell-1} \overline{A_j^{\ell-1}} R_j, \end{cases} \tag{21}$$

$$\tilde{A}_\ell = \left[\tilde{A}_{\ell-1}, 0\right] + A_\ell^\ell \left[\overline{A_{\ell-1}^{\ell-1}}, \dots, \overline{A_1^{\ell-1}}, I\right] \tag{22}$$

with $\ell = 2, \dots, n$. Similarly to the prediction coefficients $a_\ell^\ell$ from before, the factors $A_\ell^\ell$ will be the matrices of interest for the generalized transformation. To properly define this transformation, the set in which these matrices lie is investigated.

First of all, note that if all blocks in $\tilde{R}_n$ (18) are assumed to be Toeplitz matrices, we have $\overline{R_\ell} = R_\ell^H, \ell = 0, \dots, n-1$, and even stronger, $\overline{R_0} = R_0$, since this block is also a $PD$ matrix and hence Hermitian.

Next, we mention the following formula, based on the notion of the Schur complement, for the inversion of block matrices,

$$\tilde{R}_{\ell+1}^{-1} = \begin{bmatrix} \alpha_\ell & -\alpha_\ell U_\ell \tilde{R}_\ell^{-1} \\ -\tilde{R}_\ell^{-1} U_\ell^H \alpha_\ell & \tilde{R}_\ell^{-1} + \tilde{R}_\ell^{-1} U_\ell^H \alpha_\ell U_\ell \tilde{R}_\ell^{-1} \end{bmatrix}$$

with $\alpha_\ell = \left(R_0 - U_\ell \tilde{R}_\ell^{-1} U_\ell^H\right)^{-1}$. Note that $\alpha_\ell$ is a principal submatrix of the PD matrix $\tilde{R}_{\ell+1}^{-1}$ and is therefore also $PD$.

Now, the auxiliary matrix $P_\ell$ in the recursive computation $(21)$ can be written as

$$\overline{P_\ell} = R_0 + \tilde{A}_\ell U_\ell^H = R_0 - U_\ell \tilde{R}_\ell^{-1} U_\ell^H = \alpha_\ell^{-1},$$

hence $\overline{P_\ell}$( and $P_\ell$) is also a $PD$ matrix. Using the recursion expression $(22)$, an updating rule can be found for $\overline{P_\ell}$ (and consequently for $\alpha_\ell^{-1}$ ),

$$\overline{P_\ell} = \overline{P_{\ell-1}} - \Delta_\ell P_{\ell-1}^{-1} \overline{\Delta_\ell} = \left(I - A_\ell^\ell \overline{A_\ell^\ell}\right) \overline{P_{\ell-1}}, \tag{23}$$

where $\overline{P_0} = R_0$.

Finally, we show that the matrices $A_\ell^\ell$ belong to the set

$$\mathcal{D}_N = \{\Gamma \in \mathbb{C}^{N\times N} \mid I - \Gamma\bar{\Gamma} > 0\}.$$

Note that for $N = 1$, this set reduces exactly to the complex unit disk $\mathbb{D} = \{\gamma \in \mathbb{C} \mid \gamma\bar{\gamma} = \gamma\gamma^* < 1\}$. To prove that all matrix factors $A_\ell^\ell$ belong to $\mathcal{D}_N$, we start from the positive definiteness of $\overline{P_\ell}$ :

$$\overline{P_\ell} = \overline{P_{\ell-1}} - \Delta_\ell P_{\ell-1}^{-1} \overline{\Delta_\ell} > 0,$$

$$\overset{\text{congruence}}{\longrightarrow} \quad I - \overline{P_{\ell-1}^{-\frac{1}{2}}} \Delta_\ell P_{\ell-1}^{-1} \overline{\Delta_\ell} \overline{P_{\ell-1}^{-\frac{1}{2}}} > 0,$$

$$\overset{\text{similarity}}{\longrightarrow} \quad I - \Delta_\ell P_{\ell-1}^{-1} \overline{\Delta_\ell} \overline{P_{\ell-1}^{-1}} = I - A_\ell^\ell \overline{A_\ell^\ell} > 0.$$

The resulting transformation will be a mapping between the $PD\ BT$ (not $TBBT$) matrices and the new parameter space, and it is defined as

$$\begin{aligned} \mathcal{B}_{n,N}^+ \quad &\to \mathcal{P}_N \times \mathcal{D}_N^{n-1}, \\ \tilde{R}_n \quad &\mapsto (P_0, \Gamma_1, \dots, \Gamma_{n-1}), \end{aligned} \tag{24}$$

where the notation $P_0 := R_0, \Gamma_\ell := A_\ell^\ell$ is used, and $N$ denotes the size of the matrix blocks. We do not restrict the transformation to elements in $\mathcal{T}_{n,N}^+$since the inverse transformation of a random point $(P_0, \Gamma_1, \dots, \Gamma_{n-1}) \in \mathcal{P}_N \times \mathcal{D}_N^{n-1}$ does not necessarily have the Toeplitz structure in the individual blocks.

The metric. To define the generalized metric, the Kähler potential is examined as in the scalar case. Note the following possible factorization of the determinant of $\tilde{R}_n$[229]:

$$\begin{aligned} \det(\tilde{R}_n) \quad &= \det(\tilde{R}_{n-1})\det(R_0 - U_{n-1}\tilde{R}_{n-1}^{-1}U_{n-1}^H) \\ &= \det(\tilde{R}_{n-1})\det(\alpha_{n-1}^{-1}) \\ &= \det(\tilde{R}_{n-1})\det(I - A_{n-1}^{n-1}\overline{A_{n-1}^{n-1}}) \dots \det(I - A_1^1\overline{A_1^1})\det(R_0) \\ &= \det(I - A_{n-1}^{n-1}\overline{A_{n-1}^{n-1}}) \dots \det(I - A_1^1\overline{A_1^1})^{n-1} \det(R_0)^n, \end{aligned} \tag{25}$$

where the recursive updating rule $(23)$ for $\alpha_\ell^{-1}$ (and $\overline{P_\ell}$ ) is used. The resulting factorization of the Kähler potential $(14)$ becomes (in parameter space $\mathcal{P}_N \times \mathcal{D}_N^{n-1}$ )

$$\Phi(\tilde{R}_n) = -n \log(\det P_0) - \sum_{\ell=1}^{n-1} (n-\ell) \log(\det(I - \Gamma_\ell \overline{\Gamma_\ell})) - \log(\pi e),$$

where $\tilde{R}_n$ is identified with $(P_0, \Gamma_1, \dots, \Gamma_{n-1})$ under transformation $(24)$.

As before, we use complex differentiation to determine the Hessian of the Kähler potential and obtain the generalized metric

$$\mathrm{d}s^2 = n\,\mathrm{trace}(P_0^{-1}\,\mathrm{d}P_0 P_0^{-1}\,\mathrm{d}P_0) + \sum_{\ell=1}^{n-1} (n-\ell)\,\mathrm{trace}\big((I-\Gamma_\ell\bar{\Gamma}_\ell)^{-1}\,\mathrm{d}\Gamma_\ell (I-\bar{\Gamma}_\ell\Gamma_\ell)^{-1}\overline{\mathrm{d}\Gamma_\ell}\big).$$

From the metric it can be seen that the desired geometry on $\mathcal{P}_N$ is (up to a scalar $\sqrt{n}$ and $n$, respectively) given by (7) and (8). Unfortunately, the set $\mathcal{D}_N$ with the geometry described in the above metric does not correspond to any known manifold, nor does a natural distance measure present itself intuitively. However, the set $\mathcal{D}_N$ does bear a close resemblance to the set

$$\mathcal{SD}_N = \{\Omega \in \mathbb{C}^{N\times N} \mid I - \Omega\Omega^H > 0\},$$

which is (almost) the Siegel disk [228] and which has been well studied along with the Siegel upper half-plane. We discuss the slight adaptation to the transformation in order to obtain elements in the parameter space $\mathcal{P}_N \times \mathcal{SD}_N^{n-1}$ and we will also elude on the geometry of the Siegel.

We present a different generalized transformation, where the set $\mathcal{D}_N$ in transformation (24) is replaced by the Siegel disk $\mathcal{SD}_N$. Next, we show the relation between both sets and discuss how the new transformation is also a natural extension of the scalar Kähler metric. Finally, the geometry of the Siegel disk will be discussed.

The transformation. A different approach to the transformation of a $PD$ $(TB)BT$ matrix can be derived from a link with Verblunsky coefficients [235], [236] as follows.

In the previous setting of Toeplitz matrices, a one-to-one correspondence exists between a $PD$ Toeplitz matrix and a probability measure on the complex unit circle, where the elements in the Toeplitz matrix are found as the moments (or Fourier coefficients) of the corresponding probability measure [208], [212], [214], [220], [224]. The concept of orthogonality for polynomials on the unit circle is linked to the specified probability measure, and thus indirectly to the specific Toeplitz matrix. Finally, the computation of an orthonormal basis of polynomials on the unit circle can be performed using the Szegö's recursion [232], in which the Verblunsky coefficients arise. It turns out that these coefficients are equal to the prediction coefficients $a_\ell^\ell$(12) used in transformation (13) [204].

By generalizing the scalar probability measure on the complex unit circle to a nonnegative matrix measure, the collection of its moments into a matrix becomes a $PD$ $BT$ matrix [212], [213], [215]. On the other hand, constructing orthogonal matrix polynomials on the unit circle $w.r.t.$ the matrix measure results in a generalization of the Szegô recursion, with corresponding generalized Verblunsky coefficients [210], [215], [230].

We use the proposed generalization of the Verblunsky coefficients [215] to define a new transformation of a $PD$ $BT$ matrix as follows:

$$\begin{aligned} \mathcal{B}_{n,N}^{+} &\to \mathcal{P}_N \times \mathcal{SD}_N^{n-1}, \\ \tilde{R}_n &\mapsto (P_0, \Omega_1, \dots, \Omega_{n-1}), \end{aligned} \tag{26}$$

where $P_0$ is still equal to $R_0$, but now

$$\Omega_\ell := L_{\ell-1}^{-\frac{1}{2}}(R_\ell - M_{\ell-1})K_{\ell-1}^{-\frac{1}{2}}, \tag{27}$$

$$\begin{cases} L_{\ell-1} = R_0 - [R_1, \dots, R_{\ell-1}]\tilde{R}_{\ell-1}^{-1}[R_1, \dots, R_{\ell-1}]^H, \\ K_{\ell-1} = R_0 - \left[R_{\ell-1}^H, \dots, R_1^H\right]\tilde{R}_{\ell-1}^{-1}\left[R_{\ell-1}^H, \dots, R_1^H\right]^H, \\ M_{\ell-1} = [R_1, \dots, R_{\ell-1}]\tilde{R}_{\ell-1}^{-1}\left[R_{\ell-1}^H, \dots, R_1^H\right]^H \end{cases}$$

for $\ell = 1, \dots, n-1$. Comparing this transformation to the previous one, the following relations can be found for the auxiliary matrices $P_\ell$ and $\Delta_\ell$(21): $K_{\ell-1} = P_{\ell-1}, L_{\ell-1} = \overline{P_{\ell-1}}$, and $R_\ell - M_{\ell-1} = \Delta_\ell$. Hence we can also write the new transformation as

$$\Omega_\ell = \overline{P_{\ell-1}^{-1/2}} \Delta_\ell P_{\ell-1}^{-1/2},$$

which demonstrates the close relation between both transformations. The absence of the minus sign is not a problem as will become clear from the geometry of the Siegel disk (28). As for the transformation (13) of Toeplitz matrices, adaptations of (26) have also been suggested [222], [223]. However, these adaptations are again designed for more robust estimation of the correlation blocks $R_t$ in (18) from a finite number of measurements. We assume that this estimation has been performed prior to the application of the averaging operation. Moreover, as before, the different transformations result in the same parameter space $\mathcal{P}_N \times \mathcal{SD}_N^{n-1}$. Hence we will not discuss these adapted transformations any further.

It still remains to show that the coordinate matrices $\Omega_\ell$ actually are elements of the Siegel disk. In fact, this was proven for the transformation of a general $PD\ BT$ matrix by Dette and Wagener [215] and Fritzsche and Kirstein [218]. We will discuss this for the transformation of elements in the set of $PD\ TBBT$ matrices $\mathcal{T}_{n,N}^{+}$. Our interest goes specifically to $PD\ TBBT$ matrices, but we will briefly revisit the $PD\ BT$ matrices.

Suppose we have $\tilde{R}_\ell \in \mathcal{T}_{n,N}^{+}$, then by exploiting the Toeplitz structure of the blocks and $\tilde{R}_\ell = \tilde{R}_\ell$, we can show that

$$\begin{aligned}
\overline{\Delta_\ell} &= \overline{R_\ell} - \overline{M_{\ell-1}} \\
&= R_\ell^H - J_N [R_1, \dots, R_{\ell-1}]^* \tilde{R}_{\ell-1}^{-1^*} \left[R_{\ell-1}^H, \dots, R_1^H\right]^{H^*} J_N \\
&= R_\ell^H - J_N [R_1, \dots, R_{\ell-1}]^* J_{nN} \tilde{R}_{\ell-1}^{-1} J_{nN} \left[R_{\ell-1}^H, \dots, R_1^H\right]^{H^*} J_N \\
&= R_\ell^H - \left[R_{\ell-1}^H, \dots, R_1^H\right] \tilde{R}_{\ell-1}^{-1} [R_1, \dots, R_{\ell-1}]^H \\
&= \Delta_\ell^H,
\end{aligned}$$

after which we can again start from the positive definiteness of $\overline{P_\ell}$,

$$\begin{aligned}
&\overline{P_\ell} = \overline{P_{\ell-1}} - \Delta_\ell P_{\ell-1}^{-1} \overline{\Delta_\ell} > 0, \\
\xrightarrow{congruence}\quad & I - \overline{P_{\ell-1}^{-1/2}} \Delta_\ell P_{\ell-1}^{-1} \Delta_\ell^H \overline{P_{\ell-1}^{-1/2}} > 0, \\
& I - \left(\overline{P_{\ell-1}^{-1/2}} \Delta_\ell P_{\ell-1}^{-1/2}\right)\left(P_{\ell-1}^{-1/2} \Delta_\ell^H \overline{P_{\ell-1}^{-1/2}}\right) = I - \Omega_\ell \Omega_\ell^H > 0,
\end{aligned}$$

which proves $\Omega_\ell \in \mathcal{SD}_N$.

The metric. We want to define the generalized metric by starting from the Kähler potential, where we continue from (25) using the following,

$$\begin{aligned}
\det\left(I - A_\ell^\ell \overline{A_\ell^\ell}\right) &= \det\left(I - \Delta_\ell P_{\ell-1}^{-1} \overline{\Delta_\ell} \overline{P_{\ell-1}^{-1}}\right) \\
&= \det\left(I - \Delta_\ell P_{\ell-1}^{-1} \Delta_\ell^H \overline{P_{\ell-1}^{-1}}\right) \\
&= \det\left(I - \overline{P_{\ell-1}^{-1/2}} \Delta_\ell P_{\ell-1}^{-1} \Delta_\ell^H \overline{P_{\ell-1}^{-1/2}}\right) \\
&= \det\left(I - \Omega_\ell \Omega_\ell^H\right).
\end{aligned}$$

The expression for the Kähler potential and resulting generalized metric [202] are

$$\Phi(\tilde{R}_n) = -\,\mathrm{n}\log(\det P_0) - \sum_{\ell=1}^{n-1} (n-\ell)\log\big(\det(I-\Omega_\ell\Omega_\ell^H)\big) - \log(\pi e),$$

$$\mathrm{d}s^2 = n \,\mathrm{trace}\,(P_0^{-1}\,\mathrm{d}P_0 P_0^{-1}\,\mathrm{d}P_0) + \sum_{\ell=1}^{n-1} (n-\ell)\,\mathrm{trace}\,\Big(\big(I-\Omega_\ell\Omega_\ell^H\big)^{-1}\,\mathrm{d}\Omega_\ell\big(I-\Omega_\ell^H\Omega_\ell\big)^{-1}\,\mathrm{d}\Omega_\ell^H\Big). \tag{28}$$

The geometry on $\mathcal{P}_N$ remains the same as for the first transformation. For the Siegel disk $\mathcal{SD}_N$, the natural geometry can be derived from the geometry of the Siegel upper half-plane described by Siegel himself [228], using the link

$$\Omega = (B-iI)(B+iI)^{-1},$$
$$B = i(I+\Omega)(I-\Omega)^{-1},$$

where $B$ is an element of the Siegel upper half-plane ($\mathfrak{Im}(B) > 0$). We should note that this link and the Siegel disk itself are classically only defined for symmetric matrices (in order for the positive definiteness of $\mathfrak{Im}(B)$ to make sense). However, removing the symmetry restriction only disrupts the link and the definition of the Siegel upper half-plane, while the Siegel disk and its geometry remain well-defined.

The resulting (scaled) geometry on $\mathcal{SD}_N$ and a reminder of the (scaled) geometry on $\mathcal{P}_N$ are

$$\forall A,B \in \mathcal{P}_N, \forall E,F \in \mathcal{H}_N:$$
$$\langle E,F\rangle_A = n\,\mathrm{trace}(A^{-1}EA^{-1}F), \tag{29}$$
$$d_{\mathcal{P}_N}(A,B) = \sqrt{n}\big\|\log\big(A^{-1/2}BA^{-1/2}\big)\big\|_F;$$

$$\forall \Omega,\Psi \in \mathcal{SD}_N, \forall v,\omega \in \mathbb{C}^{N\times N}:$$
$$\langle v,\omega\rangle_\Omega = \frac{n-\ell}{2}\mathrm{trace}((I-\Omega\Omega^H)^{-1}v(I-\Omega^H\Omega)^{-1}\omega^H) + \frac{n-\ell}{2}\mathrm{trace}((I-\Omega\Omega^H)^{-1}\omega(I-\Omega^H\Omega)^{-1}v^H),$$
$$d^2_{\mathcal{SD}_N}(\Omega,\Psi) = \frac{n-\ell}{4}\mathrm{trace}\left(\log^2\left(\frac{I+C^{\frac{1}{2}}}{I-C^{\frac{1}{2}}}\right)\right), \tag{30}$$
$$[C = (\Psi-\Omega)(I-\Omega^H\Psi)^{-1}(\Psi^H-\Omega^H)(I-\Omega\Psi^H)^{-1},$$

where $\ell$ is chosen corresponding to the coordinate matrix ($\Omega_\ell, \ell = 1 \ldots, n-1$, from (26)) to which it relates. Note that both inner products and distance measures reduce to the scalar expressions when $N = 1$. We also point out that the distance measure $d_{\mathcal{SD}_N}$ on the Siegel disk can be written using a Frobenius norm. This is accomplished by performing the similarity transformation $(I-\Omega\Omega^H)^{-1/2}C(I-\Omega\Omega^H)^{1/2}$, which results in a Hermitian matrix (as shown below in (33)) and does not change the distance measure since only the eigenvalues of $C$ matter.

The Kähler distance $d_{BT}$ between two PD(TB)BT matrices $\tilde{T}_1$ and $\tilde{T}_2$, with transformations $\big(P_{0,1},\Omega_{1,1},\ldots,\Omega_{n-1,1}\big)$ and $\big(P_{0,2},\Omega_{1,2},\ldots,\Omega_{n-1,2}\big)$, is defined as

$$d_{BT}^2(\tilde{T}_1,\tilde{T}_2) = d_{BT}^2\left(\left(P_{0,1},\Omega_{1,1},\dots,\Omega_{n-1,1}\right),\left(P_{0,2},\Omega_{1,2},\dots,\Omega_{n-1,2}\right)\right)$$
$$= n\left\|\log\left(P_{0,1}^{-1/2}P_{0,2}P_{0,1}^{-1/2}\right)\right\|_F^2 + \sum_{\ell=1}^{n-1}\frac{n-\ell}{4}\operatorname{trace}\left(\log^2\left(\frac{I+C_\ell^{\frac{1}{2}}}{I-C_\ell^{\frac{1}{2}}}\right)\right), \quad (31)$$
$$\left[C_\ell = \left(\Omega_{\ell,2}-\Omega_{\ell,1}\right)\left(I-\Omega_{\ell,1}^H\Omega_{\ell,2}\right)^{-1}\left(\Omega_{\ell,2}^H-\Omega_{\ell,1}^H\right)\left(I-\Omega_{\ell,1}\Omega_{\ell,2}^H\right)^{-1}.\right.$$

Using the definition of a barycenter $(10)$, the generalized Kähler mean can now be found as $\mathcal{B}_{BT}$.

The distance measure discussed was proposed by Siegel as a possible natural generalization to scalar distance measure on the Poincaré disk. Other generalizations have also been investigated, and among these, the one we will refer to as the Kobayashi distance measure $d_K$ has some interesting properties.

For $\Omega, \Psi \in \mathcal{SD}_N$, it is defined as [203], [206], [217]

$$d_K(\Omega,\Psi) = \frac{1}{2}\log\left(\frac{1+\|\phi_\Omega(\Psi)\|_2}{1-\|\phi_\Omega(\Psi)\|_2}\right),$$
$$\left[\phi_\Omega(\Psi) = (I-\Omega\Omega^H)^{-\frac{1}{2}}(\Psi-\Omega)(I-\Omega^H\Psi)^{-1}(I-\Omega^H\Omega)^{\frac{1}{2}},\right. \quad (32)$$

which, up to scaling, reduces exactly to the scalar distance measure on the Poincaré disk. The 2-norm $\|\cdot\|_2$ in this expression represents the spectral norm of a matrix, given by its largest singular value.

Unfortunately, the Kobayashi distance measure is not naturally associated with the metric on the Siegel disk with which we are working. We show this by examining the differential metric at the zero matrix. By entering $\Omega = 0$ in $(28)$, our differential metric on the Siegel disk becomes $\mathrm{d}s^2 = \operatorname{trace}(\mathrm{d}\Omega\mathrm{d}\Omega^H) = \|\mathrm{d}\Omega\|_F^2$. The differential metric corresponding to the Kobayashi distance measure at the zero matrix is given by $\mathrm{d}s^2 = \|\mathrm{d}\Omega\|_2^2$ $[217]$, which is clearly not the same.

However, the main advantage of this distance measure lies in the transformation $\phi_\Omega(32)$, which acts as an automorphism on the Siegel disk. The distance between two matrices and between their transformations under $\phi_\Omega$ remains the same, for both the Siegel distance $d_{\mathcal{SD}_N}$ and the Kobayashi distance $d_K$, and this can be exploited in the computations. During each step of the optimization process, the current iteration point is translated to the origin (the zero matrix) while the original matrices of the mean are translated accordingly. Working at the origin will simplify the computation of optimization constructions such as the gradient, retractions, $etc$. In the information geometry, the idea of translation to the origin on the Poincaré and Siegel disk for barycenter computation was introduced by Barbaresco $[203]$.

We note already that this translation to the origin is no longer practical once we enforce the Toeplitz structure on the individual blocks $R_\ell, \ell = 0,\dots,n-1, i.e.$, when we go from $PD\ BT$ matrices to $PD\ TBBT$ matrices. As will be fully explained, once an iteration step $\omega$ at the translated origin is computed, the actual iteration point $\Omega_\ell$ (with respect to the original matrices) should be updated to $\phi_{(-\Omega_\ell)}(\omega)$. Imposing the Toeplitz structure on the blocks $R_\ell$ now results in a very involved condition for the step $\omega$. The process of exploiting the translation itself is further explained .

The presence of the underlying Toeplitz structure in the blocks greatly influences the computation of the generalized Kähler mean. Therefore, we first discuss the situation in

which the structure is not required, the necessary changes and resulting implications of imposing the Toeplitz condition are presented.
In the general case of $PD\ BT$ matrices, all advantages of the scalar version are still valid. The optimization of the coordinate matrices under transformation $(26)$ can be performed separately, resulting in $n$ parallel optimization processes involving $N \times N$ matrices (instead of a single process involving $nN \times nN$ matrices).
The optimization in the first coordinate matrix results in the Karcher mean $\mathcal{B}_{\mathcal{P}_N}\left(P_{0,1}, \dots, P_{0,k}\right)$ of the involved $PD$ matrices, as defined.
For the other coordinates $\left(\Omega_{\ell,i} \in \mathcal{SD}_N\right)$, the optimization at each level of $\ell(= 1, \dots, n-1)$ can be formulated in the same way, hence we omit the depen- dence on $\ell$ in the definition of the barycenter

$$\mathcal{B}_{\mathcal{SD}_N}(\Omega_1, \cdots, \Omega_\mathrm{k}) = \arg \min_{X \in \mathcal{SD}_N} \frac{1}{2} \sum_{i=1}^{k} \left\| \log \left( \frac{1 + C_i^{\frac{1}{2}}}{1 - C_i^{\frac{1}{2}}} \right) \right\|_F^2,$$

$$\left[ C_i = I - \left(I - \Omega_i \Omega_i^H\right)^{\frac{1}{2}} \left(I - \mathrm{X} \Omega_i^H\right)^{-1} (I - XX^H)(I - \Omega_i X^H)^{-1} \left(I - \Omega_i \Omega_i^H\right)^{\frac{1}{2}} \right], (33)$$

where the cost function has been rescaled and $C_i$ is written in the Hermitian form which was mentioned. The cost function in this optimization problem will be denoted as $f_{\mathcal{B}_{\mathcal{SD}_N}}(X)$.
$A$ first order optimization algorithm requires us to determine the (Riemannian) gradient of the cost function, defined for $\mathcal{SD}_N$ as

$$\mathrm{D} f_{\mathcal{B}_{\mathcal{SD}_N}}(X)[\omega_X] = \left\langle \mathrm{grad}\, f_{\mathcal{B}_{\mathcal{SD}_N}}(X), \omega_X \right\rangle_X \tag{34}$$

with the inner product $(30)$. Note that differentiating the cost function at some point requires the differentiation of the matrix inverse and matrix square root. Using the notation $g(X) = X^{-1}$ and $h(X) = X^{1/2}$, these are given by

$$\mathrm{D}g(X)[\omega] \quad = -X^{-1} \omega X^{-1}, \quad \text{inversion [211]},$$

$$\mathrm{D}h(X)[\omega] X^{\frac{1}{2}} + X^{\frac{1}{2}} \mathrm{D}h(X)[\omega], \qquad \text{square root},$$

where the latter is obtained by applying the product rule to the definition $X^{1/2} X^{1/2} = X$ and can be recognized (and solved) as a continuous Lyapunov equation $(CLE)$.
After some calculations, the emerging gradient is

$$\mathrm{grad}\, f_{\mathcal{B}_{\mathcal{SD}_N}}(X) = (I - XX^H) \sum_{i=1}^{k} \left(V_i (X - \Omega_i)(I - X^H \Omega_i)^{-1}\right)(I - X^H X), \tag{35}$$

$$\left[ \begin{array}{c} V_i = (I - \Omega_i X^H)^{-1} \left(I - \Omega_i \Omega_i^H\right)^{\frac{1}{2}} Z_i \left(I - \Omega_i \Omega_i^H\right)^{\frac{1}{2}} \left(I - X \Omega_i^H\right)^{-1}, \\ Z_i = \mathfrak{L} \left( C_i^{\frac{1}{2}}, (I - C_i)^{-1} \log \left( \frac{I + C_i^{\frac{1}{2}}}{I - C_i^{\frac{1}{2}}} \right) \right), \end{array} \right.$$

where $C_i$ is defined as in $(33)$ and $\mathfrak{L}(A, Q)$ stands for the solution $X$ of the $CLE$ $AX + XA^H = Q$. Note that the second argument in the Lyapunov operator $\mathfrak{L}$ is a Hermitian matrix, hence

the $CLE$ is well-defined. This gradient can be used to design a basic steepest descent or conjugate gradient method in order to obtain the barycenter.
Translation to the origin. Using the translation $\phi(32)$, computations can be greatly simplified. Suppose the initial guess for the barycenter $\mathcal{B}_{\mathcal{SD}_N}$ is given by a matrix $X_0$. The translation $\phi_{X_0}$ maps the matrix $X_0$ exactly onto the origin (the zero matrix) $0 \in \mathcal{SD}_N$ and by applying the same transformation to the original matrices $\Omega_i$, the distances $d_{\mathcal{SD}_N}(X_0, \Omega_i)$ and $d_{\mathcal{SD}_N}\left(0, \phi_{X_0}(\Omega_i)\right)$ remain exactly the same for all $i$. Hence the value of the barycenter cost function $f_{\mathcal{B}_{\mathcal{SD}_N}}$ does not change under this translation. The gradient of the (translated) cost function can now be computed at the origin and used in a basic descent method to obtain a new iteration point, denoted by $\Psi_1$. We can translate this new point again to the origin using the next translation $\phi_{\Psi_1}$. However, in order to keep track of the barycenter approximations with respect to the original matrices, we need to keep in mind that $\Psi_1$ is an improvement over the origin for the translated matrices $\phi_{X_0}(\Omega_i)$. The new barycenter approximation with respect to the original matrices is hence given by $X_1 = \phi_{-X_0}(\Psi_1)$ (note that $\phi_{X_0}^{-1} = \phi_{-X_0}$).
The resulting procedure is summarized in Algorithm 1. Note that $\Omega_i^{(j+1)}$ can also be computed as $\phi_{\Psi_{j+1}}\left(\Omega_i^{(j)}\right)$ [203]. However, in both this formula and the one mentioned in the algorithm, a translation needs to be performed, but by always restarting from the original matrices, the updating formula mentioned in the algorithm is less sensitive to the accumulation of roundoff errors.

**Algorithm 1** Procedure for translating to the origin

Let $\Omega_1, \ldots, \Omega_k$ be $k$ matrices in $\mathcal{SD}_N$, $X_0 \in \mathcal{SD}_N$ an initial guess

- **for** $j = 0, 1, \ldots$
  - Compute the translated matrices:
  $$(\Omega_1^{(j)}, \ldots, \Omega_k^{(j)}) = (\phi_{X_j}(\Omega_1), \ldots, \phi_{X_j}(\Omega_k));$$
  - Compute the gradient of the translated cost function at the origin (4.4):
  $$\operatorname{grad} f_{\mathcal{B}_{\mathcal{SD}_N}}(0; \Omega_1^{(j)}, \ldots, \Omega_k^{(j)}),$$
  and perform a basic descent step to obtain $\Psi_{j+1}$;
  - Obtain the next iteration point by returning to the original matrices:
  $$X_{j+1} = \phi_{-X_j}(\Psi_{j+1});$$
- **end for**

**Return:** $\mathcal{B}_{\mathcal{SD}_N}(\Omega_1, \ldots, \Omega_k)$

Finally, we present the simplified form of the gradient at the origin,

$$\operatorname{grad} f_{\mathcal{B}_{\mathcal{SD}_N}}(0; \Omega_1, \ldots, \Omega_k) = -\sum_{i=1}^{k} V_i \Omega_i, \tag{36}$$

$$\left[V_i = \mathfrak{L}\left(\left(\Omega_i \Omega_i^H\right)^{\frac{1}{2}}, \log\left(\frac{I + \left(\Omega_i \Omega_i^H\right)^{\frac{1}{2}}}{I - \left(\Omega_i \Omega_i^H\right)^{\frac{1}{2}}}\right)\right)\right.,$$

where $V_i$ is now obtained directly as the solution of a $CLE$.

As mentioned, in some applications the Toeplitz structure is not only present in the block structure, but also in the individual blocks themselves. To investigate the implications

of this restriction, we have another look at the transformation $(26)$ of the matrices, with the $n-1$ coordinate matrices in the Siegel disk given by $(27)$.
At first sight, imposing the Toeplitz structure requires the matrix $R_\ell$ in each $\Omega_\ell$ to be Toeplitz. However, the matrices $L_{\ell-1}, K_{\ell-1}$, and $M_{\ell-1}$ depend on the matrices $R_0, \dots, R_{\ell-1}$, which should also be Toeplitz matrices now. All these Toeplitz restrictions are translated in an involved way to the search space in which each $\Omega_\ell$ is located. By taking the involved connections into account, we will derive the general Kähler mean for $PD\ TBBT$ matrices. Afterwards, we present an approximation to this general Kähler mean which again allows us to perform the optimization of the coordinate matrices separately, but now sequentially in the given order of the variables as in transformation $(26)$ $(P_0 \to \Omega_1 \to \cdots \to \Omega_{n-1})$.

Instead of translating the Toeplitz restriction towards involved conditions on the coordinate matrices $(P_0, \Omega_1, \dots, \Omega_{n-1})$, we consider the barycenter cost function $f_{\mathcal{B}_{BT}}$, based on the total Kähler distance function $d_{BT}(31)$, as a function of the blocks $R_0, \dots, R_{n-1}$ of the matrix $\tilde{R}_n$. Doing so will result in a more involved gradient, but it allows us to enforce the Toeplitz structure directly onto its components.
The complexity of this differentiation "throughout" the coordinate matrices is caused by the dependence on the original blocks. While the first coordinate matrix $P_0$ only depends on $R_0$, each coordinate matrix $\Omega_\ell$ depends on the blocks $R_0, \dots, R_\ell$ for $\ell = 1, \dots, n-1$, or, reversely, $R_0$ will influence all coordinate matrices, and for each $\ell = 1, \dots, n-1$, block $R_\ell$ is present in coordinate matrices $\Omega_\ell, \dots, \Omega_{n-1}$.
The gradient. As shown in $(34)$, the gradient of the cost function depends on its derivative and the inner product on the search space. Because of the intricate connections between the variables, the gradient is now defined on the product space of the blocks as follows:

$$\begin{aligned} \mathrm{D}f_{\mathcal{B}_{BT}} \quad & \big((R_0, \dots, R_{n-1})\big)[(E_0, \omega_1, \dots, \omega_{n-1})] \\ = \quad & \big\langle \operatorname{grad} f_{\mathcal{B}_{BT}}\big((R_0, \dots, R_{n-1})\big), (E_0, \omega_1, \dots, \omega_{n-1}) \big\rangle_{(R_0, \dots, R_{n-1})} \\ := \quad & \Big\langle \operatorname{grad} f_{\mathcal{B}_{BT}}\big((R_0, \dots, R_{n-1})\big)_0, E_0 \Big\rangle_{P_0} \\ & + \sum_{\ell=1}^{n-1} \left\langle L_{\ell-1}^{-\frac{1}{2}} \operatorname{grad} f_{\mathcal{B}_{BT}}\big((R_0, \dots, R_{n-1})\big)_\ell K_{\ell-1}^{-\frac{1}{2}}, L_{\ell-1}^{-\frac{1}{2}} \omega_\ell K_{\ell-1}^{-\frac{1}{2}} \right\rangle_{\Omega_\ell}, \end{aligned} \tag{37}$$

where $(P_0, \Omega_1, \dots, \Omega_{n-1})$ is the image of $\tilde{R}_n$ under transformation $(26)$ with $L_{\ell-1}$ and $K_{\ell-1}$ the matrices formed during the transformation. The inner products $\langle .,. \rangle_{P_0}$ and $\langle .,. \rangle_{\Omega_\ell}$ are given by $(29)$ and $(30)$, respectively, and the $(\ell+1)th$ component of the gradient is represented by $\operatorname{grad} f_{\mathcal{B}_{BT}}\big((R_0, \dots, R_{n-1})\big)_\ell$. The left and right multiplication by $L_{\ell-1}^{-1/2}$ and $K_{\ell-1}^{-1/2}$ in the last inner products is a consequence of the relation between the tangent space at $R_\ell$ versus the tangent space at $\Omega_\ell$.
To demonstrate the complexity of the relations, we present the gradient below. The point at which the gradient is computed is denoted by $\tilde{R}_n$, with blocks $(R_0, \dots, R_{n-1})$ and transformation $(P_0, \Omega_1, \dots, \Omega_{n-1})$, while the $PD\ TBBT$ matrices of which the barycenter is computed will be denoted by $\tilde{R}_{n,i}$, with blocks $\big(R_{0,i}, \dots, R_{n-1,i}\big)$ and transformation $\big(P_{0,i}, \Omega_{1,i}, \dots, \Omega_{n-1,i}\big), i = 1, \dots, k$.
In the expressions, the matrices $A_j^{\ell-1}(19)-(22)$, associated with the creation of $\Delta_\ell$ and $P_{\ell-1}$ (and therefore $L_{\ell-1}, K_{\ell-1}$, and $M_{\ell-1}$ ) in the transformation of $\tilde{R}_n$, are used to increase

readability and computational efficiency. The first component of the gradient becomes the following:

$$
\begin{aligned}
&\operatorname{grad} f_{\mathcal{B}_{BT}}\big((R_0,\dots,R_{n-1})\big)_0\\
&= P_0 \sum_{i=1}^{k} \left( P_0^{-1} \log\left(P_0 P_{0,i}^{-1}\right) + \sum_{\ell=1}^{n-1} \frac{n-\ell}{2n} G_{\ell,i} \right) P_0,
\end{aligned}
$$

$$
\left[\begin{aligned}
G_{\ell,i} &= -D_{\ell,i}^{L} - D_{\ell,i}^{K} + \textstyle\sum_{j=1}^{\ell-1} \Big( -{A_j^{\ell-1}}^H D_{\ell,i}^{L} A_j^{\ell-1} - \overline{A_j^{\ell-1}}^H D_{\ell,i}^{K} \overline{A_j^{\ell-1}} \\
&\qquad + {A_j^{\ell-1}}^H L_{\ell-1}^{-\frac{1}{2}} {V_{\ell,i}^{(1)}}^H K_{\ell-1}^{-\frac{1}{2}} \overline{A_{\ell-j}^{\ell-1}} + \overline{A_{\ell-j}^{\ell-1}}^H K_{\ell-1}^{-\frac{1}{2}} V_{\ell,i}^{(1)} L_{\ell-1}^{-\frac{1}{2}} A_j^{\ell-1} \Big),\\
D_{\ell,i}^{L} &= \mathfrak{L}\left( L_{\ell-1}^{\frac{1}{2}}, \Omega_\ell V_{\ell,i}^{(1)} L_{\ell-1}^{-\frac{1}{2}} + L_{\ell-1}^{-\frac{1}{2}} {V_{\ell,i}^{(1)}}^H \Omega_\ell^H \right),\\
D_{\ell,i}^{K} &= \mathfrak{L}\left( K_{\ell-1}^{\frac{1}{2}}, \Omega_\ell^H {V_{\ell,i}^{(1)}}^H K_{\ell-1}^{-\frac{1}{2}} + K_{\ell-1}^{-\frac{1}{2}} V_{\ell,i}^{(1)} \Omega_\ell \right),\\
V_{\ell,i}^{(1)} &= (I - \Omega_{\ell,i}^H \Omega_\ell)^{-1} (\Omega_\ell^H - \Omega_{\ell,i}^H) V_{\ell,i},\\
V_{\ell,i} &= (I - \Omega_{\ell,i}\Omega_\ell^H)^{-1} (I - \Omega_{\ell,i}\Omega_{\ell,i}^H)^{\frac{1}{2}} Z_{\ell,i} (I - \Omega_{\ell,i}\Omega_{\ell,i}^H)^{\frac{1}{2}} (I - \Omega_\ell \Omega_{\ell,i}^H)^{-1},\\
Z_{\ell,i} &= \mathfrak{L}\left( C_{\ell,i}^{\frac{1}{2}}, (I - C_{\ell,i})^{-1} \log\left( \frac{I + C_{\ell,i}^{\frac{1}{2}}}{I - C_{\ell,i}^{\frac{1}{2}}} \right) \right),\\
C_{\ell,i} &= I - \Big( (I - \Omega_{\ell,i}\Omega_{\ell,i}^H)^{\frac{1}{2}} (I - \Omega_\ell \Omega_{\ell,i}^H)^{-1} (I - \Omega_\ell \Omega_\ell^H) \dots \\
&\qquad\qquad (I - \Omega_{\ell,i}\Omega_\ell^H)^{-1} (I - \Omega_{\ell,i}\Omega_{\ell,i}^H)^{\frac{1}{2}} \Big),
\end{aligned}\right.
\tag{38}
$$

where $D_{\ell,i}^L$, $D_{\ell,i}^K$, and $Z_{\ell,i}$ are obtained by solving a $CLE$. The other components of the gradient are, for $q = 1, \dots, n-1$, given by

$$
\begin{aligned}
&\operatorname{grad} f_{\mathcal{B}_{BT}}\big((R_0,\dots,R_{n-1})\big)_q\\
&= L_{q-1}^{\frac{1}{2}} (I - \Omega_q \Omega_q^H) \sum_{i=1}^{k} {V_{q,i}^{(1)}}^H (I - \Omega_q^H \Omega_q) K_{q-1}^{\frac{1}{2}}\\
&\quad + L_{q-1}^{\frac{1}{2}} (I - \Omega_q \Omega_q^H) L_{q-1}^{\frac{1}{2}} \sum_{i=1}^{k} \left( \sum_{\ell=q+1}^{n-1} \frac{n-\ell}{n-q} W_{\ell,i}^{(q)} \right) K_{q-1}^{\frac{1}{2}} (I - \Omega_q^H \Omega_q) K_{q-1}^{\frac{1}{2}},
\end{aligned}
$$

$$
\left[\begin{aligned}
W_{\ell,i}^{(q)} &= -D_{\ell,i}^{L} A_q^{\ell-1} - \overline{A_q^{\ell-1}}^H D_{\ell,i}^{K}\\
&\quad + {A_{\ell-q}^{\ell-1}}^H L_{\ell-1}^{-\frac{1}{2}} {V_{\ell,i}^{(1)}}^H K_{\ell-1}^{-\frac{1}{2}} + L_{\ell-1}^{-\frac{1}{2}} {V_{\ell,i}^{(1)}}^H K_{\ell-1}^{-\frac{1}{2}} \overline{A_{\ell-q}^{\ell-1}}\\
&\quad + \textstyle\sum_{j=q+1}^{\ell-1} \Big( -{A_{j-q}^{\ell-1}}^H D_{\ell,i}^{L} A_j^{\ell-1} - \overline{A_j^{\ell-1}}^H D_{\ell,i}^{K} \overline{A_{j-q}^{\ell-1}}\\
&\quad + {A_{j-q}^{\ell-1}}^H L_{\ell-1}^{-\frac{1}{2}} {V_{\ell,i}^{(1)}}^H K_{\ell-1}^{-\frac{1}{2}} \overline{A_{\ell-j}^{\ell-1}} + \overline{A_{\ell-j+q}^{\ell-1}}^H K_{\ell-1}^{-\frac{1}{2}} V_{\ell,i}^{(1)} L_{\ell-1}^{-\frac{1}{2}} A_j^{\ell-1} \Big),
\end{aligned}\right.
\tag{39}
$$

where $D_{\ell,i}^L$, $D_{\ell,i}^K$, and $V_{\ell,i}^{(1)}$ are the same as for the first component.

What we have done so far is to compute the gradient of $f_{\mathcal{B}_{BT}}$ as a function of the matrix blocks $(R_0, \dots, R_{n-1})$ instead of the coordinate matrices $(P_0, \Omega_1, \dots, \Omega_{n-1})$. Finally, we can impose the Toeplitz structure on the blocks.

Projection onto the Toeplitz structure. According to manifold optimization theory, computing the gradient of a cost function on some submanifold is equivalent to computing the gradient in the embedding manifold and applying the orthogonal projection onto the submanifold [21]. In our case, the embedding manifold is the set $(\mathbb{C}^{N\times N})_+^n$ containing all tuples $(R_0, \dots, R_{n-1})$ which represent the blocks of an element in $\mathcal{B}_{n,\mathrm{N}}^+$.The submanifold is given by the set $(\mathcal{T}_N)_+^n$ which contains all tuples $(R_0, \dots, R_{n-1})$ holding the blocks of an element in $\mathcal{T}_{n,N}^+$.

Above, we have computed the gradient of the cost function $f_{\mathcal{B}_{BT}}$ for the embedding manifold $(\mathbb{C}^{N\times N})_+^n$ since no additional structure was imposed on the blocks.
Hence, we need an orthogonal projection of this gradient at any point $(R_0,\dots,R_{n-1}) \in (\mathcal{T}_N)_+^n \subset (\mathbb{C}^{N\times N})_+^n$ from $T_{(R_0,\dots,R_{n-1})}(\mathbb{C}^{N\times N})_+^n$ onto $T_{(R_0,\dots,R_{n-1})}(\mathcal{T}_N)_+^n$. This projection should be orthogonal with respect to the inner product (37) and, for

$$(E_0,\omega_1,\dots,\omega_{n-1}) \in T_{(R_0,\dots,R_{n-1})}(\mathbb{C}^{N\times N})_+^n,$$

is given by

$$E_0 \mapsto \mathrm{vec}^{-1}(U_{\mathrm{H}}(U_{\mathrm{H}}^H(P_0^{-1\ T} \otimes P_0^{-1})U_{\mathrm{H}})^{-1}U_{\mathrm{H}}^H \,\mathrm{vec}(P_0^{-1}E_0P_0^{-1})), \tag{40}$$

$$\omega_\ell \mapsto \mathrm{vec}^{-1}\left(U_{\mathrm{T}}\left(U_{\mathrm{T}}^H\left(S_\ell^{K^T} \otimes S_\ell^L\right)U_{\mathrm{T}}\right)^{-1} U_{\mathrm{T}}^H \,\mathrm{vec}\left(S_\ell^L\omega_\ell S_\ell^K\right)\right), \tag{41}$$

$$\begin{cases} S_\ell^L = L_{\ell-1}^{-\frac{1}{2}}\left(I-\Omega_\ell\Omega_\ell^H\right)^{-1}L_{\ell-1}^{-\frac{1}{2}}, \\ S_\ell^K = K_{\ell-1}^{-\frac{1}{2}}\left(I-\Omega_\ell^H\Omega_\ell\right)^{-1}K_{\ell-1}^{-\frac{1}{2}} \end{cases}$$

for $\ell = 1,\dots,n-1$, where $(P_0,\Omega_1,\dots,\Omega_{n-1})$ is the transformation (26) of $\tilde{R}_n$, the $BT$ matrix containing blocks $(R_0,\dots,R_{n-1})$, with associated matrices $L_{\ell-1}$ and $K_{\ell-1}$. The vec operator is the columnwise vectorization of a matrix, and the matrices $U_{\mathrm{H}}$ and $U_{\mathrm{T}}$ are parametrization matrices for Hermitian Toeplitz and general Toeplitz matrices, respectively. Hence, $e.g.$, we write $vec\ (T_1) = U_{\mathrm{H}}t_1$ with $t_1 \in \mathbb{R}^{2N-1}$ the parametrization of $T_1 \in \mathcal{T}_N \cap \mathcal{H}_N$, and $\mathrm{vec}(T_2) = U_{\mathrm{T}}t_2$ with $t_2 \in \mathbb{C}^{2N-1}$ or $t_2 \in \mathbb{R}^{4N-2}$ a parametrization of $T_2 \in \mathcal{T}_N$.
Note that when the projection is combined with the gradient above, some cancellations occur within the vec operator of the projection. This is a consequence of the consistent use of the inner product (37) for both the Riemannian gradient and the orthogonal projection.

It is obvious that even a basic construction such as the gradient is expensive for the generalized Kähler mean with Toeplitz structure imposed on the blocks. Here we discuss an approximation to this mean which is obtained as an attempt to regain the separated optimization of the coordinate matrices.
Remember that the coordinate matrix $P_0$ only depends on the block $R_0$, coordinate matrix $\Omega_1$ depends on the blocks $R_0$ and $R_1, etc.$ The main idea of our approximation is to perform the optimization of the barycenter cost function $f_{\mathcal{B}_{BT}}$ in a greedy manner.
We start by minimizing the part of the cost function which only depends directly on $P_0$, while imposing the Toeplitz structure on $R_0$. This results in the computation of the structured geometric mean of the given coordinate matrices $\left(P_{0,1},\dots,P_{0,k}\right)$ as described by Bini et al. [20].
When this optimization process is completed, we assume $R_0$ (and $P_0$) to be fixed. Next, we continue with the optimization of $\Omega_1 = L_0^{-1/2}(R_1 - M_0)K_0^{-1/2}$, with the Toeplitz structure imposed on $R_1$. Note that since $R_0$ is assumed to be fixed, $L_0, K_0$, and $M_0$ are fixed as well, making the relation between $\Omega_1$ and $R_1$ straightforward.
When the optimization process on $R_1$ is finished, assume both $R_0$ and $R_1$ to be fixed and continue this method sequentially.
The optimization at the level of $\Omega_\ell, \ell = 1,\dots,n-1$, is performed using a combination of constructions which have already been derived. We remember the barycenter cost function $f_{\mathcal{B}_{\mathcal{SD}_N}}$ with associated gradient (35). Because of the Toeplitz restriction and the assumption that all previously optimized coordinate matrices are fixed, the tangent space at $\Omega_\ell$ is given by

$$T_{\Omega_\ell}\left(L_{\ell-1}^{-1/2}(\mathcal{T}_N - M_{\ell-1})K_{\ell-1}^{-1/2}\right) \simeq \left\{L_{\ell-1}^{-\frac{1}{2}}TK_{\ell-1}^{-\frac{1}{2}} \mid T \in \mathcal{T}_N\right\}.$$

We are now working directly on the level of $\Omega_\ell$ instead of $R_\ell$, hence the projection of the gradient onto this tangent space slightly differs from the one presented in (41) as follows:

$$\omega_\ell \mapsto \mathrm{vec}^{-1}\left(U_\mathrm{T}\left(U_\mathrm{T}^H\left(S_\ell^{K^T} \otimes S_\ell^L\right)U_\mathrm{T}\right)^{-1} U_\mathrm{T}^H \,\mathrm{vec}\left(S_\ell^L L_{\ell-1}^{\frac{1}{2}}\omega_\ell K_{\ell-1}^{\frac{1}{2}} S_\ell^K\right)\right),$$

where $U_\mathrm{T}, S_\ell^K$, and $S_\ell^L$ are the same as in (41).

This greedy Kähler mean is only an approximation to the generalized Kähler mean since by assuming the previous blocks to be fixed, the search space during the optimization of the current block is more restricted than in the general case. The approximation does allow us to partially return to the situation of separated optimization, since the optimization is performed separately on the blocks, even though they have to be computed sequentially.

When we consider the properties of the generalized Kähler mean, an intuitive approach is to start from the properties of the Kähler mean for Toeplitz matrices .

The generalized Kähler mean of $PD\ BT$ matrices and both the global and greedy version of the Kähler mean of $PD\ TBBT$ matrices will be permutation invariant, repetition invariant, and idempotent, since all of them are defined as barycenters.

As for the property of joint homogeneity, we start by discussing the change of transformation (26) when a $PD\ (TB)BT$ matrix $\tilde{R}_n$ is replaced with $\alpha\tilde{R}_n$ for any real $\alpha > 0$. We denote the transformation of $\tilde{R}_n$ by $(P_0, \Omega_1, \dots, \Omega_{n-1})$ with corresponding prediction matrices $A_j^\ell$ and auxiliary matrices $P_{\ell-1}$ and $\Delta_\ell$, and that of $\alpha R_n$ by $(P_0', \Omega_1', \Omega_{n-1}')$ $P_{\ell-1}'$ and $\Delta_\ell'$.

First, the change of the prediction matrices $A_j^{\ell'}$, and auxiliary matrices $P_{\ell-1}'$ and $\Delta_\ell', \ell = 1, \dots, n-1, j = 1, \dots, \ell$, can be found using induction. Considering (19)(22), it is clear to see that $A_1^{1'} = A_1^1, P_0' = \alpha P_0$, and $\Delta_1' = \alpha\Delta_1$. Now assuming $\widetilde{A_{\ell-1}'} = \tilde{A}_{\ell-1}$, we find $P_{\ell-1}' = \alpha P_{\ell-1}, \Delta_\ell' = \alpha\Delta_\ell$, and $A_\ell^{\ell'} = A_\ell^\ell$. As a consequence of $A_{\ell-1}' = \tilde{A}_{\ell-1}$, we find $P_{\ell-1}' = \alpha P_{\ell-1}, \Delta_\ell' = \alpha_\theta(22), \widetilde{A_\ell'} = \tilde{A}_\ell$, which closes the induction.

By writing the coordinate matrices $\Omega_\ell'$ in the form $\overline{P_{\ell-1}'^{-1/2}}\Delta_\ell' P_{\ell-1}'^{-1/2}$, we now find that $\Omega_\ell' = \Omega_\ell, \ell = 1, \dots, n-1$. Summarized, the transformation of $\alpha\tilde{R}_n$ is given by $(\alpha P_0, \Omega_1, \dots, \Omega_{n-1})$, which is consistent with the Kähler transformation of $PD$ Toeplitz matrices. Note that transformation (24) behaves in the same way for positive scaling.

Now, as for joint homogeneity, suppose we have $PD\ (TB)BT$ matrices $\tilde{T}_i, i = 1, \dots, k$, with a corresponding transformation $\left(P_{0,i}, \Omega_{1,i}, \dots, \Omega_{n-1,i}\right)$, and $k$ positive scalars $\alpha_i$. The generalized Kähler mean for $PD\ BT$ matrices is computed separately on the coordinate matrices. Combining this with the joint homogeneity of the geometric mean of $PD$ matrices (specifically, the Karcher mean) [2], [36] is sufficient to prove the property in this case.

The global version of the Kähler mean for $PD\ TBBT$ matrices can be seen to satisfy the property by studying the gradient of the cost function. If this gradient becomes the zero matrix for some matrix $\tilde{R}_n$ with given matrices $\tilde{T}_i$, $i = 1, \dots, k$, it can be checked matrices $\alpha_i\tilde{T}_i, i = 1, \dots, k$.

Moreover, the greedy approximation also satisfies the property, which can be seen as follows. We will denote the transformation of the greedy Kähler mean of the unscaled $\tilde{T}_1, \dots, \tilde{T}_k$ by $(P_0, \Omega_1, \dots, \Omega_{n-1})$. The greedy Kähler mean of the scaled matrices

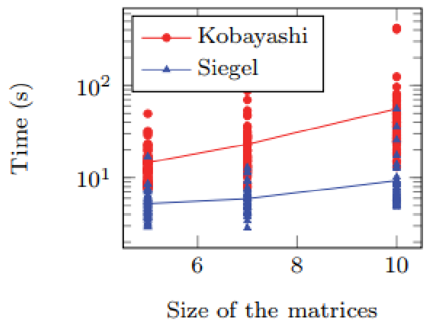
$\alpha_1\tilde{T}_1, \dots, \alpha_k\tilde{T}_k$ now starts by averaging the first coordinate matrices, resulting in $\mathcal{B}_{\mathcal{T}_N^+}(\alpha_1 P_{0,1}, \dots, \alpha_k P_{0,1}) = (\alpha_1 \cdots \alpha_k)^{1/k} P_0$ because of the joint homogeneity of the structured geometric mean for linear structures [20]. As mentioned before, the search space for the coordinates of this greedy mean is dependent on the ones that have already been computed. Hence, for the next coefficients $(\Omega_{1,1}, \dots, \Omega_{1,k})$ we still minimize the cost function $f_{\mathcal{B}_{\mathcal{SD}_N}}(X; \Omega_{1,1}, \dots, \Omega_{1,k})$. However, the search space has changed from

$$\overline{P_0^{-1/2}}\mathcal{T}_N P_0^{-1/2} \cap \mathcal{SD}_N \text{ to } (\alpha_1 \cdots \alpha_k)^{-1/k}\overline{P_0^{-1/2}}\mathcal{T}_N P_0^{-1/2} \cap \mathcal{SD}_N$$

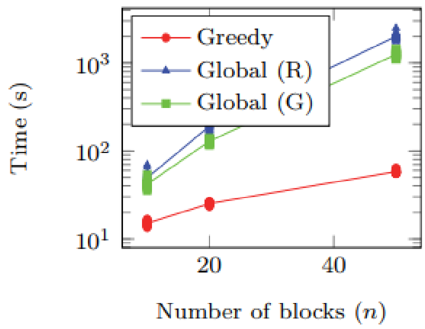
from which it can be seen that the resulting coordinate matrix $\Omega_1$ remains the same as in the unscaled setting (since a scaling of vector space $\mathcal{T}_N$ does not change the space). The other coordinate matrices $\Omega_\ell, \ell = 2, \dots, n-1$, similarly do not change.
Finally, the greedy Kähler mean of the scaled matrices is obtained with coordinate matrices $((\alpha_1 \cdots \alpha_k)^{1/k} P_0, \Omega_1, \dots, \Omega_{n-1})$, which corresponds to the correct matrix for joint homogeneity to hold.
As for the Kähler mean of $PD$ Toeplitz matrices, it is not difficult to find examples which contradict the property of monotonicity. In fact, any counterexample found for the Kähler mean of $PD$ Toeplitz matrices can again be used to contradict the property, since this mean arises as a special example of the generalized Kähler mean for blocks of size 1 .

We will analyze the various algorithms that were discussed for the generalized Kähler mean.
First of all, we will have a closer look at the Siegel disk and compare the barycenters that arise when using the Siegel distance measure $d_{\mathcal{SD}_N}$ and the Kobayashi distance measure $d_K$. Afterwards, a comparison of the global and greedy version of the generalized Kähler mean for $PD$ $TBBT$ matrices is presented, where we also combine the methods by using the greedy version as an initial guess for the global mean algorithm.

We have endowed the Siegel disk $\mathcal{SD}_N$ with the Siegel distance measure $d_{\mathcal{SD}_N}$and with the Kobayashi distance measure $d_K$. Since each distance measure




(a) Required time for the computation of the Kobayashi and Siegel barycenters $\mathcal{B}_K$ and $\mathcal{B}_{\mathcal{SD}_N}$ of 50 matrices of varying sizes.




(b) Required time for the greedy and global versions of the generalized Kähler mean for 20 PD TBBT matrices as the number of blocks varies ($n$ by $n$ blocks). The global algorithm is initiated by a random matrix (R) or the greedy approximation (G). For initialization with the greedy mean, the combined computational time of the greedy and global mean is shown.

FIG. 1. *Computational time of the barycenters and approximations.*

can be used to define a barycenter ($\mathcal{B}_{\mathcal{SD}_N}$ and $\mathcal{B}_K$, respectively) on the Siegel disk, we compare the computational time and results of both.

When we investigate the distance between the barycenters, a relative distance of the order $\mathcal{O}(10^{-1})$ can be found consistently for varying matrix sizes. Note that the diameter of the Siegel disk becomes infinite for both distance measures.
As for computational time, we display some results of both barycenters for varying sizes of matrices in Figure 1 $(a)$. The Siegel barycenter $\mathcal{B}_{\mathcal{SD}_N}$ requires less computational time, and this required time also increases more slowly as the matrices grow.
Perhaps even more interesting is the fact that when we further increase the size of the matrices, the steepest descent method to compute the Kobayashi barycenter starts exhibiting convergence problems and a lack of a unique minimizer. These problems can be ascribed to the presence of the spectral norm in the Kobayashi distance measure. This norm is given by the largest singular value of a matrix, and its derivative is only well-defined when this largest value is strictly greater than the other singular values [219]. During the computation of the barycenter $\mathcal{B}_K$, it is possible that a matrix with two or more almost equal largest singular values is entered into this derivative, causing convergence problems. Furthermore, the derivative of the spectral norm can only contribute a rank one matrix to the gradient of the barycenter cost function for each given matrix in the barycenter. Consequently, this will start causing problems when the number of matrices in the barycenter becomes too small compared to the size of the matrices.

We have suggested a steepest descent algorithm for the generalized Kähler mean of $PD\ TBBT$ matrices, followed by a greedy approximation. Here we analyze how close this approximation is to the actual mean and we investigate the computational advantage of the approximation.
First of all, in terms of computational time the greedy version has a clear advantage over the global mean, as illustrated in Figure 1( b). This was expected, since the gradient for the greedy optimization problem can be found in the gradient of

TABLE 1

*Some averaged comparative values concerning the global and greedy version of the generalized Kähler mean of 20 PD TBBT matrices. The global algorithm is initiated by a random matrix (R), one of the original matrices in the mean (O), or the greedy approximation (G).*

| Number of blocks $n$ ($n$ by $n$ blocks) | 10 | 20 | 50 |
|---|---|---|---|
| Iterations for global (R) | 24 | 24 | 23 |
| Iterations for global (O) | 25 | 23 | 23 |
| Iterations for global (G) | 13 | 12 | 13 |
| Relative distance greedy versus global | 2.28e-04 | 1.36e-04 | 8.24e-05 |
| Size global gradient at greedy | 2.34 | 2.44 | 3.14 |

the global optimization problem $(38) - (39)$ by setting the factors $G_{\ell,i}$ (for the first component) and $W_{\ell,i}^{(q)}$ (for the other components) to zero.
In fact, while the basic operations for the terms in the individual blocks of the gradient depend on the size of the matrices $(N)$, the number of terms in each block in the global gradient is dependent on the block size $(n)$ of the matrix. For the gradient in the greedy algorithm, changing the block size of the matrices from $n$ to $n+1$ corresponds to computing one additional block in the gradient, independent of all previous blocks. On the other hand, the gradient in the global algorithm will gain an additional term in each of the previous blocks of the gradient. Hence, the greedy algorithm is linearly dependent on the number of blocks $n$ in the matrices, while for the global algorithm this dependence is quadratic.

Moreover, in Table 1, the (averaged) relative distance between the global version of the generalized Kähler mean and its greedy approximation is shown for a number of block sizes. The observed relative proximity between both versions and the computational advantage of the greedy algorithm suggests that it could work well as an approximation. In fact, many applications require only a limited amount of significant digits, in which case the greedy approximation can replace the actual mean.
The greedy approximation as initial guess for the global algorithm. Next, for those applications where the global version of the generalized Kähler mean is required, we analyze the influence of the initial guess on the algorithm. Specifically, the appropriateness of the greedy version as an initial guess is investigated.
In Figure 1( b), the computational time of the global version of the mean is displayed when we use a random initial guess and the greedy mean. As can be seen, using the greedy approximation results in a faster algorithm. Note that the time to compute the greedy mean was included in these results. Table 1 also displays the advantage of the greedy initial guess, as the required number iterations of the global algorithm are reduced by half. Hence, we can conclude that the greedy approximation works well as an initializer to the global algorithm.

We have focused on a geometry for $PD$ Toeplitz matrices and a generalization thereof towards $PD$ $(TB)BT$ matrices.
In the case of Toeplitz matrices, the Kähler mean and its properties have been investigated. While this mean did not satisfy many properties relating to the ordering of matrices, such as monotonicity, it does cooperate well with the application from which it was derived [197], [23], [202], [240].
Afterwards, two possible generalizations of the Kähler transformation towards $PD$ $(TB)BT$ matrices were presented, of which the second was discussed in further detail.
Two possible geometries on the Siegel disk were investigated, where one corresponded naturally with the manifold and the other was based on a useful automorphism of the set. For $TBBT$ matrices, a global mean and a greedy approximation were derived, which were compared in numerical experiments. The greedy version of the generalized mean was a close approximation to the global mean, with a significantly lower computational cost. The greedy approximation was also shown to work well as an initializer for the global optimization algorithm, effectively reducing the number of iterations by half.

**Section (4.3): Toeplitz and Toeplitz-Block Block-Toeplitz Matrices**

The notion of geometric mean of scalars has been extended to positive definite matrices by many [2], [5], [6], [1], [36], [17]. Since the straightforward generalization $(A_1 \dots A_n)^{\frac{1}{n}}$ does not satisfy many desired properties, even it is not invariant under permutation, new definitions of geometric mean for matrices have been developed, $e.g.$ $ALM, NBMP, CHEAP,$ and Karcher mean, see [2], [3], [5], [6],[1], [17]. Because of the widespread applications of geometric mean in the many areas such as radar detection [40], [48], image processing [195], elasticity tensor analysis [10] and medical imaging [243], many researchers involved in this field. Ando, $Li,$ and Mathias [2], have suggested ten important properties, so-called $ALM$ properties, that any geometric mean should satisfy them. But most of the definitions in the literature, do not satisfy all of the $ALM$ properties, especially, they do not preserve the monotonicity. By the monotonicity property, we mean that $G(A_1, B_1) \leq G(A_2, B_2)$ whenever $A_1 \leq A_2$ and $B_1 \leq B_2$ where $A_1, A_2, B_1$ and $B_2$ are positive definite matrices and the operator $G$ stands for a given definition of 'geometric mean'.

Among the existing definitions, the Karcher mean satisfies all of the $ALM$ properties but does not preserve the structure of the matrices, see [191]. In many applications, such as designing of some radar systems, the used matrices are Toeplitz matrices [23]. A Toeplitz matrix is a matrix in which entries along their diagonals are constant, $i.e.$,

$$\begin{bmatrix} a_0 & a_1 & \cdots & & a_{n-1} \\ a_{-1} & a_0 & a_1 & \cdots & \\ \vdots & & \ddots & & \vdots \\ & & \ddots & & a_1 \\ a_{-n+1} & & \cdots & a_{-1} & a_0 \end{bmatrix}.$$

We are interested in finding a definition of a geometric mean of Toeplitz matrices that itself is Toeplitz too. Furthermore, it should satisfy some $ALM$ properties, especially the monotonicity property. Using a Riemannian structure on the manifold of all positive Hermitian $n \times n$ matrices, $D. A.$ Bini $et\ al.$ [191] have introduced a (not necessarily unique) structured geometric mean which preserves the structure of Toeplitz matrices and satisfies a few $ALM$ properties but fails to satisfy some other, especially the monotonicity property. Also, the Kähler metric mean [22],[23] preserves the structure of Toeplitz matrices but does not satisfy the monotonicity property either. Moreover, the computations are somewhat complicated and costly.

[189], introduced a geometric mean for positive semi-definite Toeplitz matrices with non-negative symbols which preserves the Toeplitz structure and many $ALM$ properties, especially monotonicity in the sense of the order (46). Recall that every Toeplitz matrix has a unique symbol function, see [194].

Now, let $a$ and $b$ are the symbol functions of Toeplitz matrices $A$ and $B$ respectively.

We say $A \prec B$ if $a \leq b$. This order is slightly more stronger than the usual ordering on matrices; see (46) and the explanations after it. We using the positive parts of the symbol functions, we extend the previous results of [189] to all positive semi-definite Toeplitz matrices, see Definition (4.3.2), Theorem (4.3.3).

The approach is based on the concept of the symbol function and its Fourier expansion. Moreover, we consider block-Toeplitz matrices with Toeplitz structured blocks, or briefly $TBBT$ matrices, $i.e.$,

$$\begin{bmatrix} A_0 & A_1 & \cdots & & A_{n-1} \\ A_{-1} & A_0 & A_1 & \cdots & \\ \vdots & & \ddots & & \vdots \\ & & \ddots & & A_1 \\ A_{-n+1} & & \cdots & A_{-1} & A_0 \end{bmatrix},$$

where

$$A_j = \begin{bmatrix} a_{j,0} & a_{j,1} & \cdots & & a_{j,N-1} \\ a_{j,-1} & a_{j,0} & a_{j,1} & \cdots & \\ \vdots & & \ddots & & \vdots \\ & & \ddots & & a_{j,1} \\ a_{j,-N+1} & & \cdots & a_{j,-1} & a_{j,0} \end{bmatrix},$$

for all $j = 0, \pm 1, \dots, \pm(n-1)$, which appear in signal processing and some other fields, see [244], Chapter 8 .

Ben Jeuris $et\,al.$ in [196], have generalized the Kähler metric mean to $TBBT$ matrices.

Their proposed definition is a mean for a specific measure which preserves the structure of the matrices but fails to satisfy the monotonicity property and its computation is somewhat complicated. We initiate the two dimensional symbol functions for $TBBT$ matrices and using two dimensional Fourier series, we introduce a definition of geometric mean for all positive semi-definite $TBBT$ matrices, Definition (4.3.4). Our proposed definition preserves the $TBBT$ structure, it satisfies the monotonicity property with respect to the order $(46)$, it is very simple to calculate and has a low cost.

The geometric mean of the positive semi-definite Toeplitz matrices and the $TBBT$ matrices are discussed, respectively.

Let $\mathbb{T} = \{e^{i\theta} \mid \theta \in \mathbb{R}\}$ be the unit circle in the complex plane. Suppose $f: \mathbb{T} \to \mathbb{C}$ be a Lebesgue measurable function and for each $1 \le p < +\infty$, let

$$\| f \|_p = \left( \frac{1}{2\pi} \int_0^{2\pi} \left| f(e^{i\theta}) \right|^p d\theta \right)^{\frac{1}{p}}$$

and

$$\| f \|_\infty = \text{ess s } \left\{ \left| f(e^{i\theta}) \right| \mid 0 \le \theta \le 2\pi \right\}.$$

As usual, we denote the set of all Lebesgue measurable functions $f: \mathbb{T} \to \mathbb{C}$ with $\| f \|_p < +\infty$ by $L^p(\mathbb{T})$ for $1 \le p \le +\infty$. It is known that $L^p(\mathbb{T})$ equipped with the norm $\|\cdot\|_p$ is a Banach space, $L^r(\mathbb{T}) \subset L^s(\mathbb{T})$ whenever $1 \le s \le r \le +\infty$, and, the space $L^2(\mathbb{T})$ is a Hilbert space with the following inner product

$$\langle f, g \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(e^{i\theta}) \overline{g(e^{i\theta})} d\theta, \ \left(f, g \in L^2(\mathbb{T})\right).$$

The Fourier coefficients of any $f \in L^1(\mathbb{T})$ are given by $\{f_n\}_{n \in \mathbb{Z}}$ where

$$f_n = \frac{1}{2\pi} \int_0^{2\pi} f(e^{i\theta}) e^{-in\theta} d\theta, \ (n \in \mathbb{Z}).$$

Notice that we can identify the interval $[0,2\pi]$ and the unit circle $T$ via the so-called exponential map $\theta \mapsto e^{i\theta}$. Let $a \in L^\infty(\mathbb{T})$ be given, the operator $M(a): L^2(\mathbb{T}) \to L^2(\mathbb{T})$ defined by $M(a)f = af$ for all $f \in L^2(\mathbb{T})$ is called the multiplication operator on $L^2(\mathbb{T})$ generated by the function $a \in L^\infty(\mathbb{T})$, $a$ is called the symbol function of $M(a)$. It is easy to see that $\| M(a) \|_{op} = \| a \|_\infty$, where

$$\| M(a) \|_{op} = sup\{ \| M(a)f \|_2 \mid f \in L^2(\mathbb{T}), \| f \|_2 = 1 \}$$

is the operator norm; for proofs and more details see [192].

The functions $\chi_n$ defined by $\chi_n(t) = t^n, (t \in \mathbb{T}, n \in \mathbb{Z})$ are an orthogonal basis for $L^2(\mathbb{T})$, and $\langle M(a)\chi_j, \chi_k \rangle = a_{k-j}$ is the $(k-j)-th$ Fourier coefficient of $a$. We can represent the operator $M(a)$ with an infinite dimensional matrix such that its entries are Fourier coefficients of $a$,

$$M(a) = \begin{bmatrix} \ddots & \ddots & & & & & \\ \ddots & a_0 & a_{-1} & & & & \\ & a_1 & a_0 & a_{-1} & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & a_1 & a_0 & a_{-1} & \\ & & & & \ddots & \ddots & \ddots \end{bmatrix}.$$

Let $V_n$ be the $n$-dimensional vector subspace of $L^2(\mathbb{T})$ generated by the set $\mathcal{B} = \{\chi_0, \cdots, \chi_{n-1}\}$. The inclusion map $\iota_n: V_n \to L^2(\mathbb{T})$ and the orthogonal projection map

$\pi_n: L^2(\mathbb{T}) \to V_n$ are well-defined linear operators and we have $\|\iota_n\| = \|\pi_n\| = 1$. Given $a \in L^\infty(\mathbb{T})$ and its associated operator $M(a)$, we consider the linear transformation $\pi_n \circ M(a) \circ \iota_n: V_n \to V_n$. The matrix $T_n(a)$ which represents the operator $\pi_n \circ M(a) \circ \iota_n$ in the ordered basis $\mathcal{B}$ is

$$T_n(a) = \begin{bmatrix} a_0 & a_1 & \dots & & a_{n-1} \\ a_{-1} & a_0 & a_1 & \dots & \\ \vdots & & \ddots & & \vdots \\ & & \ddots & & a_1 \\ a_{-n+1} & & \cdots & a_{-1} & a_0 \end{bmatrix}. \quad (42)$$

Recall that a Toeplitz matrix $A = [a_{ij}]$ is a matrix in which the entries along its diagonals are constant, $i.e.$, $a_{ij} = a_{pq}$ whenever $i - j = p - q$. Hence, the matrix $T_n(a)$ is an $n \times n$ Toeplitz matrix for each $a \in L^\infty(\mathbb{T})$.

On the other hand, let

$$A = \begin{bmatrix} a_0 & a_1 & \cdots & & a_{n-1} \\ a_{-1} & a_0 & a_1 & \cdots & \\ \vdots & & \ddots & & \vdots \\ & & \ddots & & a_1 \\ a_{-n+1} & & \cdots & a_{-1} & a_0 \end{bmatrix}$$

be a Toeplitz matrix. The symbol of $A$ is a function $a \in L^\infty(\mathbb{T})$ defined by

$$\begin{aligned} a(e^{i\theta}) &= \sum_{k=-n+1}^{n-1} a_k e^{ik\theta}, \ (\theta \in \mathbb{R}) \\ &= \sum_{k=-n+1}^{n-1} a_k \chi_k(t), \ (t \in \mathbb{T}), \end{aligned} \quad (43)$$

where $\{\chi_k\}_{k\in\mathbb{Z}}$ is the standard basis of $L^2(\mathbb{T})$, and $\chi_k(t) = t^k, (t \in \mathbb{T}, k \in \mathbb{Z})$; see [192], [194].

Notice that, by (42), we have that $T_n(a) = A$ for $a$ in (43).

As in [189], we use the following notations

$\mathcal{T}_n =$ The set of all $n \times n$ Toeplitz matrices,

$\mathcal{T}_n^+ = \{A \in \mathcal{T}_n \mid A$ is positive semi-definite $\}$, (44)

$\mathcal{T}_n^{++} = \{A \in \mathcal{T}_n \mid$ the symbol of $A$ is a non-negative function $\}$.

The relation between these sets is clarified in the following lemma.

**Lemma (4.3.1)[241]:** ([189] Lemma 3.2) If $a \in L^\infty(\mathbb{T})$ and $a \geq 0$ $a.e.$ then the matrix $T_n(a)$ for each $n \in \mathbb{N}$, is a positive semi-definite $n \times n$ Toeplitz matrix but the converse is not true.

So, we have $\mathcal{T}_n^{++} \subset \mathcal{T}_n^+ \subset \mathcal{T}_n$. Notice that the symbol function of any $A \in \mathcal{T}_n^+$ is real-valued. In fact, if $A = [a_{ij}] \in \mathcal{T}_n^+$, then $a_{ji} = \overline{a_{ij}}, a_0 \geq 0$, so

$$
\begin{aligned}
a(e^{i\theta}) &= \sum_{k=-n+1}^{n-1} a_k e^{ik\theta} \\
&= a_0 + \sum_{k=1}^{n-1} \left(a_k e^{ik\theta} + \overline{a_k} e^{-ik\theta}\right) \qquad (45) \\
&= a_0 + 2\sum_{k=1}^{n-1} \mathrm{Re}\left(a_k e^{ik\theta}\right),\ (\theta \in \mathbb{R}).
\end{aligned}
$$

We consider the cone $\mathcal{T}_n^+$of all positive semi-definite $n \times n$ Toeplitz matrices. For this aim, we will use the positive part of the symbol function. Recall that for each real valued function $f$, its positive part function is defined by

$$f^+ = max\{f, 0\}.$$

Also, we consider the following order on $\mathcal{T}_n^+$. We say $A \prec B$ for $A, B \in \mathcal{T}_n^+$if and only if $B - A \in \mathcal{T}_n^{++}$.Recall the usual ordering between two matrices $A$ and $B, i.e. A \leq B$ if and only if $B - A \in \mathcal{T}_n^+$. In the other words, if $a$ and $b$ are symbols of $A$ and $B$ respectively, then

$$A \prec B \Leftrightarrow a \leq b. \qquad (46)$$

This means that we say $B$ is greater than $A$, whenever the symbol function of $B$ is greater than or equal to the symbol function of $A$. Since $\mathcal{T}_n^{++} \subset \mathcal{T}_n^+$, so $A \prec B$ implies $A \leq B$, but the converse does not hold [189]. For example, let $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}$, then $A, B \in \mathcal{T}_n^+$and $B - A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix} \in \mathcal{T}_n^+ \setminus \mathcal{T}_n^{++}$(notice that the symbol function of this matrix is $f(e^{i\theta}) = 3 + 4\cos\theta$ which is not a positive function), so $A \leq B$ but $A \nprec B$.

Now we can introduce a definition of a geometric mean on $\mathcal{T}_n^+$as follows.

**Definition (4.3.2)[241]:** Given $m, n \geq 1$ and $A_1, A_2, \cdots, A_m \in \mathcal{T}_n^+$and let $a_1, \cdots, a_m \in L^\infty$ be their symbols, respectively. Then the geometric mean of $A_1, \cdots, A_m$ is defined by the Toeplitz matrix

$$G(A_1, \cdots, A_m) = T_n\left(\sqrt[m]{a_1^+ \cdots a_m^+}\right).$$

First we calculate the usual geometric mean of the positive parts of the symbol functions of $A_1, \cdots, A_m$; then using (42), the geometric mean of these matrices is obtained.

We reduce the geometric mean of the matrices to the geometric mean of (positive part of) their symbol functions via the Fourier transform. Hence, it does not matter that the matrices are singular or not. Also, the singularity does not affect the computation costs. As a simple example, let

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

and

$$B = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}.$$

Notice that $A, B \in \mathcal{T}_3^+$are singular matrices and their symbol functions are

$$a(\theta) = 1 + 2\cos(\theta) + 2\cos(2\theta)$$

and

$$b(\theta) = 2 - 2\cos(\theta) - 2\cos(2\theta)\ \ (\theta \in [0, 2\pi]),$$

respectively. Their geometric mean is $G(A,B) = T_3\left(\sqrt{a(\theta)^+ b(\theta)^+}\right)$.Using Simpson's rule for numerical integration with $k = 32$ subdivisions of the interval $[0,2\pi]$, we will have

$$G(A,B) = \begin{bmatrix} 0.4421 & -0.0988 & 0.1218 \\ -0.0988 & 0.4421 & -0.0988 \\ 0.1218 & -0.0988 & 0.4421 \end{bmatrix},$$

which is a positive semi-definite Toeplitz matrix.

**Theorem (4.3.3)[241]:** ($Cf.$[189] Theorem 3.7) Let $m \geq 1, A_1, \cdots, A_m, B_1, \cdots B_m \in \mathcal{T}_n^+$ and $\alpha_1, \cdots, \alpha_m \in \mathbb{R}^+$. The following properties hold:

i) Joint homogeneity. $G(\alpha_1 A_1, \cdots, \alpha_m A_m) = (\alpha_1 \cdots \alpha_m)^{\frac{1}{m}} G(A_1, \cdots, A_m)$.

ii) Permutation invariance. For any permutation $\pi$ of $\{1, \cdots, m\}$ we have

$$G(A_1, \cdots, A_m) = G\left(A_{\pi(1)}, \cdots, A_{\pi(m)}\right).$$

iii) Monotonicity. $G(A_1, \cdots, A_m) \prec G(B_1, \cdots, B_m)$ whenever $A_1 \prec B_1, \cdots, A_m \prec B_m$.

iv) Continuity. If $\{A_{k,1}\}_{k\geq 1}, \cdots, \{A_{k,m}\}_{k\geq 1}$ are monotonic decreasing sequences in ordering (46), converging to $A_1, \cdots, A_m$ respectively then the sequence $\left\{G\left(A_{k,1}, \cdots,, A_{k,m}\right)\right\}_{k\geq 1}$ converges to $G(A_1, \cdots, A_m)$.

**Proof.** i) This follows from the Definition (4.3.2), the fact that $(\alpha a)^+ = \alpha a^+$when $\alpha \geq 0$ and $a \in L^\infty(\mathbb{T})$ and the linearity of the mapping $T_n$, see [19] Corollary 3.4.

ii) By definition, this is trivial.

iii) Let $a_1, \cdots, a_m$ and $b_1, \cdots, b_m$ be the symbols of $A_1, \cdots, A_m$ and $B_1, \cdots, B_m$, respectively. By assumption $a_1 \leq b_1, \cdots, a_m \leq b_m$, so $a_1^+ \leq b_1^+, \cdots, a_m^+ \leq b_m^+$.Therefore

$$(a_1^+ \cdots a_m^+)^{\frac{1}{m}} \leq (b_1^+ \cdots b_m^+)\frac{1}{m}$$

by the Definition (4.3.2) and (46) the conclusion follows.

iv) We have

$$A_{k,1} = T_n\left(a_{k,1}\right), \cdots, A_{k,m} = T_n\left(a_{k,m}\right),$$

and

$$A_1 = T_n(a_1), \cdots, A_m = T_n(a_m).$$

Recall that in every finite dimensional vector space, all norms are equivalent, so without loss of generality, we can only consider $\|.\|_F$, the Frobenius norm. Therefore, we have $\|A_{k,i} - A_i\|_F \to 0$ as $k \to \infty$ for $0 \leq i \leq m$. On the other hand $\|a_{k,i} - a_i\|_2 = \|A_{k,i} - A_i\|_F, 0 \leq i \leq m,$ and since $\|a^+\|_\infty \leq \| a \|_\infty$ for all $a \in L^\infty$, the mapping $a \mapsto a^+$is continuous on $L^\infty$, so

$$\left\|\left(a_{k,1}^+ \cdots a_{k,m}^+\right)\frac{1}{m} - (a_1^+ \cdots a_m^+)^{\frac{1}{m}}\right\|_2 \to 0 \text{ as } k \to \infty.$$

Again, by $\| a \|_2 = \|T_n(a)\|_F$, for each $a \in L^\infty$, we have

$$\left\|G\left(A_{k,1}, \cdots, A_{k,m}\right) - G(A_1, \cdots, A_m)\right\|_F \longrightarrow 0 \text{ as } k \longrightarrow \infty.$$

Here, we introduce the notion of block-Toeplitz matrices with Toeplitz structured blocks ($TBBT$), following [196]. Fix some $n, N \in \mathbb{N}$. Let $\mathcal{T}_{n,N}$ be the vector space of all blockToeplitz matrices of $n$ by $n$ blocks in which each block is an $N \times N$ Toeplitz matrix. Formally, $\mathcal{T}_{n,N}$ includes all the following matrices

$$A=\begin{bmatrix} A_0 & A_1 & \cdots & & A_{n-1} \\ A_{-1} & A_0 & A_1 & \cdots & \\ \vdots & & \ddots & & \vdots \\ & & \ddots & & A_1 \\ A_{-n+1} & & \cdots & A_{-1} & A_0 \end{bmatrix}, \quad (47)$$

where each

$$A_j=\begin{bmatrix} a_{j,0} & a_{j,1} & \cdots & & a_{j,N-1} \\ a_{j,-1} & a_{j,0} & a_{j,1} & \cdots & \\ \vdots & & \ddots & & \vdots \\ & & \ddots & & a_{j,1} \\ a_{j,-N+1} & & \cdots & a_{j,-1} & a_{j,0} \end{bmatrix} \quad (48)$$

is an $N \times N$ Toeplitz matrix, for all $j = 0, \pm 1, \dots, \pm(n-1)$. We denote by $\mathcal{T}_{n,N}^{+}$all the positive semi-definite matrices in $\mathcal{T}_{n,N}$. Notice that $B$. Jeuris and R. Vandebril [196] denote by $\mathcal{T}_{n,N}^{+}$all the positive definite matrices in $\mathcal{T}_{n,N}$. But since our approach makes sense for semi-definite matrices as well, we consider a slightly more general case.

Along the lines of the procedure that was developed and in [189] for the Toeplitz matrices, here, we are going to introduce the concept of a symbol function for $TBBT$ matrices. For $A \in \mathcal{T}_{n,N}^{+}$using the notations (47) and (48), let $a \in L^{\infty}(\mathbb{T}^2)$ by definition be

$$a(t,s) = \sum_{|j|<n} a_j(s)\chi_j(t), \ (t,s \in \mathbb{T})$$

where $a_j \in L^{\infty}(\mathrm{T})$ is the symbol function associated to $A_j \in \mathcal{T}_N$ for every $j = 0, \pm 1, \dots, \pm(n-1)$, see (43). Thus

$$a(t,s) = \sum_{|j|<n} \sum_{|k|<N} a_{j,k}\chi_k(s)\chi_j(t), \ (t,s \in \mathbb{T})$$

or

$$a\left(e^{i\phi}, e^{i\theta}\right) = \sum_{j=-n+1}^{n-1} \sum_{k=-N+1}^{N+1} a_{j,k} e^{ik\phi} e^{ij\theta}, \ (\phi, \theta \in \mathbb{R}). \quad (49)$$

Obviously, this definition is a generalization of (2.2).

On the other hand, let $a \in L^{\infty}(\mathbb{T}^2) \subset L^2(\mathbb{T}^2)$ be an arbitrary bounded Lebesgue measurable function, it is known that ($e.g.$, see [242] )

$$a(t,s) = \sum_{j,k \in \mathbb{Z}} a_{j,k}\chi_k(s)\chi_j(t), \ (t,s \in \mathbb{T})$$

Or

$$a\left(e^{i\phi}, e^{i\theta}\right) = \sum_{j,k \in \mathbb{Z}} a_{j,k} e^{ik\phi} e^{ij\theta}, \ (\phi, \theta \in \mathbb{R})$$

where

$$a_{j,k} = \frac{1}{4\pi^2} \int_0^{2\pi} \int_0^{2\pi} a\left(e^{i\phi}, e^{i\theta}\right) e^{-ik\phi} e^{-ij\theta} d\phi d\theta, \ (j,k \in \mathbb{Z})$$

are the Fourier coefficients of $a$. Now, one can define $T_{n,N}(a) \in \mathcal{T}_{n,N}$ as in (47) and (48) where $a_{j,k}\,'s$ are the Fourier coefficients of $a$. It is easy to check that if $a$ is the symbol function of a matrix $A \in \mathcal{T}_{n,N}$ then $T_{n,N}(a) = A$.

Similar to (45), it is easy to see that the symbol function of a positive semi-definite $TBBT$ matrix is real-valued. Let $A, B \in \mathcal{T}_{n,N}^{+}$and $a, b \in L^{\infty}(\mathbb{T}^2)$ be their associated symbol functions, respectively. We say that $A \prec B$ if $a \leq b$, as in (46). Obviously, this is a partially order on the set $\mathcal{T}_{n,N}^{+}$.

Now, we can define a geometric mean in $\mathcal{T}_{n,N}^{+}$which preserves the structure of $TBBT$ matrices by its definition.

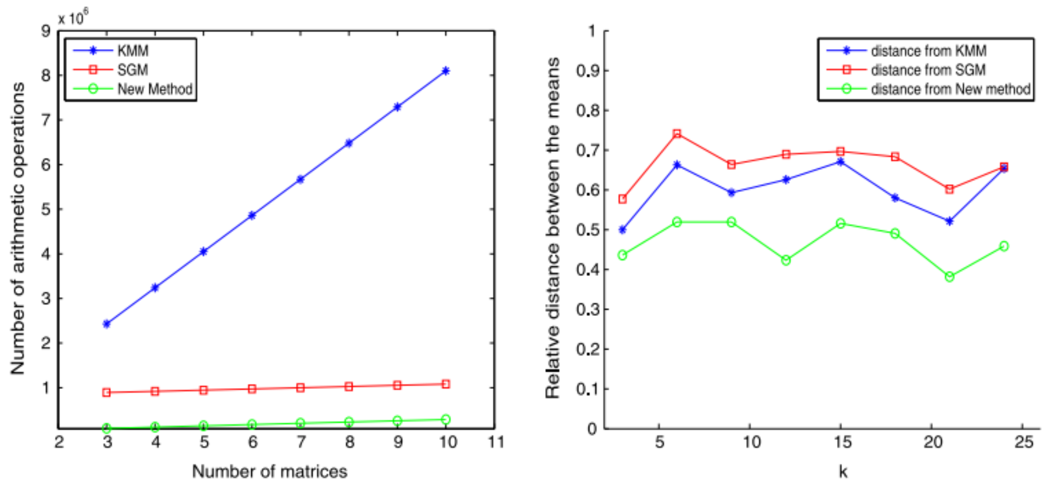
**Definition (4.3.4)[241]:** Given $m, n, N \geq 1$ and $A_1, A_2, \cdots, A_m \in \mathcal{T}_{n,N}^{+}$and let $a_1, \cdots, a_m \in L^{\infty}$ be their symbols, respectively. The geometric mean of $A_1, \cdots, A_m$ is defined by the $TBBT$ matrix

$$G(A_1, \cdots, A_m) = T_{n,N}\left((a_1^{+} \cdots a_m^{+})\frac{1}{m}\right).$$

One can prove the joint homogeneity, permutation invariance, monotonicity and continuity properties of this definition of geometric mean on $\mathcal{T}_{n,N}^{+}$quite similar to Theorem (4.3.3). In fact, the same proofs remain valid.

Consider the set of all $n \times n$ positive semi-definite Toeplitz matrices $\mathcal{T}_n^{+}$. Again, similar to [189], the number of operations equals $O(\mathrm{kmn}^2)$ where $n$ is the size of the matrices, $m$ is the number of matrices and $k$ is the number of subdivisions of the interval $[0,2\pi]$ in order to use the Simpson's rule for numerical integration. Recall that the cost of the structured geometric mean ($SGM$) in [191] for the Toeplitz matrices is equal to $O(pn^4 + pmn^3)$, where $m, n$ are the same as before and $p$ is the number of iterations in the algorithm. The Kähler metric mean in the case real Toeplitz matrices as described in the Section 5 of [191], requires $O(mn^4)$ arithmetic operations where $m, n$ are the same as before, again. In order to compare the cost of the different methods we let $n = 30, m = 3$ to $10$ , for $SGM, p = 1$ and for our method $k = 32$. The results are given in Fig. $1(a)$. Again, as this figure shows the new purposed method costs considerably less than others.

On the other hand, as mentioned in [36], because of the lack of a reference solution, the accuracy of this method is harder to verify. Here, we investigate how close our new geometric mean of five Toeplitz matrices $\mathcal{A} = \{A_1, \dots, A_5\} \subset \mathcal{T}_n^{+}$(see Definition (4.3.2)), how close to the usual geometric mean $G_0 = \sqrt[5]{A_1 \dots A_5}$; but notice that $G_0$ may not be Toeplitz.




(a) Required arithmetic operations (b) Average relative errors

**Fig. 1.** Comparison of the different methods.

For this aim, choose a subset $\mathcal{A} = \{A_1, \dots, A_5\} \subset \mathcal{T}_n^+$ randomly, with unit $(= 1)$ Frobenius norms, i.e. $\|A_1\|_F = \cdots = \|A_5\|_F = 1$, and let $G_{\mathrm{KMM}}$, $G_{\mathrm{SGM}}$ and $G_{\mathrm{new}}$ represent the Kähler metric mean, structured geometric mean and the new geometric mean of $\mathcal{A}$, respectively. One can consider the following relative errors (with Frobenius norm)

$$\mathrm{Re}_{\mathrm{KMM}}(\mathcal{A}) = \frac{\|G_{\mathrm{KMM}} - G_0\|}{\|G_0\|}, \mathrm{ResGM}(\mathcal{A}) = \frac{\|G_{\mathrm{SGM}} - G_0\|}{\|G_0\|}, \mathrm{Re}_{\mathrm{new}}\,(\mathcal{A}) = \frac{\left\|G_{\mathrm{new}} - G_0\right\|}{\|G_0\|}.$$

For $k = 3,6,9,12,15,18,21,24$, we construct $k$ subsets $\mathcal{A}_1, \dots, \mathcal{A}_k$ each of which contains five $5 \times 5$ positive definite Toeplitz matrices with unit Frobenius norm, randomly. In Fig. 1( b), the average relative errors

$$\frac{1}{k}\sum_{j=1}^{k} \mathrm{Re}_{\mathrm{KMM}}(\mathcal{A}_j),\ \frac{1}{k}\sum_{j=1}^{k} \mathrm{Re}_{\mathrm{SGM}}(\mathcal{A}_j),\ \frac{1}{k}\sum_{j=1}^{k} \mathrm{Re}_{\mathrm{new}}\,(\mathcal{A}_j)$$

are shown. The result is somewhat surprising, our proposed geometric mean $G_{\mathrm{new}}$, is even more close to the usual geometric mean $G_0 = \sqrt[5]{A_1 \dots A_5}$ than the other two previous definitions, $i.e.$, $G_{\mathrm{KMM}}$ and $G_{\mathrm{SGM}}$.

Now, consider the set of all the positive semi-definite block-Toeplitz matrices of $n$ by $n$ blocks in which each block is an $N \times N$ Toeplitz matrix, $i.e. \mathcal{T}_{n,N}^+$. Similar to the Toeplitz case, it is easy to show that the number of operations equals $O(mk^2n^3N^3)$ where $m$ is the number of matrices and $k$ is the number of subdivisions of the interval $[0,2\pi]$ in order to use Simpson's rule for numerical integration.

# Chapter 5
# Geometric Significance and Multipliers

We investigate the connection of restricted Grassmann manifolds associated to $p$-Schatten ideals and essentially commuting projections. We examine the multipliers from one model space to another. We study connections of the new order with some of the classical problems and known results. We discuss remaining problems and possible directions for further research.

## Section (5.1): Toeplitz Kernels

For $L^p$ be the usual Lebesgue spaces of complex-valued functions on the unit circle $\mathbb{T}$. The Grassmann manifold of $L^2$ is the set of all closed subspaces of $L^2$. We study the relation between geodesics on the Grassmann manifold of $L^2$ and the injectivity problem for Toeplitz operators.

To explain this relation, let $H^2$ be the Hardy space of the unit circle. Recall that the injectivity problem for Toeplitz operators consists in looking for those symbols $\varphi \in L^\infty$ such that the Toeplitz operator $T_\varphi$ is injective. We relate it to the problem of finding a geodesic on the Grassmann manifold of $L^2$ which joins two subspaces of the form $\varphi H^2$ and $\psi H^2$, where $\varphi, \psi$ are invertible functions in $L^\infty$. We will prove that such a geodesic exists if and only if the Toeplitz operator $T_{\varphi\psi^{-1}}$ and its adjoint both have trivial kernel. Furthermore, we will see that these statements are also equivalent to the existence of a minimizing geodesic joining the given subspaces.

The Grassmann manifold of an abstract Hilbert space (i.e. the set consisting of all the closed subspaces) may be identified with the bounded selfadjoint projections. It is an infinite dimensional homogeneous space which can be endowed with a Finsler metric by using the operator norm on each tangent space. Although it is complete with the corresponding rectifiable distance, there are subspaces in the same connected component that cannot be joined by a geodesic (see e.g. [246] ). This means that the Hopf-Rinow theorem fails for this manifold. Much information of its geodesics and their minimizing properties are known. The first results date back to [266], [258], [269]; both in the more general framework of selfadjoint projections in $C^*$-algebras. There has been progress about the structure of the geodesics in several Grassmann manifolds defined by imposing additional conditions on the subspaces; see [248], [249], [251] for restricted Grassmann manifolds and [250] for the Lagrangian Grassmann manifold.

Taking the Hilbert space $L^2$. This allows us to study the interplay between geodesics, functional spaces and operator theory. In contrast to the invertibility problem for Toeplitz operators, little attention has been paid in the literature to the injectivity problem until recent years. Except for the works of [257], [267], the problem remained untreated until the recent works [72], [49], [96] (see also the survey [264]). Apart from being an interesting problem in operator theory, in these there are relevant applications to harmonic analysis, complex analysis and mathematical physics.

We give classical results on Hardy spaces, Toeplitz and Hankel operators to make the article reasonably self-contained. We prove the aforementioned relation between geodesics of the Grassmann manifold of $L^2$ and the injectivity problem. Then, this result is used to derive an inequality involving the reduced minimum modulus of Toeplitz operators and the norm of a commutator.

We deal with the compact restricted Grassmannian (or Sato Grassmanian). This is a well-known Banach manifold related to KdV equations and loop groups (see [273], [274]). We need to consider the following two uniform subalgebras of $L^\infty$, the continuous functions $C$ and the usual Hardy space $H^\infty$. We show that a subspace $\varphi H^2$ belongs to the compact restricted Grassmannian if and only if $\varphi$ is an invertible function in the Sarason algebra $H^\infty + C$. This is the least nontrivial closed subalgebra lying between $H^\infty$ and $L^\infty$; it has also been studied [254], [261], [271]. The existence of geodesics in the restricted Grassmannian between two subspaces $\varphi H^2$ and $\psi H^2, \varphi, \psi$ invertible functions in $H^\infty + C$, depends only on the index of these functions. We also examine when a subspace $\varphi H^2$ can be written as $\varphi H^2 = gH^2$, where $g$ is a continuous unimodular function. These results can be carried out also in the setting of restricted Grassmannians associated to $p$-Schatten ideals by using the notion of Krein algebras defined in [253].

We focus on shift-invariant subspaces of $H^2$. Each shift-invaritant subspace can be expressed as $\varphi H^2$, where $\varphi$ is an inner function. We prove that the canonical factorization of $\varphi$ determines the class where the subspace $\varphi H^2$ belongs. Based on the results on the injectivity problem mentioned above, we provide examples showing the existence or non existence of geodesics between shift-invariant subspaces.

For $1 \leq p \leq \infty, L^p = L^p(\mathbb{T})$ denotes the usual Lebesgue spaces of functions defined on the unit circle $\mathbb{T}$. The Hardy space $H^p(1 \leq p < \infty)$ is the space of all analytic functions $f$ on the disk $\mathbb{D} = \{z \in \mathbb{C}: |z| < 1\}$ for which

$$\| f \|_{H^p} := \left( \sup_{0<r<1} \frac{1}{2\pi} \int_0^{2\pi} \left| f\left(re^{it}\right) \right|^p dt \right)^{1/p} < \infty.$$

The space of all bounded analytic functions on $\mathbb{D}$ with the norm $\| f \|_\infty = \sup_{z \in \mathbb{D}} |f(z)|$ is the Hardy space $H^\infty$. Functions in Hardy spaces have non tangencial limits a.e., a fact which is used to isometrically identify these spaces with

$$H^p = \left\{ f \in L^p : \int_0^{2\pi} f\left(e^{it}\right) \overline{\chi_n\left(e^{it}\right)} dt = 0, n < 0 \right\}.$$

Here $(\chi_k)_{k \in \mathbb{Z}}$ denotes the orthonormal basis of $L^2$ given by $\chi_k\left(e^{it}\right) = e^{ikt}$. We shall mostly use this representation of Hardy spaces as functions defined on $\mathbb{T}$ and deal with the values $p = 2, \infty$. In particular, $H^2$ is a closed subspace of the Hilbert space $L^2$ and $H^\infty$ is a closed subalgebra of $L^\infty$. For background and notational purposes, our main references are the books by Douglas, Nikol'skiĭ and Pavlović [98], [76], [261], [268].

A function $f \in H^2$ is called inner if $\left| f\left(e^{it}\right) \right| = 1$ a.e. on $\mathbb{T}$. A function $f \in H^2$ is outer if $\overline{\operatorname{span}}\{f \chi_n : n \geq 0\} = H^2$. For each $f \in H^2, f \neq 0$, there exist an inner function $f_{\text{inn}}$ and an outer function $f_{\text{out}} \in H^2$ such that $f = f_{\text{inn}} f_{\text{out}}$. This is called the inner-outer factorization, and it is unique up to a multiplicative constant.

The inner function can be further factorized. For each $a \in \mathbb{D} \setminus \{0\}$, a Blaschke factor is given by

$$b_a(z) = \frac{\bar{a}}{|a|} \frac{a - z}{1 - \bar{a}z}, \; z \in \mathbb{D}.$$

When $a = 0$, set $b_0(z) = z$. A Blaschke product is a function of the form

$$b(z) = \prod_{j=1}^{n} b_{a_j}(z), \; z \in \mathbb{D},$$

where $1 \leq n \leq \infty$. In the case where $n = \infty$, the infinite Blaschke product is convergent on compact subsets of $\mathbb{D}$ if the sequence $\{a_j\} \subseteq \mathbb{D}$ satisfies the Blaschke condition, that is, $\sum_j (1 - |a_j|) < \infty$. A finite or infinite Blaschke product is an inner function with zeros given by $\{a_j\}$. We remark that the zero set of a holomorphic function in $\mathbb{D}$ satisfies the Blaschke condition.
Let $\mu$ be a positive finite measure on $\mathbb{T}$. Suppose in addition that $\mu$ is singular with respect to the Lebesgue measure, and set

$$s_\mu(z) = \exp\left(-\int_{\mathbb{T}} \frac{\psi + z}{\psi - z} d\mu(\psi)\right),\ z \in \mathbb{D}.$$

It turns out that $s_\mu$ is an inner function and $s_\mu(z) \neq 0$ on $\mathbb{D}$. A function of this form is known as a singular inner function.
The canonical factorization of a function $f \in H^p$ states that there exists a unique factorization $f = \lambda b s_\mu f_{\text{out}}$, where $\lambda \in \mathbb{C}, |\lambda| = 1, b$ is a Blaschke product associated with the zero set of $f, s_\mu$ is a singular inner function and $f_{\text{out}}$ is the outer part of $f$.
Let $C$ denote the algebra of continuous functions on $\mathbb{T}$. The Sarason algebra is the following algebraic sum

$$H^\infty + C = \{f + g : f \in H^\infty, g \in C\}.$$

It is proved that this is indeed a closed subalgebra of $L^\infty$. The harmonic extension $\hat{\varphi}$ to $\mathbb{D}$ of a function $\varphi \in H^\infty + C$ is well-defined, and it plays a fundamental role in the characterization of invertible functions in this algebra. For $\varphi \in H^\infty + C$ and $0 < r < 1$, set $\varphi_r(e^{it}) = \hat{\varphi}(re^{it})$. Then $\varphi$ is invertible in $H^\infty + C$ if and only if there exist $\delta, \epsilon > 0$ such that $|\varphi_r(e^{it})| \geq \epsilon$ for $1 - \delta < r < 1$ and $e^{it} \in \mathbb{T}$.
This criterion allows to define the index of an invertible function in $H^\infty + C$. For a non-vanishing function $\varphi \in C$, let $ind(\varphi) \in \mathbb{Z}$ be the index (or winding number) of $\varphi$ around $z = 0$, which for differentiable $\varphi$ can be computed as

$$\text{ind}\,(\varphi) = \frac{1}{2\pi i}\oint \frac{\varphi'}{\varphi} = \frac{1}{2\pi}\int_0^{2\pi} \frac{\varphi'(e^{it})}{\varphi(e^{it})} e^{it} dt.$$

For $\varphi$ is invertible in $H^\infty + C$, set $\text{ind}\,(\varphi) = \lim_{r\to 1^-} \text{ind}\,(\varphi_r)$. This index is stable by small perturbations and it is an homomorphism of the invertible functions in $H^\infty + C$ onto the group of integers. The key property to prove these facts as well as the criterion for invertibility is that the harmonic extension is asymptotically multiplicative in $H^\infty + C$.
The largest $C^*$-algebra of $H^\infty + C$ is the set of quasicontinuous functions

$$QC = (H^\infty + C) \cap (\overline{H^\infty + C})$$

Every unimodular $\theta \in QC$ is invertible in $H^\infty + C$. In [271] Sarason proved that each unimodular function $\theta \in QC$ of index $n \in \mathbb{Z}$ can be expressed as $\theta = \chi_n e^{i(u+\tilde{v})}$, where $u, v$ are real functions in $C$ and $\tilde{v}$ stands for the harmonic conjugate of $v$ on $\mathbb{T}$.

The space of bounded linear operators on a Hilbert space $H$ to a Hilbert space $L$ is denoted by $\mathcal{B}(H, L)$ or $\mathcal{B}(H)$ if $H = L$. Let $H^2_- = \chi_{-1}\overline{H^2}$ be the orthogonal complement of the Hardy space $H^2$, and consider the orthogonal projections $P_+$and $P_-$onto $H^2$ and $H^2_-$, respectively. Three special classes of bounded operators will be used in the sequel. For $\varphi \in L^\infty$, the multiplication operator $M_\varphi \in \mathcal{B}(L^2), M_\varphi f = \varphi f$, where $f \in L^2$; the Toeplitz operator $T_\varphi \in \mathcal{B}(H^2), T_\varphi f = P_+(\varphi f)$, where $f \in H^2$; and the Hankel operator $H_\varphi \in \mathcal{B}(H^2, H^2_-), H_\varphi f = P_-(\varphi f)$, where $f \in H^2$.

Recall that the (unilateral) shift operator is given by $M_{\chi_1}$. It will be useful to state some well-known results on invariant subspaces of the shift operator.

**Theorem (5.1.1)[245]:** Suppose that $E$ is a closed subspace of $L^2$ and $M_{\chi_1}E \subseteq E$.

i) (Wiener) If $E$ is doubly invariant (i.e. $M_{\chi_1}(E) = E$ ), then $E = \chi_R L^2$ for a unique measurable subset $R \subseteq \mathbb{T}$, where $\chi_R$ is the characteristic of $R$

ii) (Beurling-Helson) If $E$ is singly invariant (i.e. $M_{\chi_1}(E) \neq E$ ), then $E = \theta H^2$ for a unique up to a constant $\theta \in L^\infty$ with $|\theta| = 1$ a.e.

iii) If $0 \neq E \subset H^2$, then $E = \theta H^2$ for some inner function $\theta$.

We will frequently use several properties of Toeplitz operators. Among the basic properties we recall that $\|T_\varphi\| = \| \varphi \|_\infty, T_\varphi^* = T_{\bar{\varphi}}$ and $T_{\varphi\psi} = T_\varphi T_\psi$ whenever $\psi \in H^\infty$. The following results will be useful.

**Theorem (5.1.2)[245]:** (Coburn's lemma) If $\varphi \in L^\infty$, then either $\ker(T_\varphi) = \{0\}$ or $\ker(T_\varphi^*) = \{0\}$, unless $\varphi \equiv 0$.

**Theorem (5.1.3)[245]:** Let $\varphi$ be a function in $L^\infty$. The following hold.

i) $T_\varphi$ is invertible if and only if it is Fredholm and has index zero.

ii) If $\varphi \in H^\infty + C$, then $T_\varphi$ is Fredholm if and only $\varphi$ is invertible in $H^\infty + C$. Furthermore, the Fredholm index of $T_\varphi$ satisfies ind $(T_\varphi) = -ind(\varphi)$.

Let $Gr$ be the Grassmann manifold of $L^2$, i.e. the set of all closed subspaces of $L^2$. Let $P_W$ denote the orthogonal projection onto a closed subspace $W \subset L^2$. In particular, we write $P_\varphi = P_{\varphi H^2}$, when $\varphi \in L^\infty$ and $\varphi H^2$ is closed. If we identify each subspace with its orthogonal projection, then

$$Gr = \{P_W : W \text{ is a closed subspace of } L^2\}.$$

As an application of Theorem (5.1.1), we determine when $\varphi H^2$ belongs to $Gr$.

**Lemma (5.1.4)[245]:** Let $\varphi$ be a nonzero function in $L^\infty$. Then $\varphi H^2$ is closed in $L^2$ if and only if $\varphi$ is invertible in $L^\infty$.

**Proof.** Clearly, if the function $\varphi$ is invertible in $L^\infty$, then the subspace $\varphi H^2$ is closed. Conversely, suppose that $\varphi H^2$ is closed. We proceed by way of contradiction and assume that the function $\varphi$ is not invertible in $L^\infty$. We need to distinguish two cases.

In the first case, we assume that there is a Borel set $S \subset \mathbb{T}$ with positive measure such that $\varphi(e^{it}) = 0$ for all $e^{it} \in S$. Moreover, we may take $S$ to be a maximal set with this property. Since $\varphi H^2$ is shift invariant, we further need to consider two cases according to whether $\varphi H^2$ is singly or doubly invariant. If $\varphi H^2$ is singly invariant, there is a function $\theta \in L^\infty$ such that $|\theta| = 1$ and $\varphi H^2 = \theta H^2$. Then, there is a function $f \in H^2$ such that $\varphi f = \theta$, which is a contradiction since $\varphi \equiv 0$ in $S$. If $\varphi H^2$ is doubly invariant, then there a Borel set $R \subset \mathbb{T}$ such that $\varphi H^2 = \chi_R L^2$. Therefore, $\varphi f = \chi_R$ for some function $f \in H^2$. Recall that for a nonzero function in $H^2$, the set $\{e^{it} \in \mathbb{T} : f(e^{it}) = 0\}$ has measure zero ([261]). Using the maximality of $S$, we find that the sets $S$ and $R^c$ must be equal with the possible exception of points in a set of measure zero. Since $\varphi \neq 0, S^c = R$ has positive measure, and we can pick a proper subset $R_1 \subset R$ such that $R \setminus R_1$ has positive measure. Again from the equation $\varphi H^2 = \chi_R L^2$, we obtain a nonzero function in $f \in H^2$ such that $\varphi f = \chi_{R_1}$. This implies that $f \equiv 0$ in $R \setminus R_1$, which contradicts the aforementioned property of functions in $H^2$.

In the second case, we suppose that $\varphi \neq 0$ a.e.. If the shift invariant subspace $\varphi H^2$ is doubly invariant, we have again that $\varphi H^2 = \chi_R L^2$ for some Borel set $R \subset \mathbb{T}$. In particular, this gives $\varphi = \chi_R g$ for $g \in L^2$, and since $\varphi \neq 0$ a.e., it follows that $\chi_R = 1$.

Thus, we get $\varphi H^2 = L^2$, which certainly cannot be possible. Next we assume that the subspace $\varphi H^2$ is single invariant. Then there is function $\theta \in L^\infty$ satisfying $|\theta| = 1$ a.e. and $\varphi H^2 = \theta H^2$. We may rewrite this as $\varphi_1 H^2 = H^2$, where $\varphi_1 = \bar{\theta}\varphi$. Note that $\varphi_1$ is not invertible in $L^\infty$, and $\varphi_1 \in H^2$, which gives $\varphi_1 \in H^\infty$. Using this fact and that $\varphi_1 \neq 0$ a.e., the Toeplitz operator $T_{\varphi_1}$ turns out to be injective. Moreover, $T_{\varphi_1} H^2 = \varphi_1 H^2 = H^2$ shows that $T_{\varphi_1}$ is invertible, and consequently, $\varphi_1$ must be invertible in $L^\infty$[261]. This gives a contradiction.

Let $\mathcal{A}$ be an abstract $C^*$-algebra. Denote by $Gr(\mathcal{A})$ the Grassmann manifold of $\mathcal{A}$, i.e. the set of all selfadjoint projections in $\mathcal{A}$. In [269], [258], Corach, Porta and Recht decribed the differential geometry of $Gr(\mathcal{A})$ in terms of projections and symmetries: one passes from projections to symmetries via the affine map

$$P \leftrightarrow \epsilon_P = 2P - 1.$$

In [258] a natural reductive structure was introduced in $Gr(\mathcal{A})$. In particular, geodesics were characterized. In [269] it was proved that these geodesics have minimal length, if one measures the length of curves by

$$L(\alpha) = \int_0^1 \| \dot{\alpha}(t) \| \, dt,$$

where $\alpha: [0,1] \to Gr(\mathcal{A})$ is a piecewise $C^1$-curve and $\|.\|$ is the norm of $\mathcal{A}$. This means that the operator norm induces a Finsler metric on Gr($\mathcal{A}$); however, note that this metric is not smooth, nor convex. We summarize these facts in the following.

This map induces a linear connection in $(\mathcal{A})$ : if $X(t)$ is a tangent field along a curve $\alpha(t) \in Gr(\mathcal{A})$,

$$\frac{DX}{dt} = E_\alpha(X).$$

The geodesics of $Gr(\mathcal{A})$ starting at $P$ with velocity $Y$ have the form $\delta(t) = e^{t\tilde{Y}} P e^{-t\tilde{Y}}$, where $\tilde{Y} = [Y, P]$ is antihermitian and co-diagonal with respect to $P$.

Let $P, Q$ be two orthogonal projections such that $\| P - Q \| < 1$. Then there exists a unique operator $X \in \mathcal{A}_h$, with $\| X \| < \pi/2$, which is co-diagonal with respect to $P$, such that $Q = e^{iX} P e^{-iX}$. The curve

$$\delta(t) = e^{itX} P e^{-itX} \qquad (1)$$

is the unique geodesic of Gr($\mathcal{A}$ ) joining $P$ and $Q$ (up to reparametrization). Moreover, this geodesic has minimal length. The exponent $X$ is an analytic function of $P$ and :

$$X = -\frac{i}{2} \log(\epsilon_p \epsilon_Q),$$

which is an analytic logarithm because $\|\epsilon_P \epsilon_Q - 1\| = \|\epsilon_P - \epsilon_Q\| = 2 \| P - Q \| < 2$.

Necessary and sufficient conditions were given for the existence of a geodesic joining two given orthogonal projections in the Grassmann manifold $Gr(H)$ of a Hilbert space $H$. This includes the case in which $\| P - Q \| = 1$. To briefly describe this result, let us recall that Halmos [263] (see also [259], [260]) proposed to understand the geometric properties of two orthogonal projections $P$ and $Q$ by considering the decomposition

$$(\mathrm{Ran}(P) \cap \ker(Q)) \oplus (\mathrm{Ran}(Q) \cap \ker(P)) \oplus (\mathrm{Ran}(P) \cap \mathrm{Ran}(Q)) \oplus (\ker(P) \cap \ker(Q)) \oplus H_0,$$

where $H_0$ is the orthogonal complement of the first four subspaces. The projections are said to be in generic position when the first four subspaces are trivial. The first two subspaces may be interpreted as an obstruction to find a geodesic between $P$ and $Q$.

**Theorem (5.1.5)[245]:** Let $\varphi, \psi$ be invertible functions in $L^\infty$. The following are equivalent.
i) $\ker(T_{\varphi\psi^{-1}}) = \ker(T_{\varphi^{-1}\psi}) = \{0\}$.
ii) There is a geodesic in $Gr$ joining $P_\varphi$ and $P_\psi$.
iii) There is unique geodesic of minimal length in $Gr$ joining $P_\varphi$ and $P_\psi$ given by

$$\delta(t) = e^{itX} P_\varphi e^{-itX},\ t \in [0,1],$$

where $X = X_{\varphi,\psi}$ is a uniquely determined selfadjoint operator such that $\| X \| \leq \pi/2$, $e^{iX} P_\varphi e^{-iX} = P_\psi$, and it is co-diagonal with respect to both $P_\varphi$ and $P_\psi$.

**Proof.** We can assume without loss of generality that $\varphi, \psi$ are unimodular functions by the argument before the statement of this theorem. Then, note that the restriction of the multiplication operator

$$M_\psi\big|_{\ker(T_{\bar{\varphi}\psi})} : \ker(T_{\bar{\varphi}\psi}) \to (\varphi H^2)^\perp \cap \psi H^2,$$

is an isomorphism. Similarly, $\ker(T_{\varphi\bar{\psi}}) \simeq \varphi H^2 \cap (\psi H^2)^\perp$. If the kernels of both $T_{\varphi\bar{\psi}}$ and $T_{\bar{\varphi}\psi}$ are trivial, then there is a geodesic joining $P_\varphi$ and $P_\psi$. Conversely, if such a geodesic exists, then $\varphi H^2 \cap (\psi H^2)^\perp$ and $(\varphi H^2)^\perp \cap \psi H^2$ have the same dimension. By Coburn's lemma, this dimension must be zero. Thus, we have shown that the first and second item are equivalent. The equivalence between the second and third item is explained.

We study in more detail the selfadjoint operator $X = X_{\varphi,\psi}$ linking the subspaces $\varphi H^2$ and $\psi H^2$ in Theorem (5.1.5). To this effect, we recall the following facts concerning Halmos' model for two orthogonal projections $P_0$ and $Q_0$ in generic position acting in a Hilbert space $H$. Under this assumption, there exists an isometric isomorphism between $H$ and a product space $K \times K$ and a positive operator $Z$ in $K$ with $\| Z \| \leq \pi/2$ and $\ker(Z) = \{0\}$. This isomorphism transforms the projections $Q_0$ and $P_0$ into

$$Q_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \text{ and } P_0 = \begin{pmatrix} C^2 & CS \\ CS & S^2 \end{pmatrix},$$

where $C = \cos(Z)$ and $S = \sin(Z)$[263]. The unique selfadjoint operator $X$ linking these projections is (see [246])

$$X = \begin{pmatrix} 0 & iZ \\ -iZ & 0 \end{pmatrix}.$$

Note that $\| X \| = \| Z \|$.
Let $\sigma(A)$ denote the spectrum of an operator $A$. Recall the definition of reduced minimum modulus $\gamma(A)$ of an operator $\neq 0$ :

$$\begin{aligned} \gamma(A) &= \inf \{\| Af \| : \| f \| = 1, f \in \ker(A)^\perp\} \\ &= \inf \sigma(|A|) \setminus \{0\}. \end{aligned}$$

**Proposition (5.1.6)[245]:** Let $\varphi, \psi$ be unimodular functions in $L^\infty$ such that

$$\ker(T_{\varphi\bar{\psi}}) = \ker(T_{\bar{\varphi}\psi}) = \{0\}.$$

Then

$$Z = M_\varphi \cos^{-1}(|T_{\varphi\bar{\psi}}|) M_{\bar{\varphi}}$$

and in particular

$$\|X_{\varphi,\psi}\| = \cos^{-1}\left(\gamma(T_{\varphi\bar{\psi}})\right).$$

**Proof.** On the non generic part of $P_\varphi$ and $P_\psi$, the operator $X = X_{\varphi,\psi}$ is trivial. Thus in order to compute its norm we restrict to the generic part, and thus they can be described by Halmos' model,

$$X = \begin{pmatrix} 0 & iZ \\ -iZ & 0 \end{pmatrix}.$$

It is elementary that, if $Q_0, P_0$ denote the reductions of $P_\varphi, P_\psi$ to the generic parts, then

$$Q_0P_0Q_0 = \begin{pmatrix} C^2 & 0 \\ 0 & 0 \end{pmatrix}.$$

Now

$$C^2 = P_\varphi P_\psi P_\varphi = M_\varphi P_+ M_{\bar{\varphi}} M_\psi P_+ M_{\bar{\psi}} M_\varphi P_+ M_{\bar{\varphi}} = M_\varphi T^*_{\varphi\bar{\psi}} T_{\varphi\bar{\psi}} M_{\bar{\varphi}} = M_\varphi \left|T_{\varphi\bar{\psi}}\right|^2 M_{\bar{\varphi}}.$$

Therefore $0 \le C = \cos(Z) = M_\varphi \left|T_{\varphi\bar{\psi}}\right| M_{\bar{\varphi}}$, and thus, $Z = M_\varphi \cos^{-1}\left(\left|T_{\varphi\bar{\psi}}\right|\right) M_{\bar{\varphi}}$. From this formula, it follows that

$$\|X_{\varphi,\psi}\| = \left\|\cos^{-1}\left(\left|T_{\varphi\bar{\psi}}\right|\right)\right\| = \cos^{-1}(\lambda_0),$$

where

$$\lambda_0 = \text{i } \sigma\left(\left|T_{\varphi\bar{\psi}}\right|\right) = \text{i } \sigma\left(\left|T_{\varphi\bar{\psi}}\right|\right) \setminus \{0\} = \gamma\left(T_{\varphi\bar{\psi}}\right).$$

The second equality can be deduced from the assumption that $T_\varphi\bar{\psi}$ is injective, which implies that 0 cannot be an isolated point of $\sigma\left(\left|T_{\varphi\bar{\psi}}\right|\right)$.

**Example (5.1.7)[245]:** Consider $\varphi = \chi_1$ and the Blaschke factor

$$\psi\left(e^{it}\right) = b_a\left(e^{it}\right) = \frac{\bar{a}}{|a|}\frac{a - e^{it}}{1 - \bar{a}e^{it}},$$

for $0 < |a| < 1$. Then by direct computation,

$$\varphi H^2 \cap (\psi H^2)^\perp = (\varphi H^2)^\perp \cap \psi H^2 = \{0\},\ (\varphi H^2)^\perp \cap (\psi H^2)^\perp = H^2_-$$

and

$$(\varphi H^2) \cap \psi H^2 = \chi_1 b_a H^2 = \chi_1(\chi_1 - a)H^2.$$

Then the generic part $H_0$ of $\varphi H^2$ and $\psi H^2$ is the two dimensional space $H^2 \ominus \chi_1(\chi_1 - a)H^2$. The reduced projections $Q_0 = P_\varphi\big|_{H_0}$ and $P_0 = P_\psi\big|_{H_0}$ are one dimensional,

$$\text{Ran}(Q_0) = H_0 \cap \chi_1 H^2 = \left\langle \frac{\chi_1}{1 - \bar{a}\chi_1} \right\rangle,\ \text{Ran}(P_0) = H_0 \cap (\chi_1 - a)H^2 = \left\langle \frac{\chi_1 - a}{1 - \bar{a}\chi_1} \right\rangle.$$

According Halmos' formulas,

$$Q_0P_0Q_0 = \begin{pmatrix} C^2 & 0 \\ 0 & 0 \end{pmatrix}.$$

Denote by $f$ and $g$ the normalizations of $\frac{\chi_1}{1-\bar{a}\chi_1}$ and $\frac{\chi_1 - a}{1 - \bar{a}_1}$, respectively. As usual, let $f_1 \otimes f_2$ be the rank one operator defined by $f_1 \otimes f_2(h) =< h, f_2 > f_1$. Then, we have another expression

$$Q_0P_0Q_0 = (f \otimes f)(g \otimes g)(f \otimes f) = |< f, g >|^2 f \otimes f.$$

Therefore,

$$\begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} = |< f, g >| f \otimes f.$$

In this case $C = \cos(Z)$ is a positive real number, and thus $Z = \cos^{-1}(|< f, g >|)$. Simple computations show that $|< f, g >| = (1 - |a|^2)^{1/2}$, which gives

$$Z = \cos^{-1}\left((1 - |a|^2)^{\frac{1}{2}}\right) = \sin^{-1}(|a|).$$

Then, the part of $X_{\varphi,\psi}$ acting on $H_0$ is

$$X_{\varphi,\psi}\big|_{H_0} = \begin{pmatrix} 0 & -i\sin^{-1}(|a|) \\ i\sin^{-1}(|a|) & 0 \end{pmatrix}.$$

The restriction of $X_{\varphi,\psi}$ to $H_0^\perp$ is trivial. Thus, $X_{\varphi,\psi}$ has rank two, and

$$\|X_{\varphi,\psi}\| = \sin^{-1}(|a|).$$

The minimality property of the geodesics in the Grassmann manifold may be used to obtain operator inequalities.

**Theorem (5.1.8)[245]:** Let $\varphi, \psi$ be unimodular functions in $L^\infty$ such that $\ker(T_{\varphi\bar{\psi}}) = \ker(T_{\bar{\varphi}\psi}) = \{0\}$. Then
$$\|M_\theta P_+ - P_+ M_\theta\| \geq \cos^{-1}\left(\gamma(T_{\varphi\bar{\psi}})\right),$$
for every real argument $\theta \in L^\infty$ of the function $\varphi\bar{\psi}$.

**Proof.** Let $\theta$ be a real function in $L^\infty$ such that $e^{i\theta} = \varphi\bar{\psi}$. Consider the curve
$$\alpha(t) = M_{e^{it\theta}} P_\varphi M_{e^{-it\theta}}.$$
Apparently, $\alpha(t)$ is a smooth curve in $Gr$ with $\alpha(0) = P_\varphi$ and $\alpha(1) = M_{\bar{\varphi}\psi} P_\varphi M_{\varphi\bar{\psi}} = P_\psi$. Then $\alpha(t)$ is longer than the (unique) minimal geodesic which joins $\varphi H^2$ and $\psi H^2$, whose length is $\|X_{\varphi,\psi}\|$. Note that
$$\dot{\alpha}(t) = iM_{e^{it\theta}} M_\theta P_\varphi - iP_\varphi M_\theta M_{e^{-it\theta}} = iM_{e^{it\theta}} M_\varphi (M_\theta P_+ - P_+ M_\theta) M_{\bar{\varphi}} M_{e^{-it\theta}}.$$
Thus, we find that $\| \dot{\alpha}(t) \| = \|M_\theta P_+ - P_+ M_\theta\|$, and using Proposition (5.1.6), we obtain
$$\cos^{-1}\left(\gamma(T_{\varphi\bar{\psi}})\right) = \|X_{\varphi,\psi}\| \leq L(\alpha) = \int_0^1 \| \dot{\alpha}(t) \| \, dt = \|M_\theta P_+ - P_+ M_\theta\|.$$
**Corollary (5.1.9)[245]:** Let $\theta$ be a real valued continuous function, then
$$\|M_\theta P_+ - P_+ M_\theta\| \geq \cos^{-1}\left(\gamma(T_{e^{i\theta}})\right).$$
**Proof.** Put $\varphi = e^{i\theta}$ and $\psi = 1$ in Theorem (5.1.8). Then, note that $\varphi$ is an invertible continuous function with zero index. Hence the operator $T_\varphi$ is Fredholm and has index zero, which implies that it is invertible.

Let $\theta_t, t \in [0,1]$, be a piecewise differentiable path of real valued functions in $C$. Then the curve $\alpha(t) = M_{e^{i\theta_t}} P_+ M_{e^{-i\theta_t}}$ is piecewise differentiable. Similarly as above, its velocity is
$$\| \dot{\alpha}(t) \| = \left\| M_{e^{i\theta_t}} \left[ M_{i\dot{\theta}_t}, P_+ \right] M_{-e^{i\theta_t}} \right\| = \left\| H_{\dot{\theta}_t} \right\| = \text{i } \left\{ \|\dot{\theta}_t - f\|_\infty : f \in H^\infty \right\}.$$
The last quantity can be regarded as the norm of $[\dot{\theta}_t]$, the class of $\dot{\theta}_t$ in the quotient $L^\infty / H^\infty$ (which is also the velocity of the curve $[\theta_t]$ in the quotient). Therefore,
$$L(\alpha) = L_{L^\infty/H^\infty}([\theta_t]).$$
Note that the curve $\theta_t$ is arbitrary between $\theta_0$ and $\theta_1$. In particular, when $\theta_t$ is a straight line, we have the following:

**Corollary (5.1.10)[245]:** Let $\theta_0, \theta_1$ be real valued continuous functions, then
$$\|\theta_0 - \theta_1\|_{L^\infty/H^\infty} \geq \left\| X_{e^{i\theta_0}, e^{i\theta_1}} \right\| = \cos^{-1}\left(\gamma(T_{e^{i(\theta_1 - \theta_o)}})\right).$$
The space $L^2$ has the orthogonal decomposition $L^2 = H^2 \oplus H^2_-$, which we now use to give the following definition. The compact restricted Grassmannian $\mathrm{Gr}_{\text{res}}$ is the manifold of closed linear subspaces $W \subset L^2$ such that

- $P_+|_W : W \to H^2 \in \mathcal{B}(W, H^2)$ is a Fredholm operator, and
- $P_-|_W : W \to H^2_- \in \mathcal{B}(W, H^2_-)$ is a compact operator.

The components of the restricted Grassmannian are parametrized by $k \in \mathbb{Z}$, where $k$ is the index of the operator $P_+|_W : W \to H^2 \in \mathcal{B}(W, H^2)$,
$$Gr^k_{\text{res}} = \left\{ W \in Gr_{\text{res}} : \text{ind}\,(P_+|_W : W \to H^2) = k \right\}.$$
In particular, since $P_+$ is the identity restricted to $H^2$, $H^2 = \text{Ran}(P_+) \in Gr^0_{\text{res}}$.

**Lemma (5.1.11)[245]:** Let $\varphi$ be an invertible function in $L^\infty$. Then the following are equivalent.

i) $\varphi H^2 \in Gr_{res}$.

ii) $\varphi$ is an invertible function in $H^\infty + C$.

iii) $\varphi H^2 = \theta H^2$ for some $\theta \in QC, |\theta| = 1$ a.e.
In this case, $\varphi H^2 \in Gr_{\text{res}}^k$ , where $k = -ind(\varphi) = -ind(\theta)$.

**Proof.** We first prove $i) \Rightarrow ii$. We claim that the Hankel operator $H_\varphi: H^2 \to H_-^2$, $H_\varphi f = P_-(\varphi f)$, is compact if and only if $P_- \mid \varphi H^2: \varphi H^2 \to H_-$is compact. In fact, note that $H_\varphi f = P_-|_{\varphi H^2}(\varphi f) = P_-|_{\varphi H^2} M_\varphi f$, for all $f \in H^2$. Since $\varphi$ is invertible in $L^\infty$, $M_\varphi: H^2 \to \varphi H^2$ is an invertible operator. Thus,

$$H_\varphi = (P_- \mid \varphi_{\varphi^2}) \left(M_\varphi\big|_{H^2}\right),\ H_\varphi \left(M_\varphi\big|_{H^2}\right)^{-1} = P_-|_{\varphi H^2},$$

which clearly implies our claim.
Suppose that $\varphi H^2 \in Gr_{\text{res.}}$ . Then, the operator $P_-|_{\varphi H^2}: \varphi H^2 \to H_-^2$ is compact, so we get that $H_\varphi$ is compact. Hartman's theorem asserts that a Hankel operator $H_\varphi$ is compact if and only if $\varphi \in H^\infty + C$ (see e.g. [76]). Thus, it follows that $\varphi \in H^\infty + C$. Since $\varphi H^2 \in Gr_{res}$, we also have that $P_+ \mid \varphi H^2: \varphi H^2 \to H^2$ is a Fredholm operator. Note that $\text{Ran}(P_+ \mid \varphi H^2) = \text{Ran}(T_\varphi)$ and $\ker(P_+ \mid \varphi H^2) = M_\varphi \ker(T_\varphi)$, where $T_\varphi$ is the Toeplitz operator with symbol $\varphi$. Therefore $T_\varphi$ is Fredholm, and thus, $\varphi$ is invertible in $H^\infty + C$.
Now we prove $) \Rightarrow i$ ). Assume that $\varphi$ is an invertible function in $H^\infty + C$. Then, we have that $T_\varphi$ is a Fredholm operator. By the same arguments as in the previous paragraph, we see that $P_+|_{\varphi H^2}: \varphi H^2 \to H^2$ is also a Fredholm operator. On the other hand, $\varphi \in H^\infty + C$ is equivalent to $H_\varphi$ compact. Hence $P_- \mid \varphi H^2: \varphi H^2 \to H_-$is compact, and consequently, $\varphi H^2 \in Gr_{\text{res}}$ .
The implication $) \Rightarrow iii$ ) is given by Theorem (5.1.1) : if $\varphi \in H^\infty + C$, then $\varphi H^2$ is singly invariant. Therefore exists a (unique up to a multiplicative constant) unimodular function $\theta$ such that $\varphi H^2 = \theta H^2$. Now $\theta = \varphi f$ for some $f \in H^2$. Since $\varphi$ is invertible in $L^\infty$, then $f \in H^\infty$. Hence, $\theta \in H^\infty + C$. Further, by the invertibility of $\varphi$, it clearly follows that $f$ is invertible in $L^\infty$. Using that $\varphi H^2 = \theta H^2 = \varphi f H^2$, we get $f H^2 = H^2$, and consequently, $f$ is an outer function. Recall that a function in $H^\infty$ is invertible if and only if it is outer and invertible in $L^\infty$. This gives $f^{-1} \in H^\infty$. Now $\bar{\theta} = \theta^{-1} = \varphi^{-1} f^{-1} \in H^\infty + C$, which proves that $\theta \in QC$.
To prove the implication $) \Rightarrow ii$ ), we observe that every unimodular $\theta \in QC$ is invertible in $H^\infty + C$. By the equivalence between $i$ ) and $ii$ ), we get $\varphi H^2 = \theta H^2 \in Gr_{\text{res}}$ , and hence $\varphi$ is invertible in $H^\infty + C$.
Suppose that $\varphi H^2 \in Gr_{\text{res.}}^k$ . To prove our claim on the index, we have pointed out that $\text{Ran}(P_+|_{\varphi H^2}) = \text{Ran}(T_\varphi)$ and $\ker(P_+|_{\varphi H^2}) = M_\varphi \ker(T_\varphi)$, where $M_\varphi$ is invertible. It follows that $k = ind(P_+ \mid \varphi H^2) = ind(T_\varphi) = -ind(\varphi)$. Moreover, $\theta = \varphi f$, and $f$ is invertible in $H^\infty$. Every invertible function in $H^\infty$ has index zero. Hence, $\text{ind}(\varphi) = ind(\theta)$.
Under the identification of each closed subspace $W \subseteq L^2$ with the orthogonal projection $P_W$, the compact restricted Grassmannian is given by

$$Gr_{\text{res}} = \{P \in \mathcal{B}(L^2): P - P_+ \text{is compact}, P = P^2 = P^*\}. \tag{2}$$

Applying the results mentioned for the algebra of compact operators, it follows that the tangent space $(TGr_{\text{res}})_P$ at some point $P \in Gr_{\text{res}}$ is given by

$$(TGr_{\text{res}})_P = \{iXP - iPX: X^* = X \text{ is compact}\}.$$

Then, using the usual operator norm, we have a Finsler metric to measure the length of curves.
On the other hand, the above presentation of $Gr_{res}$ by means of operators is related to the orthogonal projections of the $C^*$-algebra

$$\mathcal{B}_{cc} = \{T \in \mathcal{B}(L^2): [T, P_+] \text{is compact}\}. \quad (3)$$

Indeed, this algebra consists on operators with compact co-diagonal entries. Denoting by $\pi$ the projection onto the Calkin algebra, the restricted Grassmannian coincides with the class of projections $P$ such that

$$\pi(P) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

where this is a matrix decomposition with respect to $\pi(P_+)$and $\pi(P_-)$. Metric aspects of the projections in $\mathcal{B}_{cc}$ for a general Hilbert space $H$ were studied in [248]. In particular, it was proved that any pair of projections in the same connected component of $Gr_{res}$ can be joined by a geodesic of minimal length. Combining these facts and the characterization in Lemma (5.1.11), we have the following result.

**Theorem (5.1.12)[245]:** Let $\varphi, \psi$ be invertible functions in $H^\infty + C$. The following are equivalent.

i) $\text{ind}(\varphi) = ind(\psi)$.

ii) There is a geodesic in $Gr_{\text{res}}$ joining $P_\varphi$ and $P_\psi$.

iii) There is unique geodesic of minimal length in $Gr_{\text{res}}$ joining $P_\varphi$ and $P_\psi$ given by

$$\delta(t) = e^{itX} P_\varphi e^{-itX},\ t \in [0,1],$$

where $X = X_{\varphi,\psi}$ is a uniquely determined compact selfadjoint operator such that $\| X \| < \pi/2, e^{iX} P_\varphi e^{-iX} = P_\psi$, and it is co-diagonal with respect to both $P_\varphi$ and $P_\psi$.

**Proof.** We first show the equivalence between ) and $ii$ ). Suppose that ind $(\varphi) = ind(\psi)$, so we have that $P_\varphi$ and $P_\psi$ belong to the same connected component of $Gr_{res}$. According to [248] there is a (minimal) geodesic joining these projections. The converse is obvious by the characterization of the connected components of $Gr_{\text{res}}$ in terms of the index of the functions.

Similarly, to prove the equivalence between (i) and ($iii$), the only non trivial part is that $i$) implies $iii$). If $ind(\varphi) = ind(\psi)$, then ind $(\varphi\psi^{-1}) = 0$, and consequently, as we state in Theorem (5.1.3),$T_{\varphi\psi^{-1}}$ is an invertible operator. Following the same argument as in the proof of Theorem (5.1.5), but now using Lemma (5.1.11), we can assume that $\varphi, \psi$ are unimodular functions in $QC$. Therefore, $\varphi H^2 \cap (\psi H^2)^\perp \simeq \ker(T_{\varphi\bar{\psi}}) = \{0\}$ and $\psi H^2 \cap (\varphi H^2)^\perp \simeq \ker(T_{\psi\bar{\varphi}}) = \{0\}$. Under these conditions, there is a unique geodesic of minimal length joining $P_\varphi$ and $P_\psi$ of the desired form (see [248]).

Now we address the following question: when can we take the quasicontinuous function $\theta$ in Lemma (5.1.11) to be continuous? Note that this function is unique up to a multiplicative constant.

The conditions in Lemma (5.1.11) are also equivalent to have $\varphi H^2 = gH^2$, where $g \in C$ is non-vanishing. Indeed, this is easily seen from [98], which asserts that the invertibility of a function $\varphi$ in the algebra $H^\infty + C$ is equivalent to the factorization $\varphi = fg$, where $f, f^{-1} \in H^\infty$ and $g, g^{-1} \in C$. In addition, note that ind $(g) = ind(\varphi)$. However, the function $g$ is not necessary unimodular.

Assuming that the function $\varphi$ is continuous, we establish below a relation between $\theta$ and $\varphi$. Given a real valued function $u \in L^2, \tilde{u}$ is the harmonic conjugate on $\mathbb{T}$. Denote by Lip $^{\alpha}$ the Banach space of complex-valued functions on $\mathbb{T}$ satisfying a Lipschitz condition of order $\alpha(0 < \alpha \leq 1)$. We write $A = H^\infty \cap C$ for the disk algebra.

**Proposition (5.1.13)[245]:** Let $\varphi \in C$ be non-vanishing, $\theta$ denote the quasicontinuous function of Lemma (5.1.11), and set $u = -\log|\varphi|$, then

$$\theta = \frac{\varphi}{|\varphi|} e^{i\tilde{u}}.$$

In particular, $\theta \in C$, whenever $\tilde{u} \in C$. In addition, the following assertions hold.

i) If $\varphi \in \mathrm{Lip}^\alpha$ for $0 < \alpha < 1$, then $\theta \in \mathrm{Lip}^\alpha$.

ii) If $\varphi \in A$, then $\theta \in A$.

**Proof.** Recalling that $\theta H^2 = \varphi H^2$, and by the proof of $ii\,) \Rightarrow iii\,)$ in Lemma (5.1.11), one can find an invertible function $f$ in $H^\infty$ such that $\theta = f\varphi$. Since $f$ is an outer function, its harmonic extension admits a representation:

$$\hat{f}(z) = \lambda \exp\left(\frac{1}{2\pi}\int_0^{2\pi} \frac{e^{it}+z}{e^{it}-z} \log\left|f\left(e^{it}\right)\right| dt\right),\ z \in \mathbb{D},$$

for some $\lambda \in \mathbb{T}$; see [98]. We may assume that $\lambda = 1$. Note that $\hat{f} = \exp(a + ib)$ where

$$a(z) = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{Re}\left\{\frac{e^{it}+z}{e^{it}-z}\right\} \log\left|f\left(e^{it}\right)\right| dt = \log|\hat{f}(z)|,$$

since the real part of $\left(e^{it}+z\right)\left(e^{it}-z\right)^{-1}$ is the Poisson kernel. Since $|f| = 1/|\varphi|$ on $\mathbb{T}$, and $f \in H^\infty$, the following radial limit $\lim_{r\to 1^-} a\left(re^{it}\right) = \log\left|f\left(e^{it}\right)\right| = u\left(e^{it}\right)$ exists a.e. On the other hand,

$$b(z) = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{Im}\left\{\frac{e^{it}+z}{e^{it}-z}\right\} \log\left|f\left(e^{it}\right)\right| dt$$

is the harmonic conjugate of $a$ on $\mathbb{D}$ (up to a constant). By the Privalov-Plessner theorem [268], $\lim_{r\to 1^-} b\left(re^{it}\right) = \tilde{u}\left(e^{it}\right)$ a.e. Since $\theta = \varphi f$ and $f = e^u e^{i\tilde{u}} = \frac{1}{|\varphi|} e^{i\tilde{u}}$, we obtain $\theta = \frac{\varphi}{|\varphi|} e^{i\tilde{u}}$.

i) Now we assume that $\varphi \in \mathrm{Lip}^\alpha$. Since $\varphi$ is a non-vanishing continuous function, then $u = -\log|\varphi| \in \mathrm{Lip}^\alpha$. By Privalov's theorem, $\tilde{u} \in \mathrm{Lip}^\alpha$ for $\alpha < 1$ (see [268]). Clearly, $\varphi, |\varphi|^{-1} \in \mathrm{Lip}^\alpha$, which yields $\theta \in \mathrm{Lip}^\alpha$.

ii) According to [76], the outer part $\varphi_{\mathrm{out}}$ of $\varphi$ belongs to $A$. Since $\theta = \varphi f$, it follows that $|f^{-1}| = \left|\varphi_{\mathrm{out}}\right|$. Therefore, $\varphi_{\mathrm{out}} = \lambda f^{-1}$ for some $\lambda \in \mathbb{T}$. Thus, the inner part of $\varphi$ satisfies $\theta = \lambda \varphi_{\mathrm{inn}}$, and thus we obtain $\theta \in A$.

**Example (5.1.14)[245]:** In contrast to what happens with functions in $\mathrm{Lip}^\alpha$ or $A$, we now show that the class of absolutely continuous functions is not preserved in the above proposition. Let

$$u\left(e^{it}\right) = -\sum_{n\geq 2} \frac{\sin(nt)}{n\log(n)}$$

then $u \in C$; moreover $u$ is absolutely continuous on $\mathbb{T}$[276]. Let $\varphi = e^{-u}$, clearly $\varphi \in C$ is non-vanishing and absolutely continuous on $\mathbb{T}$. Since $u(\mathbb{T}) \subset \mathbb{R}$, we have $\varphi > 0$ on $\mathbb{T}$, therefore $-\log|\varphi| = u$. Let

$$v(e^{it}) = \sum_{n\geq 2} \frac{\cos(nt)}{n\log(n)},$$

and note that

$$f(z) = \sum_{n\geq 2} \frac{i}{n\log(n)} z^n = iv + u$$

is analytic, therefore $v$ is the harmonic conjugate of $u$. But $v$ is not continuous on $\mathbb{T}$, not even bounded since $\sum_{n\geq 2} \frac{1}{n\log(n)} = +\infty$, therefore $\theta = e^{iv}$ is not continuous on $\mathbb{T}$.

Let $\mathcal{K}(H, L)$ be the space of compact operators between two Hilbert spaces $H$ and $L$. Given an operator $T \in \mathcal{K}(H, L)$, we denote by $(s_n(T))_{n\geq 1}$ the sequence of its singular values. The $p$-Schatten class ($1 \leq p < \infty$) is defined by

$$\mathcal{B}_p(H, L) = \left\{ T \in \mathcal{K}(H, L) : \| T \|_p = \left( \sum_{n=1}^{\infty} s_n(T)^p \right)^{1/p} < \infty \right\}.$$

These are Banach spaces endowed with the norm $\|\cdot\|_p$. As usual, when $p = \infty$, we set $\mathcal{B}_\infty(H, L) = \mathcal{K}(H, L)$. In particular, $\mathcal{B}_p(H, H) = \mathcal{B}_p(H)$ is a bilateral ideal of $\mathcal{B}(H)$. Using the orthogonal decomposition $L^2 = H^2 \oplus H^2_-$, and the $p$-Schatten class ($1 \leq p < \infty$), one can introduce the $p$-restricted Grassmannian $Gr_{res,p}$ as the manifold of closed linear subspaces $W \subset L^2$ such that

- $P_+|_W : W \to H^2 \in \mathcal{B}(W, H^2)$ is a Fredholm operator, and
- $P_-|_W : W \to H^2_- \in \mathcal{B}_p(W, H^2_-)$.

Its connected components $Gr^k_{res,p}, k \in \mathbb{Z}$, are also described by the index of the projection $P_+|_W : W \to H^2$. The case $p = 2$ was studied in connection with loop groups [270]; it is an infinite dimensional manifold with remarkable geometric properties [252], [262], [275]. Other values of $1 \leq p \leq \infty$, or more generally restricted Grassmannians associated with symmetrically-normed ideals, were treated in [251], [256].

We denote by $B^\alpha_p$ the Besov space, where $1 \leq p < \infty$ and $0 < \alpha \leq 1$. For the definition of these spaces, and the following results we refer to Böttcher, Karlovich and Silbermann [253]. Among various generalizations of the classical Krein algebra, it was introduced the following algebra defined by means of Hankel operators:

$$K^{1/p,0}_{p,0} = \{\varphi \in L^\infty : H_\varphi \in \mathcal{B}_p(H^2, H^2_-)\},$$

where $1 \leq p \leq \infty$. It turns out to be a Banach algebra under the norm

$$\| \varphi \|_{K^{1/p,0}_{p,0}} = \| \varphi \|_{L^\infty} + \|H_\varphi\|_p.$$

In the case $p = \infty$, it simply has the usual operator norm of a compact operator. By Hartman's theorem, $K^{1/\infty,0}_{\infty,0} = H^\infty + C$, and for $1 \leq p < \infty$, one has $K^{1/p,0}_{p,0} \subseteq H^\infty + C$. Given a function $\varphi \in L^\infty$ and $1 \leq p < \infty$, Peller's theorem states that the Hankel operator $H_\varphi \in \mathcal{B}_p(H^2, H^2_-)$ if and only if $P_-\varphi \in B^{1/p}_p$ (see [98]).

Then there is an equivalent definition of $K^{1/p,0}_{p,0}$ in terms of functions instead of operators. When $1 \leq p < \infty$, it holds

$$K^{1/p,0}_{p,0} = \left\{\varphi \in L^\infty : P_-\varphi \in B^{1/p}_p\right\} = L^\infty \cap \left(H^\infty + B^{1/p}_p\right).$$

Moreover, when $p > 1$, a function $\varphi$ is invertible in $K^{1/p,0}_{p,0}$ if and only if is invertible in $H^\infty + C$.

Using the above stated results and the same arguments of Lemma (5.1.11), the following characterization can be obtained.

**Corollary (5.1.15)[245]:** Let $\varphi$ be an invertible function in $L^\infty$ and $1 \le p < \infty$. The following assertions are equivalent:

i) $\varphi H^2 \in Gr_{res,p}$.

ii) $\varphi \in K_{p,0}^{1/p,0}$ and $\varphi$ is invertible in $H^\infty + C$.

iii) $\varphi H^2 = \theta H^2$ for some $\theta \in QC \cap K_{p,0}^{1/p,0}, |\theta| = 1$ a.e.

In this case, $\varphi H^2 \in Gr_{\text{res}}^k$, where $k = -ind(\varphi) = -ind(\theta)$.

**Corollary (5.1.16)[245]:** Let $1 \le p < \infty$, and let $\varphi, \psi$ be functions in $K_{p,0}^{1/p,0}$ which are invertible in $H^\infty + C$. The following are equivalent:

i) ind $(\varphi) = ind(\psi)$.

ii) There is a geodesic in $Gr_{\text{res},\,p}$ joining $P_\varphi$ and $P_\psi$.

iii) There is unique geodesic of minimal length in $Gr_{\text{res},\,p}$ joining $P_\varphi$ and $P_\psi$ given by

$$\delta(t) = e^{itX} P_\varphi e^{-itX},\ t \in [0,1],$$

where $X = X_{\varphi,\psi}$ is a uniquely determined selfadjoint operator such that $\| X \| < \pi/2$, $e^{iX} P_\varphi e^{-iX} = P_\psi$, and it is co-diagonal with respect to both $P_\varphi$ and $P_\psi$.

Moreover, arguing as in the proof of Theorem (5.1.8) we also obtain

**Corollary (5.1.17)[245]:** Let $1 \le p < \infty$, and let $\varphi, \psi$ be functions in $K_{p,0}^{1/p,0}$ which are invertible in $H^\infty + C$, such that ind $(\varphi) =$ ind $(\psi)$. Then if $\theta \in K_{p,0}^{1/p,0}$ is such that $e^{i\theta} = \varphi\bar{\psi}$,

$$\|M_\theta P_+ - P_+ M_\theta\|_p \ge 2^{1/p} \left\|\cos^{-1}\left(\left|T_\varphi -\right|\right)\right\|_p = \text{dist}_p\left(P_\varphi, P_\psi\right).$$

For instance, if $\varphi$ and $\psi$ are $C^1$ functions (with equal index) such an argument $\theta$ exists, which is continuous and piecewise smooth.

**Proof.** Recall from Poposition (5.1.6) that

$$X_{\varphi,\psi} = \begin{pmatrix} 0 & iZ \\ -iZ & 0 \end{pmatrix}$$

and thus $(Z \ge 0)$

$$\left|X_{\varphi,\psi}\right| = \begin{pmatrix} Z & 0 \\ 0 & Z \end{pmatrix}.$$

Also $Z = M_\varphi \cos^{-1}\left(\left|T_\varphi \bar{\psi}\right|\right) M_{\bar{\varphi}}$. Then

$$\left\|X_{\varphi,\psi}\right\|_p = 2^{1/p} \| Z \|_p = 2^{1/p} \left\|\cos^{-1}\left(\left|T_{\varphi\bar{\psi}}\right|\right)\right\|_p$$

The orthogonal projections of the $C^*$-algebra $\mathcal{B}_{cc}$ defined in (3) may be classified using their image in the Calkin algebra. In addition to the restricted Grassmannian, we shall need to consider the essential class $\mathbb{E}_1$ consisting of all the orthogonal projections which have the form (in terms of $\pi(P_+)$and $\pi(P_-)$)

$$\pi(P) = \begin{pmatrix} p & 0 \\ 0 & 0 \end{pmatrix},$$

where $p \ne 0,1$ is a projection in the Calkin algebra. It was shown that the class $\mathbb{E}_1$ is connected, and in contrast to the restricted Grassmannian, there are projections which cannot be joined by a geodesic in $\mathbb{E}_1$.

Let $E$ be a closed subspace of $L^2$ such that $M_{\chi_1}(E) \subset E$. If $0 \ne E \subseteq H^2$, then $E = \varphi H^2$ for some inner function $\varphi$. We prove below that these subspaces belong to either the restricted Grassmannian or the essential class $\mathbb{E}_1$.

**Theorem (5.1.18)[245]:** Let $\varphi$ be an inner function. Then the following assertions hold:
i) $\varphi$ is a finite Blaschke product if and only if $P_\varphi \in Gr^k_{\text{res}}$, where $k$ is the number of zeros of $\varphi$.
ii) $\varphi$ is not a finite Blaschke product if and only if $P_\varphi \in \mathbb{E}_1$.
**Proof.** i) The only inner functions which are invertible in $H^\infty + C$ are the finite Blaschke products (see e.g. [272]). Therefore, the result follows from Lemma (5.1.11). The index of a Blaschke factor is equal to its number of zeros, and as we have already showed, it determines the connected component of $Gr_{\text{res}}$ where $P_\varphi$ lies.
ii) Suppose that $\varphi$ is not a finite Blaschke product. As we remarked in the preceding item, this means that $\varphi$ is not invertible in $H^\infty + C$. Therefore, $P_\varphi \notin Gr_{\text{res}}$ by Lemma (5.1.11). On the other hand, by the claim proved in the first paragraph of the same lemma, we know that $P_-P_\varphi\big|_{\varphi H^2}: \varphi H^2 \to H^2_-$ is compact, since $\varphi \in H^\infty$. Hence $a^* = P_-P_\varphi\big|_{H^2}$ is also compact, so that $P_\varphi \in \mathcal{B}_{cc}$. Similarly, we also find that $y = P_-P_\varphi\big|_{H^2_-}$ is compact. Now recall that a projection

$$P = \begin{pmatrix} x & a \\ a^* & y \end{pmatrix},$$

belongs to $G_{\text{res}}$ if and only if $a, y$ are compact operators and $x$ is Fredholm (see [248]). Applying this to $P = P_\varphi$ we obtain that $x = P_+P_\varphi\big|_{H^2}$ is not Fredholm. In order to prove that $P_\varphi \in \mathbb{E}_1$, it only remains to verify that $x$ is not compact. To this end, it suffices to show that $\dim\ker(x-1) = \infty$. But since $\varphi \in H^\infty$, we have $\ker(x-1) = H^2 \cap \varphi H^2 = \varphi H^2$, which has infinite dimension. The converse is an immediate consequence of Lemma (5.1.11) and the characterization of invertible inner functions in $H^\infty + C$.

We shall give examples of shift-invariant subspaces which can or cannot be joined by a (minimal) geodesic in the Grassmann manifold $Gr$. The simplest case is a consequence of the following result proved in [72] for Hardy spaces of the upper half-plane. It is an elementary but important step to understand Toeplitz kernels. We shall state it for the Hardy space of the circle.
**Lemma (5.1.19)[245]:** Let $\varphi, \psi$ be two inner functions. Then $\ker(T_{\varphi\bar{\psi}}) \neq \{0\}$ if and only if there exist an inner function $\theta$ and an outer function $g$ such that $\varphi\theta g = \psi\bar{g}$ on $\mathbb{T}$.
**Example (5.1.20)[245]:** Suppose that $\varphi$ divides $\psi$. This means that there is an inner function $\theta$ such that $\varphi\theta = \psi$. Thus, the equation in Lemma (5.1.19) is satisfied with $g = 1$, and consequently, $\ker(T_{\varphi\bar{\psi}}) \neq \{0\}$. Hence there is no geodesic in $Gr$ joining $\varphi H^2$ and $\psi H^2$. Note that $\ker(T_{\bar{\varphi}\psi}) = \{0\}$. In this case, it is not difficult to construct concrete examples using the following well-known description of divisors in $H^\infty$. Suppose that $\{a_j\}$ and $\{a'_j\}$ are the zero sets of $\varphi$ and $\psi$, respectively. If $\varphi = \lambda b s_\mu$ and $\psi = \lambda' b' s_{\mu'}$ are the canonical factorizations, then $\varphi$ divides $\psi$ if and only if $\{a_j\} \subseteq \{a'_j\}$ and $\mu \leq \mu'$.
The canonical factorization factorization also turns out to be relevant to give an affirmative answer to the existence of a geodesic in many concrete cases. Let $\varphi$ be an inner function. A point on $\mathbb{T}$ belongs to the support of $\varphi$ if it is a limit point of zeros of $\varphi$ or if it belongs to the support of the singular measure associated with the singular factor of $\varphi$. We write $\operatorname{supp}(\varphi)$ for the support of $\varphi$. Sarason and Lee proved the following [267].
**Theorem (5.1.21)[245]:** Let $\varphi, \psi$ be inner functions.
i) If $\operatorname{supp}(\varphi) \neq \operatorname{supp}(\psi)$, then the spectrum of $T_{\varphi\bar{\psi}}$ is the closed unit disk.

ii) If there is a point $z_0 \in \operatorname{supp}(\psi) \setminus \operatorname{supp}(\varphi)$, then $T_{\varphi\bar{\psi}} - \lambda$ has dense range for all $\lambda$.
From the above result and Theorem (5.1.5) we obtain this example.
**Example (5.1.22)[245]:** Let $\varphi, \psi$ be inner functions. Suppose that there are two points $z_0$ and $z_1$ such that $z_0 \in \operatorname{supp}(\psi) \setminus \operatorname{supp}(\varphi)$ and $z_1 \in \operatorname{supp}(\varphi) \setminus \operatorname{supp}(\psi)$. Then there is unique minimal geodesic in $Gr$ joining $P_\varphi$ and $P_\psi$ of the form stated in Theorem (5.1.5).
Now we consider the case of two inner functions with support $z = 1$. As a direct consequence of the results on Toeplitz kernels obtained by Makarov, Mitkovski and Poltoratski [72], [96] (see also the survey [264] ), one can show examples of the two inner functions of the aforementioned type such that their corresponding subspaces can or cannot be joined by a geodesic in $Gr$. These remarkable results were proved for Toeplitz operators in Hardy spaces of the upper-half plane (and other classes of functions). For this reason, we shall change to the half-plane; however by the isometry exhibited below all can be translated to the disk.
A function $F$ holomorphic on the upper half-plane $\mathbb{C}_+ = \{z: \operatorname{Im} z > 0\}$ belongs to the Hardy space $H_+^2 = H^2(\mathbb{C}_+)$if

$$\| F \|_{H_+^2} := \left( \sup_{y>0} \int_{-\infty}^{\infty} |F(x+iy)|^2 dx \right)^{1/2} < \infty.$$

As in Hardy spaces of the disk, one may consider $H_+^2$ as a Hilbert subspace of $L^2(\mathbb{R})$ since non tangencial limits exist a.e. No confusion will arise if we also denote by $P_+$the orthogonal projection of $L^2(\mathbb{R})$ onto $H_+^2$. The Toeplitz operator with symbol $U \in L^\infty(\mathbb{R})$ is defined by

$$T_U: H_+^2 \to H_+^2,\ T_U(F) := P_+(UF).$$

We write $H_+^\infty = H^\infty(\mathbb{C}_+)$for the bounded holomorphic functions on $\mathbb{C}_+$. Notice that $w = \frac{z-i}{z+i}$ is a conformal map from $\mathbb{C}_+$onto $\mathbb{D}$. Set $f(w) = F(z)$. Then, it follows that $F(z) \in H_+^\infty$ if and only if $f(w) \in H^\infty$. However, $H_+^2$ is not obtained from $H^2$ by conformal mapping. It can be shown that $f(w) \in H^2$ if and only if $\frac{\pi^{-1/2}}{(z+i)} F(z) \in H_+^2$. Taking boundary values, one sees that

$$W: H^2 \to H_+^2,\ Wf(x) = \frac{\pi^{-1/2}}{(x+i)} f\left(\frac{x-i}{x+i}\right),\ x \in \mathbb{R}.$$

is an isometry from $H^2$ onto $H_+^2$. Set $\gamma(x) = \frac{x-i}{x+i}$ and fix $\theta \in L^\infty$. Then, Toeplitz operators in the Hardy spaces of the disk and the upper half-plane are related by

$$WT_\theta = T_{\theta\circ\gamma} W.$$

The canonical factorization of functions in $H^2$ can be also derived in $H_+^2$ using the isometry $W$.
By an inner function $\Theta$ in $\mathbb{C}_+$we mean that $\Theta \in H_+^\infty$ and $|\Theta| = 1$ on $\mathbb{R}$. An inner function $\Theta(z)$ in $\mathbb{C}_+$is a meromorphic inner function if it has a meromorphic extension to C. In this case, the meromorphic extension to the lower half-plane is given by $\Theta(z) = \frac{1}{\Theta(\bar{z})}$. Each meromorphic inner function $\Theta$ admits a canonical factorization $\Theta = B_\Lambda S^a$, where $a \geq 0$ and $\Lambda$ is a discrete set in $\mathbb{C}_+$without accumulation points on $\mathbb{R}$ such that the following Blaschke condition holds

$$\sum_{\lambda\in\Lambda} \frac{\operatorname{Im} \lambda}{1 + |\lambda|^2} < \infty.$$

The function $B_\Lambda$ is the corresponding Blaschke product in $\mathbb{C}_+$, i.e.

$$B_\Lambda(z) = \prod_{\lambda\in\Lambda} \epsilon_\lambda \frac{z-\lambda}{z-\bar{\lambda}};\ |\epsilon_\lambda| = 1.$$

The other function in the factorization is given by the singular inner function $S^a(z) = e^{iaz}$. Meromorphic inner functions correspond to inner functions in $H^2$ such that $z = 1$ is the only possible accumulation point of their zeros and also the only possible singular point mass.

**Example (5.1.23)[245]:** The point spectrum of a meromorphic inner function $\Theta = B_\Lambda S^a$ is the set $\sigma(\Theta) = \{\Theta = 1\}$ or $\{\Theta = 1\} \cup \{\infty\}$. The point $\infty$ belongs to the spectrum if $\sum_{\lambda\in\Lambda} \operatorname{Im}\lambda < \infty$ and $S^a \equiv 1$ (see [72] for other equivalent conditions). Two meromorphic inner functions are said to be twins if they have the same point spectrum, possibly including infinity. The twin inner function theorem asserts that if $\Theta, J$ are twins, then $\ker\left(T_{\bar{\Theta}J}\right) = \{0\}$[72]. Thus, there is always a geodesic joining the corresponding subspaces defined by twin functions.

**Example (5.1.24)[245]:** Recall that a sequence of real numbers is separated if $|\lambda_n - \lambda_m| \geq \delta > 0$ $(n \neq m)$. A separated sequence $(\lambda_n)_{n\in\mathbb{Z}}$ is a called a Pólya sequence if every zero-type entire function bounded on $(\lambda_n)_{n\in\mathbb{Z}}$ is constant (see also [96] for a new characterization). Among several conditions, it was proved in [96] that $(\lambda_n)_{n\in\mathbb{Z}}$ is a Pólya sequence if and only if there exists a meromorphic inner function $\Theta$ with $\{\Theta = 1\} = (\lambda_n)_{n\in\mathbb{Z}}$ such that $\ker(T_{\bar{\Theta}}S^{2c}) \neq \{0\}$ for some $c > 0$. Hence there is no geodesic joining the corresponding subspaces defined by $\Theta$ and $S^{2c}$.

**Section (5.2): Multipliers between Model Spaces**

For an inner function $\Theta$, let $\mathcal{K}_\Theta := H^2 \cap (\Theta H^2)^\perp$ denote the model space of the open unit disk $\mathbb{D}$ corresponding to $\Theta$. We explore, for a pair of inner functions $u$ and $v$, the multipliers

$$\mathcal{M}(u,v) := \{\varphi \in \operatorname{Hol}(\mathbb{D}) : \varphi\mathcal{K}_u \subseteq \mathcal{K}_v\}$$

between $\mathcal{K}_u$ and $\mathcal{K}_v$.

One motivation for comes from the work of Crofoot [285] who considered a more restricted version of $\mathcal{M}(u,v)$ namely $\{\varphi \in \operatorname{Hol}(\mathbb{D}) : \varphi\mathcal{K}_u = \mathcal{K}_v\}$, in other words, the multipliers from $\mathcal{K}_u$ onto $\mathcal{K}_v$ (see also [280]). As it turns out, these onto multipliers are unique up to multiplicative constants and are outer functions. Another motivation comes from examining pre-orders on partial isometries [288], [297].

The Crofoot discussion becomes quite different if we relax the (onto) multiplier condition $\varphi\mathcal{K}_u = \mathcal{K}_v$ to just $\varphi\mathcal{K}_u \subseteq \mathcal{K}_v$. For one, as we shall see below, these (into but not necessarily onto) multipliers need not be outer functions. Secondly, unlike the onto multipliers, the into multipliers need not be unique. In fact, we give an example of when $\mathcal{M}(u,v)$ is infinite dimensional and contains unbounded functions.

After a few initial observations about $\mathcal{M}(u,v)$ we will reformulate the description of $\mathcal{M}(u,v)$ in terms of Carleson measures of model spaces and kernels of Toeplitz operators. Along the way, we will describe $\mathcal{M}(u,v)$ when $v$ is an inner multiple of $u$. We will then relate $\mathcal{M}(u,v)$ to the boundary spectra of $u$ and $v$ along with their sub-level sets.

We also consider multipliers for the model spaces of the upper half plane. In this setting we discuss a particular entire function introduced by Lyubarskii and Seip which allows us to deduce the existence of unbounded onto multipliers connecting to a question raised by Crofoot. As discussed earlier, the onto multipliers are unique (up to multiplicative constants) and thus the multipliers algebra in this case is one dimensional. In the spirit of the Lyubarskii and Seip construction above, we produce $u$ and $v$ such that $\mathcal{M}(u,v) = \mathbb{C}\varphi$, yet $\varphi$ is not an onto multiplier.

We assume we are familiar with the Hardy space $H^2$[286], [290] and model spaces $\mathcal{K}_u$[289], [98]. $\mathbb{D}$ is the open unit disk, $\mathbb{T}$ the unit circle, $m$ normalized Lebesgue measure on $\mathbb{T}$, and $L^2$ the standard Lebesgue space $L^2 := L^2(\mathbb{T}, m)$ with norm $\| f \|$ and inner product $\langle\cdot,\cdot\rangle$. The bounded analytic functions on $\mathbb{D}$ are denoted by $H^\infty$. Recall that $H^2$ is a reproducing kernel Hilbert space with kernel $k_\lambda(z) = (1 - \bar{\lambda}z)^{-1}$.
We begin with some useful observations. First notice that $\mathcal{M}(u, v) \subseteq H^2$. Indeed, if

$$k_\lambda^u(z) = \frac{1 - \overline{u(\lambda)}u(z)}{1 - \bar{\lambda}z}, \ \lambda, z \in \mathbb{D},$$

denotes the reproducing kernel for $\mathcal{K}_u$, then $k_0^u = 1 - \overline{u(0)}u \in \mathcal{K}_u$ is an invertible element of $H^\infty$. Thus if $\varphi \in \mathcal{M}(u, v)$ then $\varphi k_0^u \in \mathcal{K}_v \subseteq H^2$ from which the result follows.

Furthermore, when $\varphi \in \mathcal{M}(u, v)$, the closed graph theorem says that $M_\varphi f = \varphi f$ is a bounded operator from $\mathcal{K}_u$ to $\mathcal{K}_v$ and standard arguments show that $M_\varphi^* k_\lambda^v = \overline{\varphi(\lambda)} k_\lambda^u$. Since

$$\|k_\lambda^u\|^2 = \frac{1 - |u(\lambda)|^2}{1 - |\lambda|^2},$$

it follows that

$$|\varphi(\lambda)|^2 (1 - |u(\lambda)|^2) \lesssim (1 - |v(\lambda)|^2), \ \lambda \in \mathbb{D}. \tag{4}$$

Though this inequality will be used later on, it does not prove that $\varphi$ is always bounded. The following Proposition summarizes some basic facts which follow, or can be gleaned, from Crofoot's [285].

**Proposition (5.2.1)[277]:** Let $u$ and $v$ be inner functions.
(i) $\mathcal{M}(u, u) = \mathbb{C}$.
(ii) If $\varphi\mathcal{K}_u = \mathcal{K}_v$ then $\varphi$ is outer.
(iii) $\mathbb{C} \subseteq \mathcal{M}(u, v)$ if and only if $u$ divides $v$
(iv) Suppose $u$ divides $v$ and $u$ is not a constant multiple of $v$. Then $\mathcal{M}(v, u) = \{0\}$.
(v) If $\varphi \in \mathcal{M}(u, v)$ and $F$ is the outer factor of $\varphi$, then $F \in \mathcal{M}(u, v)$.
(vi) If $a \in \mathbb{D}$ and $u_a := \frac{u-a}{1-\bar{a}u}$, then $\frac{1}{1-\bar{a}u}\mathcal{K}_u = \mathcal{K}_{u_a}$.
The map $f \mapsto (1 - \bar{a}u)^{-1} f$ from $\mathcal{K}_u$ onto $\mathcal{K}_{u_a}$ is a constant multiple of the unitary Crofoot transform. Using operator theory techniques, Crofoot [285] showed that when the space of onto multipliers is non-empty, then $\sigma(u) = \sigma(v)$, where

$$\sigma(u) := \left\{ \xi \in \mathbb{T} : \varliminf_{z \to \xi} |u(z)| = 0 \right\}$$

is the boundary spectrum of an inner function. The following result is the $\mathcal{M}(u, v)$ analogue of this where our proof uses function theory.

**Proposition (5.2.2)[277]:** If $\mathcal{M}(u, v) \neq \{0\}$ then $\sigma(u) \subseteq \sigma(v)$.

**Proof.** Without loss of generality, we can use Proposition (5.2.1) (vi) and assume that $u(0) = 0$ (the Crofoot transform preserves the regular points in $\mathbb{T}$). Then $1 \in \mathcal{K}_u$ and so $\varphi\mathcal{K}_u \subseteq \mathcal{K}_v \Longrightarrow \varphi \in \mathcal{K}_v$. Pick $\zeta \in \mathbb{T} \setminus \sigma(v)$ (a regular point for $v$). Then [289] every function in $\mathcal{K}_v$ has an analytic continuation to a two-dimensional open neighborhood $\Omega$ of $\zeta$. In particular, $\varphi \in \mathcal{K}_v$ enjoys this property. For every $f \in \mathcal{K}_u, g := \varphi f \in \mathcal{K}_v$ has an analytic continuation to $\Omega$ and so $f = g/\varphi$ is either analytic on $\Omega$ or has a pole of order at least 1 at $\zeta$. But this second case is not possible since $f \in H^2$ must be square integrable on $\mathbb{T}$. Hence $f$ extends analytically to $\Omega$ and thus $\zeta \in \mathbb{T} \setminus \sigma(u)$.

We reformulate the description of $\mathcal{M}(u, v)$ in terms of kernels of Toeplitz operators and Carleson measures for model spaces.

**Theorem (5.2.3)[277]:** For inner $u$ and $v$ and $\varphi \in H^2$, the following are equivalent:
(i) $\varphi \in \mathcal{M}(u,v)$;
(ii) $\varphi S^*u \in \mathcal{K}_v$ and $|\varphi|^2 dm$ is a Carleson measure for $\mathcal{K}_u$, i.e.,

$$\int_{\mathbb{T}} |f|^2|\varphi|^2 dm \lesssim \| f \|^2,\ f \in \mathcal{K}_u;$$

(iii) $\varphi \in \operatorname{Ker} T_{\overline{zv}u}$ and $|\varphi|^2 dm$ is a Carleson measure for $\mathcal{K}_u$, where $T_{\overline{zv}u} f = P_+(\overline{zv}uf)$ is the standard Toeplitz operator on $H^2$.
Furthermore, the following are equivalent:
(iv) $\varphi \in \mathcal{M}(u,v) \cap H^\infty$;
(v) $\varphi S^*u \in \mathcal{K}_v \cap H^\infty$.
(vi) $\varphi \in \operatorname{Ker} T_{\overline{zv}u} \cap H^\infty$.
**Proof.** Recall that $\operatorname{Ker} T_{\bar{u}} = \mathcal{K}_u$ [289] and that $T_{fg} = T_f T_g$ if either $\bar{f} \in H^\infty$ or $g \in H^\infty$[289]. Also observe that $T_{\bar{z}} = S^*$ and that $T_{1-u(0)\bar{u}}$ is invertible. Using these facts, along with the identity (on $\mathbb{T}$),

$$\varphi S^* u = \varphi \bar{z}(u - u(0)) = \varphi \bar{z} u(1 - u(0)\bar{u}), \tag{5}$$

it follows that $\varphi S^*u \in \mathcal{K}_v \Leftrightarrow \varphi \in \operatorname{Ker} T_{\overline{zv}u}$. This yields (ii) $\Leftrightarrow$ (iii) and $(vi) \Longrightarrow (v)$. The implication $(v) \Longrightarrow (vi)$ needs an additional argument. Indeed, suppose that $\varphi S^*u \in \mathcal{K}_v \cap H^\infty$. Then the above equivalences yield $\varphi \in \operatorname{Ker} T_{\overline{zv}u}$, and we just have to check that $\varphi$ is bounded. We already know that $\varphi \in H^2$. Thus in order to verify $\varphi \in H^\infty$, it suffices to prove that $\varphi|_{\mathrm{T}} \in L^\infty$ (Smirnov's theorem [286]). By assumption, $\varphi S^* u = g \in H^\infty$ and (5) shows that $\varphi|_{\mathrm{T}} \in L^\infty$.
The implications $(i) \Longrightarrow (ii)$ and $(iv) \Longrightarrow (v)$ are automatic. The implication $(v) \Longrightarrow (iv)$ becomes automatic once we have shown $(ii) \Longrightarrow$ (i). So it remains to prove $(ii) \Longrightarrow (i)$. Observe that $f \in \mathcal{K}_v$ if and only if $v\overline{zf} \in \mathcal{K}_v$ (see [289]). We know that $\varphi S^*u \in K_v$ which means, via (5) that $v\overline{u\varphi} \in H^2$. Since $|\varphi|^2 dm$ is a Carleson measure for $\mathcal{K}_u$ (i.e., $\varphi \in \mathcal{M}(\mathcal{K}_u, H^2)$) it suffices to show that $\varphi g \in \mathcal{K}_v$ for all $g \in \mathcal{K}_u \cap H^\infty$ (which is dense in $\mathcal{K}_u$ ). Indeed,

$$v\overline{z\varphi g} = v\overline{u\varphi} \cdot u\overline{zg} \in H^2 \cdot H^\infty \subseteq H^2.$$

**Corollary (5.2.4)[277]:** $\operatorname{Ker} T_{\overline{zv}u} \cap H^\infty = \mathcal{M}(u,v) \cap H^\infty \subseteq \mathcal{M}(u,v) \subseteq \operatorname{Ker} T_{\overline{zv}u}$.
We will see in Example (5.2.7) below that, in general, $\mathcal{M}(u,v) \subsetneq \operatorname{Ker} T_{\overline{zv}u}$.
**Corollary (5.2.5)[277]:** Suppose $u$ and $v$ are inner and $v = uI$. Then the following are equivalent:
(i) $\varphi \in \mathcal{M}(u,v)$;
(ii) $\varphi \in \mathcal{K}_{zI}$ and $|\varphi|^2 dm$ is a Carleson measure for $\mathcal{K}_u$.
Furthermore, the following are equivalent:
(iii) $\varphi \in \mathcal{M}(u,v) \cap H^\infty$;
(iv) $\varphi \in \mathcal{K}_{zI} \cap H^\infty$.
If $I$ is a finite Blaschke product then $\mathcal{M}(u,v) \cap H^\infty = \mathcal{M}(u,v) = \mathcal{K}_{zI}$.
Our next result uses analytic continuation and the boundary spectrum to construct a class of inner functions $u$ and $v$, with $v = uI$, such that the Carleson condition on $|\varphi|^2 dm$ is automatic as soon as $\varphi \in \mathcal{K}_{zI}$.
**Theorem (5.2.6)[277]:** Let $u$ and $v$ be inner functions with and $v = uI$ for some inner function $I$. Suppose further that $\sigma(u) \cap \sigma(I) = \emptyset$. Then $\mathcal{M}(u,v) = \mathcal{K}_{zI}$. Furthermore, if $I$ is not a finite Blaschke product then $\mathcal{M}(u,v)$ contains unbounded functions.

**Proof.** By Corollary (5.2.5), we just need to check that $|\varphi|^2 dm$ is a Carleson measure for $\mathcal{K}_u$ for every $\varphi \in \mathcal{K}_{zI}$. Let $V$ be a two dimensional neighborhood of $\sigma(I)$ that is far from $\sigma(u)$. By [289] $\varphi$ extends analytically outside $V$ (i.e., $\overline{\mathbb{D}} \setminus V$ ) and thus can be assumed to be bounded outside $V$. Similarly, every $f \in \mathcal{K}_u$ extends analytically to $V$ and can be assumed to be bounded there. From here it follows that $\varphi f \in H^2$. By the Closed Graph Theorem, $\varphi \in \mathcal{M}(\mathcal{K}_u, H^2)$, equivalently, $|\varphi|^2 dm$ is a Carleson measure for $\mathcal{K}_u$.
For the last part, note that if $I$ is not a finite Blaschke product then $\mathcal{K}_{zI}$ is infinite dimensional [289] and thus, via a well-known theorem of Grothendieck [295], contains unbounded functions.
We now construct an example of when $\operatorname{Ker} T_{zI} = \operatorname{Ker} T_{\overline{zv}u}$ contains functions which do not define Carleson measures for $\mathcal{K}_u$ and thus $\mathcal{M}(u,v) \subsetneq \operatorname{Ker} T_{\overline{zv}u}$. Hence the Carleson condition is important in Theorem (5.2.3).
**Example (5.2.7)[277]:** Set $\lambda_n = 1 - 2^{-n}, n \geqslant 1$, and note this is the zero sequence of an interpolating Blaschke product $I$. With $w_n = n^{-1}$, notice that $\sum_{n\geqslant 1} w_n^2 < \infty$. By an interpolation theorem from [98], there is a $\varphi \in \mathcal{K}_I \subseteq \mathcal{K}_{zI}$ such that

$$\varphi(\lambda_n) = \frac{w_n}{(1 - |\lambda_n|^2)^{1/2}} \asymp \frac{2^{n/2}}{n} \to \infty.$$

Now take $u(z) = \exp((z+1)/(z-1))$ and observe that since $\lambda_n \to 1$ on $(0,1)$ we have $u(\lambda_n) \to 0$. If $v = uI$ then $\varphi \in \mathcal{K}_I \subseteq \mathcal{K}_{zI} = \operatorname{Ker} T_{\overline{zv}u}$. However, $\varphi \notin \mathcal{M}(u,v)$ since, if it were, (4) would imply that

$$|\varphi(\lambda_n)|^2 (1 - |u(\lambda_n)|^2) \lesssim 1 - |v(\lambda_n)|^2 \lesssim 1.$$

The above discussion now yields a contradiction. Thus we have $\mathcal{M}(u,v) \subsetneq \operatorname{Ker} T_{\overline{zv}u} = \mathcal{K}_{zI}$.

We now consider finite dimensional model spaces. For an inner $u$, the degree of $u$ is $n$ if $u$ is a finite Blaschke product with $n$ zeros and equal to $\infty$ otherwise. When $u$ is a finite Blaschke product with $n$ zeros $\{\lambda_1, \cdots, \lambda_n\}$, we have

$$\mathcal{K}_u = \left\{ \frac{p(z)}{\prod_{j=1}^n (1 - \bar{\lambda}_j z)} : p \in \mathcal{P}_{n-1} \right\}. \tag{6}$$

where $\mathcal{P}_{n-1}$ are the polynomials of degree at most $n-1$.
**Theorem (5.2.8)[277]:** If $u$ is a finite Blaschke product with zeros $\{a_1, \dots, a_m\}$ and $v$ is a finite Blaschke product with zeros $\{b_1, \dots, b_n\}$ where $m \leqslant n$, and the zeros are repeated according to their multiplicity, then

$$\mathcal{M}(u,v) = \mathcal{M}(u,v) \cap H^\infty = \left\{ q(z) \frac{\prod_{i=1}^m (1 - \bar{a}_i z)}{\prod_{j=1}^n (1 - \bar{b}_j z)} : q \in \mathcal{P}_{n-m} \right\}.$$

**Proof.** The $\supseteq$ containment follows essentially from (6). For the $\subseteq$ containment, notice from Theorem (5.2.3) that $\varphi \in \mathcal{M}(u,v) \Longrightarrow \varphi \in \operatorname{Ker} T_{\overline{zv}u}$ which is equivalent to

$$u\varphi \in \operatorname{Ker} T_{\overline{zv}} = \mathcal{K}_{zv} = \left\{ \frac{p(z)}{\prod_{j=1}^n (1 - \bar{b}_j z)} : p \in \mathcal{P}_n \right\} \subseteq H^\infty.$$

The result now follows.
**Theorem (5.2.9)[277]:** If $u$ is a finite Blaschke product and $v$ is any inner function with infinite degree, then $\mathcal{M}(u,v) \cap H^\infty \neq \{0\}$.
**Proof.** By [290] there is an $a \in \mathbb{D}$ (in fact "most" ) such that the Frostman shift $v_a = \frac{v-a}{1-\bar{a}v}$ of $v$ is a Blaschke product of infinite degree. Factor $v_a = IJ$, where $I$ and $J$ are Blaschke products with the degree of $I$ equal to the degree of $u$, and use [98] to obtain $\mathcal{K}_I \subseteq \mathcal{K}_{v_a}$.

From Theorem (5.2.8) there is a rational $\varphi \in H^\infty$ such that $\varphi\mathcal{K}_u \subseteq \mathcal{K}_I \subseteq \mathcal{K}_{v_a}$. Proposition (5.2.1)
(vi) now yields $(1-\bar{a}v)\varphi\mathcal{K}_u \subseteq \mathcal{K}_v$

We discuss some results using sub-level sets of inner functions. We start with a "maximum principle" result of Cohn [284].

**Theorem (5.2.10)[277]:** Suppose $\Theta$ is inner and $f \in \mathcal{K}_\Theta$ is bounded on $\{|\Theta| < \epsilon\}$ for some $\epsilon \in (0,1)$. Then $f \in H^\infty$.

This result can be used to show that under certain circumstances, all multipliers must be bounded.

**Corollary (5.2.11)[277]:** Let $u$ and $v$ be inner. If, for some $\epsilon_1, \epsilon_2 \in (0,1), \{|v| < \epsilon_2\} \subseteq \{|u| < \epsilon_1\}$, then $\mathcal{M}(u,v) = \operatorname{Ker} T_{\overline{zv}u} \cap H^\infty$.

**Proof.** Let $\varphi \in \mathcal{M}(u,v)$. The estimate in (4) says that when $\lambda \in \{|v| < \epsilon_2\} \subseteq \{|u| < \epsilon_1\}$ we have $|\varphi(\lambda)|^2 \lesssim (1-\epsilon_1^2)^{-1}$ and thus $\varphi$ is bounded on $\{|v| < \epsilon_2\}$. Since $k_0^u = 1 - \overline{u(0)}u \in \mathcal{K}_u$ and bounded on $\mathbb{D}$, we see that $k_0^u\varphi \in \mathcal{K}_v$ and bounded on $\{|v| < \epsilon_2\}$. Apply Theorem (5.2.10) to obtain $k_0^u\varphi \in H^\infty$. Since $k_0^u$ is invertible in $H^\infty$, we get $\varphi \in H^\infty$. Now apply Corollary (5.2.4).

**Example (5.2.12)[277]:** Let $u$ be any singular inner function and $v = u^\alpha$ for some $\alpha > 1$ (or perhaps $u$ a Blaschke product, or any inner function, and $\in \mathbb{N}$ ). Notice that $u$ divides $v$ and so $\mathcal{M}(u,v) \neq \{0\}$ (Corollary (5.2.5)). Furthermore if $\epsilon_2 \in (0,1)$ and $z \in \{|v| < \epsilon_2\}$ then $|u(z)|^{1/\alpha} \leqslant \epsilon_2^{1/\alpha}$. Setting $\epsilon_1 = \epsilon_2^{1/\alpha}$ we see that $\{|v| < \epsilon_2\} \subseteq \{|u| < \epsilon_1\}$. Corollary (5.2.11) yields $\mathcal{M}(u,v) \subseteq H^\infty$. Combine this with Corollary (5.2.5) to see that $\mathcal{M}(u,v) = \mathcal{K}_{zu^{\alpha-1}} \cap H^\infty$.

Carleson measure results of Cohn [282], [283] allow us, in the special case where $u$ satisfies the connected level set condition (i.e., $\{|u| < \varepsilon\}$ is connected for some $\varepsilon > 0$), to replace the condition that $|\varphi|^2 dm$ is a Carleson measure in Theorem (5.2.3) and Corollary (5.2.5) with

$$\sup_{\lambda \in \mathbb{D}} (1-|u(\lambda)|^2) \int_{\mathbb{T}} \frac{1-|\lambda|^2}{|\xi-\lambda|^2} |\varphi(\xi)|^2 dm(\xi) < \infty.$$

We will now turn to the upper half plane which in certain situations is a more appropriate setting. If $\mathbb{C}_+$denotes the upper-half plane, we set $\mathcal{H}^2$ to be the corresponding Hardy space. There is a natural unitary operator $\mathcal{U}$ from $H^2$ onto $\mathcal{H}^2$ given by

$$(\mathcal{U}f)(z) := \frac{1}{\sqrt{\pi}(z+i)} f(\omega(z)),$$

where $\omega(z) := \frac{z-i}{z+i}$ maps $\mathbb{C}_+$onto $\mathbb{D}$ and $\mathbb{R} \cup \{-\infty, \infty\}$ onto $\mathbb{T}$. As with $H^2$, one can define, for $\Psi \in L^\infty(\mathbb{R})$, the Toeplitz operator $T_\Psi$ on $\mathcal{H}^2$.

For an inner function $U$ on $\mathbb{C}_+$, we define the model space

$$\mathcal{K}_U := \mathcal{H}^2 \cap (U\mathcal{H}^2)^\perp.$$

The corresponding reproducing kernel function for $\mathcal{K}_U$ is

$$K_\lambda^U(z) := \frac{i}{2\pi} \frac{1-\overline{U(\lambda)}U(z)}{z-\bar{\lambda}},\ \lambda, z \in \mathbb{C}_+.$$

Note that if $u$ is an inner function on $\mathbb{D}$ and $U = u \circ \omega$, then $U$ is an inner function on $\mathbb{C}_+$(and vice versa). Furthermore, $\mathcal{U}\mathcal{K}_u = \mathcal{K}_U$.

We need the elementary Blaschke factor on $\mathbb{C}_+$with zero at :

$$b_i^+(z) := \frac{z-i}{z+i},$$

and
$$k_i(z)=\frac{1}{\sqrt{\pi}}\frac{1}{z+i},$$
the corresponding kernel at $i$. Observe that $\mathcal{U}f=k_i\times(f\circ\omega), f\in H^2$.
We begin with some elementary but useful facts. The proofs are straightforward.
**Lemma (5.2.13)[277]:** Let $\psi\in L^\infty(\mathbb{T})$ and $\Psi=\psi\circ\omega$. Then
$$f\in \operatorname{Ker} T_\psi \Leftrightarrow F:=\mathcal{U}f\in \operatorname{Ker} T_\Psi.$$
**Lemma (5.2.14)[277]:** $\varphi\in\mathcal{M}(u,v)$ if and only if $\Phi=\varphi\circ\omega\in\mathcal{M}(U,V)$.
**Corollary (5.2.15)[277]:** With the notation from above, the following are equivalent for $\Phi$ analytic on $\mathbb{C}_+$:
(i) $\Phi\in\mathcal{M}(U,V)$;
(ii) $\Phi k_i\in \operatorname{Ker}T\overline{b_i^+VU}$ and $|\Phi|^2dx$ is a Carleson measure for $\mathcal{K}_U$.
We now discuss a situation when the Carleson condition becomes more tractable. We begin with a result from Baranov [279].
**Theorem (5.2.16)[277]:** Let $U$ be an inner function in $\mathbb{C}_+$such that $|U'(x)|\asymp 1, x\in\mathbb{R}$. For a positive Borel measure $\mu$ on $\mathbb{R}$, the following are equivalent:
(i) $\mu$ is a Carleson measure for $\mathcal{K}_U$.
(ii) We have $M:=\sup_{x\in\mathbb{R}}\mu([x,x+1])<\infty$.
**Theorem (5.2.17)[277]:** Let $U$ and $V$ be inner functions with $|U'(x)|\asymp 1, x\in\mathbb{R}$.
Then
$$\mathcal{M}(U,V)=\left\{\Phi\in(z+i)\operatorname{Ker}T_{\overline{b_i^+VU}}: M\coloneqq\sup_{x\in\mathbb{R}}\int_x^{x+1}|\Phi(t)|^2dt<\infty\right\}.$$
**Proof.** Observe that $\Phi k_i\in\operatorname{Ker}T_{\overline{b_i^+VU}}\Leftrightarrow\Phi\in(z+i)\operatorname{Ker}T_{\overline{b_i^+VU}}$ and apply Corollary (5.2.15) and Theorem (5.2.16).
**Lemma (5.2.18)[277]:** We have
$$F\in\operatorname{Ker}T_{\bar{V}U}\Leftrightarrow F\in\left((z+i)\operatorname{Ker}T_{\overline{b_i^+VU}}\right)\cap\mathcal{H}^2.$$
**Proof.** The function $F$ belongs to Ker $T_{U\bar{V}}$ if and only if there is a $\psi\in\mathcal{H}^2$ such that $\bar{V}UF=\bar{\psi}$. A calculation shows that
$$\overline{V(x)b_i^+(x)}U(x)F(x)k_i(x)=\overline{(\psi k_i)(x)},\ x\in\mathbb{R}.$$
Hence $Fk_i\in\operatorname{Ker}T_{\overline{b_i^+VU}}$ and so $F\in(z+i)\operatorname{Ker}T\overline{b_i^+VU}\cdot$
The converse argument is in the same spirit. Indeed, when
$$F\in(z+i)\operatorname{Ker}T_{\overline{b_i^+VU}}\backsim\mathcal{H}^2,$$
we get $F(x)(\overline{V(x)}U(x))=\overline{\psi(x)}(x-i)=\overline{\psi(x)(x+\iota)}$. Since $F\in\mathcal{H}^2$, and $U\bar{V}$ is bounded, we deduce that $\psi(z+i)\in\mathcal{H}^2$, and so $F\in\operatorname{Ker}T_{\bar{V}U}$.
**Corollary (5.2.19)[277]:** Let $U$ and $V$ be inner functions with $|U'(x)|\asymp 1, x\in\mathbb{R}$. Then $\mathcal{M}(U,V)\cap\mathcal{H}^2=\operatorname{Ker}T_{U\bar{V}}$.

We notice that an example constructed in [293] answers a question of Crofoot [285]. We will state the result for the model spaces $\mathcal{K}_\Theta$ of the upper-half plane and then use Lemma (5.2.14).
The construction is based on the relationship between the model subspaces generated by meromorphic inner functions and the de Branges spaces of entire functions [55].
First we define the Paley-Wiener class

$$PW = \left\{F \in \text{Hol}(\mathbb{C}): \frac{F}{e^{-i\pi z}}, \frac{F^*}{e^{-i\pi z}} \in \mathcal{H}^2\right\},\ F^*(z) := \overline{F(\bar{z})}.$$

Let $E$ be an entire function which belongs to the Hermite-Biehler class $HB$, i.e.,

$$|E(z)| \geqslant |E(\bar{z})|,\ \Im z > 0$$

and $E$ does not have any zeros in $\mathbb{C}_+^-$(the closed upper half plane). With $E \in HB$, define the de Branges space

$$\mathcal{H}(E) := \left\{F \in \text{Hol}(\mathbb{C}) : \frac{F}{E}, \frac{F^*}{E} \in \mathcal{H}^2\right\}. \tag{7}$$

The norm in $\mathcal{H}(E)$ is defined by

$$\| F \|_E = \left\|\frac{F}{E}\right\|_{L^2(\mathrm{R})},\ F \in \mathcal{H}(E).$$

If $E \in HB$, then $\Theta = E^*/E$ is a meromorphic inner function in $\mathbb{C}_+$, meaning that $\Theta$ is an inner function and that $\Theta$ has an analytic continuation to an open neighborhood of $\mathbb{C}_+^-$. Conversely, each meromorphic inner function $\Theta$ admits a representation $\Theta = E^*/E$ for some entire function $E \in HB$. One can see from the identity $\mathcal{K}_U = \mathcal{H}^2 \cap U\overline{\mathcal{H}^2}$ that when $\Theta = E^*/E$, the operator $F \mapsto F/E$ is unitary from $\mathcal{H}(E)$ onto the model space $\mathcal{K}_\Theta$, that is to say,

$$\mathcal{K}_\Theta = \frac{1}{E}\mathcal{H}(E). \tag{8}$$

When $E(z) = e^{-i\pi z}$, one can check that $E \in HB, \Theta = E^*/E$ satisfies $\Theta(z) = e^{2i\pi z}$, and $\mathcal{K}_\Theta = e^{i\pi z}\mathcal{H}(E) = e^{i\pi z}PW$.

**Theorem (5.2.20)[277]:** There are two inner functions $B$ and $\Theta$ on $\mathbb{C}_+$and an unbounded analytic function $\Psi$ on $\mathbb{C}_+$such that $\Psi\mathcal{K}_B = \mathcal{K}_\Theta$.

**Proof.** Fix $\delta \in \left(0, \frac{1}{4}\right)$ and set

$$E_\delta(z) = (z+i)\prod_{k=1}^{\infty}\left(1 - \frac{z}{k-\delta-ik^{-4\delta}}\right)\left(1 - \frac{z}{-k+\delta-ik^{-4\delta}}\right).$$

It is shown in [293] that $E_\delta \in HB$,

$$\mathcal{H}(E_\delta) = PW, \tag{9}$$

with equivalent norms, and

$$|E_\delta(x)| \simeq (1+|x|)^{2\delta}\text{dist}(x, \Lambda_\delta),\ x \in \mathbb{R}, \tag{10}$$

where

$$\begin{aligned}\Lambda_\delta &= E_\delta^{-1}(\{0\})\\ &= \left\{k-\delta-ik^{-4\delta}: k \geqslant 1\right\} \cup \left\{-k+\delta-ik^{-4\delta}: k \geqslant 1\right\} \cup \{-i\}.\end{aligned}$$

If we define $I_\delta = E_\delta^*/E_\delta$, then $I_\delta$ is a meromorphic inner function on $\mathbb{C}_+$.

Define $\Psi_\delta(z) = e^{i\pi z}E_\delta(z)$ and use (8) and (9) to obtain

$$\Psi_\delta\mathcal{K}_{I_\delta} = e^{i\pi z}E_\delta\mathcal{K}_{I_\delta} = e^{i\pi z}\mathcal{H}(E_\delta) = e^{i\pi z}PW = \mathcal{K}_\Theta,$$

where $\Theta(z) = e^{2\pi iz}$. Hence $\Psi_\delta$ is a multiplier from $\mathcal{K}_{I_\delta}$ onto $\mathcal{K}_\Theta$. We now argue that $\Psi_\delta$ is unbounded. Indeed, the zero set $\Lambda_\delta$ of $E_\delta$ contains

$$z_k = (k-\delta) - ik^{-4\delta},\ k \geqslant 1.$$

For each interval $(k-\delta, k+1-\delta)$, the zeros $z_k$ and $z_{k+1}$ lie just below the respective endpoints $k-\delta$ and $k+1-\delta$. If $x_k$ is the midpoint of $(k-\delta, k+1-\delta)$, one can see that $\text{dist}(x_k, \Lambda_\delta) \geqslant \frac{1}{2}$. From (10) we conclude that

$$|E_\delta(x_k)| \simeq (1+x_k)^{2\delta}\text{dist}(x_k, \Lambda_\delta) \gtrsim (1+x_k)^{2\delta} \simeq k^{2\delta}$$

which goes to infinity as $k \to \infty$. The fact that $\Psi_\delta$ is unbounded now follows.

This example can be transferred to the disk via $= I_\delta \circ \omega^{-1}, v = \Theta \circ \omega^{-1}, \varphi = \Psi_\delta \circ \omega^{-1}$, and applying Lemma (5.2.14).

Crofoot proved that $\varphi\mathcal{K}_u = \mathcal{K}_v \Longrightarrow \mathcal{M}(u,v) = \mathbb{C}\varphi$. A natural question to ask is whether or not $\mathcal{M}(u,v) = \mathbb{C}\varphi \Longrightarrow \varphi\mathcal{K}_u = \mathcal{K}_v$? The answer, in general, is no. Similar to Theorem (5.2.20), we construct our example in the upper-half plane setting.

**Theorem (5.2.21)[277]:** There are two inner functions $B$ and $\Theta$ on $\mathbb{C}_+$such that $\mathcal{M}(B,\Theta) = \mathbb{C}\Psi$, with $\Psi \not\equiv 0$, but $\Psi\mathcal{K}_B \subsetneq \mathcal{K}_\Theta$.

**Proof.** Let $\Theta(z) = e^{i2\pi z}$, so that $\mathcal{K}_\Theta = e^{i\pi z}PW$, and let $E(z)$ be the canonical product associated with the sequence $\Lambda = \{-i + n + \operatorname{sign}(n)\delta\}_{n\in\mathbb{Z}}$ where we now choose the limit case in the Ingham-Kadets theorem: $\delta = 1/4$. As before, set $B = E^*/E$.

It is known that the family $\mathcal{F} = \{e^{i\lambda_n x} : n \in \mathbb{Z}\}$ is complete and minimal in $L^2(-\pi,\pi)$[291], from which it can also be deduced that $E$ is of exponential type $\pi$ (see some standard computations in [291] along with a more general result [298] ). This yields the following two properties: (i) $\mathcal{H}(E) \subseteq PW$; (ii) $\operatorname{Ker} T_{\bar{\Theta}B} = \{0\}$. To see (i), observe first that on $\mathbb{R}$ we have $E(x) \simeq (1+|x|)^{-2\delta} = (1+|x|)^{-1/2}$[291] so that when $f \in \mathcal{H}(E)$ (see (7)), then

$$\int_{\mathbb{R}} \frac{|f|^2}{|E|^2} dm \simeq \int_{\mathbb{R}} |f|^2(1+|x|)dm < \infty,$$

implying that $f \in L^2(\mathbb{R})$. Moreover, since $E$ is of exponential type $\pi$, if $f \in \mathcal{H}(E)$, then $f$ is also of exponential type $\pi$. So, by an alternate definition of the Paley-Wiener space, we conclude that $f \in PW$. Property (ii) follows from the completeness of $\mathcal{F}$ which means that $\Lambda$ is a uniqueness sequence for $PW$. This is equivalent to $\operatorname{Ker} T_{\bar{\Theta}B} = \{0\}$.

We are now in a position to prove our claim. By (i), as in the proof of Theorem (5.2.20), define $\Psi(z) = e^{i\pi z}E(z)$ and use (8) and (9) to obtain $\Psi\mathcal{K}_B = e^{i\pi z}E\mathcal{K}_B = e^{i\pi z}\mathcal{H}(E) \subseteq e^{i\pi z}PW = \mathcal{K}_\Theta$, and so $\Psi \in \mathcal{M}(B,\Theta)$. By Corollary (5.2.15), the dimension of the multiplier space is bounded by that of $Ker\, T_{\overline{b_i^+ \ominus B}}$ By (ii), we have $\operatorname{Ker} T_{\bar{\Theta}B} = \{0\}$. Now $T_{\overline{b_i^+}\Theta B} = T_{\overline{b_i^+}} T_{\bar{\Theta}B}$, and $\dim \operatorname{Ker} T_{\bar{b}_i^+} = 1$, so, by injectivity of $T_{\bar{\Theta}B}$, at most one function can be sent to 0 by $T_{\overline{b_i^+}\Theta \mathrm{B}}$. So the multiplier algebra is at most one dimensional, and, since $\varphi$ already belongs to this algebra, its dimension is precisely one. Finally it is clear that the weight $(1 + |x|)$ appearing in the norm of $\mathcal{H}(E)$ does not produce an equivalent norm to that in $PW$ (one could for instance consider the family $f_n(z) = \frac{\sin(\pi(z-n))}{\pi(z-n)}\Big)$ so that $\mathcal{H}(E) \subsetneq PW$.

When $\mathcal{M}(u,v) \neq \{0\}$ we know from Proposition (5.2.2) that $\sigma(u) \subseteq \sigma(v)$. Is it the case that the boundary behavior in $\mathcal{K}_u$ is the same as in $\mathcal{K}_v$ ? To discuss this further, we need the following result of Ahern and Clark [278]: For an inner function $u$, every $f \in \mathcal{K}_u$ has a non-tangential limit at $\zeta$ if and only if

$$\underline{\lim}_{z\to\zeta} \frac{1-|u(z)|}{1-|z|} < \infty.$$

The last equivalent condition says that $u$ has a finite angular derivative at $\zeta$ and $\zeta$ is called an Ahern-Clark point for $\mathcal{K}_u$.

In the upper-half plane case note that $\infty$ is an Ahern-Clark point for a model space $\mathcal{K}_U$ precisely when $U \circ \omega^{-1}$ has a finite angular derivative at $z = 1$ (equivalently $U$ has an angular derivative at $\infty$ ). When $U$ is a Blaschke product with zeros $\mu_n$, this happens precisely when

$$\sum_{n\geqslant 1} \Im\mu_n < \infty. \tag{11}$$

**Proposition (5.2.22)[277]:** There exists two inner functions $U$ and $V$ in the upper half plane such that $\mathcal{M}(U,V)$ is non trivial, $\sigma(U)=\sigma(V)=\{\infty\}$, and $V$ has an angular derivative at $\infty$ while $U$ does not.

**Proof.** Let

$$E_1(z)=\prod_{n=1}^{\infty}\left(1+\frac{z}{2^n i}\right),\ E_2(z)=\prod_{n=1}^{\infty}\left(1-\frac{z}{2^n-2^{-2n}i}\right).$$

Standard estimates from canonical products yield

$$\left|\frac{E_1(z)}{E_2(z)}\right| \asymp \left|\frac{z+2^m i}{z-2^m+2^{-2m}i}\right|,\ |z|\in[2^m-2^{m-2},2^m+2^{m-1}].$$

Observe that this fraction is largest when $z$ is close to $2^m$ where it behaves like $2^{3m}$. Setting $\tilde{E}_2 := \left(z+\frac{i}{2}\right)^3 E_2$, we get that $E_1/\tilde{E}_2$ is bounded on $\mathbb{C}_+$ and for any $F\in\mathcal{H}(E_1)$ we have

$$\frac{F}{\tilde{E}_2}=\frac{F}{E_1}\cdot\frac{E_1}{\tilde{E}_2}.$$

Thus $F\in\mathcal{H}(\tilde{E}_2)$. Hence $E_1/\tilde{E}_2$ is a multiplier from $\mathcal{K}_U$ to $\mathcal{K}_V$ for the inner functions $U=E_1^*/E_1$ and $V=\tilde{E}_2^*/\tilde{E}_2$. The assertions about the Ahern-Clark properties follow from (11).

If $u$ is inner, we can associate [286] a unique positive finite measure $\sigma_u$ on $\mathbb{T}$, called the Clark measure, via the identity

$$\frac{1-|u(z)|^2}{|1-u(z)|^2}=\int_{\mathbb{T}}\frac{1-|z|^2}{|z-\xi|^2}d\sigma_u(\xi),\ z\in\mathbb{D}. \tag{12}$$

Note that $\sigma_u\perp m$ and that $u(0)=0$ if and only if $\sigma_u$ is a probability measure. This process can be reversed [281], [100].

We now exploit these measures to obtain additional information about multipliers. Using straightforward arguments from the theory of reproducing kernel Hilbert spaces, one obtains the following.

**Lemma (5.2.23)[277]:** Let $u,v$ be two inner functions and $\varphi\in H^2$. Then $\varphi\in\mathcal{M}(u,v)$ if and only if there exists a bounded linear operator $L_\varphi:\mathcal{K}_v\to\mathcal{K}_u$ satisfying $L_\varphi(k_\lambda^v)=\overline{\varphi(\lambda)}k_\lambda^u,\lambda\in\mathbb{D}$.

Here is the rephrasing of the lemma above in terms of Clark measures.

**Theorem (5.2.24)[277]:** Let $u,v$ be two inner functions and $\sigma_u,\sigma_v$ be their associated Clark measures. For $\varphi\in H^2$, the following are equivalent:

(i) $\varphi\in\mathcal{M}(u,v)$;

(ii) there exists a bounded linear operator $\mathfrak{L}_\varphi: L^2(\sigma_v)\to L^2(\sigma_u)$ satisfying

$$\mathfrak{L}_\varphi(k_\lambda)=\overline{\varphi(\lambda)}\frac{1-\overline{u(\lambda)}}{1-\overline{v(\lambda)}}k_\lambda,\ \lambda\in\mathbb{D}. \tag{13}$$

**Proof.** Assume that $\varphi\in\mathcal{M}(u,v)$. By Lemma (5.2.23), the (bounded) operator $L_\varphi:\mathcal{K}_v\to\mathcal{K}_u$ satisfies $L_\varphi k_\lambda^v=\overline{\varphi(\lambda)}k_\lambda^u,\lambda\in\mathbb{D}$. Define $\mathfrak{L}_\varphi := V_u^{-1}L_\varphi V_v: L^2(\sigma_u)\to L^2(\sigma_v)$, where the Clark operator $V_u: L^2(\sigma_u)\to\mathcal{K}_u$ is defined by $V_u k_\lambda=(1-\overline{u(\lambda)})^{-1}k_\lambda^u,\lambda\in\mathbb{D}$. A result of Poltoratski [294] says that every $f\in\mathcal{K}_u$ has radial limits $\sigma_u$-almost everywhere and $V_u^{-1}(f)=f$ on the carrier of $\sigma_u$. The identity in (13) now follows.

It is easy to see that the above argument can be reversed.

**Corollary (5.2.25)[277]:** Let $u, v$ be inner with associated Clark measures $\sigma_u$ and $\sigma_v$ satisfying $\sigma_u \ll \sigma_v$. If $\varphi = (1-v)/(1-u)$ and $h = d\sigma_u/d\sigma_v$, the following are equivalent: (i) $\varphi \in \mathcal{M}(u, v)$; (ii) $h \in L^\infty(\sigma_v)$.

**Proof.** $(ii) \Longrightarrow (i)$ : Using Theorem (5.2.24), $\varphi \in \mathcal{M}(u, v)$ if and only if there exists a bounded linear operator $\mathfrak{L}_\varphi: L^2(\sigma_v) \longrightarrow L^2(\sigma_u)$ such that

$$\mathfrak{L}_\varphi(k_\lambda) = \overline{\varphi(\lambda)} \frac{1-\overline{u(\lambda)}}{1-\overline{v(\lambda)}} k_\lambda = k_\lambda,\ \lambda \in \mathbb{D}.$$

For every $f \in L^2(\sigma_v)$, we have

$$\int_{\mathbb{T}} |f(\xi)|^2 d\sigma_u(\xi) = \int_{\mathbb{T}} |f(\xi)|^2 h(\xi) d\sigma_v(\xi) \leqslant \| h \|_{L^\infty(\sigma_v)} \| f \|_{L^2(\sigma_v)}^2.$$

Hence if we define $\mathfrak{L}_\varphi(f) = f$ for $f \in L^2(\sigma_v)$, then $\mathfrak{L}_\varphi$ is bounded from $L^2(\sigma_v)$ into $L^2(\sigma_u)$, which proves $(1-v)/(1-u) \in \mathcal{M}(u, v)$.

$(i) \Longrightarrow (ii)$ : Again using Theorem (5.2.24), the map $\mathfrak{L}_\varphi(k_\lambda) = k_\lambda$ extends linearly to a bounded operator from $L^2(\sigma_v)$ into $L^2(\sigma_u)$. In particular, for any $f$ in the linear span of $\{k_\lambda: \lambda \in \mathbb{D}\}$, we have

$$\int_{\mathbb{T}} |f|^2 h d\sigma_v = \int_{\mathbb{T}} |f|^2 d\sigma_u \lesssim \int_{\mathbb{T}} |f|^2 d\sigma_v.$$

Since the linear span of $\{k_\lambda: \lambda \in \mathbb{D}\}$ is dense in $L^2(\sigma_v)$ (use $\sigma_v \perp m$ along with [290]), we get $h \in L^\infty(\sigma_v)$.

It was shown in [296] that if $\sigma_u \ll \sigma_v$ and $h := d\sigma_u/d\sigma_v$, then $h \in L^2(\sigma_v)$ if and only if $(1-v)/(1-u) \in H^2$.

**Example (5.2.26)[277]:** If $v(z) = \exp((z+1)/(z-1))$, one can show [289] that the Clark measure $\sigma_v$ is discrete and given by

$$\sigma_v = \sum_{n=-\infty}^{\infty} c_n \delta_{z_n},\ z_n = \frac{2\pi i n - 1}{2\pi i n + 1},\ c_n = \frac{2}{4\pi^2 n^2 + 1}.$$

Now pick $c_n'$ satisfying $0 \leqslant c_n' \leqslant M c_n$ for some $M \geqslant 1$ and define $\mu' = \sum_{n \geqslant 1} c_n' \delta_{z_n}$. See [289]. In other words, we have $d\mu' = h d\sigma_v$, where $0 \leqslant h \leqslant M$. Then there is a unique inner function $u$ such that its associated Clark measure is precisely $\mu'$. Corollary (5.2.25) says that $(1-v)/(1-u) \in \mathcal{M}(u, v)$. This construction can be done more generally starting from any finite measure $\sum_{n \geqslant 1} c_n \delta_{z_n}$ on $\mathbb{T}$ and its associated inner function $v$. See also [288].

## Section (5.3): Toeplitz Order

Toeplitz operator $T_U$ with symbol $U \in L^\infty(\mathbb{R})$ on the Hardy space $H^2$ in the upper half-plane $\mathbb{C}_+$ is defined as

$$T_U f = P_+ U f,$$

where $P_+$ denotes the orthogonal projection from $L^2(\mathbb{R})$ onto $H^2$ (For further discussion). This standard definition can be extended to larger function spaces and more general symbols to accommodate various applications of Toeplitz-type operators in Complex and Harmonic analysis. A recently developed approach based on the use of Toeplitz operators brought new progress to the area of Uncertainty Principle in Harmonic Analysis (UP), see [72], [49], [311]. This note is devoted to further development of the Toeplitz approach.

One of the cases of the Toeplitz operator which appears most often in applications is the operator with the symbol $U = \bar{I} J$ where $I$ and $J$ are inner functions. Recall that a bounded analytic function in the upper half-plane is called inner if its boundary values are unimodular almost everywhere with respect to Lebesgue measure on the boundary.

Inner functions constitute arguably the most important collection of functions in the standard one-dimensional complex function theory. Starting with the seminal result by Beurling, which says that all closed invariant subspaces of the shift operator $Sf: f \mapsto zf$ in the Hardy space $H^2$ in the unit disk have the form $\theta H^2$ where $\theta$ is inner, these functions became a focal point of research for complex analysts. Beurling's result implies that the invariant subspaces for the operator adjoint to $S$, the backward shift operator $S^*f: f \mapsto (f - f(0))/z$, have the form $K_\theta = (\theta H^2)^\perp = H^2 \ominus \theta H^2$. This property of the spaces $K_\theta$ put them into the foundation of the famous Nagy-Foias functional model theory which says that any completely non-unitary contractive operator $T$ in a Hilbert space $H$, satisfying $\|(T^*)^n x\| \to 0$ for all $x \in H$, is unitarily equivalent to a compression of multiplication by $z$ on one of such $K_\theta$ spaces for a properly chosen inner function $\theta$ (in general, such spaces are vector-valued, see [98]).
These fundamental results demonstrated the importance of inner functions and related spaces in function theoretic problems stemming from functional analysis. Such problems became the main stream of complex function theory in the last several decades of the 20 th century. At present, inner functions are firmly established as a key ingredient of complex analysis and appear in most of its applications, including Harmonic Analysis, Control Theory, Spectral Theory of differential operators, Signal Processing and Mathematical Physics. Via the same connections, Toeplitz operators of the type $T_{\bar{I}J}$ where $I$ and $J$ are inner functions, appear in many of such applications.
Problems on injectivity and invertibility of Toeplitz operators with symbols $\bar{I}J$ have been known to play crucial role in the study of Riesz bases, frames and completeness in various function spaces, see for instance [61], [72]. As was mentioned before, recently such operators have become a central object in the Toeplitz approach to UP [72], [49], [311]. Via the Toeplitz approach, these and similar operators apply to many fields of analysis including questions in Fourier analysis and spectral problems for differential operators, see for instance [72], [311], [306], [307].
Intuitively, the property that the Toeplitz operator $T_{\bar{I}J}$ has a non-trivial kernel means that $I$ is, in some sense, larger than $J$. Similarly, invertibility of such an operator indicates that $I$ and $J$ are 'equivalent' or have roughly the same 'size'. However, as we will discuss, neither of these properties can yield a formal definition of order or equivalence, since they lack axiomatic properties of transitivity and reflexivity correspondingly.
We attempt to fix this problem and 'lift' these intuitive notions to the level of formal order and equivalence. Via the Toeplitz approach the new order encompasses a variety of problems and applications mentioned above. It reveals relations between problems of Complex and Harmonic analysis and helps to systematize some of the well-known questions from the area of UP and its applications. We present the basic definitions and properties of Toeplitz order, outline its connections with known problems, and to suggest further directions for research.

We will mostly concern ourselves with inner functions in the upper halfplane $\mathbb{C}_+$. Such functions can be represented as a product

$$I = B_\Lambda J_\mu,$$

where $B_\Lambda$ is the Blaschke product corresponding to the sequence $\Lambda = \{\lambda_n\} \subset \mathbb{C}_+$of zeros of $I$ and $J_\mu$ is a singular inner function corresponding to a positive singular measure $\mu$ on $\hat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$. The measure can be represented as $\mu = \nu + c\delta_\infty$ where $\nu$ is Poisson-finite on $\mathbb{R}$, i.e.,

$$\int \frac{d\nu(x)}{1+x^2} < \infty,$$
and $c \geq 0$ is the mass at infinity. The singular function $J_\mu$ is defined as
$$J_\mu = e^{-\mathcal{S}\mu} = e^{-\mathcal{S}\nu + icz},$$
where $\mathcal{S}\mu$ is the Schwarz integral of $\mu$ :
$$\mathcal{S}\mu(z) = \frac{1}{\pi i}\int \left[\frac{1}{t-z} - \frac{t}{1+t^2}\right] d\mu(t).$$
The Blaschke Product $B_\Lambda$ for $\Lambda = \{\lambda_n\}$ is defined as
$$B_\Lambda = \prod c_n \frac{z-\lambda_n}{z-\bar{\lambda}_n},$$
where $c_n$ are unimodular constants chosen so that $c_n \frac{i-\lambda_n}{i-\bar{\lambda}_n} > 0$. If $\Lambda$ is an infinite sequence then the necessary and sufficient condition for the normal convergence of the partial products of $B_\Lambda$ is that $\Lambda$ satisfies the Blaschke condition
$$\sum \frac{\Im\lambda_n}{1+|\lambda_n|^2} < \infty.$$
We will use the notation $S^a(z) = e^{iaz}$ for the complex exponential function, which is the singular inner function corresponding to the pointmass $a > 0$ at infinity. Using our notations $S^a = J_{a\delta_\infty}$.

Similar statements and formulas are true for the case of the unit disk, see for instance [56], [65].

A special role in our notes will be played by meromorphic inner functions (MIF) which are inner functions in the upper half-plane that can be extended meromorphically to the whole plane. The above formulas imply that an inner function is a MIF if and only if its Blaschke factor corresponds to a discrete sequence $\Lambda \subset \mathbb{C}$ (a sequence without finite accumulation points) and the measure in the singular factor is a point mass at infinity, i.e. $J_\mu = S^a = e^{iaz}$ for some non-negative $a$.

For each inner function $\theta(z)$ one may consider a model subspace
$$K_\theta = H^2 \ominus \theta H^2$$
of the Hardy space $H^2 = H^2(\mathbb{C}_+)$.Here ' $\ominus$ ' stands for the orthogonal difference, i.e., $K_\theta$ is the orthogonal complement of the space $\theta H^2 = \{\theta f \mid f \in H^2\}$ in $H^2$. As was mentioned in these subspaces play an important role in complex and harmonic analysis, as well as in operator theory, see [98].

Each inner function $\theta(z)$ defines a positive harmonic function
$$\Re\frac{1+\theta(z)}{1-\theta(z)}$$
and, by the Herglotz representation, a positive measure $\sigma$ such that
$$\Re\frac{1+\theta(z)}{1-\theta(z)} = py + \frac{1}{\pi}\int \frac{y\,d\sigma(t)}{(x-t)^2+y^2},\ z = x+iy. \qquad (14)$$
for some $p \geq 0$. The number $p$ can be viewed as a point mass at infinity. The measure $\sigma$ is a singular Poisson-finite measure, supported on the set where non-tangential limits of $\theta$ are equal to 1 . The measure $\sigma + p\delta_\infty$ on $\hat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ is called the Clark measure for $\theta(z)$. We will sometimes denote the Clark measure defined in (14) by $\sigma_1$. If $\alpha \in \mathbb{C}, |\alpha| = 1$, then $\sigma_\alpha$ is the measure defined by (14) with $\theta$ replaced by $\bar{\alpha}\theta$. In some settings it is convenient to call the measure $\sigma_{-1}$ the 'Clark dual' of the measure $\sigma_1$.

Conversely, for every positive Poisson-finite singular measure $\sigma$ and a number $p \geq 0$, there exists an inner function $\theta(z)$ satisfying (14).
Every function $f \in K_\theta$ has non-tangential boundary values $\sigma_1$-a.e. and can be recovered from these values via the formula

$$f(z) = \frac{p}{2\pi i}(1-\theta(z)) \int f(t)\overline{(1-\theta(t))}dt + \frac{1-\theta(z)}{2\pi i}\int \frac{f(t)}{t-z}d\sigma(t) \quad (15)$$

see [309]. If the Clark measure does not have a point mass at infinity, the formula is simplified to

$$f(z) = \big(1-\theta(z)\big)Kf\sigma \quad (16)$$

where $Kf\sigma$ stands for the Cauchy integral

$$Kf\sigma(z) = \frac{1}{2\pi i}\int \frac{f(t)}{t-z}d\sigma(t). \quad (17)$$

This gives an isometry of $L^2(\sigma)$ onto $K_\theta$. The Clark measure $\sigma_1$ has a point mass at infinity if and only if $1-\theta(t) \in L^2(\mathbb{R})$.
Similar formulas can be written for any $\sigma_\alpha$ corresponding to $\theta$. For any $\alpha, |\alpha| = 1$ and any $f \in K_\theta, f$ has non-tangential boundary values $\sigma_\alpha$-a.e. on $\hat{\mathbb{R}}$. Those boundary values can be used in (15) or (16) to recover $f$.
In the case of meromorphic $\theta(z)$(MIF), every function $f \in K_\theta$ also has a meromorphic extension in $\mathbb{C}$, and it is given by the formula (15).
Each meromorphic inner function $\theta(z)$ can be written as $\theta(t) = e^{i\phi(t)}$ on $\mathbb{R}$, where $\phi(t)$ is a real analytic and strictly increasing function. The function $\phi(t) = \arg\theta(t)$ is a continuous branch of the argument of $\theta(z)$.
For any inner function $\theta$ in the upper half-plane we define its spectrum spec $_\theta$ as the closure of the set $\{\theta = 1\}$, the set of points on the line where the non-tangential limit of $\theta$ is equal to 1, plus the infinite point if the corresponding Clark measure has a point mass at infinity, i.e. if $p$ in (14) is positive. If $\mathrm{spec}_\theta \subset \mathbb{R}$, then $p$ in (14) is 0 .
Recall that a sequence of real points is discrete if it has no finite accumulation points. Note that $\mathrm{spec}_\theta$ is discrete if and only if $\theta$ is meromorphic. The corresponding Clark measure is discrete with masses at the points of the set $\{\theta = 1\}$ given by

$$\sigma(\{x\}) = \frac{2\pi}{|\theta'(x)|},$$

plus possibly a point mass at infinity (related similarly to the derivative at infinity).
If $\Lambda \subset \mathbb{R}(\hat{\mathbb{R}})$ is a given discrete sequence, one can easily construct a meromorphic inner function $\theta$ satisfying $\{\theta = 1\} = \Lambda$ by considering a positive Poisson-finite measure concentrated on $\Lambda$ and then choosing $\theta$ to satisfy (14). One can prescribe the derivatives of $\theta$ at $\Lambda$ with a proper choice of pointmasses.
The same construction shows that an arbitrary continuous growing function $\gamma$ on $\mathbb{R}$ can be approximated, up to a bounded function, by the argument of a meromorphic inner function. If $\Lambda = \{\gamma = 2\pi n\}$ then $\theta$, constructed as above with $\{\theta = 1\} = \Lambda$, satisfies $|\gamma - \arg\theta| < 2\pi$ on $\mathbb{R}$. Furthermore, if $\Gamma = \{\gamma = (2n+1)\pi\}$ one can easily construct $\theta$ so that $\{\theta = 1\} = \Lambda$ and $\{\theta = -1\} = \Gamma$ and achieve an even better approximation $|\gamma - \arg\theta| < \pi$.
For more information and further references on Clark measures see [85],[302] or [100].

Recall that the Toeplitz operator $T_U$ with a symbol $U \in L^\infty(\mathbb{R})$ is the map

$$T_U : H^2 \to H^2,\ F \mapsto P_+(UF),$$

where $P_+$is the Riesz projection, i.e. the orthogonal projection from $L^2(\mathbb{R})$ onto the Hardy space $H^2$. Passing from a function in $H^2$ to its non-tangential boundary values on $\mathbb{R}, H^2$ can be identified with a closed subspace of $L^2(\mathbb{R})$ formed by functions $f \in L^2(\mathbb{R})$ whose Fourier transform $\hat{f}$ is supported on $[0, \infty)$, which makes the Riesz projection correctly defined.
We will use the following notation for kernels of Toeplitz operators (or Toeplitz kernels) in $H^2$ :

$$N[U] = \ker T_U.$$

An important observation is that $N[\bar{\theta}] = K_\theta$ if $\theta$ is an inner function. Along with $H^2$-kernels, one may consider Toeplitz kernels $N^p[U]$ in other Hardy classes $H^p$, the kernel $N^{1,\infty}[U]$ in the 'weak' space $H^{1,\infty} = H^p \cap L^{1,\infty}, 0 < p < 1$, or the kernel in the Smirnov class $\mathcal{N}^+(\mathbb{C}_+)$, defined as

$$N^+[U] = \left\{f \in \mathcal{N}^+ \cap L^1_{loc}(\mathbb{R}) : \bar{U}\bar{f} \in \mathcal{N}^+\right\}$$

for $\mathcal{N}^+$and similarly for other spaces.
If $\theta$ is a meromorphic inner function, $K_\theta^+ = N^+[\bar{\theta}]$ can also be considered. For more on such kernels see [311].

Recall that an entire function $F(z)$ is said to be of exponential type at most $a > 0$ if

$$|F(z)| = O\left(e^{a|z|}\right)$$

as $z \to \infty$. The infimum of such $a$ is the exponential type of $F$. We denote by $\Pi$ the Poisson measure on $\mathbb{R}, d\Pi(x) = dx/(1 + x^2)$.
A classical theorem of Krein gives a connection between the Smirnov class $\mathcal{N}^+(\mathbb{C}_+)$ and the Cartwright class $C_a$ consisting of all entire functions $F(z)$ of exponential type $\leq a$ which satisfy

$$\log|F(t)| \in L^1_\Pi.$$

An entire function $F(z)$ belongs to the Cartwright class $C_a$ if and only if

$$\frac{F(z)}{S^{-a}(z)} \in N^+(\mathbb{C}_+), \text{and } \frac{F^\#(z)}{S^{-a}(z)} \in N^+(\mathbb{C}_+),$$

where $F^\#(z) = \overline{F(\bar{z})}$.
Recall that a Paley-Wiener space $PW_a$ is defined as a space of entire functions of exponential type at most $a$ which belong to $L^2(\mathbb{R})$. Equivalently, $PW_a = C_a \cap L^2(\mathbb{R})$. As an immediate consequence one obtains a connection between the Hardy space $H^2(\mathbb{C}_+)$ and the Paley-Wiener space $\mathrm{PW}_a$. Namely, an entire function $F(z)$ belongs to the PaleyWiener class $\mathrm{PW}_a$ if and only if

$$\frac{F(z)}{S^{-a}(z)} \in H^2(\mathbb{C}_+),\ \frac{F^\#(z)}{S^{-a}(z)} \in H^2(\mathbb{C}_+).$$

The definition of the de Branges spaces of entire functions may be viewed as a generalization of the last definition of the Paley-Wiener spaces with $S^{-a}(z)$ replaced by a more general entire function. Consider an entire function $E(z)$ satisfying the inequality

$$|E(z)| > |E(\bar{z})|,\ z \in \mathbb{C}_+.$$

Such functions are usually called de Branges functions. The de Branges space $B(E)$ associated with $E(z)$ is defined to be the space of entire functions $F(z)$ satisfying

$$\frac{F(z)}{E(z)} \in H^2(\mathbb{C}_+),\ \frac{F^\#(z)}{E(z)} \in H^2(\mathbb{C}_+).$$

It is a Hilbert space equipped with the norm $\| F \|_E = \| F/E \|_{L^2(\mathbb{R})}$. If $E(z)$ is of exponential type then all the functions in the de Branges space $B(E)$ are of exponential type not greater

then the type of $E(z)$ (see, for example, the last part in the proof of Lemma (5.3.16). 5 in [91]). A de Branges space is called short (or regular) if together with every function $F(z)$ it contains $(F(z)-F(a))/(z-a)$ for any $a\in\mathbb{C}$.
One of the most important features of de Branges spaces is that they admit a second, axiomatic, definition. Let $H$ be a Hilbert space of entire functions that satisfies the following axioms:

- (A1) If $F\in H, F(\lambda)=0$, then $\frac{F(z)(z-\bar{\lambda})}{z-\lambda}\in H$ with the same norm;
- (A2) For any $\lambda\notin\mathbb{R}$, point evaluation at $\lambda$ is a bounded linear functional on $H$;
- (A3) If $F\in H$ then $F^{\#}\in H$ with the same norm.

Then $H=B(E)$ for a suitable de Branges function $E$. In [55].
Usually, for a given Hilbert space of entire functions it is not difficult to check the above axioms and conclude that the space is a de Branges space (if the axioms do hold). It is however a challenging problem in many situations to find a generating function $E$. This problem can be viewed as a deep and abstract generalization of the inverse spectral problem for second order differential operators.
Every de Branges space $B(E)$ is a reproducing kernel Hilbert space, i.e., for each point $\lambda\in\mathbb{C}$ there exists a function $k_\lambda\in B(E)$ such that

$$F(\lambda)=<F,k_\lambda>$$

for any $F\in B(E)$. The reproducing kernel $k_\lambda$ is given by the formula

$$k_\lambda(z)=\frac{E(z)\bar{E}(\lambda)-E^{\#}(z)E\left(\bar{\lambda}\right)}{2\pi i\left(\bar{\lambda}-z\right)}.$$

It is not difficult to show that for any de Branges function $E$ sequences of reproducing kernels $\{k_\lambda\}_{\lambda\in\Lambda}$, where $\Lambda\subset\mathbb{R},\Lambda=\{E^{\#}/E=\alpha\}$ for some constant $\alpha,|\alpha|=1$, form orthogonal bases of $B(E)$. Moreover, these are the only orthogonal bases of reproducing kernels.
De Branges spaces possess the so called nesting property, which makes Krein-de Branges theory especially suitable to study spectral problems for differential operators. It says that for any two de Branges spaces $B(E_1)$ and $B(E_2)$ isometrically embedded into a third de Branges space, either $B(E_1)\subset B(E_2)$ or $B(E_2)\subset B(E_1)$. It follows that any space $B(E)$ admits a unique chain of subspaces $B(E_t), 0\le t\le 1$ monotone by inclusion with $E_0=$ const and $E_1=E$ (in the case of so-called jump intervals the parameter $t$ may not take all values from 0 to 1). Moreover, for any positive Poisson-finite measure $\mu$ on $\mathbb{R}$ there is a unique regular chain of de Branges spaces isometrically embedded into $L^2(\mu)$.
Every de Branges function $E(z)$ gives rise to a MIF

$$\theta(z)=\theta_E(z)=E^{\#}(z)/E(z)$$

and a model space $K_\theta$ that this inner function generates. There exists a well known isometric isomorphism between $B(E)$ and $K_\theta$ given by $F\to F/E$. Conversely, for every MIF $\theta$ there exists a de Branges function $E$ such that $\theta=\theta_E$. Such a function $E$ is unique up to a multiplication by a real entire function without zeros in $\mathbb{C}\setminus\mathbb{R}$ (an entire function is called real if it is real on $\mathbb{R}$ ). We call a de Branges function $E$ an Hermite-Biehler (HB) function if it has no zeros on the real line. For a given MIF $\theta$ one can always choose the corresponding de Branges function $E$ to be an HB function.

As was mentioned above, if $\theta$ is a MIF then all Clark measures $\sigma_\alpha$ of $\theta$ are discrete and their point masses can be computed by $\sigma_\alpha(\lambda)=2\pi/|\theta'(\lambda)|$ for $\lambda\in\{\theta=\alpha\}$. We will call the measures $|E|^2\sigma_\alpha$, where $\sigma_\alpha$ is a Clark measure for $\theta(z)=E^{\#}(z)/E(z)$, spectral measures of the corresponding de Branges space. It is well known (and follows from a similar property for Clark measures) that for any spectral measure $\nu$ of a de Branges space $B(E)$ the natural embedding gives an isometric isomorphism between $B(E)$ and $L^2(\nu)$. This isomorphism generalizes the Parseval theorem.

On the real line each inner $\theta(z)$ coming from a de Branges function can be written as $\theta(t) = e^{i\phi(t)}, t \in \mathbb{R}$, where $\phi(t)$ is real analytic strictly increasing function, a continuous branch of the argument of $\theta(z)$ on $\mathbb{R}$. The phase function of the corresponding de Branges space is defined by $\psi(t) = \phi(t)/2$ and is equal to $-\arg E$.

If $E$ is an HB function we will denote by $\theta_E$ the corresponding MIF $\theta_E = E^{\#}/E$. If $\mu$ is a positive singular Poisson-finite measure on $\hat{\mathbb{R}}$ we denote by $\theta_\mu$ the inner function with the Clark measure equal to $\mu$. Even though for a MIF $\theta$ the function $E$ such that $\theta = \theta_E$ is not unique, we will use the notation $E_\theta$ for one of such functions. The reader may think of a function with lowest order and type among all such HB functions $E$.

We use Toeplitz operators to define partial order and equivalence on the set of inner functions in the upper half-plane. Our definitions can be naturally extended to broader classes of functions and measures, however we choose to concentrate on the inner case. Moreover, in most applications discussed in the rest, the inner functions are meromorphic (MIFs).

We begin with the following definition. Recall that $N[U]$ denotes the $H^2$ −kernel of the Toeplitz operator with symbol $U$.

**Definition (5.3.1)[299]:** If $\theta$ is an inner function we define its (Toeplitz) dominance set $\mathcal{D}(\theta)$ as

$$\mathcal{D}(\theta) = \left\{ I \text{ inner } \mid N[\bar{\theta} I] \neq 0 \right\}.$$

Every collection of sets admits natural partial ordering by inclusion. In our case, we consider dominance sets $\mathcal{D}(\theta)$ as subsets of the set of all inner functions in the upper half-plane and the partial order $\subset$ on this collection. This partial order induces a preorder on the set of all inner functions in $\mathbb{C}_+$. Proceeding in a standard way, we can modify this preorder into a partial order by introducing equivalence classes of inner functions. The details of this definition are as follows.

**Definition (5.3.2)[299]:** We will say that two inner functions $I$ and $J$ are Toeplitz equivalent, writing $I \overset{\mathrm{T}}{\sim} J$, if $\mathcal{D}(I) = \mathcal{D}(J)$. This equivalence relation divides the set of all inner functions in $\mathbb{C}_+$into equivalence classes. We call this relation Toeplitz equivalence (TE).

Further, we introduce a partial order on these equivalence classes defining it as follows.

**Definition (5.3.3)[299]:** We write $I \leqslant J$ (meaning that the equivalence class of $I$ is 'less or equal' than the equivalence class of $J$) if $\mathcal{D}(I) \subset \mathcal{D}(J)$. We call this partial order on the set of inner functions in $\mathbb{C}_+$Toeplitz order (TO).

The following simple examples illustrate our definitions.

**Example (5.3.4)[299]:** Let $B_n$ and $B_k$ be Blaschke products of degree $n$ and $k$ correspondingly. Then $B_n \overset{\mathrm{T}}{\sim} B_k$ iff $n = k$ and $B_n \overset{\mathrm{T}}{<} B_k$ iff $n < k$.

If $J_\mu$ and $J_\nu$ are two singular functions, $J_\mu \overset{\mathrm{T}}{\leqslant} J_\nu$ if $\nu - \mu$ is a non-negative measure. However, there exist $\mu$ and $\nu$ such that $\mu \perp \nu$ but $J_\mu \leqslant J_\nu$, as follows from an example given in [300]. It is a good exercise on Toeplitz kernels to establish the statements of the above example.

As was explained, Clark theory provides a natural one-to-one correspondence between inner functions in $\mathbb{C}_+$and positive singular Poisson-finite measures on $\hat{\mathbb{R}}$. Via this

correspondence one may introduce Toeplitz equivalence and order on the set of all such measures. I.e., for any two positive singular Poisson-finite measures $\mu$ and $\nu, \mu \overset{T}{\sim} \nu$ if $I_\mu \overset{T}{\sim} I_\nu$ and $\mu \overset{T}{\leqslant} \nu$ if $I_\mu \overset{T}{\leqslant} I_\nu$. Similarly, Toeplitz order on inner functions induces an order on Hermite-Biehler functions, de Branges spaces, canonical systems, model spaces, model contractions, etc.
In the same way one can order the set of all unimodular functions on the real line. If $U = e^{i\phi}$ is a unimodular function on $\mathbb{R}$ (and $\phi$ is a measurable real function) then one can defined its dominance set $\mathcal{D}(U)$ as the set of inner functions $\theta$ such that $N[\bar{U}\theta]$ is non-trivial. After that, once again using the ordering of dominance sets by inclusion, one can introduce equivalence classes on the set of unimodular functions and partial order on those classes. In a slightly different way, one may view the ordering described above as an ordering of equivalence classes of real measurable functions $\phi$ on $\mathbb{R}$ defined as $\phi \leqslant \psi$ if $e^{i\phi} \leqslant e^{i\psi}$. Analogously, TO can be moved from the upper half-plane to the unit disk or even more general domains. Without any changes in the above definitions, TO can be extended to bounded analytic functions in $\mathbb{C}_+$or even unbounded functions if one is willing to deal with unbounded Toeplitz operators.
Using quadratic forms one can consider Toeplitz operators with distributional symbols. If $m$ is a distribution on $\mathbb{R}$ then $\mathcal{D}(m)$ can be defined as the set of inner functions such that $T_{\theta\bar{m}}$ exists and has a non-trivial kernel. After finding a way to overcome obvious technical difficulties in this definition, one can proceed with an extension of TO to this class. In particular, one obtains a different way to extend TO to the set of measures and it may be interesting to study relations with the extension outlined above.
Perhaps the simplest way to order inner functions is by division, i.e., to say that $I \leq J$ if $I$ divides $J$ (if $J/I$ is an inner function). The main flaw of the order by division is that most pairs of inner functions remain incomparable. It is easy to see that TO is an extension of the order by division since $I \leqslant J$ whenever $I$ divides $J$. While for two functions to be comparable in the order by division the zero set of one has to be a subset of the zero set of the other, in TO one zero set only needs to be 'near' the other.
Another way to define an order on inner functions is to say that $I > J$ if $N^\infty[\bar{I}J] \neq 0$ or if $N^+[\bar{I}J] \neq 0$ (the kernel in the Smirnov class $\mathcal{N}^+$). These orders are different from ours. The $N^+$-order is related to (and used implicitly in) the Beurling-Malliavin theory. This order is meaningful, but less relevant to problems discussed in these notes. As follows from Lemma (5.3.13), TO is a proper extension of the $H^\infty$-version of the above order.
While all versions of Toeplitz order mentioned seem to be interesting, we will concentrate on the inner version of TO in $\mathbb{C}_+$as defined.

We study the dominance set $\mathcal{D}(\theta)$, the key element of Toeplitz order. We will identify two important subsets of $\mathcal{D}(\theta)$, the sets of base and total elements, and discuss their relations with adjacent questions.

Let $I, J$ be two MIFs and let us denote by $\phi = \phi(I, J)$ the difference of arguments $\frac{1}{2}(\arg I - \arg J)$. Recall that the argument of a MIF on the real line can be chosen as a real analytic function and therefore the last expression makes sense. If $\phi(I, J)$ is Poisson-summable then its harmonic conjugate $\tilde{\phi}$ exists and we will denote by $h(I, J)$ the outer function $\exp(\tilde{\phi} - i\phi)$. Note that then $h(I, J) = 1/h(J, I)$. Clearly, a sufficient condition for $I \overset{T}{\sim} J$ is that $\phi$ has a bounded harmonic conjugate, i.e.,

$$0 < c < h < C < \infty$$

on $\mathbb{R}$ for some constants $c$ and $C$. Indeed, in that case $f \in N[\bar{I}L]$ iff $hf \in N[\bar{J}L]$, which implies that $\mathcal{D}(I) = \mathcal{D}(J)$. However, this condition is not necessary.

**Example (5.3.5)[299]:** Let $I = B_\Lambda$, where $\lambda_n = 2^n(1+i), n \in \mathbb{N}$. Notice that then $|I'| = O(1/x)$ on $\mathbb{R}$ as $x \to \infty$. Let us construct $J$ in the following way. For a rare subsequence $n_k = 2^k, k \in \mathbb{N}$, pull the zeros closer to the real line, i.e., define

$$J = I \prod \left( d_k \frac{z - (2^{n_k} + ic_k)}{z - \lambda_{n_k}} \cdot \frac{z - \bar{\lambda}_{n_k}}{z - (2^{n_k} - ic_k)} \right),$$

where $c_k > 0$ are positive constants tending to zero and $d_k$ are convergence constants. Then $\phi = \phi(J, I)$ satisfies $0 < c < \exp(\tilde{\phi})$ on $\mathbb{R}$, because

$$\exp(\tilde{\phi}) = \prod q_k \frac{1 - z/\bar{\lambda}_{n_k}}{1 - z/(2^{n_k} - ic_k)}$$

with proper convergence constants $q_k$. One can show that if $c_k$ tend to zero slow enough (say, $c_k = 1/k$ ) we also have

$$\exp(\tilde{\phi}) \in L^2(\mathbb{R}, |I'|dx) \setminus L^\infty(\mathbb{R}).$$

If $f \in N[\bar{I}L]$ for some MIF $L$ then $hf \in N[\bar{J}L]$ because $f$ belongs to $K_I$ and hence is bounded on $\mathbb{R}$ by $C|I'|^{1/2}$. Conversely, if $f \in N[\bar{J}L]$ then $f/h \in N[\bar{I}L]$ because $1/h$ is bounded. Therefore $I \overset{\mathrm{T}}{\sim} J$ even though $\tilde{\phi}$ is unbounded.

We say that an inner function $I \in \mathcal{D}(\theta)$ is a base element if it does not divide any other element of $\mathcal{D}(\theta)$. In other words, base elements are the maximal elements of $\mathcal{D}(\theta)$ with respect to the order by division. We will denote by $\mathcal{D}_B(\theta)$ the set of all base elements of $\mathcal{D}(\theta)$.

We denote by $b_a$ the Blaschke factor with zero at $a \in \mathbb{C}_+$:

$$b_a = \frac{\bar{a}}{a} \frac{z - a}{z - \bar{a}}.$$

If $\theta(a) = 0$ for some $a \in \mathbb{C}_+$then $\theta_a = \theta/b_a$ is a base element of $\mathcal{D}(\theta)$. More generally, one can show that if $\theta(c) = a$ for some $c \in \mathbb{C}_+$, then $\mathbf{b}_a(\theta)/b_c$ is a base element, where $\mathbf{b}_a$ is the Möbius transform of the unit disk with zero at $a$,

$$\mathbf{b}_a = \frac{z - a}{1 - \bar{a}z}.$$

A general description of the set $\mathcal{D}_B(\theta)$ in terms of $\theta$ is an important but difficult problem. It generalizes the problem of describing complete and minimal sequences of reproducing kernels in model and de Branges spaces.

Let $I$ and $J$ be two inner functions. We say that $f \in N[\bar{I}J]$ is purely outer if $f$ is outer and

$$\bar{I}Jf = \bar{g}$$

where $g$ is outer. Note that then automatically $f = g$.

We call an element $I$ of $\mathcal{D}(\theta)$ total, if $N[\bar{\theta}I]$ contains a purely outer function. We chose this name for such elements because, in a sense, each total element represents a total inner component of one of the functions from $N[\bar{\theta}] = K_\theta$. Indeed, if $If \in N[\bar{\theta}]$ for some inner $I$ and outer $f$, then

$$\bar{\theta}If = \bar{J}\bar{f}$$

for some inner $J$. Then

$$\bar{\theta}IJf = \bar{f}$$

and therefore $N[\bar{\theta}IJ]$ contains a purely outer function and $IJ$ is a total element of $\mathcal{D}(\theta)$. Moreover, every total element can be obtained this way, i.e., it is a total inner component of

a function from $N[\bar{\theta}]$, combining inner components in both half-planes. We denote by $\mathcal{D}_T(\theta)$ the subset of all total elements of $\mathcal{D}(\theta)$.
One can show that together with each function $I$ the set $\mathcal{D}_T(\theta)$ it contains every $J$ such that $I/J$ is a finite Blaschke product. It follows that $\mathcal{D}_T(\theta)$ contains the set of all inner divisors $I$ of $\theta$ such that $\theta/I$ is a finite Blaschke product. Finite products can be replaced is this statement with all Blaschke products whose arguments $\psi$ satisfy $\psi/2 \in \log|H^2|$. Here we denote by $\log|H^2|$ the set of functions

$$\{f|f = \ln|g|, g \in H^2(\mathbb{C}_+)\}.$$

In other words, $\log|H^2|$ consists of real functions $f$ such that $f \in L^1_\Pi$ and $\exp(f_+) \in L^2$, where $f_+ = \max(f, 0)$.

**Proposition (5.3.6)[299]:** Every element of $\mathcal{D}(\theta)$ is a factor of a total element.

To prove the last statement just notice that if $\bar{\theta}IJf = \bar{L}\bar{f}$ for some inner $J, L$ and outer $f \in H^2$, then $JLI$ is the desired total element.

In regard to relations between our new sets we have

**Proposition (5.3.7)[299]:** For every inner $\theta$

$$\mathcal{D}_B(\theta) \subset \mathcal{D}_T(\theta) \subset \mathcal{D}(\theta).$$

The sets $\mathcal{D}_T$ and $\mathcal{D}_B$ are equal iff $\theta$ is a Blaschke factor (in which case they are also equal to $\mathcal{D}(\theta)$ and consist of constants).

**Proof.** If $I$ is a base element then the relation $\bar{\theta}If = \bar{J}\bar{f}$ implies that $f$ is outer and $J$ is constant: otherwise $I \div JI \in \mathcal{D}(\theta)$, which contradicts that $I$ is a base element. Hence, $f \in N[\bar{\theta}I]$ is purely outer and $I$ is also a total element. The second statement follows from the fact $\mathcal{D}_T$ contains base elements divided by any finite Blaschke sub-products. Notice that $\mathcal{D}_B$ cannot consist of singular functions only.

Since together with every element $\mathcal{D}(\theta)$ contains all of its inner divisors, Proposition (5.3.6) implies that $\mathcal{D}(\theta)$ is determined by the set of its total elements. The inverse statement follows from Theorem (5.3.22):

**Corollary (5.3.8)[299]:** If $I \overset{\mathrm{T}}{\sim} J$ then $\mathcal{D}_T(I) = \mathcal{D}_T(J)$.

To deduce the above corollary note that total elements of $\mathcal{D}$ are the total inner components of the de Branges space. The spaces $B(E_I)$ and $B(E_J)$ must coincide as sets by Theorem (5.3.22).

It is not difficult to describe elements of $\mathcal{D}_T(\theta)$ in terms of arguments. Let us start with $\mathcal{D}_T(\theta)$ in the case when $\theta$ is a MIF. In this case all functions from $\mathcal{D}_T(\theta)$ are MIFs and their arguments are real-analytic functions on $\mathbb{R}$ (defined uniquely up to $2\pi n$ ). Recall the notation $\phi(I, J) = \frac{1}{2}(\arg I - \arg J)$.

**Proposition (5.3.9)[299]:**

$$I \in \mathcal{D}_T(\theta) \Leftrightarrow \tilde{\phi}(\theta, I) \in \log|H^2|.$$

As to $\mathcal{D}(\theta)$, recall that it consists of all divisors of functions from $\mathcal{D}_T(\theta)$.

**Corollary (5.3.10)[299]:** I belongs to $\mathcal{D}(\theta)$ iff $\phi(\theta, I) = \tilde{h} + \frac{1}{2}\alpha$ where $h \in \log|H^2|$ and $\alpha$ is an argument of an inner function.

To establish the above statement, simply notice that $\phi - \frac{1}{2}\alpha$ for some argument of a MIF $\alpha$ is the argument of a purely outer element of $N[\bar{\theta}I]$.

With some additional effort one can find analogs of statements for general (non-MIF) inner functions.

A well known theorem by L. de Branges, Theorem (5.3.28)6 from [55], page 271, is an important result in the area of UP. One can find a discussion of this result and its applications in [312].
More general versions of this theorem from [84], [310] played important roles in the study of the Gap and Type problems, see also [311]. Here we present the statement from [84] in the settings of TO.
**Theorem (5.3.11)[299]:** Let $I, \theta$ be inner functions in $\mathbb{C}_+, \theta \in \mathcal{D}(I)$.
Then there exists an inner function $J$ in $\mathbb{C}_+$such that $\text{spec}_J \subset \text{spec}_I$ and $\theta \in \mathcal{D}_T(J)$.
The function $J$ can be chosen so that the purely outer $f \in N[\bar{J}\theta]$ is also zero-free on $\mathbb{R}$. If $\theta$ is a meromorphic function, then $J$ can be chosen as a meromorphic function.
If $I$ is a MIF, then $f$ in the statement is analytic through $\mathbb{R}$ and the term 'zero-free' can be understood in the usual sense. In the general case, a function $f \in H^2$ has a zero at $x \in \mathbb{R}$ if $f/(z-x) \in H^2$, and a zero-free function has no such points.
Let us finish with the following problem. Given a collection of inner functions, we will call the minimal $\mathcal{D}(\theta)$ containing these functions the Toeplitz hull (TH) of our collection. It seems to be an interesting question to find TH for a given collection. In view of our discussion, versions of this problem are equivalent to finding the minimal de Branges space or model space for a given collection of zero sets, etc.

As was mentioned before, another natural way to introduce a partial order on the set of inner functions is by division. We say that an inner function $I$ divides another inner function $J$ if $J/I$ is an inner function. The relation 'divides' satisfies the axioms of a partial order. Toeplitz order introduced above is an extension of the order by division, i.e. if $I$ divides $J$ then $I \overset{\mathrm{T}}{\leqslant} J$.
TO is a proper extension because one can easily construct a pair of inner functions $I$ and $J$ such that $I \overset{\mathrm{T}}{\leqslant} J$ but $J$ does not divide $I$. Indeed, choose any pair such that $J$ divides $I$ and $J$ has at least one zero. Then that zero has also to be a zero of $I$. Take that zero of $I$ and move it by a finite distance in $\mathbb{C}_+$.It is not difficult to show (an exercise on Toeplitz kernels) that then we still have $I \overset{\mathrm{T}}{\leqslant} J$, although $J$ no longer divides $I$.

Intuitively, when $N[\bar{I}J] \neq 0$ for two inner functions $I$ and $J$ it means that $I$ is 'larger' than $J$. This relation between $I$ and $J$ starts to resemble a strict order even more after one recalls that by a lemma of Coburn $N[\bar{I}J]$ and $N[\bar{J}I]$ cannot be non-trivial simultaneously. Formally, however, this relation does not constitute an order due to the lack of transitivity: $N[\bar{I}J] = N[\bar{J}L] = 0$ does not imply $N[\bar{I}L] = 0$.

Accordingly, the relation $I \asymp J$, which can be defined to mean that $N[\bar{I}J] = 0$ and $N[\bar{J}I] = 0$, fails to produce a formal equivalence. An interesting geometric connection for this relation is observed in [245]. It is shown that for two inner functions $I \asymp J$ holds if and only if the subspaces $IH^2$ and $JH^2$, viewed as points in the Grassmanian manifold of all closed subspaces of $L^2$, are connected by a geodesic. Lack of transitivity for this relation can be illustrated by the following example.
**Example (5.3.12)[299]:** Let us construct three MIFs $I, J$ and $L$ such that $I \asymp J$ and $J \asymp L$ but $I \not\asymp L$, where the relation ' $\asymp$ ' is defined as above.
Let $C > 0$ be a large number and let $I$ be a Blaschke product with zeros at $n + iC$, $n \in \mathbb{Z}$.
Let $J$ be the Blaschke product with zeros at $n + iC$ for $n < 0$ and at $\left(n + \frac{1}{2}\right) + iC$ for $n \geq 0$.
Finally, let $L$ be the Blaschke product with zeros at $n + iC, n \in \mathbb{Z}, n \neq 0$.

Then $\psi = 2\phi(J, I) = \arg J - \arg I$ tends to 0 as $x \to -\infty$. For large positive $x$, $\left|\psi(x) + \frac{\pi}{2}\right| < \varepsilon$, where $\varepsilon = \varepsilon(C)$ is a small number, $\varepsilon(C) \to 0$ as $C \to \infty$. From basic properties of Toeplitz kernels, since

$$\limsup_{x\to-\infty} \psi(x) - \liminf_{x\to\infty} \psi(x) < \pi \text{ and}$$
$$\limsup_{x\to-\infty} -\psi(x) - \liminf_{x\to\infty} -\psi(x) < \pi,$$

both $N[\bar{I}J] = 0$ and $N[\bar{J}I] = 0$, i.e., $I \asymp J$. Similarly, $\psi = \arg L - \arg J$ tends to 0 as $x \to -\infty$ and $\left|\psi(x) + \frac{\pi}{2}\right| < \varepsilon$ near $\infty$, which implies $J \asymp L$.

It is left to notice that $N[\bar{I}L] = N\left[\bar{b}_{Ci}\right]$, where $b_{Ci} = -\frac{z+Ci}{z-Ci}$ is the Blaschke factor, and the kernel contains an $H^2$-function $\frac{1}{z-Ci}$. Hence $I \not\asymp L$.

To study the relations between TO and triviality of kernels further let us formulate the following statement, showing in particular that TO is an extension of the order mentioned.

**Lemma (5.3.13)[299]:** Let $I_1$ and $I_2$ be non-constant inner functions such that $N^\infty[\bar{I}_1 I_2] \neq 0$. Then $I_1 \geqslant I_2$.

**Proof.** If $\theta \in \mathcal{D}(I_2)$ then for $f \in N[\bar{I}_2\theta]$ and $g \in N^\infty[\bar{I}_1 I_2]$ we have

$$\bar{I}_1\theta g f = (\bar{I}_1 I_2 g)(\bar{I}_2\theta f) \in \bar{H}^2.$$

Therefore $\theta \in \mathcal{D}(I_1)$.

Following [72], we will call two MIFs $I$ and $J$ twins if $\text{spec}_I = \text{spec}_J$. This relation naturally appears in applications to spectral problems involving isospectral differential operators.

Clearly, twin relation is an equivalence relation on the set of all MIFs, which is different from the Toeplitz equivalence. It is obvious that $I \overset{\mathrm{T}}{\sim} J$ does not imply that $I$ is a twin of $J$. The opposite implication fails in general as well. However, we do have the following statement. We use the notation $f \asymp g$ for two functions $f$ and $g$ if $c|f| < |g| < C|f|$ for some positive constants $c$ and $C$ and all values of the argument.

**Lemma (5.3.14)[299]:** Let $I$ and $J$ be two MIF twins with the common spectrum $\sigma \subset \hat{\mathbb{R}}$. Then $I \overset{\mathrm{I}}{\sim} J$ iff $I' \asymp J'$ on $\sigma$.

**Proof.** Let $\mu = \alpha_\infty\delta_\infty + \sum \alpha_n \delta_{x_n}$ and $\nu = \beta_\infty\delta_\infty + \sum \beta_n\delta_{x_n}$ be the Clark measures of $I$ and $J$ respectively.

Suppose first that $I' \not\asymp J'$ on $\sigma$. Then by the formula for pointmasses of Clark measures given, $\alpha_n \not\asymp \beta_n$. WLOG, we can assume that there exists $f \in K_I (f \in L^2(\mu))$ such that $f\mu \neq h\nu$ for any $h \in L^2(\nu)$. Let $\theta \in \mathcal{D}(I)$ be the total inner component of $f \in K_I$. We can assume that it is the inner component of $f$ in $\mathbb{C}_+$. Since $f = (1 - I)Kf\mu, \theta$ is the inner factor of the Cauchy integral $Kf\mu$. If $I \overset{\mathrm{T}}{\sim} J$ then $\theta \in \mathcal{D}(J)$ and there exists $g \in K_J$ such that $Kg\nu$ is divisible by $\theta$ in $\mathbb{C}_+$.Moreover, by Corollary (5.3.8) , $g$ can be chosen so that $\theta$ is its total inner component. Then $Kg\nu/Kf\mu$ is an entire function of exponential type zero without zeros. Hence it is a constant, which implies $f\mu = \text{const} \cdot g\nu$ and we have a contradiction.

It is left to notice that if $I' \asymp J'$ then $L^2(\mu) = L^2(\nu)$, which implies $\mathcal{D}(I) = \mathcal{D}(J)$ and $I \overset{\mathrm{T}}{\sim} J$.

Another important relation between inner functions, which resembles equivalence, comes from invertibility of the Toeplitz operator with the symbol $\bar{I}J$. Due to the work of Hruschev, Nikolski, and Pavlov [61], this condition became one of the main tools in the study of basis properties for systems of reproducing kernels, including the classical problem

on exponential bases as a particular case. Up to some technical details, a system of reproducing kernels $\{k_\lambda\}_{\lambda\in\Lambda}$ forms a Riesz basis in a model space $K_I$ if and only if $T_{\bar{I}B_\Lambda}$ is invertible.
Intuitively, the condition that $T_{\bar{I}J}$ is invertible also tells us that the functions $I$ and $J$ are similar. This relation is reflexive as $T_{\bar{I}J}$ is invertible iff $T_{\bar{J}I}$ is. Our next goal is to show that Toeplitz equivalence is not the same as the invertibility of $T_{\bar{I}J}$. As a matter of fact, unlike Toeplitz equivalence, invertibility is not a formal equivalence since, once again, it lacks transitivity.
**Example (5.3.15)[299]:** Similar to Example (5.3.12), construct $I_1, I_2, I_3$ so that the difference of arguments

$$\arg I_{n+1} - \arg I_n, n = 1,2,$$

is smooth and close to $\pi/3$ at $\infty$ and to $-\pi/3$ at $-\infty$. Then $T_{\bar{I}_n I_{n+1}}, n = 1,2,$ is invertible but $T_{\bar{I}_1 I_3}$ is not, as follows from a theorem by Devinatz and Widom. Thus invertibility does not induce an equivalence relation.

While we do not see a reasonable 'if and only if' condition which describes TE in terms of the arguments or other requisites of inner functions, here we give some simple 'one-sided' conditions for two MIFs to be equivalent. Recall that for two inner functions $I$ and $J$ we denote by $\phi = \phi(I,J)$ the function $\phi = \frac{1}{2}(\arg I - \arg J)$. If $I \overset{T}{\sim} J$ then $\tilde{\phi}$ is Poisson summable and $h = h(I,J)$ stands for the outer function $h = e^{\tilde{\phi} - i\phi}, |h| = \exp\tilde{\phi}$.
As was mentioned before, if $\tilde{\phi}(I,J)$ is bounded, i.e., $|h(I,J)|$ is bounded and separated from zero on $\mathbb{R}$, then $I \overset{T}{\sim} J$. This condition is not necessary as was shown in Example (5.3.5).
**Lemma (5.3.16)[299]:** Let I and $J$ be two MIFs, $I \overset{T}{\sim} J$. Then

$$\frac{|J'|}{|I'|}\exp 2\tilde{\phi} \asymp 1$$

on $\mathbb{R}$.
**Proof.** By Theorem (5.3.24) multiplication by $h = h(J,I)$ is a bounded operator $K_J \to K_I$. Hence,

$$\begin{aligned}\left|h(x)k_x^J(x)\right| &= \left|< k_x^I, hk_x^J >_{K_I}\right| \le \|k_x^I\|_{K_I}\|hk_x^J\|_{K_I} \le \\ &\le C|I'(x)|^{1/2}\|k_x^J\|_{K_J} = \\ &= C|I'(x)|^{1/2}|J'(x)|^{1/2} \le C|J'(x)|\left(\frac{|I'|}{|J'|}\right)^{1/2} = C\left|k_x^J(x)\right|\left(\frac{|I'|}{|J'|}\right)^{1/2},\end{aligned}$$

for all $x \in \mathbb{R}$, which implies one of the two estimates. Applying similar argument to the operator $K_I \to K_J$ we obtain the other.
Further metric properties of $h$ give the following conditions.
**Theorem (5.3.17)[299]:** Let $I, J$ be $MIFs, \phi = \phi(I,J)$.
If the functions $|J'|^{1/2}\exp(-\tilde{\phi})$ and $|I'|^{1/2}\exp(\tilde{\phi})$ belong to $L^2(\mathbb{R})$ then $I \overset{T}{\sim}$.
If $I \overset{T}{\sim} J$ then $\tilde{\phi} - \log(1 + |x|) \in \log|H^2|$.
**Proof.** By Theorem (5.3.24), $I \overset{T}{\sim} J$ iff multiplication by $h(I,J)$ is a bounded invertible operator from $K_I$ to $K_J$. Note that since every $f \in K_I$ satisfies $|f| \le\| f \|_2 \, |I'|^{1/2}$ the conditions in the statement imply that $hf \in H^2$ and therefore $hf \in K_J$. Similarly, for every $f \in K_J, f/h \in K_I$.

Note that two singular MIFs cannot be equivalent unless they are constant multiples of each other. Hence, if $I \overset{T}{\sim} J$ then one of them, say $I$, has a zero. If $I(a) = 0$ then $I/b_a$ is a base element of $\mathcal{D}(I)$, and therefore a base element of $\mathcal{D}(J)$. By Proposition (5.3.7), it is a total element of $\mathcal{D}(J)$ and by Proposition (5.3.9), $\tilde{\phi}(J, I/b_a) \in \log|H^2|$. It is left to notice that $\tilde{\phi}(J, I/b_a) \sim \tilde{\phi}(I, J) - \log(1 + |x|)$ as $x \to \pm\infty$
of each other?

The condition of comparability for the derivatives of the inner functions appearing is worth exploring a bit further. Such conditions appear in applications. Inner functions corresponding to Schrödinger equations with regular potentials, as well as to other similar classes of canonical systems, will satisfy this condition. Completeness problems for various families of special functions also lead to MIFs with comparable derivatives, see [72]. Let us provide the following description of Toeplitz equivalence pertaining to this case.

**Lemma (5.3.18)[299]:** Consider two MIFs I and $J$ such that $I' \asymp J'$ on $\mathbb{R}$. Then $I \overset{T}{\sim} J$ iff $\phi(I, J) = \frac{1}{2}(\arg I - \arg J)$ has a bounded harmonic conjugate.

**Proof.** If $\tilde{\phi}$ is bounded then $I \overset{\mathrm{T}}{\sim} J$. Assume now that $I \overset{\mathrm{T}}{\sim} J$ but $\tilde{\phi}$ is unbounded. This contradicts Lemma (5.3.16).

In terms of the model space $K_\theta$, the set of dominance $\mathcal{D}(\theta)$ has a natural meaning. It is the set of all inner components of functions from $K_\theta$.

In case of MIFs, $K_\theta$ is directly related to the de Branges space $B(E)$ via the isometric isomorphism $EK_\theta = B(E)$. Hence, $\mathcal{D}(\theta)$ is also the set of all inner components of functions $f/E, f \in B(E)$, in the upper half-plane.

The set $\mathcal{D}_T(\theta)$ can be similarly identified with the subset of all total inner components of functions from $K_\theta$ or $B(E)$ as was discussed.

If $\theta$ is a MIF and $I \in \mathcal{D}_B(\theta)$ then $I = B_\Lambda S^a$ for some Blaschke sequence $\Lambda, \lambda_n \to \infty$ and $a \geq 0$. In the case of pure Blaschke product, $a = 0$, the sequence $\Lambda$ satisfies $\Lambda = \{f = 0\}$ (with multiplicities) for some function from $K_\theta$( or $B(E)$). In the case $a > 0$, for any Möbius transform $\mathbf{b}_w$ of the unit disk, $\mathbf{b}_w(S^a)B_\Lambda$ is a Blaschke product from $\mathcal{D}_B(\theta)$. Hence, $\Lambda \cup \left\{\frac{2\pi n}{a} + iC\right\}, \Re C > 0$, is again equal to $\{f = 0\}$ for some $f \in K_\theta(B(E))$.

We will return to the discussion of zero sets.

Sets of inner components of functions from $K_\theta$ have been studied by other authors, see for instance [303], [300]. As follows from our discussion above, $I \in \mathcal{D}(\theta)$ iff ' $I$ lurks within $K_\theta$, using the terminology of [300]. In [277] the authors study the set of multipliers $\mathcal{M}(I, J)$ between model spaces $K_I$ and $K_J$, i.e., the set of all $H^\infty$-functions $\phi$ such that $\phi K_I \subset K_J$. In relation to TO, $\mathcal{M}(I, J) \neq \{0\}$ implies $I \overset{\mathrm{T}}{\leqslant} J$. The reader may find additional properties of $\mathcal{D}(\theta)$.

**Lemma (5.3.19)[299]:** Let $\theta$ be a MIF, $f \in K_\theta, f(iy) \neq o\left(y^{-3/2}\right)$ as $y \to \infty$. Then the total inner component of $f$ is a base element of $\mathcal{D}(\theta)$.

**Proof.** Suppose that the total inner component $I$ of $f$ is not a base element. Then there exists inner $J$ and outer $g$ such that $I$ properly divides $J$ and $Jg \in K_\theta$. Let $h$ be an outer component of $f$. Then the argument of the outer function $g/h$ is $-\frac{1}{2}\arg(J/I)$, i.e., it is a continuous decreasing function on $\mathbb{R}$ which decreases by at least $\pi$. By Claim 1 below and the asymptotics of $f$ this implies that $g(iy) \neq o\left(y^{-1/2}\right)$ as $y \to \infty$. This contradicts $g \in H^2$.

The following can be easily established.

**Claim 1.** Let $h$ be an outer function in $\mathbb{C}_+$whose argument $\psi$ on $\mathbb{R}$ satisfies

$$\liminf_{x\to-\infty}\psi(x)-\limsup_{x\to\infty}\psi(x)\geq\pi.$$

Then

$$y=O\big(h(iy)\big)\text{ as }y\to\infty.$$

Our next statement combined with Lemma (5.3.19) shows that functions whose total inner components are base elements of $\mathcal{D}(\theta)$ are dense in $K_\theta$.

**Proposition (5.3.20)[299]:** For every inner $\theta$, the space $K_\theta$ contains a dense subset of functions $f$ satisfying

$$|f(iy)|\sim\frac{1}{y}\text{ as }y\to\infty.$$

**Proof.** Let $C(z)$ be the Cayley transform from the unit disc to the upper half-plane. Then $\Phi(z)=\theta(C(z))$ is an inner function in the unit disc. Recall that $K_\theta$ is obtained from $K_\Phi$ via the map $f(z)\mapsto(C^{-1}(w)-1)f(C^{-1}(w))$. Now the statement is equivalent to the statement that functions with finite non-zero limits $\lim_{r\to1-}f(r)$ are dense in $K_\Phi$.

Let $\Phi_n$ be a sequence of divisors of $\Phi$ such that $\Phi_n\to\Phi$ point-wise in $\mathbb{D}$ and each $\Phi_n$ can be analytically continued through 1 . Then $\cup K_{\Phi_n}$ is dense in $K_\Phi$. But in each $K_{\Phi_n}$ all functions can be continued through 1 and a dense subset have non-zero values there.

Let $E$ be a de Branges function and let $\theta=\theta_E$ be a corresponding MIF. If $F\in B(E)$ then $F=I_1fE$ in $\mathbb{C}_+$, where $I_1$ is inner and $f\in H^2$ is outer. Similarly, in $\mathbb{C}_-$, $F=\bar{I}_2\bar{f}E^\#$. An important property of $B(E)$ is that the inner components can be moved from one half-plane to the other, i.e., if $I_3I_4=I_1$ then the function $G$ defined as $I_3fE$ in $\mathbb{C}_+$and as $\bar{I}_4\bar{I}_2\bar{f}E^\#$ in $\mathbb{C}_-$also belongs to $B(E)$. Similarly one can move inner factors from $\mathbb{C}_-$to $\mathbb{C}_+$.

The set of all inner functions $I_1(I_2)$ appearing this way for a fixed $B(E)$ is exactly the dominance set $\mathcal{D}(\theta)$.

If $F$ is an entire function defined as above in $\mathbb{C}_\pm$, we will call the inner function $I_1I_2$ the total inner component of $F$. If $I$ is a total inner component for a function from a de Branges space then the argument of $fE$ on $\mathbb{R}$ is determined by the argument of $I$ up to $\pi n$. The argument of MIF $I$ is a real-analytic function on $\mathbb{R}$, while the argument of $fE$ is piece-wise real analytic, making a jump of $-\pi$ at each real zero of $fE$. All in all we have

$$\arg fE=\frac{1}{2}\arg I(\mathrm{mod}\pi)\qquad(18)$$

Note that total inner components of functions from $B(E)$ are exactly the elements of $\mathcal{D}_T(\theta_E)$.

Denote by $\mathcal{D}_T^*(\theta_E)$ the set of exact total elements, the total elements corresponding to functions from $B(E)$ which have no zeros on the real line. If $f\in B(E)$ is such a function and $I\in\mathcal{D}_T^*$ is its inner component in $\mathbb{C}_+$then the last equation holds exactly, i.e., $\arg fE=\frac{1}{2}\arg I$ on $\mathbb{R}$ for a proper choice of arguments on both sides.

Let $I$ and $J$ be two MIFs such that $N[\bar{I}J]\neq0$. Notice that

$$\bar{I}Jf=\frac{E_I}{E_I^\#}\frac{E_J^\#}{E_J}f=\bar{g}$$

which shows that an $H^2$-function $f$ belongs to $N[\bar{I}J]$ iff $\frac{E_I}{E_J} f$ can be continued to the lower half-plane as an entire function (the formula for the continuation is $\frac{E_I^{\#}}{E_J^{\#}} \bar{g}$). Consider the space of entire functions $B = \frac{E_I}{E_J} N[\bar{I}J]$ equipped with the norm

$$\|f E_I / E_J\| = \| f \|_{H^2}.$$

By verifying the axioms one can conclude that $B = B(E)$ is a de Branges space for some HB function $E$. We will denote this HB function by $E_{I,J}$.

To summarize, to each pair of MIFs $I, J$ such that $N[\bar{I}J] \neq 0$ there corresponds an HB function $E_{I,J}$. Our construction implies the following important property:

**Proposition (5.3.21)[299]:** The set $J\mathcal{D}\left(\theta_{E_{I,J}}\right)$ is the set of all functions from $\mathcal{D}(I)$ divisible by $J$

While model spaces $K_\theta$ are equal as sets if and only if the corresponding inner functions are equal up to a constant multiple, de Branges spaces $B(E)$ and $B(\tilde{E})$ can be equal as sets for two different functions $E$ and $\tilde{E}$.

Equality of two de Branges spaces as sets of functions, with (possibly) different norms, is an important aspect of spectral theory for differential equations. The so-called GelfandLevitan theory which treats spectral problems for regular Schrödinger equations and Dirac systems utilizes the fact that the corresponding de Branges spaces are equal to Paley-Wiener spaces as sets. This property becomes the key ingredient of the theory allowing one to use the structure of Paley-Wiener spaces to study relations between the potential of the differential operator and the Fourier transform of its spectral measure.

An extension of Gelfand-Levitan techniques to more general classes of Krein's canonical systems, see for instance [307], requires further understanding of properties of HB functions $E$ and $\tilde{E}$ which produce equal, as sets, spaces $B(E)$ and $B(\tilde{E})$. Such questions are also equivalent to problems on sampling measures.

Although total description of such pairs of HB functions presents an important open problem, intuitively such functions must be similar to each other, which raises a natural question on the correspondence of this relation and Toeplitz equivalence for the MIFs $\theta$ and $\tilde{\theta}$. Our next theorem connects this problem to TO.

We will use the notation $B(E) \doteq B(\tilde{E})$ for the two de Branges spaces equal as sets. Note that if $B(E) \doteq B(\tilde{E})$ then norms in the spaces are automatically equivalent.

**Theorem (5.3.22)[299]:** Let $E$ and $\tilde{E}$ be HB functions such that $E/\tilde{E} \in \mathcal{N}(\mathbb{C}_+)$.Then $\theta \overset{\sim}{\sim} \tilde{\theta}$ for the corresponding MIFs iff $B(E) \doteq B(\tilde{E})$.

Conversely, if $\theta \overset{T}{\sim} \tilde{\theta}$ for two MIFs $\theta$ and $\tilde{\theta}$ then the corresponding HB functions can be chosen to satisfy $E/\tilde{E} \in \mathcal{N}(\mathbb{C}_+)$and $B(E) \doteq B(\tilde{E})$.

**Proof.** Suppose first that $B(E) \doteq B(\tilde{E})$. Since $\mathcal{D}(\theta)$ and $\mathcal{D}(\tilde{\theta})$ are the sets of inner components of $F/E$ for the elements $F$ of the corresponding space in $\mathbb{C}_+, \mathcal{D}(\theta) = \mathcal{D}(\tilde{\theta})$ and $\theta \overset{T}{\sim} \tilde{\theta}$.

Conversely, let $\mathcal{D}(\theta) = \mathcal{D}(\tilde{\theta})$. Then the subsets of base elements, $\mathcal{D}_B$, coincide as well. If $I \in \mathcal{D}_B(\theta) = \mathcal{D}_B(\tilde{\theta})$ then $I$ is a total inner component for some $F \in B(E)$ and for some $G \in B(\tilde{E})$. Note that then $F/G$ is a zero-free entire function. Indeed, since the total zero components of $F$ and $G$ coincide, $F/G$ may only have zeros on the real line. Then $F$ has zeros on the real line, say at $a \in \mathbb{R}$. But then $(z - i)\frac{F}{z-a}$ is an element of $B(E)$ with total

inner component $b_i I$ which contradicts the property that $I$ is a base element. Hence $F/G$ is zero-free. It must be outer in both half-planes because otherwise $I$ is not a base element in one of the $\mathcal{D}$ sets. Hence, $F/G = \text{const}$. We obtain that the sets of functions in $B(E)$ and $B(\tilde{E})$ whose total inner components are base elements coincide.
Let now $F \in B(E) \setminus B(\tilde{E})$. By Proposition (5.3.20) , there exists $H \in B(E)$ such that $(H/E)(iy) \neq o\left(y^{-3/2}\right)$ as $y \to \infty$. By Lemma (5.3.19), the total inner component of $H$ is a base element and therefore $H \in B(\tilde{E})$ by the argument above.
Notice that $(F/E)(iy) = o\left(y^{-3/2}\right)$, because otherwise its total inner component would have been a base element which would imply $F \in B(\tilde{E})$. Hence, $(F+H)/E \neq o\left(y^{-3/2}\right)$, which implies that $F + H$ has a base total inner component and therefore belongs to both de Branges spaces. Since $H$ also belongs to both spaces, so does $F$ and we arrive at a contradiction.
In the process of the last proof we have established the following useful property.
**Proposition (5.3.23)[299]:** If the total inner component I of a function $F$ from $B(E)$ is a base element of $\mathcal{D}(\theta_E)$ then $F$ has no real zeros. Equivalently, $\mathcal{D}_B(\theta) \subset \mathcal{D}_T^*(\theta)$ for any $MIF\theta$.

We formulate our result for general inner functions. In order to do that we will need to extend the notations $\phi(I,J)$ and $h(I,J)$ introduced from the case of MIFs to the general case.
To make sense of the definition $\phi(I,J) = \frac{1}{2}(\arg I - \arg J)$ in the general case we understand $\arg I(J)$ as a measurable function on $\mathbb{R}$ such that $I/e^{i\arg I}$ is positive a.e. on $\mathbb{R}$. It is not difficult to show that if $I \overset{\mathrm{T}}{\sim} J$ then their arguments can be chosen in such a way that $\tilde{\phi}$ exists and $h(I,J) = e^{\tilde{\phi} - i\phi}$ is an outer function. In what follows we will assume that $\phi$ and $h$ correspond to the arguments of $I$ and $J$ chosen in such a way.

**Theorem (5.3.24)[299]:** $I \overset{\mathrm{T}}{\sim} J$ iff multiplication by $h(I,J)$ is a bounded and invertible operator $K_I \to K_J$.
For general $I$ and $J$ this means that if $I \overset{\mathrm{T}}{\sim} J$ then their arguments can be chosen so that the outer function $h(I,J)$ exists and multiplication by $h(I,J)$ is a bounded and invertible operator $K_I \to K_J$. Conversely, if the arguments can be chosen in such way, then $I \overset{T}{\sim} J$.
**Proof.** Suppose first that $I$ and $J$ are MIFs. Then $h = E_I/E_J$ and the equivalence of $I \overset{\mathrm{T}}{\sim} J$ and $B(E_I) \doteq B\left(E_J\right)$ gives the statement.
In the general case, if multiplication by $h(I,J)$ is a bounded and invertible operator $K_I \to K_J$ then the sets of all inner components of functions from $K_I$ and $K_J$ coincide because $h$ is outer. Hence $\mathcal{D}(I) = \mathcal{D}(J)$ and $I \overset{\mathrm{T}}{\sim} J$. In the opposite direction, if $I \overset{\mathrm{T}}{\sim} J$ one can reduce the proof to the case of MIFs via a limiting argument.

We call a sequence $\Lambda \subset \mathbb{C}$ a zero set of a de Branges space $B(E)$ iff there exists $f \in B(E), f \not\equiv 0$, such that $f = 0$ on $\Lambda$ (with multiplicities). We call $\Lambda$ an exact zero set if there exists $f \in B(E)$ such that $\{f = 0\} = \Lambda$ (with multiplicities). A maximal zero set is a sequence $\Lambda$ of points such that there exists a non-zero function from the space vanishing on $\Lambda$, but there is no such function for any set properly containing $\Lambda$.
A maximal zero set is exact but not vice versa. Blaschke products corresponding to maximal zero sets are base elements from $\mathcal{D}(\theta)$ and those corresponding to exact zero sets are total

elements, see Lemma (5.3.25) below. Maximal zero sets are also related to complete and minimal sequences.
We say that $\Lambda$ is a complete and minimal sequence for $B(E)$ iff the system of reproducing kernels $\{k_\lambda\}_{\lambda\in\Lambda}$ is complete and minimal (i.e., any proper subsequence is incomplete) in $B(E)$. Note that a sequence is complete and minimal iff the same sequence minus any one of its points is a maximal zero set.

For sequences $\Lambda \in \mathbb{C}_+$similar definitions can be given for the model spaces $K_\theta$. We will now establish relations between zero sets and the subsets of the dominance set.
Recall that, as was defined, the spectrum $\mathrm{spec}_I$ of a MIF $I$ is the sequence of points from $\hat{\mathbb{R}}$ where the function is equal to 1. If $\Lambda \subset \mathbb{C} \setminus \mathbb{R}$ is a sequence of complex points we denote by $B_\Lambda$ the Blaschke product with zeros at the points of $\Lambda$ in $\mathbb{C}_+$and at the points conjugate to the points of $\Lambda$ in $\mathbb{C}_-$, assuming the Blaschke condition holds.
**Lemma (5.3.25)[299]:** Let $E$ be an HB function, $\theta = \theta_E$. Let $\Lambda \subset \mathbb{C} \setminus \mathbb{R}$ and $\Gamma \subset \mathbb{R}$ be sequences of points.

(i) $\Lambda \cup \Gamma$ is a zero set of $B(E)$ iff there exists an inner $I$ such that spec $_I = \Gamma$ and $B_\Lambda I \in \mathcal{D}(\theta)$;

(ii)$\Lambda \cup \Gamma$ is an exact zero set of $B(E)$ iff there exists an inner $I$ such that $\mathrm{spec}_I = \Gamma$ and $B_\Lambda I \in \mathcal{D}_T^*(\theta)$;

(iii) $\Lambda \cup \Gamma$ is a maximal zero set of $B(E)$ iff there exists an inner $I$ such that $\mathrm{spec}_I = \Gamma$ and $B_\Lambda I \in \mathcal{D}_B(\theta)$.

**Proof.** (i) Suppose that $F = 0$ on $\Lambda \cup \Gamma$ for some $F \in B(E)$. Then there exists a finite positive measure $\mu$ concentrated on $\Gamma$,

$$\mu = \sum a_n \delta_{\gamma_n},$$

such that $a_n > 0$ are small enough to satisfy $(F/E)K\mu \in H^2$. Then for $I = I_\mu$ (the inner function whose Clark measure is $\mu$) we have $(F/(E(1-I)) \in H^2$ and $= F/(1-I) \in B(E)$. For the function $G/E \in K_{\theta_E}$ we have

$$\bar{\theta}_E \frac{G}{E} = \bar{\theta}_E \frac{F}{E}\frac{1}{1-I} = \bar{h}\frac{\bar{I}}{1-\bar{I}}$$

a.e. on $\mathbb{R}$ for some $h \in H^2, \bar{h} = \bar{\theta}_E F/G$. Here we use the fact that $F/E \in K_{\theta_E}$ and the observation that $\bar{I}(1-I)/(1-\bar{I}) > 0$ a.e. on $\mathbb{R}$. Since $F$ vanishes on $\Lambda$, the inner component of $G/E$ is divisible by $B_\Lambda$. According to the last equation, the inner component of $\theta_E \bar{G}/\bar{E}$ is divisible by $I$. Hence the total inner component of $G/E$ is divisible by $B_\Lambda I$.
Conversely, let $B_\Lambda I \in \mathcal{D}(\theta)$ for some inner $I$ such that $\mathrm{spec}_I = \Gamma$. Then $B(E)$ contains a function equal to $B_\Lambda I f E$ in $\mathbb{C}_+$, where $f$ is outer from $H^2$. Then $B(E)$ also contains a function equal to $B_\Lambda f E$ in $\mathbb{C}_+$. Subtracting we obtain a function in $B(E)$ equal to $B_\Lambda(1-I)fE$ in $\mathbb{C}_+$, which vanishes on $\Lambda \cup \Gamma$.
(ii) and (iii) can be proved similarly.
**Theorem (5.3.26)[299]:** Every element of $\mathcal{D}(\theta)$ is a divisor of a base element.
Before we prove the last statement let us note that each de Branges space possesses a large collection of maximal zero sets (complete and minimal sequences, minus one point). For instance, if one takes an orthogonal basis of reproducing kernels described and deletes one point from the corresponding sequence, the remaining sequence is a maximal zero set. By 'perturbing' this real sequence one can obtain a maximal zero set in $\mathbb{C}_\pm$.Note that maximal zero sets $\Lambda$ in $\mathbb{C}_+$, as any zero sets of a de Branges space $B(E)$ in $\mathbb{C}_+$, satisfy the Blaschke

condition. The corresponding Blaschke products $B_\Lambda$ are exactly the base elements of $\mathcal{D}(\theta_E)$ which have no singular divisor.
**Proof.** First let us assume that $\theta$ is a MIF and let $J \in \mathcal{D}(\theta)$. Let $\Lambda \subset \mathbb{C}_+$ be a maximal zero set of $B\big(E_{\theta,J}\big)$. Then $JB_\Lambda$ is a base element of $\mathcal{D}(\theta)$. Indeed, if $b_a J B_\Lambda \in \mathcal{D}(\theta)$ for some Blaschke factor $b_a, a \in \mathbb{C}_+$then by Proposition (5.3.21) , $\Lambda \cup \{a\}$ is a zero set of $B\big(E_{I,J}\big)$, which contradicts maximality of $\Lambda$. Hence $J$ divides a base element of $\mathcal{D}(I)$.
Finally, in the case of non-MIF $\theta$, notice that $\theta$ is a normal limit of MIFs and apply a limiting argument.
Considering the case when $J$ is a Blaschke product in the statement of Theorem (5.3.26) and using Lemma (5.3.25), we obtain the following result by Yu. Belov. In fact, our last proof is a variation of the proof in [301].
**Corollary (5.3.27)[299]:** ([301]). Any incomplete sequence of reproducing kernels in a de Branges space is contained in a complete and minimal sequence of reproducing kernels.
(Note that any incomplete sequence of reproducing kernels is automatically minimal, which is used implicitly in the above statement.)
Denote by $\mathcal{Z}(B(E))$ the collection of all zero sets for the space $B(E)$ and let $\mathcal{Z}_e$ stand for the exact zero sets. Then Theorem (5.3.26) becomes the following statement.
**Theorem (5.3.28)[299]:** The collection of zero sets $\mathcal{Z}(B(E))$ determines the space $B(E)$ uniquely within the regularity class of $E$, i.e., if $\mathcal{Z}(B(E)) = \mathcal{Z}(B(\tilde{E}))$ and $E/\tilde{E} \in \mathcal{N}(\mathbb{C}_+)$then $B(E) \doteq B(\tilde{E})$.
**Proof.** The collection of sets $\Lambda \setminus \mathbb{R}, \Lambda \in \mathcal{Z}(B(E))$ determines the set of Blaschke products from $\mathcal{D}(\theta), \theta = \theta_E$. For the non-Blaschke elements we have the simple observation that whenever $BS^a \in \mathcal{D}(\theta)$, the Blaschke product $B\mathbf{b}_w(S^a)$ belongs to $\mathcal{D}(\theta)$ as well for all $w \in \mathbb{D}$, which implies that the sets $\Lambda \setminus \mathbb{R}, \Lambda \in \mathcal{Z}_e(B(E))$ determine $\mathcal{D}(\theta)$ uniquely. It follows that $\mathcal{Z}_e(B(E))$ determines $\mathcal{D}(\theta)$ and the statement follows from Theorem (5.3.26).
Note that since $\mathcal{Z}(B(E))$ determines $B(E)$, it also determines $\mathcal{Z}_e(B(E))$. Conversely, since zero sets are subsets of exact zero sets, $\mathcal{Z}_e(B(E))$ determines $\mathcal{Z}(B(E))$. Even easier one can establish the same connection between $\mathcal{Z}$ and $\mathcal{Z}_m$, the collection of maximal zero sets, as each maximal zero set is a maximal element of $\mathcal{Z}$ with respect to inclusion. Hence either of the sets $\mathcal{Z}_e$ or $\mathcal{Z}_m$ can be substituted into the last statement instead of $\mathcal{Z}$. However, the statement with $\mathcal{Z}(B(E))$ is the strongest of the three.

The last statement raises a natural question: if TE is equivalent to equality of the corresponding de Branges spaces, are the relations $\overset{\mathrm{T}}{\leqslant}$ and $\geqslant^{\mathrm{T}}\geqslant$ equivalent to inclusions of the spaces? If the answer were positive we would obtain an equivalent definition of TO.
The relation does hold in one direction:
**Proposition (5.3.29)[299]:** If $B(E) \subset B(\tilde{E})$ then $\theta \overset{\mathrm{T}}{\leqslant} \tilde{\theta}$.
The statement follows from the fact that the corresponding dominance set consists of all inner components of $F/E, F \in B(E) (F/\tilde{E}, F \in B(\tilde{E}))$ in $\mathbb{C}_+$.
However, as shown by the example below, the opposite direction fails.
**Example (5.3.30)[299]:** Consider a sequence $\lambda_n = \left(2^{|n|}\operatorname{sign} n + \frac{1}{2}\right)\pi + \varepsilon_{|n|} i, n \in \mathbb{Z}$, where $\varepsilon_n \downarrow 0$, and the corresponding Blaschke product $B_\Lambda$. Denote $I = B_\Lambda S$ (where, once again, $S(z) = e^{iz}$ ) and consider the corresponding Cartwright HB function $E_I$.

Let $s(z) = \sin z/z$ be the sinc function. Then, if $\varepsilon_n$ decays to 0 fast enough, $s/E_I \notin L^2$. Hence $s \notin B(E_I)$ and $B(E_S) \not\subset B(E_I)$ because $s \in B(E_S) = PW_1$. At the same time, since $S$ is a divisor of $I, S \overset{\mathrm{T}}{<} I$.

Recall that for a partial order a chain is a subset where every pair of elements is comparable. On the other hand, every de Branges space, or every Poisson finite measure on the real line, gives rise to a chain of de Branges spaces of entire functions. Although the term 'chain' is given different meanings in these two situations, we note the following simple connections between de Branges chains and chains in Toeplitz order.

It follows from Proposition (5.3.29) that de Branges chains produce chains in Toeplitz order: if $\{B(E_t)\}$ is a de Branges chain then $\theta_{E_t}$ is a chain in TO. Clearly, such chains do not present all possible chains in TO since, for instance, not all such chains consist of MIFs. Even if we restrict our attention to all TO chains in the subset of MIFs, de Branges chains do not produce all such chains as follows from Example (5.3.30). Finding a way to determine if a chain in TO is a de Branges chain seems like another interesting problem.

We look at connections of Toeplitz order with some of the classical problems of Harmonic Analysis. We provide only a brief overview of such connections without going into deeper technical details.

We start with two completeness problems for families of complex exponentials, the Beurling-Malliavin (BM) problem and its extensions studied in [58], [72], [49], and the Type problem recently considered in [310]. We then discuss sampling problems in Paley-Wiener and de Branges spaces with some remarks on the two-weight Hilbert problem, see [77] and [119], [305], [103]

Let $\Lambda = \{\lambda_n\}$ be a sequence of distinct points in the complex plane and let

$$E_\Lambda = \left\{e^{i\lambda_n x}\right\}$$

be a sequence of complex exponential functions on $\mathbb{R}$ with frequencies from $\Lambda$. For any complex sequence $\Lambda$ its radius of completeness is defined as

$$R(\Lambda) = \sup\ \left\{a \mid E_\Lambda \text{ is complete in } L^2(0,a)\right\}.$$

The famous BM problem which was solved in [50], [51], asks to find a formula for $R(\Lambda)$ for an arbitrary $\Lambda \subset \mathbb{C}$.

It is well-known in the theory of completeness that the general problem can be easily reduced to the case of real sequences $\Lambda$. If $\Lambda$ is a general complex sequence then $E_\Lambda$ is complete in $L^2([0,a])$ if and only if $E_{\Lambda'}$ is complete in the same space, where $\Lambda'$ is the real sequence defined as $\lambda'_n = 1/\Re\frac{1}{\lambda_n}$ (WLOG $\Lambda$ has no purely imaginary points), see for instance [66]. Also, as will be explained below, one can always assume that $\Lambda$ is a discrete sequence, i.e. has no finite accumulation points.

A system of complex exponentials $E_\Lambda$ is incomplete in $L^2([0,a])$ if and only if there exists a non-zero $f \in L^2([0,a])$ such that $f \perp e^{i\lambda_n x}$ for all $\lambda_n \in \Lambda$. Taking the Fourier transform of $f$ we arrive at the equivalent reformulation that $E_\Lambda$ is incomplete in $L^2([0,2a)])$ if and only if $\Lambda$ is a zero for $PW_a$.

One immediate consequence of this connection is that if $\Lambda$ has a finite accumulation point then $R(\Lambda) = \infty$. Also since any zero set $\Lambda \subset \mathbb{C}_+$of a $PW$-space must satisfy the Blaschke condition, $R(\Lambda) = \infty$ for any non-Blaschke $\Lambda \subset \mathbb{C}_+$.

To give the formula for $R(\Lambda)$ we will need the following definitions.

If $\{I_n\}$ is a sequence of disjoint intervals on $\mathbb{R}$, we call it short if

$$\sum \frac{|I_n|^2}{1+\text{dist}^2(0,I_n)} < \infty$$

and long otherwise.

If $\Lambda$ is a sequence of real points define its exterior BM density (effective BM density) as

$$D^*(\Lambda) = \sup \{d \mid \exists \text{ long } \{I_n\} \text{ such that } \#(\Lambda \cap I_n) \geqslant d|I_n|\}, \forall n\}$$

For a complex sequence define $D^*(\Lambda) := D^*(\Lambda')$.

**Theorem (5.3.31)[299]:** (Beurling and Malliavin, around 1961, [50], [51]). Let $\Lambda$ be a discrete sequence. Then

$$R(\Lambda) = D^*(\Lambda).$$

In regard to Toeplitz order, BM theorem is equivalent to the following statements.

Recall that any MIF $I$ has the form $I = B_\Lambda S^a$ where $B$ is a Blaschke product and $S^a = e^{iaz}$ is the exponential function. Put $r(I) = D^*(\Lambda) + a$.

The most direct equivalent of Theorem (5.3.31) is in terms of the dominance set.

**Theorem (5.3.32)[299]:** $I \in \mathcal{D}(S^b)$ if $r < b$ and $I \notin \mathcal{D}(S^b)$ if $r > b$.

The statement can be equivalently reformulated in terms of TO.

**Theorem (5.3.33)[299]:** For any MIF $I$,

$$I \overset{\text{T}}{\leqslant} S^b \Rightarrow r(I) \leq b$$

and

$$r(I) < b \Rightarrow \text{I} <^{\text{T}} S^b.$$

Note that equivalence of the last two statement no longer holds if $S$ is replaced with a general inner function. Finding a broader set of functions for which the equivalence does hold is an open problem.

**Proof.** The general case can be trivially reduced to the case $I = B_\Lambda$. Suppose first that $r = D^*(\Lambda) > b$. Let $a \in \Lambda$ be a zero of $I$. Then $D^*(\Lambda \setminus \{a\}) > b$ and $I/b_a \notin \mathcal{D}(S^b)$ by BM theorem (Theorem (5.3.32)). Since $I/b_a \in \mathcal{D}(I)$, the relation $I \overset{\text{T}}{\leqslant} S^b$ does not hold.

To establish the second statement, suppose that $r = D^*(\Lambda) < b$. If $J \in \mathcal{D}(I)$ then there exists $f \in N[\bar{I}J], f \not\equiv 0$. Also, since $D^*(\Lambda) < b$, by Theorem (5.3.32) there exists $g \in N[\bar{S}^b I], g \not\equiv 0$. Note that $Ig \in S^{b/2}PW_{b/2}$ which implies $g \in H^\infty$. Then

$$\bar{S}^b Jgf = (\bar{S}^b Ig)(\bar{I}Jf) \in \bar{H}^2,$$

which means that $J \in \mathcal{D}(S^b)$. Hence $\mathcal{D}(I) \subset \mathcal{D}(S^b)$ and $I \overset{\text{T}}{<} S^b$.

As we can see, the Beurling-Malliavin formula gives a metric condition for the relation of TO in the very specific case when one of the functions to be compared is the exponential function. Similar descriptions for more general classes of inner functions, especially those appearing in applications to completeness problems and spectral analysis remain mostly open. Below we present one of such extensions found in [49].

Reformulations of the BM theorem given present a clear direction for generalizations of the BM theory. While a statement analogous to with a general inner function in place of $S^a$ may be out of reach at the moment, one can attempt to replace the exponential function with an inner function from a larger class. To determine the right classes of inner functions to study in these settings one may look at a variety of applications of the Toeplitz Approach in Harmonic analysis and Spectral Theory.

One of such extensions was recently studied in [72], [49]. As was shown in [72] the class of MIFs with polynomially growing arguments appears naturally in a number of applications including completeness problems for Airy and Bessel functions, spectral problems for regular Schrödinger operators and Dirac systems, etc. An analog of Theorem (5.3.32) proved in [49] can be applied to some of such problems. Here we present an equivalent reformulation similar to Theorem (5.3.33).

Let $\gamma: \mathbb{R} \to \mathbb{R}$ be a continuous function such that $\gamma(\mp\infty) = \pm\infty$, i.e.,

$$\lim_{x\to-\infty} \gamma(x) = +\infty, \ \lim_{x\to+\infty} \gamma(x) = -\infty.$$

Define $\gamma^*$ to be the smallest non-increasing majorant of  :

$$\gamma^*(x) = \max_{t\in[x,+\infty)} \gamma(t).$$

The family of intervals $BM(\gamma) = \{I_n\}$ is defined as the collection of the connected components of the open set

$$\{x \in \mathbb{R} \mid \gamma(x) \neq \gamma^*(x)\}.$$

Let $\kappa \geq 0$ be a constant. We say that $\gamma$ is $\kappa$-almost decreasing if

$$\sum_{I_n\in BM(\gamma)} (\operatorname{dist}(I_n, 0) + 1)^{\kappa-2}|I_n|^2 < \infty. \qquad (19)$$

As before, an argument of a MIF $I$ on $\mathbb{R}$ is a real analytic function $\psi$ such that $I = e^{i\psi}$.

**Theorem (5.3.34)[299]:** Let $U$ be a MIF with $|U'| \asymp x^\kappa, \kappa \geq 0, \gamma = \arg U$ on $\mathbb{R}$. Let $J$ be another $MIF, \sigma = \arg J$ on $\mathbb{R}$. Then

I) If $\sigma - (1-\varepsilon)\gamma$ is $\kappa$-almost decreasing, then $J \in \mathcal{D}(U)$;

II) If $\sigma - (1+\varepsilon)\gamma$ is not $\kappa$-almost decreasing, then $J \notin \mathcal{D}(U)$.

Let us point out that even finding an analog for the above statement for $\kappa < 0$ presents an open problem. Such MIFs appear in some of the applications mentioned in [49].

Let us reformulate the last theorem using the relations of TO.

**Theorem (5.3.35)[299]:** In the conditions of Theorem (5.3.34),

I) If $\sigma - (1-\varepsilon)\gamma$ is $\kappa$-almost decreasing, then $J\ ^{\mathrm{T}}\ ^{\mathrm{T}}U$;

II) If $J \overset{\mathrm{T}}{\leqslant} U$ then $\sigma - (1+\varepsilon)\gamma$ is $\kappa$-almost decreasing.

**Proof.** Once again, The general case can be reduced to the case $= B_\Lambda$ : otherwise replace the singular factor of $J$ with its Frostman transform $\mathbf{b}_w(S^a)$.

I) If $\sigma - (1-\varepsilon)\gamma$ is $\kappa$-almost decreasing, then by Theorem (5.3.34) there exists a non-trivial function $f \in N[\bar{U}JB]$, where $B$ is any finite Blaschke product. Denote the zeros of $B$ by $a_1, \dots a_n$. Note that then

$$h = \frac{f}{(z-a_1)(z-a_2)\dots(z-a_n)} \in N[\bar{U}J].$$

If $n = n(\kappa)$ is large enough, $h$ is bounded because $|f| < C|U'|^{1/2}$. Suppose now that $I \in \mathcal{D}(J)$, i.e., there exists non-trivial $g \in N[\bar{J}I]$. Then

$$\bar{U}Ihg = (\bar{U}Jh)(\bar{J}Ig) \in \bar{H}^2,$$

which implies that $I \in \mathcal{D}(U)$. Hence $\mathcal{D}(J) \subset \mathcal{D}(U)$ and $J \overset{\mathrm{T}}{<} U$.

II) If $\sigma - (1+\varepsilon)\gamma$ is not $\kappa$-almost decreasing then $\sigma^* - (1+\varepsilon)\gamma$ is not $\kappa$-almost decreasing where $\sigma^* = \arg J/b_a$ for some zero $a$ of $J$. By Theorem (5.3.34) it means that $J/b_a \notin \mathcal{D}(U)$, while $J/b_a \in \mathcal{D}(J)$. Hence the relation $J \overset{\mathrm{T}}{\leqslant} U$ does not hold.

Like the Beurling-Malliavin problem, the Type problem concerns completeness of complex exponentials in $L^2$-spaces. This time one considers $L^2(\mu)$ for a general finite positive measure $\mu$ on $\mathbb{R}$ and studies completeness of families of exponential functions with frequencies from a fixed interval. We define the type of $\mu$ as

$$\mathcal{T}_\mu = \inf \left\{a \mid e^{ist}, s \in [-a, a], \text{are complete in } L^2(\mu)\right\}.$$

The problem is to find $\mathcal{T}_\mu$ in terms of $\mu$. This problem was considered by N. Wiener (in an equivalent form, [313]) A. Kolmogorov and M. Krein, see [304], [94] or [311]. Using the Toeplitz approach, a formula for $\mathcal{T}_\mu$ was recently found in [310], see also [311]. The idea of the Toeplitz approach to the Type problem can be expressed in terms of Toeplitz order in the following form. Recall that for a positive singular Poisson-finite measure $\mu$ we denote by $\theta_\mu$ the inner function with Clark measure $\mu$. The general case of the Type problem can be easily reduced to the singular case.

**Theorem (5.3.36)[299]:** Let $\mu$ be a positive singular Poisson-finite measure. Then

$$\mathcal{T}_\mu = \sup\ \{a \mid S^a \in \mathcal{D}(\theta_\mu)\}.$$

We say that an inner function $\theta$ divides a Cauchy integral $K\mu$ for some finite complex measure $\mu$ if $K\mu/\theta \in H^p$ for some $p > 0$. Note that then $K\mu/\theta = K\nu$ where $\nu$ is another finite complex measure, $\nu = \bar{\theta}\mu$[309].

**Proof.** Recall that according to the Clark formula every function from $K_\theta, \theta = \theta_\mu$ can be represented in the form $f = (1-\theta)Kf\mu$. Since $1-\theta$ is an outer function in $\mathbb{C}_+$,

$$\begin{aligned} \mathrm{s}\ \{a \mid S^a \in \mathcal{D}(\theta_\mu)\} &= \sup\ \{a \mid S^a \text{ divides } f \in K_\theta\} = \\ &= \sup\ \{a \mid S^a \text{ divides } Kf\mu, f \in L^2(\mu)\}. \end{aligned}$$

By a theorem of Aleksandrov [85] $S^a$ divides $Kf\mu$ iff $f \perp e^{ist}, s \in [-a,a]$. Such an $f$ exists iff the family of exponentials $e^{ist}, s \in [-a,a]$ is incomplete in $L^2(\mu)$.

Utilizing the Beurling-Malliavin multiplier theorem one can deduce the following statement.

**Theorem (5.3.37)[299]:** Let $\mu$ be a positive singular Poisson-finite measure. Then

$$\mathcal{T}_\mu = \sup\ \left\{a \mid S^a \overset{\mathrm{T}}{\leqslant} \theta_\mu\right\}.$$

Let $\mu, \nu$ be two positive Poisson-finite measures such that the Hilbert (Cauchy) transform is bounded from $L^2(\mu)$ to $L^2(\nu)$. Initially one can understand this property in the sense that for a dense family of functions $f \in L^2(\mu)$ the Cauchy integral $Kf\mu$ in the upper half-plane has non-tangential boundary values $f^*(x)$ at $\nu$-a.e. point $x$ and the norm estimate $\|f^*\|_{L^2(\nu)} \le C \parallel f \parallel_{L^2(\mu)}$ holds for all $f$ from that family with a uniform $C$. It follows from a theorem by Aleksandrov [85] that then $f^*$ actually exists $\nu$-a.e. for all $f \in L^2(\nu)$ (and the same norm estimate holds). The general two-weight Hilbert problem asks to describe pairs of measures with this property.

Extensive studies of the 'Tauberian' version of the two-weight Hilbert problem were started in [119] and recently completed in [305], [103]. These important results produced a real analytic description of pairs $\mu$ and $\nu$. We connect this problem with TO.

Once again, if $\mu$ is a positive singular Poisson-finite measure on $\hat{\mathbb{R}}$ we denote by $\theta_\mu$ the corresponding inner function, i.e., the function whose Clark measure is $\mu$. By a theorem from [309], every function $f$ from the model space $K_{\theta_\mu}$ has non-tangential boundary values a.e. with respect to $\mu$. The operator of embedding $K_{\theta_\mu} \to L^2(\mu)$ is a unitary operator. As was mentioned before, this statement generalizes the Parseval theorem from $K_S$ and the counting measure of $\mathbb{Z}$, which is the Clark measure for $S$, to an arbitrary model space and the corresponding Clark measure. The function $f \in K_\theta$ can be recovered from its boundary values in $L^2(\mu)$ via the formula $f = (1-\theta)Kf\mu$.

Some of these connections have already been used in our discussion of TO. To summarize these relations let us recall that the dominance set of $\theta = \theta_\mu$ is the set of all inner divisors of functions from $K_\theta$. As was discussed,

$$\mathcal{D}(\theta_\mu) = \{I \mid I \text{ is an inner divisor of } Kf\mu, f \in L^2(\mu)\}.$$

Let us now return to a pair of Poisson-finite measures $\mu$ and $\nu$ such that the Cauchy transform is bounded from $L^2(\mu)$ to $L^2(\nu)$. In view of the above, this is equivalent to saying that $K_{\theta_\mu}$ is embedded (via passing from a function to its non-tangential boundary values) into $L^2(\eta), \eta = |K\mu|^2\nu$ (or $\left|1-\theta_\mu\right|^{-2}\nu$ ). Note that under the condition of boundedness of the Cauchy transform, the integral $K\mu$, or equivalently the inner function $\theta_\mu$, have non-tangential boundary values $\nu$-a.e. and the above definition of $\eta$ makes sense.

In the case when the measures $\mu$ and $\nu$ are discrete the condition of boundedness of the Cauchy transform can be reformulated in terms of de Branges spaces. Recall that we denote by $E_\mu$ an Hermite-Biehler function such that $E_\mu K_{\theta_\mu} = B(E_\mu)$. The boundedness of the Cauchy transform is equivalent to the boundedness of the natural embedding of $B(E_\mu)$ into $L^2(\gamma), \gamma = \left|E_\mu K_\mu\right|^2\nu$. Note that if $E_\mu = A_\mu + iB_\mu$ is the standard representation of $E_\mu(A_\mu, B_\mu$ are real entire functions, $2A_\mu = E_\mu + E_\mu^\#, 2iB_\mu = E_\mu - E_\mu^\#)$ then $\left|E_\mu K_\mu\right| = \left(A_\mu^2 + B_\mu^2\right)/\left|B_\mu\right|$.

We say that a positive measure $\nu$ on $\hat{\mathbb{R}}$ is sampling for a Banach space $H$ of analytic functions in $\mathbb{C}_+$if the non-tangential limits $f^*(x)$ exist $\nu$-a.e. for a dense family of $f \in H$ and

$$\| f \|_H \asymp \|f^*\|_{L^2(\nu)}.$$

An important case of the two-weight Hilbert problem is when

$$\| f \|_{L^2(\mu)} \asymp \| Kf\mu \|_{L^2(\nu)}.$$

In view of our discussion above, this is equivalent to the property that $\eta = |K\mu|^2\nu$ is a sampling measure for $K_{\theta_\mu}$. The general property, when the Cauchy transform is only norm-bounded from above, can be reduced to the sampling case by adding the Clark measure $\mu_{-1}$ to $\eta$. Namely, if $\mu = \sigma_1$ is the Clark measure for $\theta$, let us denote by $\mu_{-1} = \sigma_{-1}$ the Clark dual measure. The Cauchy transform is bounded from $L^2(\mu)$ to $L^2(\nu)$ iff $\tau = \eta + \mu_{-1}$ is a sampling measure for $K_\theta$.

Reformulating Clark theory for MIFs in terms the corresponding de Branges spaces, we may notice that for any Poisson-finite positive discrete measure $\mu$ on $\mathbb{R}$ there exists a unique regular de Branges space $B(E)$ such that $B(E) = L^2(\mu)$ and supp $\mu = \{E = \bar{E}\}$. We will denote the corresponding HB function by $E^\mu$ and the MIF$(E^\mu)^\#/E^\mu$ by $I^\mu$. The measure $\mu$ is called a de Branges measure for $B(E^\mu)$. Note the following clear connection with the Clark measure $\sigma$ for $I^\mu$ :

$$\mu = \sigma/|E^\mu|^2.$$

Other Clark measures $\sigma_\alpha, \alpha \in \mathbb{T}$ produce other de Branges measures to form the family of de Branges measures for the given space.

As we saw above, the two-weight Hilbert problem is directly related to the problem of description of sampling measures for model spaces $K_\theta$. If $\theta$ is a MIF and $\nu$ is a discrete Poisson-finite measure on $\mathbb{R}$ then $\nu$ is sampling for $K_\theta$ if and only if $\nu/|E_\theta|$ is sampling for $B(E_\theta)$, where $E_\theta$ is any HB function such that $(E_\theta)^\#/E_\theta = \theta$. Thus, in the case of discrete measures, the two-weight problem connects to the description of sampling measures for de Branges spaces.

For the last problem we have the following reformulation in terms of TO. Any measure satisfying

$$\| f \|_{B(E)} = \| f \|_{L^2(\mu)}$$

is called a spectral measure for $B(E)$. Any de Branges space $B(E)$ possesses an infinite family of spectral measures with the de Branges measure defined above being one of them. The spectral measures for a given de Branges space are de Branges measures for the space, de Branges measures for larger de Branges spaces in the chain which contains the given space, and limits of such measures along the chain. The set of spectral measures of a given de Branges space is quite well understood in Krein-de Branges theory. Those measures are spectral measures for the corresponding Krein canonical systems of differential equations, see [55], [91], [308].

The following statement follows from Theorem (5.3.22).

**Theorem (5.3.38)[299]:** Let $\mu, \nu$ be two positive discrete Poisson-finite measures on $\mathbb{R}. TFAE$

(i) The Hilbert (Cauchy) transform is bounded from $L^2(\mu)$ to $L^2(\nu)$.

(ii) The measure $\eta = (\nu + \mu_{-1})/|E_\mu|^2$ is a spectral measure for some $B(F)$ such that $\theta_F \overset{\mathrm{T}}{\sim} \theta_\mu$.

Note that the condition $\theta_F \overset{x}{\sim} \theta_\mu$ means that the inner factors of functions from $B(F)$ in $\mathbb{C}_+$are the same as inner factors of Cauchy integrals $Kf\mu, f \in L^2(\mu)$.

In [77] a theorem by de Branges from [55] was applied to describe sampling sequences for the Paley-Wiener space. Recall that the Paley-Wiener space is a de Branges space with $E = S^{-1}$. Using the same ideas we can formulate the following statement in terms of TO.

**Theorem (5.3.39)[299]:** $\nu$ is a sampling measure for $B(E)$ iff

$$P\nu = \Re \frac{F + F^{\#}\phi}{F - F^{\#}\phi}$$

for some HB function $F$ such that $\theta_F \overset{\sim}{\sim} \theta_E$ and some $\phi \in H^\infty, \| \phi \| \leq 1$.

# Chapter 6
# Cesàro Theorem and Hadamard Space with Convergence of Inductive Means

We study the Ces`aro theorem for the case of positive-order. We give a sharp asymptotic estimate for the limit of (non-weighed) inductive means for the $p$-Schatten class. We study ` an asymptotic property for the Hansen's inductive geometric mean and as an application of the Toeplitz lemma, we show a convergence of the Hansen's inductive mean.

## Section (6.1): Higher Order Multi–Dimensional Extensions

It is known that the Lévy Laplacian [331] is an infinite dimensional Laplacian that can be defined as a Cesàro mean of order 1 of the second derivatives along the elements of an orthonormal basis of a Hilbert space. Motivated by [316] this construction was generalized in [322] to higher order extension of Cesàro means, leading to the notion of exotic Laplacian. The solution of the heat equation for the hierarchy of exotic Laplacians was first obtained in [317].

In [318] it was proved that all exotic Laplacians can be realized in appropriate completions of subspaces of the Hida distribution space, thus showing that this space plays a fundamental role not only for the Lévy Laplacian, but for the whole exotic hierarchy. Finally, in [319] the Markov process generated by the exotic Laplacian of order $2a$ was identified to the Brownian motion associated of the $a$-th distribution derivative of the standard white noise, thus providing a natural probabilistic interpretation for the exotic Laplacians. In fact, after this result, the term exotic seems no longer justified, since these are natural expressions of a fundamental mathematical object such as the since these are natural standard white noise.

The above mentioned identifications were made possible, on one side by a generalization to higher order means, of the known Cesàro's theorems on the arithmetic mean, on the other side, by the formulation and proof of the inverses of these results. This generalization was achieved in successive steps in increasing order of generality: the first result, obtained in [324], concerned sequences (as in the original Cesàro theorem) and means of integer order. The second result in [324] concerns the converse of the mfirst one: this seems to be a new type of Cesàro theorems, Both results played a crucial role in the construction, given in [318] of a similarity relation among exotic Laplacians of order $\geq 1$. This was extended in [319] to sequences and means of arbitrary real order. Finally, sequences are replaced by arbitrary functions on $\mathbb{R}^d$. Such an extension is required in order to bring white noise theory nearer to the quantum field theory formalism and constitutes a first non-trivial step in this direction.

We establish a continuous multidimensional extension of Cesàro theorem for positive higher order. We prove one of our main result concerning the higher order, multi-dimensional and continuous extension of Cesàro theorem. We prove a converse version of the higher order Cesàro theorem studied. We introduce a construction that allows to reduce all Cesarro type theorems to the corresponding results in the discrete 1-dimensional case for sequences.

We study positive order Cesàro theorems for functions on multidimensional Euclidean space.

The Euclidean distance in $\mathbb{R}^d$ is denoted by $\operatorname{dist}(\cdot,\cdot)$ and, for any $r > 0$,

$$B(x,r) := \{y \in \mathbb{R}^d : \operatorname{dist}(x,y) < r\}$$

is the open ball in $\mathbb{R}^d$, centered in $x \in \mathbb{R}^d$, with radius $r$ and $B(x,r)^{\mathrm{c}}$ its complement. We define the generalized Cesàro means as linear functionals defined on some vector subspaces of $L^1_{\text{loc}}(\mathbb{R}^d)$, the space of all locally integrable functions on $\mathbb{R}^d$, and we prove some properties of these means. We start with the Cesàro mean of order 1 , correpsonding to the original version of Cesàro theorem. For a given $p \in \mathbb{R}_+$, the linear functional $C_p$ defined by:

$$\text{Dom}\left(C_p\right) := \left\{ g \in L^1_{\text{loc}}(\mathbb{R}^d) : \lim_{r\to\infty} \frac{1}{|B(0,r)|^p} \int_{B(0,r)} g(k)dk \text{ exists} \right\} \quad (1)$$

$$C_p(g) := \lim_{r\to\infty} \frac{1}{|B(0,r)|^p} \int_{B(0,r)} g(k)dk \quad (2)$$

is called the Cesàro mean of order $p$ of $g \in \text{Dom}\left(C_p\right)$. The Cesàro mean of order 1 is simply called the Cesaro mean.

**Remark (6.1.1)[314]:** In the case $d = 1$,

$$B(0,r) = (-r, r),$$

hence (2) becomes, for $p = 1$ :

$$C_1(g) := \lim_{r\to\infty} \frac{1}{2r} \int_{-r}^{r} g(t)dt \quad (3)$$

for $g \in \text{Dom}(C_1)$, which is slightly different from the usual definition of Cesàro mean

$$C_1(g) := \lim_{r\to\infty} \frac{1}{r} \int_{0}^{r} g(t)dt \quad (4)$$

for $g \in \text{Dom}(C_1)$ that only considers the interval $(0, r)$. For the continuous and multidimensional extensions, the symmetric formulation has some advantages.

**Theorem (6.1.2)[314]:** Let $g \in L^1_{\text{loc}}(\mathbb{R}^d)$ be such that the limit

$$\lim_{|t|\to\infty} \frac{g(t)}{(1 + |t|^2)^{pd}} = \lim_{|t|\to\infty} \frac{g(t)}{|t|^{2pd}} = C \quad (5)$$

exists for some $p \geq 0$. Then:

$$C_{2p+1}(g) = \frac{1}{|B(0,1)|^{2p}} \frac{C}{(2p + 1)} \quad (6)$$

in the sense that the left hand side exists and is equal to the right hand side.

**Proof.** It is clear that the two limits in (5) are equal in the sense that, one exists if and only if the other one does and in this case equality holds. Assumption (5) implies that, for any $\epsilon > 0$, there exists $r_\epsilon \in \mathbb{R}_+$such that, if $|t| \geq r_\epsilon$, then

$$C - \epsilon \leq \frac{g(t)}{|t|^{2pd}} \leq C + \epsilon \text{ or equivalently } (C - \epsilon)|t|^{2pd} \leq g(t) \leq (C + \epsilon)|t|^{2pd}$$

Therefore, for all $r > r_\epsilon$ we obtain that

$$\begin{aligned}(C - \epsilon) \frac{1}{|B(0,r)|^{2p+1}} &\int_{B(0,r)\setminus B(0,r_e)} |t|^{2pd}dt \\ &\leq \frac{1}{|B(0,r)|^{2p+1}} \int_{B(0,r)\setminus B(0,r_e)} g(t)dt \\ &\leq (C + \epsilon) \frac{1}{|B(0,r)|^{2p+1}} \int_{B(0,r)\setminus B(0,r_e)} |t|^{2pd}dt \end{aligned} \quad (7)$$

Clearly

$$\frac{1}{|B(0,r)|^{2p+1}} \int_{B(0,r_\epsilon)} |t|^{2pd}dt \leq \frac{1}{|B(0,r)|^{2p+1}} r_\epsilon^{2pd} \int_{B(0,r_\epsilon)} dt = \frac{r_\epsilon^{2pd}|B(0,r_\epsilon)|}{|B(0,r)|^{2p+1}}$$

which tends to 0 as $r \to \infty$. Put

$$c_\epsilon(r) := \frac{1}{|B(0,r)|^{2p+1}} \int_{B(0,r_e)} |t|^{2pd} dt$$

Then from the known formula:

$$|B(0,r)| = \frac{\pi^{d/2} r^d}{\Gamma\left(\frac{d}{2}+1\right)} = |B(0,1)| r^d$$

we obtain that

$$\begin{aligned} \frac{1}{|B(0,r)|^{2p+1}} \int_{B(0,r)\backslash B(0,r_\epsilon)} |t|^{2pd} dt &= \frac{1}{|B(0,1)|^{2p+1}} \frac{1}{r^{(2p+1)d}} \int_{B(0,r)\backslash B(0,r_\epsilon)} |t|^{2pd} dt \\ &= \frac{1}{|B(0,1)|^{2p+1}} \frac{1}{r^{(2p+1)d}} \int_{B(0,r)} |t|^{2pd} dt - c_\epsilon(r) \\ &= \frac{d}{|B(0,1)|^{2p}} \frac{1}{r^{(2p+1)d}} \int_0^r s^{d-1} s^{2pd} ds - c_\epsilon(r) \end{aligned}$$

which implies that

$$\frac{1}{|B(0,r)|^{2p+1}} \int_{B(0,r)\backslash B(0,r_\epsilon)} |t|^{2pd} dt = \frac{1}{|B(0,1)|^{2p}} \frac{1}{(2p+1)} - c_\epsilon(r)$$

From this it follows that, since $\lim_{r\to\infty} c_\epsilon(r) = 0$, by taking limit as $r \to \infty$ in (7),

$$(C-\epsilon) \frac{1}{|B(0,1)|^{2p}} \frac{1}{(2p+1)} \le \liminf_{r\to\infty} \frac{1}{|B(0,r)|^{2p+1}} \int_{B(0,r)\backslash B(0,r_e)} g(t) dt$$
$$\le \limsup_{r\to\infty} \frac{1}{|B(0,r)|^{2p+1}} \int_{B(0,r)\backslash B(0,r_e)} g(t) dt \le (C+\epsilon) \frac{1}{|B(0,1)|^{2p}} \frac{1}{(2p+1)}$$

Since $\epsilon > 0$ is arbitrary and $\lim_{r\to\infty} \frac{1}{\sqrt{B(0,r)|^{2p+1}}} \int_{B(0,r_e)} g(t) dt = 0$, it follows that

$$\lim_{r\to\infty} \frac{1}{|B(0,r)|^{2p+1}} \int_{B(0,r)} g(t) dt = \frac{1}{|B(0,1)|^{2p}} \frac{C}{(2p+1)}$$

in the sense that the limit on the left hand side exists and the identity holds. This proves (6). By taking $p = 0$ in Theorem (6.1.2), we have the following multi-dimensional continuous Cesàro theorem.

**Theorem (6.1.3)[314]:** If $g \in L^1_{\text{loc}}(\mathbb{R}^d)$ is such that the limit

$$\lim_{|t|\to\infty} =: g_\infty \tag{8}$$

exists in $\mathbb{C}$, then

$$C_1(g) = \lim_{r\to+\infty} \frac{1}{|B(0,r)|} \int_{B(0,r)} g(s) ds =: g_\infty \tag{9}$$

in the sense that the limit exists and the equality holds.

We study higher order Cesarro theorems for functions on multidimensional Euclidean space.

**Theorem (6.1.4)[314]:** If, for some $p \in \mathbb{R}_+$and some $g \in L^1_{\text{loc}}(\mathbb{R}^d)$, the limit

$$\lim_{r\to\infty} \frac{1}{|B(0,r)|^p} \int_{B(0,r)} g(t) dt = C_p(g) \tag{10}$$

exists in $\mathbb{C}$, then for any $a \in \mathbb{R}_+^* = \mathbb{R}_+ \setminus \{0\}$, it holds that

$$\lim_{r\to\infty} \frac{1}{|B(0,r)|^{p+a}} \int_{B(0,r)} |t|^{ad} g(t) dt = \frac{1}{|B(0,1)|^a} \frac{p}{p+a} C_p(g) \tag{11}$$

in the sense that the limit exists and the equality holds.

**Proof.** Let $g \in \mathrm{Dom}(C_p)$ and $p \in \mathbb{R}_+$. If $p > 0$, for some $R > 0$ we consider the function

$$g_R(t) = \begin{cases} g(t) & \text{if } |t| \geq R, \\ 0 & \text{if } |t| < R. \end{cases}$$

If $p = 0$ we regard $g_R(t)$ as $g(t)$. Then we can check that $g_R \in L^1_{\mathrm{loc}}(\mathbb{R}^d), C_p(g) = C_p(g_R)$ and

$$\lim_{r\to\infty} \frac{1}{|B(0,r)|^{p+a}} \int_{B(0,r)} |t|^{ad} g(t)dt = \lim_{r\to\infty} \frac{1}{|B(0,r)|^{p+a}} \int_{B(0,r)} |t|^{ad} g_R(t)dt$$

in the sense that, if there exists one side of the equality, then there exists another side. Therefore we may prove (11) for $g_R$. We also have

$$\frac{1}{|B(0,r)|^{p+a}} \int_{B(0,r)} |t|^{ad} g_R(t)dt = \frac{1}{|B(0,r)|^{p+a}} r^{ad} \int_{B(0,r)} g_R(t)dt - \frac{1}{|B(0,r)|^{p+a}} \int_{B(0,r)} (r^{ad} - |t|^{ad}) g_R(t)dt \quad (12)$$

On the one hand, by using the identity $|B(0,r)| = |B(0,1)|r^d$, we obtain that

$$\begin{aligned} \frac{1}{|B(0,r)|^{p+a}} r^{ad} \int_{B(0,r)} g_R(t)dt &= \frac{1}{|B(0,1)|^{p+a} r^{pd}} \int_{B(0,r)} g_R(t)dt \\ &= \frac{1}{|B(0,1)|^{a}} \frac{1}{|B(0,r)|^{p}} \int_{B(0,r)} g_R(t)dt \end{aligned}$$

and assumption (11) implies that the limit of the right hand side for $r \to \infty$ exists and is equal to

$$\frac{1}{|B(0,1)|^{a}} C_p(g_R) \quad (13)$$

On the other hand, in the second term of the difference in (12), by using $d$-dimensional spherical coordinate representation, we have

$$\int_{B(0,r)} (r^{ad} - |t|^{ad}) g_R(t)dt = \int_0^r (r^{ad} - s^{ad}) s^{d-1} \left( \int_\Theta \hat{g_R}(s,\hat{\theta})d\hat{\theta} \right) ds$$

for some function $\hat{g_R}$ induced by $g_R$ via $d$-dimensional spherical coordinate representation, where $\Theta = [0,\pi]^{d-2} \times [0,2\pi]$. Then, by using the identity $|B(0,r)| = |B(0,1)|r^d$ again, we obtain that

$$\begin{aligned} &\frac{1}{|B(0,r)|^{p+a}} \int_{B(0,r)} (r^{ad} - |t|^{ad}) g_R(t)dt \\ &= \frac{1}{|B(0,1)|^{p+a} r^{(p+a)d}} \int_0^r (r^{ad} - s^{ad}) s^{d-1} \left( \int_\Theta \hat{g_R}(s,\hat{\theta})d\hat{\theta} \right) ds \\ &= \frac{ad}{|B(0,1)|^{p+a} r^{(p+a)d}} \int_0^r s^{d-1} \left( \int_0^r \tau^{ad-1} d\tau \right) \left( \int_\Theta \hat{g_R}(s,\hat{\theta})d\hat{\theta} \right) ds \\ &= \frac{ad}{|B(0,1)|^{p+a} r^{(p+a)d}} \int_0^r \left[ \int_0^\tau s^{d-1} \left( \int_\Theta \hat{g_R}(s,\hat{\theta})d\hat{\theta} \right) ds \right] \tau^{ad-1} d\tau \qquad (14) \end{aligned}$$

which becomes

$$= \frac{ad}{|B(0,1)|^a r^{(p+a)d}} \int_0^r \tau^{pd+ad-1} \left( \frac{1}{|B(0,1)|^p \tau^{pd}} \int_{B(0,\tau)} g_R(t)dt \right) d\tau$$
$$= \frac{ad}{|B(0,1)|^a r^{(p+a)d}} \int_0^r \tau^{pd+ad-1} \left( \frac{1}{|B(0,\tau)|^p} \int_{B(0,\tau)} g_R(t)dt \right) d\tau$$
$$= \frac{ad}{|B(0,1)|^a r^{(p+a)d}} \int_0^r \tau^{pd+ad-1} \left( \frac{1}{|B(0,\tau)|^p} \int_{B(0,\tau)} g_R(t)dt - C_p(g_R) \right) d\tau$$
$$+ \frac{ad}{|B(0,1)|^a r^{(p+a)d}} \left( \int_0^r \tau^{pd+ad-1} d\tau \right) C_p(g_R) \qquad (15)$$

The second term of (15) is equal to

$$\frac{ad}{|B(0,1)|^a r^{(p+a)d}} \frac{1}{(p+a)d} r^{pd+ad} C_p(g_R) = \frac{1}{|B(0,1)|^a} \frac{a}{p+a} C_p(g_R) \qquad (16)$$

and the first term of (15) is majorized, in modulus, by

$$\frac{ad}{|B(0,1)|^a r^{(p+a)d}} \int_0^r \tau^{(p+a)d-1} \left| \frac{1}{|B(0,\tau)|^p} \int_{B(0,\tau)} g_R(t)dt - C_p(g_R) \right| d\tau \qquad (17)$$

Let $\varepsilon > 0$ be given. Then since by assumption

$$\lim_{\tau \to \infty} \left| \frac{1}{|B(0,\tau)|^p} \int_{B(0,\tau)} g_R(t)dt - C_p(g_R) \right| = 0$$

there exists $t_\varepsilon$ such that, for any $\tau \geq t_\varepsilon$

$$\left| \frac{1}{|B(0,\tau)|^p} \int_{B(0,\tau)} g_R(t)dt - C_p(g_R) \right| \leq \varepsilon$$

Moreover, since $g_R$ is locally integrable, the map

$$\tau \in \mathbb{R}_+ \longmapsto \frac{1}{|B(0,\tau)|^p} \int_{B(0,\tau)} g_R(t)dt$$

is continuous, therefore there exists a constant $C > 0$ such that for all $\tau \leq t_\varepsilon$

$$\left| \frac{1}{|B(0,\tau)|^p} \int_{B(0,\tau)} g_R(t)dt - C_p(g_R) \right| \leq C$$

Therefore, by splitting the integral in $\tau$ in (17) into the two pieces, (17) is majorized by

$$\frac{ad}{|B(0,1)|^a r^{(p+a)d}} \left( C \int_0^{t_\varepsilon} \tau^{(p+a)d-1} d\tau + \varepsilon \int_{t_\varepsilon}^r \tau^{(p+a)d-1} d\tau \right)$$
$$\leq \frac{ad}{|B(0,1)|^a} \left( \frac{1}{(p+a)d} \left( \frac{t_\varepsilon}{r} \right)^{(p+a)d} C + \varepsilon \frac{1}{r^{(p+a)d}} \int_0^r \tau^{(p+a)d-1} d\tau \right)$$
$$\leq \frac{a}{|B(0,1)|^a (p+a)} \left( \left( \frac{t_\varepsilon}{r} \right)^{(p+a)d} C + \varepsilon \right)$$

and so

$$\lim_{r\to\infty}\frac{ad}{|B(0,1)|^a r^{(p+a)d}}\int_0^r \tau^{(p+a)d-1}\left|\frac{1}{|B(0,\tau)|^p}\int_{B(0,\tau)} g_R(t)dt - C_p(g_R)\right| d\tau$$
$$\leq \lim_{r\to\infty}\frac{a}{|B(0,1)|^a (p+a)}\left(\left(\frac{t_\varepsilon}{r}\right)^{(p+a)d} C+\varepsilon\right)$$
$$\leq \frac{a}{|B(0,1)|^a(p+a)}\varepsilon$$

from which, since $\varepsilon > 0$ is arbitrary,

$$\lim_{r\to\infty}\frac{ad}{|B(0,1)|^a r^{(p+a)d}}\int_0^r \tau^{pd+ad-1}\left(\frac{1}{|B(0,\tau)|^p}\int_{B(0,\tau)} g_R(t)dt - C_p(g_R)\right) d\tau = 0 \quad (18)$$

Hence by (14), (15), (16) and (18), we have

$$\lim_{r\to\infty}\frac{1}{|B(0,r)|^{p+a}}\int_{B(0,r)} (r^{ad}-|t|^{ad})g_R(t)dt = \frac{1}{|B(0,1)|^a}\frac{a}{p+a}C_p(g_R).$$

Therefore, by (12) and (13), we have

$$\lim_{r\to\infty}\frac{1}{|B(0,r)|^{p+a}}\int_{B(0,r)} |t|^{ad}g_R(t)dt = \frac{1}{|B(0,1)|^a}C_p(g_R) - \frac{1}{|B(0,1)|^a}\frac{a}{p+a}C_p(g_R)$$
$$= \frac{1}{|B(0,1)|^a}\frac{p}{p+a}C_p(g_R),$$

which implies (11) as desired.

The following theorem gives the converse of Theorem (6.1.4).

**Theorem (6.1.5)[314]:** If, for some $p \in \mathbb{R}_+^*$, some a $\in \mathbb{R}_+$ and some $f \in L^1_{\text{loc}}(\mathbb{R}^d)$, the limit

$$C_{p+a}(f) := \lim_{r\to\infty}\frac{1}{|B(0,r)|^{p+a}}\int_{B(0,r)} f(t)dt \quad (19)$$

exists in $\mathbb{C}$ and $|\cdot|^{-ad}f(\cdot) \in L^1_{\text{loc}}(\mathbb{R}^d)$, then it holds that

$$\lim_{r\to\infty}\frac{1}{|B(0,r)|^p}\int_{B(0,r)} |t|^{-ad}f(t)dt = \frac{p+a}{p}|B(0,1)|^a C_{p+a}(f) \quad (20)$$

in the sense that the limit exists and the equality holds.

**Proof.** The proof is similar to the one Let $p \in \mathbb{R}_+^*, a \in \mathbb{R}_+$ and $f \in L^1_{\text{loc}}(\mathbb{R}^d)$ be such that (19) holds. Then we have

$$\frac{1}{|B(0,r)|^p}\int_{B(0,r)} |t|^{-ad}f(t)dt = \frac{1}{|B(0,r)|^p}r^{-ad}\int_{B(0,r)} f(t)dt$$
$$-\frac{1}{|B(0,r)|^p}\int_{B(0,r)} (r^{-ad}-|t|^{-ad})f(t)dt \quad (21)$$

On the one hand, with the same notations-

$$-\frac{1}{|B(0,r)|^p}\int_{B(0,r)}(r^{-ad}-|t|^{-ad})f(t)dt$$

$$=-\frac{1}{|B(0,1)|^p r^{pd}}\int_0^r (r^{-ad}-s^{-ad})s^{d-1}\left(\int_\Theta \hat{f}(s,\hat{\theta})d\hat{\theta}\right)ds$$

$$=\frac{ad}{|B(0,1)|^p r^{pd}}\int_0^r s^{d-1}\left(\int_s^r \tau^{-ad-1}d\tau\right)\left(\int_\Theta \hat{f}(s,\hat{\theta})d\hat{\theta}\right)ds$$

$$=\frac{ad}{|B(0,1)|^p r^{pd}}\int_0^r \left[\int_0^\tau s^{d-1}\left(\int_\Theta \hat{f}(s,\hat{\theta})d\hat{\theta}\right)ds\right]\tau^{-ad-1}d\tau = \tag{22}$$

$$=\frac{ad|B(0,1)|^a}{r^{pd}}\int_0^r \tau^{pd-1}\left(\frac{1}{|B(0,\tau)|^{p+a}}\int_{B(0,\tau)} f(t)dt\right)d\tau$$

$$=\frac{ad|B(0,1)|^a}{r^{pd}}\int_0^r \tau^{pd-1}\left(\frac{1}{|B(0,\tau)|^{p+a}}\int_{B(0,\tau)} f(t)dt - C_{p+a}(f)\right)d\tau$$

$$+\frac{ad|B(0,1)|^a}{r^{pd}}\left(\int_0^r \tau^{pd-1}d\tau\right)C_{p+a}(f). \tag{23}$$

Therefore, by (21), (22) and (23) the following identity holds:

$$\frac{1}{|B(0,r)|^p}\int_{B(0,r)}|t|^{-ad}f(t)dt = k_1(r)+k_2(r)+k_3(r),$$

where

$$k_1(r)=\frac{1}{|B(0,r)|^p}r^{-ad}\int_{B(0,r)} f(t)dt, \tag{24}$$

$$k_2(r)=\frac{ad|B(0,1)|^a}{r^{pd}}\left(\int_0^r \tau^{pd-1}d\tau\right)C_{p+a}(f), \tag{25}$$

$$k_3(r)=\frac{ad|B(0,1)|^a}{r^{pd}}\int_0^r \tau^{pd-1}\left(\frac{1}{|B(0,\tau)|^{p+a}}\int_{B(0,\tau)} f(t)dt - C_{p+a}(f)\right)d\tau \tag{26}$$

Then by (24) and the assumption we obtain that

$$\lim_{r\to\infty}k_1(r)=|B(0,1)|^a\lim_{r\to\infty}\frac{1}{|B(0,r)|^{p+a}}\int_{B(0,r)} f(t)dt=|B(0,1)|^a C_{p+a}(f)$$

and from (25), we have

$$k_2(r)=\frac{ad|B(0,1)|^a}{r^{pd}}\left(\int_0^r \tau^{pd-1}d\tau\right)C_{p+a}(f)=\frac{a}{p}|B(0,1)|^a C_{p+a}(f)$$

Also, by the same arguments used for the proof of (18), we prove that $\lim_{t\to\infty}k_3(t)=0$. Therefore, from the above arguments it follows that

$$\begin{aligned}\lim_{r\to\infty}\frac{1}{|B(0,r)|^p}\int_{B(0,r)}|t|^{-ad}f(t)dt &= \lim_{r\to\infty}(k_1(r)+k_2(r)+k_3(r))\\ &= |B(0,1)|^a C_{p+a}(f)+\frac{a}{p}|B(0,1)|^a C_{p+a}(f)\\ &= \frac{p+a}{p}|B(0,r)|^a C_{p+a}(f)\end{aligned}$$

which is the desired result.

By reducing from continuous to discrete case, we show how, from the previous theorems, one can obtain the corresponding statements for sequences.

Sequences of complex numbers, i.e. elements of $\mathbb{C}^\infty$, are identified with functions in $L^1(\mathbb{R})$ which are constant in the intervals $[k,k+1)(k \in \mathbb{Z})$. We define the embedding $c\colon \mathbb{C}^\infty \ni a \mapsto c_a \in L^1_{\text{loc}}(\mathbb{R})$ by

$$c_a(t) := \begin{cases} a_n, & n \in \mathbb{N}, t \in [n,n+1), \\ 0, & t < 1. \end{cases}$$

The action of the functionals $\tilde{C}_p$ on $\mathbb{C}^\infty$ is defined by

$$\tilde{C}_p(a) := 2^p C_p(c_a) := \lim_{n\to+\infty} \frac{1}{n^p} \sum_{k=1}^{n} a_k$$

$$\operatorname{Dom}(\tilde{C}_p) := \left\{ a \in \mathbb{C}^\infty : \lim_{n\to\infty} \frac{1}{n^p} \sum_{k=1}^{n} a(k) \text{ exists in } \mathbb{C} \right\}$$

Then clearly for any $p \geq 0$,

$$a \in \operatorname{Dom}(\tilde{C}_p) \Leftrightarrow c_a \in \operatorname{Dom}(C_p)$$

Let $Q$ be the locally integrable function on $\mathbb{R}$ defined by

$$Q(t) = \begin{cases} \dfrac{[|t|]}{|t|}, & t \geq 1 \\ 0, & t < 1 \end{cases}$$

where $[|t|]$ is the largest integer smaller than $|t|$. For our convenience, we understand that $\infty \cdot 0 = 0$.

**Lemma (6.1.6)[314]:** Let $a \in \operatorname{Dom}(\tilde{C}_p)$. Then for any $\beta \in \mathbb{R}$, it holds that

$$\lim_{r\to\infty} \frac{1}{r^p} \int_1^r \left[ \left( \frac{[|t|]}{|t|} \right)^\beta - 1 \right] c_a(t)\,dt = 0$$

**Proof.** Let $\beta \in \mathbb{R}$ be given. Then we obtain that

$$\lim_{r\to\infty} \frac{1}{r^p} \int_1^r \left[ \left( \frac{[|t|]}{|t|} \right)^\beta - 1 \right] c_a(t)\,dt = \lim_{r\to\infty} \frac{1}{r^p} \sum_{k=1}^{[r]-1} a_k \left( \int_k^{k+1} \left[ \left( \frac{k}{t} \right)^\beta - 1 \right] dt \right) + \lim_{r\to\infty} \frac{a_{[r]}}{r^p} \int_{[r]}^r \left[ \left( \frac{[r]}{t} \right)^\beta - 1 \right] dt \tag{27}$$

Then we can easily see that the second term of the right hand side of (27) is zero and the first term of the right hand side of (27) coincides with the following:

$$\lim_{N\to\infty} \frac{1}{N^p} \sum_{k=1}^{N} a_k b_k,\ b_k = \int_k^{k+1} \left[ \left( \frac{k}{t} \right)^\beta - 1 \right] dt$$

Therefore, we have

$$\lim_{r\to\infty} \frac{1}{r^p} \int_1^r \left[ \left( \frac{[|t|]}{|t|} \right)^\beta - 1 \right] c_a(t)\,dt = \lim_{N\to\infty} \frac{1}{N^p} \sum_{k=1}^{N} a_k b_k \tag{28}$$

Then the Abel identity implies that

$$\lim_{N\to\infty}\frac{1}{N^p}\sum_{k=1}^{N} a_k b_k = \lim_{N\to\infty}\frac{1}{N^p}\left(\sum_{k\le N} a_k\right) b_N - \lim_{N\to\infty}\frac{1}{N^p}\sum_{k\le N-1} (b_{k+1}-b_k)\left(\sum_{r\le k} a_r\right)$$
$$= -\lim_{N\to\infty}\frac{1}{N^p}\sum_{k\le N-1} (b_{k+1}-b_k)k^p\left(\frac{1}{k^p}\sum_{r\le k} a_r\right)$$
On the other hand, by direct computation we can prove that $\lim_{k\to\infty} k(b_{k+1}-b_k)=0$. Let $\{d_n\}_{n=1}^{\infty}$ be a sequence given by
$$d_n = -n(b_{n+1}-b_n)\left(\frac{1}{n^p}\sum_{r\le n} a_r\right),\ n\in\mathbb{N}$$
Then we have $\lim_{n\to\infty} d_n = 0\tilde{C}_p(a)=0$, i.e., for any $\epsilon>0$ there exists a number $N_0$ such that $|d_n|<\epsilon$ if $n\ge N_0$. Using this sequence (28) is represented as follows:
$$\lim_{r\to\infty}\frac{1}{r^p}\int_1^r \left[\left(\frac{[|t|]}{|t|}\right)^{\beta}-1\right] c_a(t)dt = \lim_{N\to\infty}\frac{1}{N^p}\sum_{k=1}^{N} a_k b_k$$
$$= -\lim_{N\to\infty}\frac{1}{N^p}\sum_{k\le N-1} (b_{k+1}-b_k)k^p\left(\frac{1}{k^p}\sum_{r\le k} a_r\right)$$
$$= \lim_{N\to\infty}\frac{1}{N^p}\sum_{k\le N-1} k^{p-1}d_k$$
Therefore we obtain that
$$\left|\lim_{r\to\infty}\frac{1}{r^p}\int_1^r \left[\left(\frac{[|t|]}{|t|}\right)^{\beta}-1\right] c_a(t)dt\right| \le \lim_{N\to\infty}\frac{1}{N^p}\sum_{k\le N-1} k^{p-1}|d_k|$$
$$= \lim_{N\to\infty}\frac{1}{N^p}\sum_{k=0}^{N_0-1} k^{p-1}|d_k| + \lim_{N\to\infty}\frac{1}{N^p}\sum_{k=N_0}^{N-1} k^{p-1}|d_k| \le \lim_{N\to\infty}\frac{1}{N^p}\sum_{k=N_0}^{N-1} k^{p-1}|d_k| \le C\epsilon$$
with a constant $C$ independent on $N$. Since $\epsilon$ is arbitrary, this gives the proof of the assertion.

**Lemma (6.1.7)[314]:** Let $a\in \mathrm{Dom}(\tilde{C}_p)$. Then for any $\beta\in\mathbb{R}, Q^{\beta}c_a\in\mathrm{Dom}(C_p)$ and
$$C_p(Q^{\beta}c_a)=C_p(c_a)=2^{-p}\tilde{C}_p(a).$$
**Proof.** For given $\beta\in\mathbb{R}$, by applying Lemma (6.1.6), we obtain that
$$C_p(Q^{\beta}c_a) = \lim_{r\to\infty}\frac{1}{|B(0,r)|^p}\int_{B(0,r)}\left(\frac{[|t|]}{|t|}\right)^{\beta} c_a(t)dt$$
$$= \lim_{r\to\infty}\frac{1}{|B(0,r)|^p}\int_{B(0,r)}\left[\left(\frac{[|t|]}{|t|}\right)^{\beta}-1\right] c_a(t)dt + C_p(c_a)$$
$$= C_p(c_a)$$
which gives the proof.

Denote $q$ the identity function on $\mathbb{R}, q(t)=t$ for any $t\in\mathbb{R}$.

**Lemma (6.1.8)[314]:** Let $a\in\mathrm{Dom}(\tilde{C}_p)$ and $\beta\in\mathbb{R}$. Then $|q|^{\beta}Q^{\beta}c_a\in\mathrm{Dom}(C_p)$ and
$$C_p(|q|^{\beta}Q^{\beta}c_a)=2^{-p}\tilde{C}_p\left(q_d^{\beta}a\right)$$
where $q_d$ is the discretization of the multiplication operator $q$, i.e., $\left(q_d^{\beta}a\right)(n)=n^{\beta}a_n$ for $a\in\mathbb{C}^{\infty}$.

**Proof.** We obtain that

$$\begin{aligned}
C_p\left(|q|^{\beta}Q^{\beta}c_a\right) &= \lim_{r\to\infty}\frac{1}{(2r)^p}\int_{-r}^{r}|q|(t)^{\beta}Q(t)^{\beta}c_a(t)dt\\
&= \lim_{r\to\infty}\frac{1}{(2r)^p}\int_{-r}^{r}|t|^{\beta}\left(\frac{[|t|]}{|t|}\right)^{\beta}c_a(t)dt\\
&= \lim_{n\to\infty}\frac{1}{(2n)^p}\int_{1}^{n}[|t|]^{\beta}c_a(t)dt\\
&= 2^{-p}\lim_{n\to\infty}\frac{1}{n^p}\sum_{k=1}^{n}k^{\beta}a_k\\
&= 2^{-p}\tilde{C}_p\left(q_d^{\beta}a\right)
\end{aligned}$$

which implies the proof.

By Lemma (6.1.7), the continuous higher order Cesàro theorem reduces the higher order Cesàro theorem for sequences. Then the following theorems follow from Theorems (6.1.4) and (6.1.5).

**Theorem (6.1.9)[314]:** ([324], [319]) Let the sequence $a=(a_n)_{n=1}^{\infty}\in\mathbb{C}^{\infty}$ be such that, for some $p>0$ the limit

$$\lim_{N\to\infty}\frac{1}{N^p}\sum_{n=1}^{N}a_n=:C_p(a)$$

exists. Then for each $\alpha\in\mathbb{R}_+$, one has

$$\lim_{N\to\infty}\frac{1}{N^{p+\alpha}}\sum_{n=1}^{N}n^{\alpha}a_n=\frac{p}{p+\alpha}C_p(a)$$

in the sense that the limit on the left hand side exists and the equality holds.

**Theorem (6.1.10)[314]:** ([324],[319]) Let $p>0,\alpha\geq 0$ and let $a=(a_n)_{n=1}^{\infty}$ be a sequence in $\mathbb{C}^{\infty}$ such that the limit

$$\lim_{N\to\infty}\frac{1}{N^{p+\alpha}}\sum_{n=1}^{N}a_n=:C_{p+\alpha}(a)$$

exists. Then

$$\lim_{N\to\infty}\frac{1}{N^p}\sum_{n=1}^{N}n^{-\alpha}a_n=\frac{p+\alpha}{p}C_{p+\alpha}(a).$$

**Corollary (6.1.11)[350]:** (See [314]). Let $g^m\in L^1_{\mathrm{loc}}(\mathbb{R}^d)$ be such that the limit

$$\lim_{|t|\to\infty}\sum_m\frac{g^m(t)}{(1+|t|^2)^{(1+\epsilon)d}}=\lim_{|t|\to\infty}\sum_m\frac{g^m(t)}{|t|^{2(1+\epsilon)d}}=C_m \qquad (29)$$

exists for some $\epsilon\geq -1$. Then:

$$C^m_{3+2\epsilon}(g^m)=\frac{1}{|B(0,1)|^{2(1+\epsilon)}}\frac{C_m}{(3+2\epsilon)} \qquad (30)$$

in the sense that the left hand side exists and is equal to the right hand side.

**Proof.** It is clear that the two limits in (29) are equal in the sense that, one exists if and only if the other one does and in this case equality holds. Assumption (29) implies that, for any $\epsilon>0$, there exists $1+\epsilon\in\mathbb{R}_+$such that, if $|t|\geq 1+\epsilon$, then

$$C_m - \epsilon \leq \sum_m \frac{g^m(t)}{|t|^{2(1+\epsilon)d}} \leq C_m + \epsilon \text{ or equivalently } (C_m - \epsilon)|t|^{2(1+\epsilon)d} \leq \sum_m g^m(t) \leq (C_m + \epsilon)|t|^{2(1+\epsilon)d}$$

Therefore, for all $\epsilon > 0$ we obtain that

$$(C_m - \epsilon)\frac{1}{|B(0,1+2\epsilon)|^{3+2\epsilon}}\int_{B(0,1+2\epsilon)\setminus B(0,1+\epsilon)} |t|^{2(1+\epsilon)d} dt$$
$$\leq \frac{1}{|B(0,1+2\epsilon)|^{3+2\epsilon}}\int_{B(0,1+2\epsilon)\setminus B(0,1+\epsilon)} \sum_m g^m(t) dt$$
$$\leq (C_m + \epsilon)\frac{1}{|B(0,1+2\epsilon)|^{3+2\epsilon}}\int_{B(0,1+2\epsilon)\setminus B(0,1+\epsilon)} |t|^{2(1+\epsilon)d} dt \qquad (31)$$

Clearly

$$\frac{1}{|B(0,1+2\epsilon)|^{3+2\epsilon}}\int_{B(0,1+\epsilon)} |t|^{2(1+\epsilon)d} dt \leq \frac{1}{|B(0,1+2\epsilon)|^{3+2\epsilon}}(1+\epsilon)^{2(1+\epsilon)d}\int_{B(0,1+\epsilon)} dt$$
$$= \frac{(1+\epsilon)^{2(1+\epsilon)d}|B(0,1+\epsilon)|}{|B(0,1+2\epsilon)|^{3+2\epsilon}}$$

which tends to 0 as $\epsilon \to \infty$. Put

$$c_\epsilon(1+2\epsilon) := \frac{1}{|B(0,1+2\epsilon)|^{3+2\epsilon}}\int_{B(0,1+\epsilon)} |t|^{2(1+\epsilon)d} dt$$

Then from the known formula:

$$|B(0,1+2\epsilon)| = \frac{\pi^{d/2}(1+2\epsilon)^d}{\Gamma\left(\frac{d}{2}+1\right)} = |B(0,1)|(1+2\epsilon)^d$$

we obtain that

$$\frac{1}{|B(0,1+2\epsilon)|^{3+2\epsilon}}\int_{B(0,1+2\epsilon)\setminus B(0,1+\epsilon)} |t|^{2(1+\epsilon)d} dt$$
$$= \frac{1}{|B(0,1)|^{3+2\epsilon}}\frac{1}{(1+2\epsilon)^{(3+2\epsilon)d}}\int_{B(0,1+2\epsilon)\setminus B(0,1+\epsilon)} |t|^{2(1+\epsilon)d} dt$$
$$= \frac{1}{|B(0,1)|^{3+2\epsilon}}\frac{1}{(1+2\epsilon)^{(3+2\epsilon)d}}\int_{B(0,1+2\epsilon)} |t|^{2(1+\epsilon)d} dt - c_\epsilon(1+2\epsilon)$$
$$= \frac{d}{|B(0,1)|^{2(1+\epsilon)}}\frac{1}{(1+2\epsilon)^{(3+2\epsilon)d}}\int_0^{1+2\epsilon} s^{d-1}s^{2(1+\epsilon)d} ds - c_\epsilon(1+2\epsilon)$$

which implies that

$$\frac{1}{|B(0,1+2\epsilon)|^{3+2\epsilon}}\int_{B(0,1+2\epsilon)\setminus B(0,1+\epsilon)} |t|^{2(1+\epsilon)d} dt$$
$$= \frac{1}{|B(0,1)|^{2(1+\epsilon)}}\frac{1}{(3+2\epsilon)} - c_\epsilon(1+2\epsilon)$$

From this it follows that, since $\lim_{\epsilon\to\infty} c_\epsilon(1+2\epsilon) = 0$, by taking limit as $\epsilon \to \infty$ in (31),

$$
\begin{aligned}
(C_m-\epsilon)\frac{1}{|B(0,1)|^{2(1+\epsilon)}}\frac{1}{(3+2\epsilon)}
&\le \liminf_{\epsilon\to\infty}\frac{1}{|B(0,1+2\epsilon)|^{3+2\epsilon}}\int_{B(0,1+2\epsilon)\setminus B(0,1+\epsilon)}\sum_m g^m(t)dt\\
&\le \limsup_{\epsilon\to\infty}\frac{1}{|B(0,1+2\epsilon)|^{3+2\epsilon}}\int_{B(0,1+2\epsilon)\setminus B(0,1+\epsilon)}\sum_m g^m(t)dt\\
&\le (C_m+\epsilon)\frac{1}{|B(0,1)|^{2(1+\epsilon)}}\frac{1}{(3+2\epsilon)}
\end{aligned}
$$

Since $\epsilon>0$ is arbitrary and $\lim_{\epsilon\to\infty}\frac{1}{\sqrt{B(0,1+2\epsilon)|^{3+2\epsilon}}}\int_{B(0,1+\epsilon)}\sum_m g^m(t)dt=0$, it follows that

$$\lim_{\epsilon\to\infty}\frac{1}{|B(0,1+2\epsilon)|^{3+2\epsilon}}\int_{B(0,1+2\epsilon)}\sum_m g^m(t)dt=\frac{1}{|B(0,1)|^{2(1+\epsilon)}}\frac{C_m}{(3+2\epsilon)}$$

in the sense that the limit on the left hand side exists and the identity holds. This proves (30). By taking $\epsilon=-1$ in Corollary (6.1.11), we have the following multi-dimensional continuous Cesàro theorem (see [314]).

**Corollary (6.1.12)[250]:** (See [314]). If, for some $1+\epsilon\in\mathbb{R}_+$and some $g^m\in L^1_{\text{loc}}(\mathbb{R}^d)$, the limit

$$\lim_{\epsilon\to\infty}\frac{1}{|B(0,1+2\epsilon)|^{1+\epsilon}}\int_{B(0,1+2\epsilon)}\sum_m g^m(t)dt=\sum_m C^m_{1+\epsilon}(g^m) \tag{32}$$

exists in $\mathbb{C}$, then for any $a\in\mathbb{R}_+^*=\mathbb{R}_+\setminus\{0\}$, it holds that

$$
\begin{aligned}
&\lim_{\epsilon\to\infty}\frac{1}{|B(0,1+2\epsilon)|^{1+\epsilon+a}}\int_{B(0,1+2\epsilon)}\sum_m |t|^{ad}g^m(t)dt\\
&\quad=\frac{1}{|B(0,1)|^a}\frac{1+\epsilon}{1+\epsilon+a}\sum_m C^m_{1+\epsilon}(g^m)
\end{aligned} \tag{33}
$$

in the sense that the limit exists and the equality holds.

**Proof.** Let $g^m\in\text{Dom}(C^m_{1+\epsilon})$ and $1+\epsilon\in\mathbb{R}_+$. If $\epsilon\ge0$, for some $R>0$ we consider the function

$$g^m_R(t)=\begin{cases}g^m(t) & \text{if } |t|\ge R,\\ 0 & \text{if } |t|<R.\end{cases}$$

If $\epsilon=-1$ we regard $g^m_R(t)$ as $g^m(t)$. Then we can check that $g^m_R\in L^1_{\text{loc}}(\mathbb{R}^d)$, $C^m_{1+\epsilon}(g^m)=C^m_{1+\epsilon}(g^m_R)$ and

$$
\begin{aligned}
&\lim_{\epsilon\to\infty}\frac{1}{|B(0,1+2\epsilon)|^{1+\epsilon+a}}\int_{B(0,1+2\epsilon)}\sum_m |t|^{ad}g^m(t)dt\\
&\quad=\lim_{\epsilon\to\infty}\frac{1}{|B(0,1+2\epsilon)|^{1+\epsilon+a}}\int_{B(0,1+2\epsilon)}\sum_m |t|^{ad}g^m_R(t)dt
\end{aligned}
$$

in the sense that, if there exists one side of the equality, then there exists another side. Therefore we may prove (33) for $g^m_R$. We also have

$$\frac{1}{|B(0,1+2\epsilon)|^{1+\epsilon+a}}\int_{B(0,1+2\epsilon)}\sum_{m}|t|^{ad}g_R^m(t)dt$$
$$=\frac{1}{|B(0,1+2\epsilon)|^{1+\epsilon+a}}(1+2\epsilon)^{ad}\int_{B(0,1+2\epsilon)}\sum_{m}g_R^m(t)dt$$
$$-\frac{1}{|B(0,1+2\epsilon)|^{1+\epsilon+a}}\int_{B(0,1+2\epsilon)}\sum_{m}((1+2\epsilon)^{ad}-|t|^{ad})g_R^m(t)dt \quad (34)$$

On the one hand, by using the identity $|B(0,1+2\epsilon)|=|B(0,1)|(1+2\epsilon)^d$, we obtain that

$$\frac{1}{|B(0,1+2\epsilon)|^{1+\epsilon+a}}(1+2\epsilon)^{ad}\int_{B(0,1+2\epsilon)}\sum_{m}g_R^m(t)dt$$
$$=\frac{1}{|B(0,1)|^{1+\epsilon+a}(1+2\epsilon)^{(1+\epsilon)d}}\int_{B(0,1+2\epsilon)}\sum_{m}g_R^m(t)dt$$
$$=\frac{1}{|B(0,1)|^{a}}\frac{1}{|B(0,1+2\epsilon)|^{1+\epsilon}}\int_{B(0,1+2\epsilon)}\sum_{m}g_R^m(t)dt$$

and assumption (33) implies that the limit of the right hand side for $\epsilon\to\infty$ exists and is equal to

$$\frac{1}{|B(0,1)|^{a}}C_{1+\epsilon}^m(g_R^m) \quad (35)$$

On the other hand, in the second term of the difference in (34), by using $d$-dimensional spherical coordinate representation, we have

$$\int_{B(0,1+2\epsilon)}\sum_{m}((1+2\epsilon)^{ad}-|t|^{ad})g_R^m(t)dt$$
$$=\int_0^{1+2\epsilon}\sum_{m}((1+2\epsilon)^{ad}-s^{ad})s^{d-1}\left(\int_\Theta \hat{g}_R^m(s,\hat{\theta})d\hat{\theta}\right)ds$$

for some function $\hat{g}_R^m$ induced by $g_R^m$ via $d$-dimensional spherical coordinate representation, where $\Theta=[0,\pi]^{d-2}\times[0,2\pi]$. Then, by using the identity $|B(0,1+2\epsilon)|=|B(0,1)|(1+2\epsilon)^d$ again, we obtain that

$$\frac{1}{|B(0,1+2\epsilon)|^{1+\epsilon+a}}\int_{B(0,1+2\epsilon)}\sum_{m}((1+2\epsilon)^{ad}-|t|^{ad})g_R^m(t)dt$$
$$=\frac{1}{|B(0,1)|^{1+\epsilon+a}(1+2\epsilon)^{(1+\epsilon+a)d}}\int_0^{1+2\epsilon}\sum_{m}((1+2\epsilon)^{ad}-s^{ad})s^{d-1}\left(\int_\Theta \hat{g}_R^m(s,\hat{\theta})d\hat{\theta}\right)ds$$
$$=\frac{ad}{|B(0,1)|^{1+\epsilon+a}(1+2\epsilon)^{(1+\epsilon+a)d}}\int_0^{1+2\epsilon}\sum_{m}s^{d-1}\left(\int_0^{1+2\epsilon}\tau^{ad-1}d\tau\right)\left(\int_\Theta \hat{g}_R^m(s,\hat{\theta})d\hat{\theta}\right)ds$$
$$=\frac{ad}{|B(0,1)|^{1+\epsilon+a}(1+2\epsilon)^{(1+\epsilon+a)d}}\int_0^{1+2\epsilon}\left[\int_0^{\tau}s^{d-1}\left(\int_\Theta\sum_{m}\hat{g}_R^m(s,\hat{\theta})d\hat{\theta}\right)ds\right]\tau^{ad-1}d\tau \quad (36)$$

which becomes

$$= \frac{ad}{|B(0,1)|^a (1+2\epsilon)^{(1+\epsilon+a)d}} \int_0^{1+2\epsilon} \sum_m \tau^{(1+\epsilon)d+ad-1} \left( \frac{1}{|B(0,1)|^{1+\epsilon} \tau^{(1+\epsilon)d}} \int_{B(0,\tau)} g_R^m(t)dt \right) d\tau$$
$$= \frac{ad}{|B(0,1)|^a (1+2\epsilon)^{(1+\epsilon+a)d}} \int_0^{1+2\epsilon} \sum_m \tau^{(1+\epsilon)d+ad-1} \left( \frac{1}{|B(0,\tau)|^{1+\epsilon}} \int_{B(0,\tau)} g_R^m(t)dt \right) d\tau$$
$$= \frac{ad}{|B(0,1)|^a (1+2\epsilon)^{(1+\epsilon+a)d}} \int_0^{1+2\epsilon} \sum_m \tau^{(1+\epsilon)d+ad-1} \left( \frac{1}{|B(0,\tau)|^{1+\epsilon}} \int_{B(0,\tau)} g_R^m(t)dt \right.$$
$$\left. - C_{1+\epsilon}^m(g_R^m) \right) d\tau + \frac{ad}{|B(0,1)|^a (1+2\epsilon)^{(1+\epsilon+a)d}} \left( \int_0^{1+2\epsilon} \sum_m \tau^{(1+\epsilon)d+ad-1} d\tau \right) C_{1+\epsilon}^m(g_R^m) \quad (37)$$

The second term of (37) is equal to

$$\frac{ad}{|B(0,1)|^a (1+2\epsilon)^{(1+\epsilon+a)d}} \frac{1}{(1+\epsilon+a)d} (1+2\epsilon)^{(1+\epsilon)d+ad} \sum_m C_{1+\epsilon}^m(g_R^m)$$
$$= \frac{1}{|B(0,1)|^a} \frac{a}{1+\epsilon+a} \sum_m C_{1+\epsilon}^m(g_R^m) \quad (38)$$

and the first term of (37) is majorized, in modulus, by

$$\frac{ad}{|B(0,1)|^a (1+2\epsilon)^{(1+\epsilon+a)d}} \int_0^{1+2\epsilon} \tau^{(1+\epsilon+a)d-1} \left| \frac{1}{|B(0,\tau)|^{1+\epsilon}} \int_{B(0,\tau)} \sum_m \left( g_R^m(t)dt \right.\right.$$
$$\left.\left. - C_{1+\epsilon}^m(g_R^m) \right) \right| d\tau \quad (39)$$

Let $\varepsilon > 0$ be given. Then since by assumption

$$\lim_{\tau \to \infty} \left| \frac{1}{|B(0,\tau)|^{1+\epsilon}} \int_{B(0,\tau)} \sum_m \left( g_R^m(t)dt - C_{1+\epsilon}^m(g_R^m) \right) \right| = 0$$

there exists $t_\varepsilon$ such that, for any $\tau \geq t_\varepsilon$

$$\left| \frac{1}{|B(0,\tau)|^{1+\epsilon}} \int_{B(0,\tau)} \sum_m \left( g_R^m(t)dt - C_{1+\epsilon}^m(g_R^m) \right) \right| \leq \varepsilon$$

Moreover, since $g_R^m$ is locally integrable, the map

$$\tau \in \mathbb{R}_+ \longmapsto \frac{1}{|B(0,\tau)|^{1+\epsilon}} \int_{B(0,\tau)} \sum_m g_R^m(t)dt$$

is continuous, therefore there exists a constant $C_m > 0$ such that for all $\tau \leq t_\varepsilon$

$$\left| \frac{1}{|B(0,\tau)|^{1+\epsilon}} \int_{B(0,\tau)} \sum_m \left( g_R^m(t)dt - C_{1+\epsilon}^m(g_R^m) \right) \right| \leq C_m$$

Therefore, by splitting the integral in $\tau$ in (39) into the two pieces, (39) is majorized by

$$\frac{ad}{|B(0,1)|^a(1+2\epsilon)^{(1+\epsilon+a)d}}\left(C_m\int_0^{t_\varepsilon}\tau^{(1+\epsilon+a)d-1}d\tau+\varepsilon\int_{t_\varepsilon}^{1+2\epsilon}\tau^{(1+\epsilon+a)d-1}d\tau\right)$$
$$\le\frac{ad}{|B(0,1)|^a}\left(\frac{1}{(1+\epsilon+a)d}\left(\frac{t_\varepsilon}{1+2\epsilon}\right)^{(1+\epsilon+a)d}C_m\right.$$
$$\left.+\varepsilon\frac{1}{(1+2\epsilon)^{(1+\epsilon+a)d}}\int_0^{1+2\epsilon}\tau^{(1+\epsilon+a)d-1}d\tau\right)$$
$$\le\frac{a}{|B(0,1)|^a(1+\epsilon+a)}\left(\left(\frac{t_\varepsilon}{1+2\epsilon}\right)^{(1+\epsilon+a)d}C_m+\varepsilon\right)$$
and so
$$\lim_{\epsilon\to\infty}\frac{ad}{|B(0,1)|^a(1+2\epsilon)^{(1+\epsilon+a)d}}\int_0^{1+2\epsilon}\tau^{(1+\epsilon+a)d-1}\left|\frac{1}{|B(0,\tau)|^{1+\epsilon}}\int_{B(0,\tau)}\sum_m\left(g_R^m(t)dt\right.\right.$$
$$\left.\left.-C_{1+\epsilon}^m(g_R^m)\right)\right|d\tau\le\lim_{\epsilon\to\infty}\frac{a}{|B(0,1)|^a(1+\epsilon+a)}\left(\left(\frac{t_\varepsilon}{1+2\epsilon}\right)^{(1+\epsilon+a)d}C_m+\varepsilon\right)$$
$$\le\frac{a}{|B(0,1)|^a(1+\epsilon+a)}\varepsilon$$
from which, since $\varepsilon>0$ is arbitrary,
$$\lim_{\epsilon\to\infty}\frac{ad}{|B(0,1)|^a(1+2\epsilon)^{(1+\epsilon+a)d}}\int_0^{1+2\epsilon}\tau^{(1+\epsilon)d+ad-1}\left(\frac{1}{|B(0,\tau)|^{1+\epsilon}}\int_{B(0,\tau)}\sum_m\left(g_R^m(t)dt\right.\right.$$
$$\left.\left.-C_{1+\epsilon}^m(g_R^m)\right)\right)d\tau=0 \tag{40}$$
Hence by (36), (37), (38) and (40), we have
$$\lim_{\epsilon\to\infty}\frac{1}{|B(0,1+2\epsilon)|^{1+\epsilon+a}}\int_{B(0,1+2\epsilon)}\sum_m\left((1+2\epsilon)^{ad}-|t|^{ad}\right)g_R^m(t)dt$$
$$=\frac{1}{|B(0,1)|^a}\frac{a}{1+\epsilon+a}\sum_m C_{1+\epsilon}^m(g_R^m).$$
Therefore, by (34) and (35), we have
$$\lim_{\epsilon\to\infty}\frac{1}{|B(0,1+2\epsilon)|^{1+\epsilon+a}}\int_{B(0,1+2\epsilon)}\sum_m|t|^{ad}g_R^m(t)dt$$
$$=\frac{1}{|B(0,1)|^a}\sum_m C_{1+\epsilon}^m(g_R^m)-\frac{1}{|B(0,1)|^a}\frac{a}{1+\epsilon+a}\sum_m C_{1+\epsilon}^m(g_R^m)$$
$$=\frac{1}{|B(0,1)|^a}\frac{1+\epsilon}{1+\epsilon+a}\sum_m C_{1+\epsilon}^m(g_R^m),$$
which implies (33) as desired.

**Corollary (6.1.13)[250]:** If, for some $(1+\epsilon)\in\mathbb{R}_+^*$, some a $\in\mathbb{R}_+$and some $f^m\in L_{\text{loc}}^1(\mathbb{R}^d)$, the limit
$$C_{1+\epsilon+a}^m(f^m):=\lim_{\epsilon\to\infty}\frac{1}{|B(0,1+2\epsilon)|^{1+\epsilon+a}}\int_{B(0,1+2\epsilon)}\sum_m f^m(t)dt \tag{41}$$

exists in $\mathbb{C}$ and $|\cdot|^{-ad}f^m(\cdot) \in L^1_{\text{loc}}(\mathbb{R}^d)$, then it holds that

$$\lim_{\epsilon\to\infty} \frac{1}{|B(0,1+2\epsilon)|^{1+\epsilon}} \int_{B(0,1+2\epsilon)} \sum_m |t|^{-ad} f^m(t)dt$$
$$= \frac{1+\epsilon+a}{1+\epsilon} |B(0,1)|^a \sum_m C^m_{1+\epsilon+a}(f^m) \qquad (42)$$

in the sense that the limit exists and the equality holds.

**Proof.** The proof is similar Let $(1+\epsilon) \in \mathbb{R}^*_+, a \in \mathbb{R}_+$ and $f^m \in L^1_{\text{loc}}(\mathbb{R}^d)$ be such that (41) holds. Then we have

$$\frac{1}{|B(0,1+2\epsilon)|^{1+\epsilon}} \int_{B(0,1+2\epsilon)} \sum_m |t|^{-ad} f^m(t)dt$$
$$= \frac{1}{|B(0,1+2\epsilon)|^{1+\epsilon}} (1+2\epsilon)^{-ad} \int_{B(0,1+2\epsilon)} \sum_m f^m(t)dt$$
$$- \frac{1}{|B(0,1+2\epsilon)|^{1+\epsilon}} \int_{B(0,1+2\epsilon)} \sum_m \left((1+2\epsilon)^{-ad} - |t|^{-ad}\right) f^m(t)dt \qquad (43)$$

On the one hand, with the same notations

$$- \frac{1}{|B(0,1+2\epsilon)|^{1+\epsilon}} \int_{B(0,1+2\epsilon)} \sum_m \left((1+2\epsilon)^{-ad} - |t|^{-ad}\right) f^m(t)dt$$
$$= - \frac{1}{|B(0,1)|^{1+\epsilon}(1+2\epsilon)^{(1+\epsilon)d}} \int_0^{1+2\epsilon} \sum_m \left((1+2\epsilon)^{-ad} - s^{-ad}\right) s^{d-1} \left(\int_\Theta \hat{f}^m(s,\hat{\theta})d\hat{\theta}\right) ds$$
$$= \frac{ad}{|B(0,1)|^{1+\epsilon}(1+2\epsilon)^{(1+\epsilon)d}} \int_0^{1+2\epsilon} s^{d-1} \left(\int_s^{1+2\epsilon} \tau^{-ad-1} d\tau\right) \left(\int_\Theta \sum_m \hat{f}^m(s,\hat{\theta})d\hat{\theta}\right) ds$$
$$= \frac{ad}{|B(0,1)|^{1+\epsilon}(1+2\epsilon)^{(1+\epsilon)d}} \int_0^{1+2\epsilon} \left[\int_0^\tau s^{d-1} \left(\int_\Theta \sum_m \hat{f}^m(s,\hat{\theta})d\hat{\theta}\right) ds\right] \tau^{-ad-1} d\tau = \qquad (44)$$
$$= \frac{ad|B(0,1)|^a}{(1+2\epsilon)^{(1+\epsilon)d}} \int_0^{1+2\epsilon} \tau^{(1+\epsilon)d-1} \left(\frac{1}{|B(0,\tau)|^{1+\epsilon+a}} \int_{B(0,\tau)} \sum_m f^m(t)dt\right) d\tau$$
$$= \frac{ad|B(0,1)|^a}{(1+2\epsilon)^{(1+\epsilon)d}} \int_0^{1+2\epsilon} \tau^{(1+\epsilon)d-1} \left(\frac{1}{|B(0,\tau)|^{1+\epsilon+a}} \int_{B(0,\tau)} \sum_m \left(f^m(t)dt - C^m_{1+\epsilon+a}(f^m)\right)\right) d\tau$$
$$+ \frac{ad|B(0,1)|^a}{(1+2\epsilon)^{(1+\epsilon)d}} \left(\int_0^{1+2\epsilon} \tau^{(1+\epsilon)d-1} d\tau\right) \sum_m C^m_{1+\epsilon+a}(f^m). \qquad (45)$$

Therefore, by (43), (44) and (45) the following identity holds:

$$\frac{1}{|B(0,1+2\epsilon)|^{1+\epsilon}} \int_{B(0,1+2\epsilon)} \sum_m |t|^{-ad} f^m(t)dt$$
$$= k_1(1+2\epsilon) + k_2(1+2\epsilon) + k_3(1+2\epsilon),$$

Where

$$k_1(1+2\epsilon) = \frac{1}{|B(0,1+2\epsilon)|^{1+\epsilon}} (1+2\epsilon)^{-ad} \int_{B(0,1+2\epsilon)} \sum_m f^m(t)dt, \qquad (46)$$

$$k_2(1+2\epsilon) = \frac{ad|B(0,1)|^a}{(1+2\epsilon)^{(1+\epsilon)d}} \left(\int_0^{1+2\epsilon} \tau^{(1+\epsilon)d-1} d\tau\right) \sum_m C^m_{1+\epsilon+a}(f^m), \qquad (47)$$

$$k_3(1+2\epsilon) = \frac{ad|B(0,1)|^a}{(1+2\epsilon)^{(1+\epsilon)d}} \int_0^{1+2\epsilon} \tau^{(1+\epsilon)d-1} \left( \frac{1}{|B(0,\tau)|^{1+\epsilon+a}} \int_{B(0,\tau)} \sum_m \left(f^m(t)dt \right.\right.$$
$$\left.\left. - C^m_{1+\epsilon+a}(f^m)\right) \right) d\tau \tag{48}$$

Then by (46) and the assumption we obtain that

$$\lim_{\epsilon\to\infty} k_1(1+2\epsilon) = |B(0,1)|^a \lim_{\epsilon\to\infty} \frac{1}{|B(0,1+2\epsilon)|^{1+\epsilon+a}} \int_{B(0,1+2\epsilon)} \sum_m f^m(t)dt$$
$$= |B(0,1)|^a \sum_m C^m_{1+\epsilon+a}(f^m)$$

and from (47), we have

$$k_2(1+2\epsilon) = \frac{ad|B(0,1)|^a}{(1+2\epsilon)^{(1+\epsilon)d}} \left( \int_0^{1+2\epsilon} \tau^{(1+\epsilon)d-1} d\tau \right) \sum_m C^m_{1+\epsilon+a}(f^m)$$
$$= \frac{a}{1+\epsilon} |B(0,1)|^a \sum_m C^m_{1+\epsilon+a}(f^m)$$

Also, by the same arguments used for the proof of (40), we prove that $\lim_{t\to\infty} k_3(t) = 0$. Therefore, from the above arguments it follows that

$$\lim_{\epsilon\to\infty} \frac{1}{|B(0,1+2\epsilon)|^{1+\epsilon}} \int_{B(0,1+2\epsilon)} \sum_m |t|^{-ad} f^m(t)dt$$
$$= \lim_{\epsilon\to\infty} \left(k_1(1+2\epsilon) + k_2(1+2\epsilon) + k_3(1+2\epsilon)\right)$$
$$= |B(0,1)|^a \sum_m C^m_{1+\epsilon+a}(f^m) + \frac{a}{1+\epsilon} |B(0,1)|^a \sum_m C^m_{1+\epsilon+a}(f^m)$$
$$= \frac{1+\epsilon+a}{1+\epsilon} |B(0,1+2\epsilon)|^a \sum_m C^m_{1+\epsilon+a}(f^m)$$

which is the desired result.

**Corollary (6.1.14)[250]:** (See [314]). Let $a^m \in \mathrm{Dom}\left(\tilde{C}^m_{1+\epsilon}\right)$. Then for any $\beta \in \mathbb{R}$, it holds that

$$\lim_{\epsilon\to\infty} \frac{1}{(1+2\epsilon)^{1+\epsilon}} \int_1^{1+2\epsilon} \sum_m \left[ \left( \frac{[|t|]}{|t|} \right)^\beta - 1 \right] c_{a^m}(t)dt = 0$$

**Proof.** Let $\beta \in \mathbb{R}$ be given. Then we obtain that

$$\lim_{\epsilon\to\infty} \frac{1}{(1+2\epsilon)^{1+\epsilon}} \int_1^{1+2\epsilon} \sum_m \left[ \left( \frac{[|t|]}{|t|} \right)^\beta - 1 \right] c_{a^m}(t)dt$$
$$= \lim_{\epsilon\to\infty} \frac{1}{(1+2\epsilon)^{1+\epsilon}} \sum_{k=1}^{[1+2\epsilon]-1} \sum_m a^m_k \left( \int_k^{k+1} \left[ \left( \frac{k}{t} \right)^\beta - 1 \right] dt \right)$$
$$+ \lim_{\epsilon\to\infty} \sum_m \frac{a^m_{[1+2\epsilon]}}{(1+2\epsilon)^{1+\epsilon}} \int_{[1+2\epsilon]}^{1+2\epsilon} \left[ \left( \frac{[1+2\epsilon]}{t} \right)^\beta - 1 \right] dt \tag{49}$$

Then we can easily see that the second term of the right hand side of (49) is zero and the first term of the right hand side of (49) coincides with the following:

$$\lim_{N\to\infty}\frac{1}{N^{1+\epsilon}}\sum_{k=1}^{N}\sum_{m} a_k^m b_k,\ b_k=\int_k^{k+1}\left[\left(\frac{k}{t}\right)^{\beta}-1\right]dt$$

Therefore, we have

$$\lim_{\epsilon\to\infty}\frac{1}{(1+2\epsilon)^{1+\epsilon}}\int_1^{1+2\epsilon}\left[\left(\frac{[|t|]}{|t|}\right)^{\beta}-1\right]c_{a^m}(t)dt=\lim_{N\to\infty}\frac{1}{N^{1+\epsilon}}\sum_{k=1}^{N}\sum_{m} a_k^m b_k \qquad (50)$$

Then the Abel identity implies that

$$\lim_{N\to\infty}\frac{1}{N^{1+\epsilon}}\sum_{k=1}^{N}\sum_{m} a_k^m b_k$$

$$=\lim_{N\to\infty}\frac{1}{N^{1+\epsilon}}\left(\sum_{k\le N}\sum_{m} a_k^m\right)b_N$$

$$-\lim_{N\to\infty}\frac{1}{N^{1+\epsilon}}\sum_{k\le N-1}(b_{k+1}-b_k)\left(\sum_{1+2\epsilon\le k}\sum_{m} a_{1+2\epsilon}^m\right)$$

$$=-\lim_{N\to\infty}\frac{1}{N^{1+\epsilon}}\sum_{k\le N-1}(b_{k+1}-b_k)k^{1+\epsilon}\left(\frac{1}{k^{1+\epsilon}}\sum_{1+2\epsilon\le k}\sum_{m} a_{1+2\epsilon}^m\right)$$

On the other hand, by direct computation we can prove that $\lim_{k\to\infty}k(b_{k+1}-b_k)=0$. Let $\{d_n\}_{n=1}^{\infty}$ be a sequence given by

$$d_n=-n(b_{n+1}-b_n)\left(\frac{1}{n^{1+\epsilon}}\sum_{1+2\epsilon\le n}\sum_{m} a_{1+2\epsilon}^m\right),\ n\in\mathbb{N}$$

Then we have $\lim_{n\to\infty}d_n=0, \tilde{C}_{1+\epsilon}^m(a^m)=0$, i.e., for any $\epsilon>0$ there exists a number $N_0$ such that $|d_n|<\epsilon$ if $n\ge N_0$. Using this sequence (50) is represented as follows:

$$\lim_{\epsilon\to\infty}\frac{1}{(1+2\epsilon)^{1+\epsilon}}\int_1^{1+2\epsilon}\sum_{m}\left[\left(\frac{[|t|]}{|t|}\right)^{\beta}-1\right]c_{a^m}(t)dt=\lim_{N\to\infty}\frac{1}{N^{1+\epsilon}}\sum_{k=1}^{N}\sum_{m} a_k^m b_k$$

$$=-\lim_{N\to\infty}\frac{1}{N^{1+\epsilon}}\sum_{k\le N-1}(b_{k+1}-b_k)k^{1+\epsilon}\left(\frac{1}{k^{1+\epsilon}}\sum_{1+2\epsilon\le k}\sum_{m} a_{1+2\epsilon}^m\right)$$

$$=\lim_{N\to\infty}\frac{1}{N^{1+\epsilon}}\sum_{k\le N-1}k^{1+\epsilon-1}d_k$$

Therefore we obtain that

$$\left|\lim_{\epsilon\to\infty}\frac{1}{(1+2\epsilon)^{1+\epsilon}}\int_1^{1+2\epsilon}\sum_{m}\left[\left(\frac{[|t|]}{|t|}\right)^{\beta}-1\right]c_{a^m}(t)dt\right|\le\lim_{N\to\infty}\frac{1}{N^{1+\epsilon}}\sum_{k\le N-1}k^{\epsilon}|d_k|$$

$$=\lim_{N\to\infty}\frac{1}{N^{1+\epsilon}}\sum_{k=0}^{N_0-1}k^{\epsilon}|d_k|+\lim_{N\to\infty}\frac{1}{N^{1+\epsilon}}\sum_{k=N_0}^{N-1}k^{\epsilon}|d_k|\le\lim_{N\to\infty}\frac{1}{N^{1+\epsilon}}\sum_{k=N_0}^{N-1}k^{\epsilon}|d_k|\le\sum_{m} C_m\epsilon$$

with a constant $C_m$ independent on $N$. Since $\epsilon$ is arbitrary, this gives the proof of the assertion.

**Corollary (6.1.15)[250]:** (See [314]). Let $a^m\in\mathrm{Dom}\left(\tilde{C}_{1+\epsilon}^m\right)$. Then for any $\beta\in\mathbb{R}, Q^{m\beta}c_{a^m}\in\mathrm{Dom}(C_{1+\epsilon}^m)$ and

$$C_{1+\epsilon}^m\left(Q^{m\beta}c_{a^m}\right)=C_{1+\epsilon}^m(c_{a^m})=2^{-(1+\epsilon)}\tilde{C}_{1+\epsilon}^m(a^m).$$

**Proof.** For given $\beta \in \mathbb{R}$, by applying Corollary (6.1.14), we obtain that

$$C_{1+\epsilon}^{m}\left(Q^{m\beta}c_{a^m}\right) = \lim_{\epsilon\to\infty}\frac{1}{|B(0,1+2\epsilon)|^{1+\epsilon}}\int_{B(0,1+2\epsilon)}\sum_{m}\left(\frac{[|t|]}{|t|}\right)^{\beta}c_{a^m}(t)dt$$
$$= \lim_{\epsilon\to\infty}\frac{1}{|B(0,1+2\epsilon)|^{1+\epsilon}}\int_{B(0,1+2\epsilon)}\sum_{m}\left[\left(\frac{[|t|]}{|t|}\right)^{\beta}-1\right]c_{a^m}(t)dt+\sum_{m}C_{1+\epsilon}^{m}(c_{a^m})$$
$$=\sum_{m}C_{1+\epsilon}^{m}(c_{a^m})$$

which gives the proof.

Denote $q$ the identity function on $\mathbb{R}, q(t) = t$ for any $t \in \mathbb{R}$.

**Corollary (6.1.16)[250]:** (See [314]). Let $a^m \in \operatorname{Dom}\left(\tilde{C}_{1+\epsilon}^{m}\right)$ and $\beta \in \mathbb{R}$. Then $|q|^{\beta}Q^{m\beta}c_{a^m} \in \operatorname{Dom}(C_{1+\epsilon}^{m})$ and

$$C_{1+\epsilon}^{m}\left(|q|^{\beta}Q^{m\beta}c_{a^m}\right) = 2^{-(1+\epsilon)}\tilde{C}_{1+\epsilon}^{m}\left(q_d^{\beta}a^m\right)$$

where $q_d$ is the discretization of the multiplication operator $q$, i.e., $\left(q_d^{\beta}a^m\right)(n) = n^{\beta}a_n^m$ for $a^m \in \mathbb{C}^{\infty}$.

**Proof.** We obtain that

$$C_{1+\epsilon}^{m}\left(|q|^{\beta}Q^{m\beta}c_{a^m}\right) = \lim_{\epsilon\to\infty}\frac{1}{(2(1+2\epsilon))^{1+\epsilon}}\int_{-(1+2\epsilon)}^{1+2\epsilon}\sum_{m}|q|(t)^{\beta}Q^{m}(t)^{\beta}c_{a^m}(t)dt$$
$$= \lim_{\epsilon\to\infty}\frac{1}{(2(1+2\epsilon))^{1+\epsilon}}\int_{-(1+2\epsilon)}^{1+2\epsilon}\sum_{m}|t|^{\beta}\left(\frac{[|t|]}{|t|}\right)^{\beta}c_{a^m}(t)dt$$
$$= \lim_{n\to\infty}\frac{1}{(2n)^{1+\epsilon}}\int_{1}^{n}\sum_{m}[|t|]^{\beta}c_{a^m}(t)dt$$
$$= 2^{-(1+\epsilon)}\lim_{n\to\infty}\frac{1}{n^{1+\epsilon}}\sum_{k=1}^{n}\sum_{m}k^{\beta}a_k^m$$
$$= 2^{-(1+\epsilon)}\sum_{m}\tilde{C}_{1+\epsilon}^{m}\left(q_d^{\beta}a^m\right)$$

which implies the proof.

## Section (6.2): Law of Large Numbers for Weighted Inductive Means

The law of large numbers plays an important role in probability theory, which is concerned with the convergence of $(S_n - b_n)/n$, where $S_n = \sum_{i=1}^{n}X_i, \{X_i\}_{i=1}^{n}$ is a sequence of independent real random variables, and $\{b_n\}$ is a sequence of real numbers. If $(S_n - b_n)/n$ converges almost surely to zero, then the convergence theorem is called the strong law of large numbers, and if $(S_n - b_n)/n$ converges in probability to zero, then the convergence theorem is referred to the weak law of large numbers. The law of large numbers for real-valued random variables can be extended to a general metric space valued random variables. We study the law of large numbers for random variables valued in a Hadamard space. A Hadamard space is a complete metric space of which the metric satisfies the semiparallelogram law, and it gives geometric structures (like nonpositive curvature). In

[231], Sturm proved a law of large numbers for a (non-weighted) inductive mean in Hadamard spaces. Sturm also defined a expectation (or mean, or barycenter) of random variables taking values in a Hadamard space as a unique minimizer. However, there is another way to define expectations of random variables by using the law of large numbers in a Hadamard space

A weighted sum of a sequence $\{X_i\}$ of real-valued random variables is of the form

$$S_n := \sum_{i=1}^{n} a_{ni} X_i, \tag{51}$$

where the weighted sequence $\{a_{ni} \mid 1 \leq i \leq n\}$ is a triangular array. On the other hand, in a Hadamard space $N$, we can define a weighted sum with a positive weighted sequence by a weighted inductive mean. In the case of real-valued random variables with the usual Euclidean metric, the weighted inductive mean is exactly equal to the weighted sum given as in (51) (see [41]). The law of large numbers for weighted sums of random variables has been studied by [336], [337], [339], [340], [242], [243], [244], etc.

We study the law of large numbers for weighted inductive means with a positive weighted sequence $\{a_{ni} \mid 1 \leq i \leq n\}$ for Hadamard space valued random variables $\{X_i\}$, and then we prove the law of large numbers under certain independence and some conditions for a weighted sequence. In fact, Lim and Pálfia [225] proved that the deterministic weighted inductive mean on Hadamard spaces converges to the least squares mean.

We recall elementary notions of the probability measure on a Hadamard space and the variance inequality in a Hadamard space. We prove the law of large numbers for weighted inductive means of independent identically distributed random variables valued in a Hadamard space. We prove an asymptotic property for (non-weighed) inductive means in the $p$-Schatten class.

We recall basic facts of a Hadamard space (i.e., a complete non-positively curved metric space) and probability measures on Hadamard spaces. Let $(N, d)$ be a nonempty complete metric space. If for any $x, y \in N$, there exists a point $m \in N$ such that $d(x, m) = d(m, y) = \frac{1}{2} d(x, y)$, which is called a midpoint of $x$ and $y$, then $(N, d)$ is called a geodesic space. We call $(N, d)$ a Hadamard space if for any $x, y \in N$, there exists a point $m \in N$ such that

$$d(z, m)^2 \leq \frac{1}{2} d(z, x)^2 + \frac{1}{2} d(z, y)^2 - \frac{1}{4} d(x, y)^2 \text{ for any } z \in N. \tag{52}$$

Indeed, the point $m$ in (52) is the midpoint of $x$ and $y$ with the property $d(x, m) = d(m, y) = \frac{1}{2} d(x, y)$. We note that any Hadamard space becomes a geodesic space. Equation (52) is called the semiparallelogram law, since if $(N, d)$ is a Hilbert space then (52) becomes the parallelogram law by replacing the inequality with the equality (see [15]). It is satisfied by the length metric in any simply connected nonpositively curved Riemannian manifold [39]. Therefore, the metric inequality (52) in a Hadamard space yields a metric generalization of non-positive curvature.

For a positive real constant $D > 0$, a geodesic of speed $D$ in $N$ is a map $p: [0,1] \to N$ with the property that for any $t \in [0,1]$, there exists a positive constant $\alpha$ such that

$$d\big(p(t_1), p(t_2)\big) = D|t_1 - t_2| \text{ for all } t_1, t_2 \in [0,1] \text{ with } t - \alpha \leq t_1, t_2 \leq t + \alpha.$$

A map $p$ is said to be a geodesic if it is a geodesic of some speed $D$.

**Proposition (6.2.1)[335]:** ([231]). Let $(N,d)$ be a Hadamard space. For $x,y \in N$, there exists a unique geodesic $p:[0,1] \to N$ of speed $D = d(x,y)$ with $p(0) = x$ and $p(1) = y$. Furthermore, for any $z \in N$ and $t \in [0,1]$,

$$d(z,p(t)) \leq (1-t)d(z,x) + td(z,y) \quad (53)$$

and

$$d(z,p(t))^2 \leq (1-t)d(z,x)^2 + td(z,y)^2 - t(1-t)d(x,y)^2. \quad (54)$$

Throughout, $(N,d)$ and $(\Omega,\mathcal{F},P)$ denote a Hadamard space and a probability space, respectively, unless specified otherwise. The phase "almost surely," abbreviated a.s., is often used.

A function $X:\Omega \to N$ is an $N$-valued random variable (or simply, random variable) if $X$ is a Borel measurable function. For a Borel subset $B$ of $N$, the distribution of $X$ is defined by $P_X(B) := P(X^{-1}(B))$. For $1 \leq p < \infty$, let $L^p(\Omega,N)$ be the set of all random variables $X$ such that

$$\int_\Omega [d(z,X(\omega))]^p dP(\omega) = \int_N [d(z,x)]^p dP_X(x) < \infty \text{ for some } z \in N,$$

and $L^\infty(\Omega,N)$ be the set of all random variables $X$ such that

$$d(z,X(\omega)) < R \text{ a.s. for some } z \in N \text{ and } R > 0.$$

Let $\varphi$ be a real-valued function defined on $N$. If there exists a point $x \in N$ such that $\varphi(x) = \inf_{z\in N} \varphi(z)$, then $x$ is called a minimizer and denoted by $x := \operatorname{argmin}_{z\in N} \varphi(z)$. We define the expectation (or barycenter) of $X$ in $L^1(\Omega,N)$ [231] as follows: for each fixed $y \in N$,

$$\mathbf{E}X := \underset{z\in N}{\operatorname{argmin}} E[d(z,X)^2 - d(y,X)^2] = \underset{z\in N}{\operatorname{argmin}} \int_N [d(z,x)^2 - d(y,x)^2] dP_X(x).$$

For any $X \in L^1(\Omega,N)$, we define the variance of $X$ by

$$\mathbf{V}(X) := \inf_{z\in N} E[d(z,X)^2].$$

**Remark (6.2.2)[335]:** Let $X$ be a Hadamard space $N$-valued random variable. Then we have the following properties (see [231]).

(i) For a real-valued random variable $X$ with the Euclidean metric, the expectation is the same as the usual expectation: $\mathbf{E}X = E[X] = \int_{\mathbb{R}} x dP_X(x)$.

(ii) The expectation $\mathbf{E}X$ is uniquely determined and is independent of $y \in N$.

(iii) If we restrict to $X \in L^2(\Omega,N)$, then $\mathbf{E}X$ is the unique minimizer of $z \mapsto d(z,X)^2$, i.e.,

$$\mathbf{E}X = \underset{z\in N}{\operatorname{argmin}} E[d(z,X)^2] = \underset{z\in N}{\operatorname{argmin}} \int_N d(z,x)^2 dP_X(x). \quad (55)$$

Hence, $\mathbf{V}(X) = E[d(\mathbf{E}X,X)^2] < \infty$

The following proposition is the variance inequality in a Hadamard space, which is one of important properties.

**Proposition (6.2.3)[335]:** ([231]). For any random variable $X$ in $L^2(\Omega,N)$ and for all $z \in N$,

$$E[d(z,X)^2] \geq d(z,\mathbf{E}X)^2 + E[d(\mathbf{E}X,X)^2]. \quad (56)$$

Let $\{X_i\}$ be a sequence of independent identically distributed $N$-valued random variables and $\{a_{ni}\}$ be a positive weighted sequence with $\sum_{i=1}^n a_{ni} = 1$ for any $n \in \mathbb{N}$. We define a new weighted sequence $\{y_{(n-\ell)i}\}_{i=1}^{n-\ell}$ for $\ell = 1, \ldots, n-1$ as follows:

$$y_{(n-\ell)i} := \frac{a_{ni}}{\sum_{k=1}^{n-\ell} a_{nk}} \text{ for all } i = 1, \ldots, n-\ell. \quad (57)$$

We see that

$$y_{(n-\ell)(n-\ell)} = \frac{a_{n(n-\ell)}}{\sum_{k=1}^{n-\ell} a_{nk}} \text{ for } \ell = 1, \dots, n-1.$$

We also define a sequence $\{S_n\}$ of random variables, which is called the weighted inductive mean as follows [225]:

$$S_1 = X_1 \text{ and } S_n = Y_{n-1}\#_{a_{nn}}X_n \ (n \geq 2), \tag{58}$$

where $Y_1 = X_1$ and $Y_{n-k} := Y_{n-(k+1)}\#y_{(n-k)(n-k)}X_{n-k}$ for $k = 1, \dots, n-2$. Here, $A\#_tB$ denotes the $t$-weighted geometric mean which is the point $p(t)$ on the geodesic $p: [0,1] \to N$ connecting $p(0) = A$ and $p(1) = B$. For our purpose, we first prove the which is the point $p(t)$ ont following two lemmas.

**Lemma (6.2.4)[335]:** Let $\ell = 1, \dots, n-1$ and $\{y_{(n-\ell)i}\}_{i=1}^{n-\ell}$ be the weighted sequence given in (57). Then the following equality holds:

$$\sum_{i=1}^{n} a_{ni}^2 = (1-a_{nn})^2 \sum_{k=1}^{n-2} \left[\left(\prod_{i=1}^{k} \left(1 - y_{(n-i)(n-i)}\right)^2\right) y_{(n-(k+1))(n-(k+1))}^2\right] \tag{59}$$
$$+(1-a_{nn})^2 y_{(n-1)(n-1)}^2 + a_{nn}^2.$$

**Proof.** We see that $1 - a_{nn} = \sum_{i=1}^{n-1} a_{ni}$ and

$$1 - y_{(n-i)(n-i)} = \frac{\sum_{k=1}^{n-(i+1)} a_{nk}}{\sum_{k=1}^{n-i} a_{nk}} \text{ and } y_{(n-i)(n-i)} = \frac{a_{n(n-i)}}{\sum_{k=1}^{n-i} a_{nk}} \ (i = 1, \dots, n-1). \tag{60}$$

Then by a direct computation, we obtain

$$(1-a_{nn})^2 \prod_{i=1}^{k} \left(1 - y_{(n-i)(n-i)}\right)^2 y_{(n-(k+1))(n-(k+1))^2}$$

$$= \left(\sum_{i=1}^{n-1} a_{ni}\right)^2 \left(\frac{\sum_{i=1}^{n-2} a_{ni}}{\sum_{i=1}^{n-1} a_{ni}}\right)^2 \cdots \left(\frac{\sum_{i=1}^{n-(k+1)} a_{ni}}{\sum_{i=1}^{n-k} a_{ni}}\right)^2 \left(\frac{a_{n(n-(k+1))}}{\sum_{i=1}^{n-(k+1)a_{nk}}}\right)^2 = a_{n(n-(k+1))}^2.$$

Therefore, we have that

$$(1-a_{nn})^2 \sum_{k=1}^{n-2} \left[\left(\prod_{i=1}^{k} \left(1 - y_{(n-i)(n-i)}\right)^2\right) y_{(n-(k+1))(n-(k+1))}^2\right] + (1-a_{nn})^2 y_{(n-1)(n-1)}^2$$
$$+ a_{nn}^2 = \sum_{k=1}^{n-2} a_{n(n-(k+1))}^2 + a_{n(n-1)}^2 + a_{nn}^2 = \sum_{i=1}^{n} a_{ni}^2,$$

which completes the proof.

**Lemma (6.2.5)[335]:** Let $X, Y$ and $Z$ be independent identically distributed random variables in $L^2(\Omega, N)$. Then for any $0 \leq \lambda \leq 1$, it holds:

$$E[d(\mathbf{E}Z, X\#, Y)^2] \leq (1-\lambda)^2 E[d(\mathbf{E}Z, X)^2] + \lambda^2 E[d(\mathbf{E}Z, Y)^2]. \tag{61}$$

**Proof.** For any $0 \leq \lambda \leq 1$, we have that

$$\begin{aligned}
E\left[d\left(\mathbf{E}Z, X\#\lambda_\lambda^2\right]\right. \leq\ & (1-\lambda)E[d(\mathbf{E}Z, X)^2] + \lambda E[d(\mathbf{E}Z, Y)^2] - (1-\lambda)\lambda E[d(X, Y)^2] \\
\leq\ & (1-\lambda)E[d(\mathbf{E}Z, X)^2] + \lambda E[d(\mathbf{E}Z, Y)^2] \\
& -(1-\lambda)\lambda\left[E[d(X, \mathbf{E}Y)^2] + E[d(\mathbf{E}Y, Y)^2]\right] \\
=\ & (1-\lambda)^2 E[d(\mathbf{E}Z, X)^2] + \lambda^2 E[d(\mathbf{E}Z, Y)^2],
\end{aligned}$$

where the first inequality follows from the inequality (54) and the second inequality follows from the inequality (61) in Proposition (6.2.3). Therefore, the proof is completed.

The following theorem is the weak law of large numbers for weighted inductive means in a Hadamard space.

**Theorem (6.2.6)[335]:** Let $\{X_i\}$ be a sequence of independent identically distributed random variables in $L^2(\Omega, N)$ and $\{a_{ni}\}$ be a positive weighted sequence with $\sum_{i=1}^{n} a_{ni} = 1$. If $\{a_{ni}\}$ satisfies the condition

$$\lim_{n\to\infty} \sum_{i=1}^{n} a_{ni}^2 = 0, \tag{62}$$

then $S_n \to \mathbf{E}X_1$ in probability, where $S_n$ is the weighted inductive mean given in (58).

**Proof.** The proof can be obtained by some modification of the proof of Theorem (6.2.20) in [231], but we give the proof for completeness. For $\ell = 1, \dots, n-1$, let $\{y_{(n-\ell)i}\}_{i=1}^{n-\ell}$ be the weighted sequence given in (57) associated with $\{a_{ni}\}$. By Lemma (6.2.5), we obtain that

$$\begin{aligned} E[d(\mathbf{E}X_1, S_n)^2] &= E\left[d\left(\mathbf{E}X_1, Y_{n-1}\#_{a_{nn}}X_n\right)^2\right] \\ &\le (1-a_{nn})^2 E[d(\mathbf{E}X_1, Y_{n-1})^2] + a_{nn}{}^2 E[d(\mathbf{E}X_1, X_n)^2], \end{aligned} \tag{63}$$

and that for $i = 1, \dots, n-2$

$$\begin{aligned} &E\left[d\left(\mathbf{E}X_1, Y_{(n-i)}\right)^2\right] \\ &\qquad \le \left(1 - y_{(n-i)(n-i)}\right)^2 E\left[d\left(\mathbf{E}X_1, Y_{n-(i+1)}\right)^2\right] \\ &\qquad + y_{(n-i)(n-i)}^2 E[d(\mathbf{E}X_1, X_{n-i})^2]. \end{aligned} \tag{64}$$

By the inequalities (63), (64) and Lemma (6.2.4), we have

$$\begin{aligned} E[d(\mathbf{E}X_1, S_n)^2] \le\ & (1-a_{nn})^2 \prod_{i=1}^{n-2} \left(1 - y_{(n-i)(n-i)}\right)^2 E[d(\mathbf{E}X_1, Y_1)^2] \\ & +(1-a_{nn})^2 \prod_{i=1}^{n-3} \left(1 - y_{(n-i)(n-i)}\right)^2 y_{22}^2 E[d(\mathbf{E}X_1, X_2)^2] \\ & + \cdots + (1-a_{nn})^2 \left(1 - y_{(n-1)(n-1)}\right)^2 y_{(n-2)(n-2)}{}^2 E\left[d\left(\mathbf{E}X_1, X_{(n-2)}\right)^2\right] \\ & +(1-a_{nn})^2 y_{(n-1)(n-1)}{}^2 E[d(\mathbf{E}X_1, X_{n-1})^2] + a_{nn}{}^2 E[d(\mathbf{E}X_1, X_n)^2] \\ =\ & \left\{(1-a_{nn})^2 \sum_{k=1}^{n-2} \left[\left(\prod_{i=1}^{k} \left(1 - y_{(n-i)(n-i)}\right)^2\right) y_{(n-(k+1))(n-(k+1))}{}^2\right]\right. \\ & \left. +(1-a_{nn})^2 y_{(n-1)(n-1)}{}^2 + a_{nn}{}^2\right\} \mathbf{V}(X_1) \\ =\ & \left(\sum_{i=1}^{n} a_{ni}^2\right) \mathbf{V}(X_1), \end{aligned} \tag{65}$$

where the last equality follows from Equation (59) in Lemma (6.2.4). Therefore, by the equalities (62) and (65) we have that

$$E[d(\mathbf{E}X_1, S_n)^2] \le \left(\sum_{i=1}^{n} a_{ni}^2\right) \mathbf{V}(X_1) \longrightarrow 0 \text{ as } n \to \infty.$$

Since the $L^2$-convergence implies the convergence in probability, we are done.

**Example (6.2.7)[335]:** We now consider an example of a weighted sequence $\{a_{ni}\}$ satisfying the condition in Theorem (6.2.6). Let $0 < r < 2$ be given and let $\{b_i\}$ be a sequence of positive real numbers such that $\sum_{i=1}^{n} b_i = n^{1/r}$ and $\sum_{i=1}^{n} b_i^2 =$

$o\left(n^{2/r}\right). Put a_{ni} = b_i/n^{1/r}$ for all $i = 1, ...,$ Since $\sum_{i=1}^{n} a_{ni} = 1$ and $\sum_{i=1}^{n} b_i^2 = o\left(n^{2/r}\right)$, we have

$$\sum_{i=1}^{n} a_{ni}^2 = \frac{1}{n^{2/r}} \sum_{i=1}^{n} b_i^2 \longrightarrow 0$$

as $n \to \infty$. Such a weighted sequence $\{a_{ni}\}$ satisfies the conditions in Theorem (6.2.6).

**Corollary (6.2.8)[335]:** Let $\{X_i\}$ be a sequence of independent identically distributed random variables in $L^2(\Omega, N)$. If $\{a_{ni}\}$ is a positive weighted sequence with $\sum_{i=1}^{n} a_{ni} = 1$ such that for $1 \le r < 2$,

$$\max_{1\le i\le n} a_{ni} = O\left(1/n^{1/r}\right), \tag{66}$$

then $S_n \to \mathbf{E}X_1$ in probability, where $S_n$ is given as in (58).

**Proof.** It is easy to see that

$$\lim_{n\to\infty} \sum_{i=1}^{n} a_{ni}^2 \le C^2 \lim_{n\to\infty} n^{1-\frac{2}{r}} = 0$$

for some constant $C > 0$, since $1 \le r < 2$. The conclusion follows from Theorem (6.2.6).

**Remark (6.2.9)[335]:** Let $\{X_i\}$ be a sequence of independent identically distributed random variables in $L^2(\Omega, N)$ and $\{a_{ni}\}$ be a positive weighted sequence with $\sum_{i=1}^{n} a_{ni} = 1$. By Lemma (6.2.5) and using the proof in Lemma (6.2.4), we obtain

$$\begin{aligned}
E[d(Y_{n-i}, \mathbf{E}X_1)^2] = \; & E\left[d\left(Y_{n-(i+1)} \# y_{(n-i)(n-i)} X_{n-i}, \mathbf{E}X_1\right)^2\right] \\
\le \; & \left(1 - y_{(n-i)(n-i)}\right)^2 E\left[d\left(Y_{n-(i+1)}, \mathbf{E}X_1\right)^2\right] + y_{(n-i)(n-i)}^2 E[d(X_{n-i}, \mathbf{E}X_1)^2] \\
\le \; & \left(1 - y_{(n-i)(n-i)}\right)^2 \left(1 - y_{(n-(i+1))(n-(i+1))}\right)^2 E\left[d\left(Y_{n-(i+2)}, \mathbf{E}X_1\right)^2\right] \\
& + \left(1 - y_{(n-i)(n-i)}\right)^2 y_{(n-(i+1))(n-(i+1))}^2 E\left[d\left(X_{n-(i+1)}, \mathbf{E}X_1\right)^2\right] \\
& + y_{(n-i)(n-i)}^2 E[d(X_{n-i}, \mathbf{E}X_1)^2] \\
& \vdots \\
& \le \prod_{j=0}^{n-i-2} \left(1 - y_{(n-(i+j))(n-(i+j))}\right)^2 E[d(Y_1, \mathbf{E}X_1)^2] \\
& + \prod_{j=0}^{n-i-3} \left(1 - y_{(n-(i+j))(n-(i+j))}\right)^2 y_{22}{}^2 E[d(X_2, \mathbf{E}X_1)^2] \\
& + \cdots + \left(1 - y_{(n-i)(n-i)}\right)^2 y_{(n-(i+1))(n-(i+1))}^2 E\left[d\left(X_{n-(i+1)}, \mathbf{E}X_1\right)^2\right] \\
& + y_{(n-i)(n-i)}^2 E[d(X_{n-i}, \mathbf{E}X_1)^2] \\
= \; & \frac{\sum_{k=1}^{n-i} a_{nk}^2 E[d(X_k, \mathbf{E}X_1)^2]}{\left(\sum_{k=1}^{n-i} a_{nk}\right)^2},
\end{aligned} \tag{67}$$

where $\{Y_{n-i}\}_{i=1}^{n-1}$ is given as in (58). The inequality (67) is useful for the proof of the strong law of large numbers (see Theorem (6.2.11)).

The following theorem is the strong law of large numbers for weighted inductive means in a Hadamard space.

To prove the strong law of large numbers, we need the following lemma.

**Lemma (6.2.10)[335]:** Under the assumptions of Theorem (6.2.11), for any $z \in N$, there exists $R > 0$ such that for all $n \in \mathbb{N}$ and $i = 1,2, \dots, n-1$,

$$d(S_n, z) \le \left( \sum_{k=1}^{n} a_{nk} d(X_k, z) \right) < R \text{ and } d(Y_{n-i}, z) \le \left( \frac{\sum_{k=1}^{n-i} a_{nk} d(X_k, z)}{\sum_{k=1}^{n-i} a_{nk}} \right) < R, (68)$$

where $S_n$ and $\{Y_{n-i}\}_{i=1}^{n-1}$ are given as in (58).

**Proof.** By the convexity of the metric d and Equation (60), we obtain

$$\begin{aligned} d(S_n, z) = d\left(Y_{n-1} \#_{a_{nn}} X_n, z\right) &\le (1 - a_{nn}) d(Y_{n-1}, z) + a_{nn} d(X_n, z) \\ &\le (1 - a_{nn})\left(1 - y_{(n-1)(n-1)}\right) d(Y_{n-2}, z) + (1 - a_{nn}) y_{(n-1)(n-1)} d(X_{n-1}, z) \\ &+ a_{nn} d(X_n, z) \\ &\vdots \\ \le (1 - a_{nn}) &\left[ \prod_{i=1}^{n-2} \left(1 - y_{(n-i)(n-i)}\right) d(Y_1, z) + \prod_{i=1}^{n-3} \left(1 - y_{(n-i)(n-i)}\right) y_{22} d(X_2, z) + \cdots \right. \\ &\left. + \left(1 - y_{(n-1)(n-1)}\right) y_{(n-2)(n-2)} d(X_{n-2}, z) + y_{(n-1)(n-1)} d(X_{n-1}, z) \right] \\ &+ a_{nn} d(X_n, z) \\ &= \sum_{k=1}^{n} a_{nk} d(X_k, z). \qquad (69) \end{aligned}$$

Since $X_i \in L^\infty(\Omega, N) (i = 1, \dots, n)$, there exist $z \in N$ and $R > 0$ such that $d(X_i(\omega), z) < R$ a.s. Then, by the inequalities (69), it holds that $d(S_n, z) < R$ a.s. for all $n \in \mathbb{N}$. By similar arguments, we have the second inequality given as in (68).

**Theorem (6.2.11)[335]:** Let $\{X_i\}$ be a sequence of independent identically distributed random variables in $L^\infty(\Omega, N)$. If $\{a_{ni}\}$ is a positive weighted sequence satisfying the following conditions:

(i) There exists a constant $C \ge 1$ such that $\max_{1 \le i \le n} a_{ni} \le C \min_{1 \le i \le n} a_{ni}$,

(ii) $\sum_{i=1}^{n} a_{ni} = 1$ for all $n \in \mathbb{N}$,

then

$$S_n \longrightarrow \mathbf{E} X_1 \text{ a.s. as } n \to \infty, \qquad (70)$$

where $S_n$ is given as in (58).

**Proof.** For the proof, it suffices to show that $d(S_n, \mathbf{E}X_1) \to 0$ a.s. as $n \to \infty$. Note that for all $k, n \in \mathbb{N}$

$$d(S_n, \mathbf{E}X_1) \le d(S_n, Y_{k^2}) + d(Y_{k^2}, \mathbf{E}X_1).$$

Now, we assume that $k, n$ are positive integers such that $k^2 < n \le (k+1)^2$. We first prove that $d(S_n, Y_{k^2}) \to 0$ a.s. as $k \to \infty$. Since $X_i \in L^\infty(\Omega, N)$, there exist $z \in N$ and $R > 0$ such that $d(X_i(\omega), z) < R$ a.s. By (68) we have that d $(S_n, z) < R$ and $d(Y_{n-i}, z) < R$ a.s. for all $n \in \mathbb{N}$ and $i = 1, \dots, n-1$. By the convexity of $x \longmapsto d(x, z)$, we have that

$$\begin{aligned} d(S_n, Y_{n-1}) &= d\left(Y_{n-1} \#_{a_{nn}} X_n, Y_{n-1}\right) \\ &\le (1 - a_{nn}) d(Y_{n-1}, Y_{n-1}) + a_{nn} d(X_n, Y_{n-1}) = a_{nn} d(X_n, Y_{n-1}) \\ &\le 2 a_{nn} R \text{ a.s.} \qquad (71) \end{aligned}$$

By similar arguments, we also obtain that

$$
\begin{aligned}
d\big(Y_{n-i}, Y_{n-(i+1)}\big) &= d\left(Y_{n-(i+1)} \#_{y_{(n-i)(n-i)}} X_{n-i}, Y_{n-(i+1)}\right) \\
&\le y_{(n-i)(n-i)} d\big(X_{n-i}, Y_{n-(i+1)}\big) \\
&\le 2 y_{(n-i)(n-i)} R \ \text{ a.s.} \qquad (72)
\end{aligned}
$$

for $i = 1, \dots, n-2$. By inequalities (71), (72) and the condition (i), we have that for all $k, n \in \mathbb{N}$ with $k^2 < n \le (k+1)^2$

$$
\begin{aligned}
d(S_n, Y_{k^2}) &\le 2\big(y_{(k^2+1)(k^2+1)} + y_{(k^2+2)(k^2+2)} + \cdots + a_{nn}\big) R \\
&\le \frac{\sum_{\ell=k^2+1}^{n} \max\limits_{1\le i\le n} a_{ni}}{\sum_{\ell=1}^{k^2+1} \min\limits_{1\le i\le n} a_{ni}} 2R \le \frac{(n-k^2) C \min\limits_{1\le i\le n} a_{ni}}{(k^2+1) \min\limits_{1\le i\le n} a_{ni}} 2R \le \frac{2k+1}{k^2+1} 2RC \longrightarrow 0,
\end{aligned}
$$

so that $d(S_n, Y_{k^2}) \to 0$ a.s. as $k \to \infty$.

Our second claim is that $Y_{k^2} \to \mathbf{E}X_1$ a.s. as $k \to \infty$. Indeed, by Chebyshev's inequality and the condition (i), we have that for any $\epsilon > 0$,

$$
\begin{aligned}
\sum_{k=1}^{\infty} P(d(Y_{k^2}, \mathbf{E}X_1) > \epsilon) &\le \frac{1}{\epsilon^2} \sum_{k=1}^{\infty} E(d(Y_{k^2}, \mathbf{E}X_1)^2) \\
&\le \frac{1}{\epsilon^2} \sum_{k=1}^{\infty} \frac{k^2 C^2 \left(\min\limits_{1\le i\le n} a_{ni}\right)^2}{k^4 \left(\min\limits_{1\le i\le n} a_{ni}\right)^2} \mathbf{V}(X_1) \\
&= \frac{C^2}{\epsilon^2} \sum_{k=1}^{\infty} \frac{1}{k^2} \mathbf{V}(X_1) < \infty.
\end{aligned}
$$

Therefore, by Borel-Cantelli Lemma, we have that $Y_{k^2} \to \mathbf{E}X_1$ a.s. as $k \to \infty$, which completes the proof.

The following examplé gives the strong law of lārge numbërs for wéighed inductive meañs of real-valued random variablēs.

**Example (6.2.12)[335]:** Let $\left(\mathbb{R}_+, \tilde{d}\right)$ be a complete metric space with the metric given by

$$\tilde{d}(x, y) = |\log x - \log y|.$$

It is clear that $\left(\mathbb{R}_+, \tilde{d}\right)$ is a Hadamard space, where the midpoint of $x$ and $y$ is $\sqrt{xy}$ (see [3]). If $\{X_i\}$ is a sequence of independent identically distributed and positive real-valued random variables and if $\{a_{ni}\}$ is a positive weighted sequence, then the weighted inductive mean $S_n$ is the same as in (58), and we have that

$$\mathbf{E}X = \underset{z\in\mathbb{R}}{\operatorname{argmin}} E[d(z, X)^2] = e^{E[\log X]}.$$

Therefore, if a positive weighted sequence $\{a_{ni}\}$ satisfies the conditions $(i)$ and ( ii ) in Theorem (6.2.11), then we have that $S_n \to \mathbf{E}X_1$ a.s. In particular, if we choose $a_{ni} = 1/n$ for $i = 1, \dots, n$, then we obtain that

$$(X_1 X_2 \cdots X_n)^{1/n} \longrightarrow \mathbf{E}X_1 \ \text{ a.s.}$$

On the other hands, if $(\mathbb{R}, d)$ is the Euclidean space with usual metric given by $d(x, y) = |x - y|$, $(\mathbb{R}, d)$ is a Hadamard space ( see [41], [231]). For a sequence $\{X_i\}$ of real valued independent identically distributed random variables and for a positive weighted sequence $\{a_{ni}\}$, the weighted inductive mean is given by $S_n = \sum_{i=1}^{n} a_{ni} X_i$( see [41]). Therefore, if the weighted sequence $\{a_{ni}\}$ satisfies the conditions (i) and (ii) in Theorem (6.2.11), then we see that

$$S_n \longrightarrow \mathbf{E}X_1 \ \text{ a.s. } \ \text{as } n \to \infty,$$

where $\mathbf{E}X = \int_{\mathbb{R}} x dP_X(x)$, which gives the strong law of large numbers for weighted sums of real-valued random variables. In particular, if we choose $a_{ni} = 1/n$ for $i = 1, \dots, n$, then we have

$$\frac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow \mathbf{E}X_1 \ \text{ a.s. } \ \text{as } n \to \infty.$$

The following example is the law of large numbers for weighed inductive means which are converging to the least squares mean.

**Example (6.2.13)[335]:** Let $\{X_n\}$ be a sequence of independent identically distributed random variables having the distribution $\sum_{i=1}^{n} \omega_i \delta_{x_i}$ for $x_1, \dots, x_n \in N$. If a positive weighted sequence $\{a_{ni}\}$ satisfies the conditions (i) and (ii) in Theorem (6.2.11), then we have

$$S_n \longrightarrow \mathbf{E}X_1 = \Lambda(\omega_1, \dots, \omega_n; x_1, \dots, x_n) \ \text{ a.s. } \ \text{as } n \to \infty, \qquad (73)$$

where $S_n$ is the weighted inductive mean given in (58) and $\Lambda(\omega_1, \dots, \omega_n; x_1, \dots, x_n)$ is the weighted least squares mean given by

$$\Lambda(\omega_1, \dots, \omega_n; x_1, \dots, x_n) = \operatorname*{argmin}_{z \in N} \sum_{i=1}^{n} \omega_i d(z, x_i)^2.$$

Indeed, the right hand side of (73) is obtained by taking $P_X = \sum_{i=1}^{n} \omega_i \delta_{x_i}$ in (55), i.e.,

$$EX = \operatorname*{argmin}_{z \in N} \int_N d(z, x)^2 dP_X(x) = \operatorname*{argmin}_{z \in N} \sum_{i=1}^{n} \omega_i d(z, x_i)^2 = \Lambda(\omega_1, \dots, \omega_n; x_1, \dots, x_n).$$

Let $\mathcal{B}(H)$ be the Banach space of all bounded linear operators on a separable Hilbert space $H$ equipped with the operator norm, $\mathcal{B}(H)_{sa}$ be the set of all self-adjoint elements in $\mathcal{B}(H)$ and $\mathcal{P}$ be the set of all positive invertible elements in $\mathcal{B}(H)_{sa}$. For $1 \le p < \infty$, the $p$-Schatten class $S_p(H)$ of $\mathcal{B}(H)$ is defined by

$$\mathcal{S}_p(H) = \{x \in \mathcal{B}(H) \mid x \text{ is a compact operator, } \| x \|_p < \infty\}.$$

The class $S_p(H)$ is a Banach space with respect to the norm

$$\| x \|_p := \left( \sum_j [s_j(x)]^p \right)^{1/p} = (\operatorname{Tr} |x|^p)^{1/p},$$

where $s_j(x)$ is the sequence of singular values of $x$ with decreasing order, $|x| = (x^* x)^{1/2}$ and Tr is the usual trace on $\mathcal{B}(H)$ (see [241]).

On the $p$-Schatten class $S_p(H)$, we define the norm $\|\cdot\|_{p,b}$ associated with $b \in \mathcal{P}$ by

$$\| a \|_{p,b} = \|b^{-1/2} a b^{-1/2}\|_p \ \text{ for } a \in S_p(H).$$

We denote by $S_{p,b}$ the $p$-Schatten class $S_p(H)$ equipped with the norm $\|\cdot\|_{p,b}$, that is, $S_{p,b} = (S_p(H), \|\cdot\|_{p,b})$. For $1 < p < \infty$, let $\Delta_p := \{I + a \in \mathcal{P} \mid a \in S_{p,b}\}$ be the positive cone of the operator algebra that is obtained by adjoining the unit to the ideal of compact $p$-Schatten operators. We define the geodesic distance between two points $x, y \in \Delta_p$ as follows [338]:

$$d_p(x, y) = \inf\{L(\gamma) \mid \gamma \text{ is a } \Delta_p - \text{valued smooth curve on } [0,1], \gamma(0) = x, \gamma(1) = y\}$$
$$= \|\log(x^{-1/2} y x^{-1/2})\|_p,$$

where $L(\gamma) = \int_0^1 \|\gamma'(t)\|_{p,\gamma(t)} dt = \int_0^1 \|\gamma(t)^{-1/2} \gamma'(t) \gamma(t)^{-1/2}\|_p dt$. Note that $( \Delta_p, d_p )$ is a complete metric space (see [338]).

Now, we recall the useful facts of a geodesic and properties of metric $d_p$.

**Theorem (6.2.14)[335]:** ([338]). Let $x, y \in \Delta_p$ and $\gamma_{x,y}(t) := x^{1/2}\left(x^{-1/2} y x^{-1/2}\right)^t x^{1/2}$. Then $\gamma_{x,y}$ is a unique short piecewise smooth curve joining $x$ to $y$ in $\Delta_p$. Furthermore, $t \mapsto d_p\left(\gamma_{x_1,y_1}(t), \gamma_{x_2,y_2}(t)\right)$ is a convex function for any $x_1, x_2, y_1, y_2 \in \Delta_p$.

Note that for any $x, y \in \Delta_p$, since $\gamma_{x,y}\left(\frac{1}{2}\right)$ is a mid point of $x$ and $y$, $\left(\Delta_p, d_p\right)$ is a geodesic space. The following theorem is a weak semiparallelogram law in $\Delta_p$.

**Theorem (6.2.15)[335]:** ([338]). Let $z \in \Delta_p$ for $1 < p \le 2$ and $\gamma: [0,1] \to \Delta_p$ be a geodesic. Then for $0 < \alpha_p := p - 1 \le 1$, we have

$$d_p(z, \gamma(t))^2 \le (1-t) d_p(z, \gamma(0))^2 + t d_p(z, \gamma(1))^2 - t(1-t)\alpha_p d_p(\gamma(0), \gamma(1))^2. \quad (74)$$

In particular, for any $x, y \in \Delta_p$, we have

$$d_p\left(z, \gamma_{x,y}(t)\right)^2 \le (1-t) d_p(z, x)^2 + t d_p(z, y)^2 - t(1-t)\alpha_p d_p(x, y)^2. \quad (75)$$

If $p = 2$, then $\alpha_p = 1$ and so the inequality (74) is the semiparallelogram law. Thus, $(\Delta_2, d_2)$ is a Hadamard space. In the remaining, we denote by $\left(\Delta_p, d_p\right)$ the complete metric space given in the beginning and let $\alpha_p = p - 1$ with $1 < p \le 2$.

**Definition (6.2.16)[335]:** ([231]). Let $(X, d)$ be a complete geodesic space. A function $\varphi: X \to \mathbb{R}$ is uniformly convex if there exists a strictly increasing function $\omega: \mathbb{R}_+ \to \mathbb{R}_+$ such that for any geodesic $\gamma: [0,1] \to X$,

$$\varphi\left(\gamma\left(\frac{1}{2}\right)\right) \le \frac{1}{2}\left[\varphi\left(\gamma(0)\right) + \varphi\left(\gamma(1)\right)\right] - \omega\left(d\left(\gamma(0), \gamma(1)\right)\right).$$

**Proposition (6.2.17)[335]:** ([231]). Let $(X, d)$ be a complete geodesic space. If $\varphi: X \to \mathbb{R}$ is a uniformly convex and lower semicontinuous function on $X$, then there exists a unique minimizer $x \in X$, i.e., a unique point $x \in X$ with $\varphi(x) = \inf_{z \in X} \varphi(z)$.

We now consider $\Delta_p$-valued random variables, their distributions, the independence and the space $L^q\left(\Omega, \Delta_p\right)$ for $1 < p < \infty$ and $1 \le q < \infty$, etc., by using arguments .

**Theorem (6.2.18)[335]:** For a fixed element $y \in \Delta_p$ and any $X \in L^1\left(\Omega, \Delta_p\right)$, there exists a unique element $x \in \Delta_p$ which minimizes the real-valued function on $\Delta_p$ given by $z \mapsto E\left[d_p(z, X)^2 - d_p(y, X)^2\right]$.

**Proof.** Let $F_y(z) = E\left[d_p(z, X)^2 - d_p(y, X)^2\right]$. Then

$$\begin{aligned} \left|F_y(z)\right| &= \left|E\left[\left(d_p(z, X) - d_p(y, X)\right)\left(d_p(z, X) + d_p(y, X)\right)\right]\right| \\ &\le d_p(z, y)\left(E\left[d_p(z, X)\right] + E\left[d_p(y, X)\right]\right), \end{aligned}$$

so that $\left|F_y(z)\right| < \infty$. For any two elements $z_0, z_1 \in \Delta_p$, let $\gamma(t)$ be a geodesic joining $z_0$ and $z_1$. Then by using the inequality (74) for $d_p$, we obtain that

$$\begin{aligned} F_y(\gamma(t)) &= E\left[d_p(\gamma(t), X)^2 - d_p(y, X)^2\right] \\ &\le E\left[(1-t) d_p(z_0, X)^2 + t d_p(z_1, X)^2 - t(1-t)\alpha_p d_p(z_0, z_1)^2 \right. \\ &\quad \left. - d_p(y, X)^2\right] \\ &= (1-t) E\left[d_p(z_0, X)^2 - d_p(y, X)^2\right] + t E\left[d_p(z_1, X)^2 - d_p(y, X)^2\right] - t(1 \\ &\quad - t)\alpha_p d_p(z_0, z_1)^2 \\ &= (1-t) F_y(z_0) + t F_y(z_1) - t(1-t)\alpha_p d_p(z_0, z_1)^2, \end{aligned} \quad (76)$$

which implies that $F_y$ is a uniformly convex function, i.e.,

$$F_y\left(\gamma\left(\frac{1}{2}\right)\right) \le \frac{1}{2}\left(F_y(z_0) + F_y(z_1)\right) - \omega\left(d_p(z_0, z_1)\right),$$

where $\omega(t) = \alpha_p t^2/4$ with $t \in \mathbb{R}_+$.Moreover, the continuity of $z \mapsto F_y(z)$ immediately follows from the inequality

$$\begin{aligned}\left|F_y(z) - F_y(z')\right| &= \left|E\left[d_p(z,X)^2 - d_p(y,X)^2\right] - E\left[d_p(z',X)^2 - d_p(y,X)^2\right]\right| \\ &\le E\left[\left|d_p(z,X)^2 - d_p(z',X)^2\right|\right].\end{aligned}$$

Therefore, $F_y$ is uniformly convex and continuous, so that by applying Proposition (6.2.17), we get the existence and uniqueness of minimizer.

We see that the case $p = 2$ in Theorem (6.2.18) can be obtained from the results in [231] since $(\Delta_2, d_2)$ is a Hadamard space. For $X \in L^1(\Omega, \Delta_p)$, the minimizer of the function

$$z \in \Delta_p \longmapsto E\left[d_p(z,X)^2 - d_p(y,X)^2\right] \in \mathbb{R}$$

is called the expectation of $X$ and denoted by $\mathbf{E}X$. The following theorem gives a variance inequality in $(\Delta_p, d_p)$ with $1 < p \le 2$.

**Theorem (6.2.19)[335]:** Let $1 < p \le 2$. For any $X \in L^1(\Omega, \Delta_p)$ and $z \in \Delta_p$, we have

$$E\left[d_p(z,X)^2 - d_p(\mathbf{E}X,X)^2\right] \ge \alpha_p d_p(z,\mathbf{E}X)^2. \qquad (77)$$

**Proof.** Given a random variable $X \in L^1(\Omega, \Delta_p)$ and $z \in \Delta_p$, put $z_1 = z$, and $z_0 = \mathbf{E}X$. We denote by $\gamma(t)$ the geodesic joining $\mathbf{E}X$ and $z$ in (76). Since $\mathbf{E}X$ is the minimizer, we have

$$\begin{aligned}0 \le F_{\mathbf{E}X}(\gamma(t)) &\le (1-t)F_{\mathbf{E}X}(\mathbf{E}X) + tF_{\mathbf{E}X}(z) - t(1-t)\alpha_p d_p(z,\mathbf{E}X)^2 \\ &= tF_{\mathbf{E}X}(z) - t(1-t)\alpha_p d_p(z,\mathbf{E}X)^2.\end{aligned}$$

Therefore, we obtain

$$(1-t)\alpha_p d_p(z,\mathbf{E}X)^2 \le F_{\mathbf{E}X}(z) = E\left[d_p(z,X)^2 - d_p(\mathbf{E}X,X)^2\right].$$

By taking the limit $t \to 0$, we get the inequality (77).

Now, we consider a limiting procedure of random variables taking values in $(\Delta_p, d_p)$ for $1 < p \le 2$. The followings are key lemmas for proving Theorem (6.2.22).

**Lemma (6.2.20)[335]:** Let $0 < \alpha \le 1$ and $C \ge 0$ be given. If $\{a_n\}$ is a nonnegative sequence satisfying the following recurrence relation

$$a_1 \le C \text{ and } a_n \le \frac{(n-1)(n-\alpha^2)}{n^2} a_{n-1} + \frac{(n-(n-1)\alpha)}{n^2} C \ (n \ge 2),$$

then we have

$$a_n \le \left(\frac{(\alpha+1)n + (1-\alpha)\alpha n^2}{\alpha(\alpha^2+1)n^2} + \frac{(\alpha-1)(\alpha+1)^2\Gamma(n+1-\alpha^2)}{\alpha(\alpha^2+1)\Gamma(2-\alpha^2)nn!}\right) C$$

and

$$\limsup_{n\to\infty} a_n \le \frac{1-\alpha}{1+\alpha^2} C.$$

**Proof.** By the induction on $n$, we can get the proof.

**Lemma (6.2.21)[335]:** Let $1 < p \le 2$. If $X, Y$ and $Z$ are independent identically distributed random variables in $L^2(\Omega, \Delta_p)$, then we have that for any $0 \le \lambda \le 1$,

$$\begin{aligned}&E[d(\mathbf{E}Z, X\#_\lambda Y)^2] \\ &\le (1-\lambda)\left(1-\lambda\alpha_p^2\right)E[d(\mathbf{E}Z,X)^2] + \lambda\left(1-(1-\lambda)\alpha_p\right)E[d(\mathbf{E}Z,Y)^2]. \quad (78)\end{aligned}$$

**Proof.** For any $0 \le \lambda \le 1$, we have

$$E[d(\mathbf{E}Z, X\#_\lambda Y)^2]$$

$$\le (1-\lambda)E[d(\mathbf{E}Z,X)^2]+\lambda E[d(\mathbf{E}Z,Y)^2]-(1-\lambda)\lambda\alpha_p E[d(X,Y)^2]$$
$$\le (1-\lambda)E[d(\mathbf{E}Z,X)^2]+\lambda E[d(\mathbf{E}Z,Y)^2]-(1-\lambda)\lambda\alpha_p\left[\alpha_p E[d(X,\mathbf{E}Y)^2]+E[d(\mathbf{E}Y,Y)^2]\right]$$
$$=(1-\lambda)(1-\lambda\alpha_p^2)E[d(\mathbf{E}Z,X)^2]+\lambda(1-(1-\lambda)\alpha_p)E[d(\mathbf{E}Z,Y)^2],$$
where the first inequality follows from the inequality (74) and the second inequality follows from the inequality (77) in Theorem (6.2.19). This completes the proof.

The following theorem gives an asymptotic upper bound of $E[d_p(\mathbf{E}X_1,Z_n)^2]$, where $Z_n$ is the (non-weighted) inductive mean.

**Theorem (6.2.22)[335]:** If $\{X_i\}$ is a sequence of independent identically distributed random variables in $L^2(\Omega,\Delta_p)$ with $1<p\le 2$, then
$$\limsup_{n\to\infty} E[d_p(\mathbf{E}X_1,Z_n)^2]\le \frac{1-\alpha_p}{1+\alpha_p^2}\mathbf{V}(X_1)$$
where $Z_n=Z_{n-1}\#_{1/n}X_n$ for $n\ge 2$ and $Z_1=X_1$. In particular, if each random variable $X_i$ takes values in $L^2(\Omega,\Delta_2)$, then the weak law of large numbers holds, i.e.,
$$Z_n\longrightarrow \mathbf{E}X_1 \text{ in probability.} \tag{79}$$

**Proof.** By Lemma (6.2.21), we obtain
$$E[d_p(\mathbf{E}X_1,Z_n)^2]$$
$$=E\left[d_p\left(\mathbf{E}X_1,Z_{n-1}\#_{1/n}X_n\right)^2\right]$$
$$\le\left(\frac{(n-1)(n-\alpha_p^2)}{n^2}\right)E[d_p(\mathbf{E}X_1,Z_{n-1})^2]+\left(\frac{n-(n-1)\alpha_p}{n^2}\right)\mathbf{V}(X_1).$$
Since $E[d_p(\mathbf{E}X_1,Z_1)^2]=\mathbf{V}(X_1)$, by Lemma (6.2.20), we have for $n\ge 3$,
$$E[d_p(\mathbf{E}X_1,Z_n)^2]$$
$$\le\left(\frac{(\alpha_p+1)n+(1-\alpha_p)\alpha_p n^2}{\alpha_p(\alpha_p^2+1)n^2}+\frac{(\alpha_p-1)(\alpha_p+1)^2\Gamma(n+1-\alpha_p^2)}{\alpha_p(\alpha_p^2+1)\Gamma(2-\alpha_p^2)nn!}\right)\mathbf{V}(X_1)$$
$$\longrightarrow\frac{1-\alpha_p}{1+\alpha_p^2}\mathbf{V}(X_1)$$
as $n\to\infty$. Since $\alpha_2=1$, we see that the weak law of large numbers (79) holds.

## Section (6.3): Toeplitz Lemma in Geodesic Metric Space

For $\{x_n\}$ be a sequence of real numbers with $\lim_{n\to\infty}x_n=x$. Then the well-known Toeplitz lemma says that the weighted mean:
$$S_n:=\sum_{i=1}^{n}a_{ni}x_i \tag{80}$$
converges to same limit $x$, where $\{a_{ni}\}$ is a weighted sequence of positive real numbers with $\lim_{n\to\infty}a_{ni}=0$ for any fixed $i$ and $\sum_{i=1}^{n}a_{ni}=1$ for any $n\in\mathbb{N}$. The Toeplitz lemma, considered as a generalization of the Cesàro theorem, is a useful tool for the study of convergence theorems in probability theory and has been applied many convergence results,

e.g., the law of large numbers and the limits of sequences. [131], proved that the Toeplitz lemma also holds in the sense of the mean convergence and constructed an example for which the Toeplitz lemma does not hold in the sense of the convergence in probability. Also, a complete convergence version of Toeplitz lemma has been studied in [153]. In [314], the authors studied the higher order, multi-dimensional and continuous version of the Cesàro theorem which was proposed for the study of continuous analogue of the Lèvy Laplacian as an infinite dimensional Laplacian.

In a geodesic metric space, the notion of geodesic allow us to define a weighted inductive mean with a positive weighted sequence. In the case of real-valued sequence with the (usual) Euclidean metric, the weighted inductive mean is exactly equal to the weighted mean (see (80)). Thus it is a very natural and interesting question is whether the Toeplitz lemma holds for the weighted inductive means in a geodesic metric space.

Another type of weighted inductive means has been introduced by Hansen [142] (also see [347]) in 2014 and so is called the Hansen's inductive mean. [347], proved the Hansen's inductive mean is a contractive weighted geometric mean for the trace matric. But, for the Hansen's inductive mean, it is not known whether it converges or not (see Problem 1 in [347]), and so it is also one of interesting questions that any convergence of the Hansen's inductive mean.

We study the Toeplitz lemma for inductive means in a geodesic metric space and by using the Toeplitz lemma, we prove the Cesàro theorem for inductive means. Also, we study an asymptotic property for the Hansen's inductive geometric mean and as an application of the Toeplitz lemma, we prove a convergence of the Hansen's inductive mean.

We recall the basic notions for a $p$-uniformly convex metric space and we prove the Toeplitz lemma for weighted inductive means in a $p$ uniformly convex metric space. Also, we prove the the Cesàro theorem for a non-weighted inductive means. We recall the notion of random variables in a metric space and then we study some inequalities for weighted inductive means. Using this inequality we first prove an asymptotic property for the Hansen's inductive geometric mean of independent identically distributed (i.i.d.) random variables valued in a Hadamard space. Secondly, by using the Toeplitz lemma, we prove a convergence of the Hansen's inductive geometric mean of i.i.d. random variables valued in a Hadamard space. Finally, we consider two interesting examples of metric spaces satisfying the Toeplitz Lemma for the matrix geometric mean and the geometric mean of positive operators, respectively.

For $(M,d)$ be a metric space. For any $x,y\in M$, a continuous map $\gamma:[0,1]\to M$ is called a geodesic joining $x$ and $y$ if it satisfies the following properties: for any $s,t\in[0,1]$,

$$d\big(\gamma(s),\gamma(t)\big)\leq|s-t|d\big(\gamma(0),\gamma(1)\big) \text{ and } \gamma(0)=x,\gamma(1)=y.$$

A metric space $(M,d)$ is called a geodesic metric space if for any $x,y\in M$, there exists a geodesic $\gamma:[0,1]\to M$ such that $\gamma(0)=x$ and $\gamma(1)=y$.

For any fixed $2\leq p<\infty$, let $(M,d)$ be a $p$-uniformly convex metric space, i.e., $M$ is a geodesic metric space and for any $z\in M$ and any geodesic $\gamma:[0,1]\to M$ with $\gamma(0)=x$, $\gamma(1)=y$, there exists a constant $0<c_M\leq 1$ such that

$$d(z,\gamma(t))^p\leq(1-t)d(z,x)^p+td(z,y)^p-c_Mt(1-t)d(x,y)^p.$$

For each $n\in\mathbb{N}$ and each finite sequence $\boldsymbol{w}:=\{w_{ni}\}$ of positive real numbers, $\boldsymbol{w}$ is called a positive weighted sequence if $\sum_{i=1}^{n}w_{ni}=1$. For a given positive weighted sequence $\boldsymbol{w}:=\{w_{ni}\}$,

we define a new positive weighted sequence $\widehat{\boldsymbol{w}}_\ell:=\{\widehat{w}_{\ell i}\}_{i=1}^{\ell}$ for $\ell=1,\dots,n$ as follows:

$$\hat{w}_\ell = \frac{w_{ni}}{\sum_{j=1}^{\ell} w_{nj}} \text{ for all } i = 1, \dots \ell. \tag{81}$$

In fact, $\hat{\boldsymbol{w}}_n = \boldsymbol{w}$.

For each $n \in \mathbb{N}$, positive weighted sequence $\boldsymbol{w} = \{w_{ni}\}$ and (finite) sequence $\boldsymbol{x} := \{x_i\}_{i=1}^n$ of elements in $(M, d)$, we define a sequence $\{S_n\}$ of elements in $M$, which is called the weighted inductive mean, as follows [225]:

$$S_1(1; x_1) = x_1 \text{ and } S_n(\boldsymbol{w}; \boldsymbol{x}) = z_{n-1} \#_{w_{nn}} x_n \ (n \geq 2). \tag{82}$$

where $z_1 = x_1$ and $z_\ell := z_{\ell-1} \#_{\hat{w}_\ell t} x_\ell$ for $\ell = 2,3, \cdots, n-1$. Here $x \#_t y$ denotes the $t$-weighted geometric mean which is the point $\gamma(t)$ on the geodesic $\gamma: [0,1] \to M$ with $\gamma(0) = x$ and $\gamma(1) = y$. The Hansen's (weighted) inductive (geometric) mean $H_n$ is defined as follows [347]:

$$H_1(1; x_1) = x_1,$$
$$H_n(\boldsymbol{w}; \boldsymbol{x}) = H_{n-1}\left(\hat{\boldsymbol{w}}_{n-1}; x_1 \#_{w_{nn}} x_n, x_2 \#_{w_{nn}} x_n, \cdots, x_{n-1} \#_{w_{nn}} x_n\right). \tag{83}$$

Note that $S_2(x_1, x_2) = H_2(x_1, x_2)$, but, in general, $S_n(\boldsymbol{w}; \boldsymbol{x}) \neq H_n(\boldsymbol{w}; \boldsymbol{x})$ for $n \geq 3$ (see [347]).

**Lemma (6.3.1)[345]:** Let $(M, d)$ be a p-uniformly convex metric space. Let $\boldsymbol{x} := \{x_i\}_{i=1}^n$ be a sequence of elements in $(M, d)$ and $\boldsymbol{w} := \{w_{ni}\}$ be a positive weighted sequence. Then we have for any $z$ in $M$,

$$d(S_n(\boldsymbol{w}; \boldsymbol{x}), z)^p \leq \sum_{k=1}^{n} w_{nk} d(x_k, z)^p, \tag{84}$$

$$d(H_n(\boldsymbol{w}; \boldsymbol{x}), z)^p \leq \sum_{k=1}^{n} w_{nk} d(x_k, z)^p. \tag{85}$$

**Proof.** Using the convexity of $x \mapsto d(x, z)^p$ (it is clear by the definition of a $p$-uniformly convex metric space) and the mathematical induction, we can prove the desired assertion.

Let $(M, d)$ be a metric space and let

$$\Delta_n := \{\boldsymbol{w} = (w_{n1}, w_{n2}, \cdots, w_{nn}) \in \mathbb{R}^n \mid \boldsymbol{w} \text{ is a positive weighted sequence}\}$$

Let $G: \Delta_n \times M^n \to M$ be a contractive mean for the given metric $d$, i.e,

$$d(G(\boldsymbol{w}, \boldsymbol{x}), G(\boldsymbol{w}, \boldsymbol{y})) \leq \sum_{i=1}^{n} w_{ni} d(x_i, y_i)$$

for $\boldsymbol{x} = \{x_i\}_{i=1}^n, \boldsymbol{y} = \{y_i\}_{i=1}^n \in M^n$ and $\boldsymbol{w} := \{w_{ni}\} \in \Delta_n$. Then for any $p \geq 1$, since the function $t^p$ $(t \geq 0)$ is convex, by using the Jensen's inequality we have that

$$\begin{aligned} d(G(\boldsymbol{w}; \boldsymbol{x}), z)^p &\leq \left(\sum_{i=1}^{n} w_{ni} d(x_i, z)\right)^p \\ &\leq \sum_{i=1}^{n} w_{ni} d(x_i, z)^p \end{aligned}$$

for any $\boldsymbol{x} = \{x_i\}_{i=1}^n, \boldsymbol{y} = \{y_i\}_{i=1}^n \in M^n, z \in M$ and $\boldsymbol{w} := \{w_{ni}\} \in \Delta_n$, which is compared to Lemma (6.3.1).

The following theorem is well-known as Toeplitz Lemma for real valued sequences and so its proof is also well-known.

**Theorem (6.3.2)[345]:** Let $\{x_n\}$ be a sequence of real numbers with $\lim_{n\to\infty} x_n = 0$. Let $\{w_{ni}\}(n \in \mathbb{N})$ be a weighted sequence of real numbers with $\lim_{n\to\infty} w_{ni} = 0$ for any fixed $i \geq 1$ and $\sum_{i=1}^{n} |w_{ni}| < \infty$. Then it holds that

$$\lim_{n\to\infty} \sum_{i=1}^{n} w_{ni} x_i = 0.$$

The following theorem is a geodesic metric space version of Toeplitz Lemma.

**Theorem (6.3.3)[345]:** (Toeplitz Lemma for inductive means). Let $(M, d)$ be a p-uniformly convex metric space. Let $\boldsymbol{x} := \{x_i\}_{i=1}^{n}$ be a sequence of elements in $(M, d)$ and $\boldsymbol{w} := \{w_{ni}\}$ be a positive weighted sequence. Suppose that there exists a element $z$ in $M$ such that $x_n \to z$, i.e., $\lim_{n\to\infty} d(x_n, z) = 0$ and $\lim_{n\to\infty} w_{ni} = 0$ for any fixed $i \geq 1$. Then we have

$$S_n(\boldsymbol{w}; \boldsymbol{x}) \to z \text{ and } H_n(\boldsymbol{w}; \boldsymbol{x}) \to z, \tag{86}$$

as $n \to \infty$. In particular, if $b_n := \sum_i^n a_i, a_i > 0$ with $b_n \to \infty$ as $n \to \infty$, then $x_n \to z$ implies that

$$S_n(\boldsymbol{w}; \boldsymbol{x}) \to z \text{ and } H_n(\boldsymbol{w}; \boldsymbol{x}) \to z,$$

where $\boldsymbol{w} := \left\{w_{ni} = \frac{a_i}{b_n}\right\}$ for $n \in \mathbb{N}$ and $i = 1,2, \cdots, n$.

**Proof.** By Lemma (6.3.1), we have

$$d(S_n(\boldsymbol{w}, \boldsymbol{x}), z)^p \leq \sum_{i=1}^{n} w_{ni} d(x_i, z)^p,\ d(H_n(\boldsymbol{w}; \boldsymbol{x}), z)^p \leq \sum_{i=1}^{n} w_{ni} d(x_i, z)^p.$$

Therefore, by applying Theorem (6.3.2) to the right-hand side in above inequalities, the proof is completed. For the second part, by putting $w_{ni} = \frac{a_i}{b_n}$ with $a_i > 0$ for $i = 1,2, \cdots, n$ in (86), the proof is immediate.

If we take $w_{ni} = 1/n$ for all $i = 1, \dots, n$ in the weighted inductive mean (82), then we obtain that $\widehat{w}_{\ell i} = 1/\ell$ for all $i = 1, \dots, \ell$ and $\ell = 1, \dots, n-1$. Hence, we have

$$S_1(\boldsymbol{x}) = x_1,$$
$$S_n(\boldsymbol{x}) = S_{n-1} \#_{1/n} x_n,\ n \geq 2,$$

and

$$H_1(x) = x_1,$$
$$H_n(x) = H_{n-1}\left(\frac{1}{n-1}; x_1 \#_{1/n} x_n, x_2 \#_{1/n} x_{nh}, \dots, x_{n-1} \#_{1/n} x_n\right),\ n \geq 2,$$

which are called $a$ non-weighted inductive mean and a non-weighted Hansen's inductive mean, respectively.

By using Theorem (6.3.3), we have the following Cesàro theorem for a non-weighted inductive mean and a non-weighted Hansen's inductive mean, respectively.

**Corollary (6.3.4)[345]:** (Cesaro Theorem for inductive means). Let $x := \{x_1\}_{i=1}^{n}$ be a sequence of elements in $M$. Suppose that there exists a element zin $M$ such that $x_n \to z$ Then

$$S_n(x) \to z \text{ and } H_n(x) \to z, \tag{87}$$

as $n \to \infty$, where $S_n$ is a non-weighted inductive mean and $H_n$ is a non-weighted Hansen's inductive mean given, respectively.

We always assume that $(M, d)$ is a Hadamard space, i.e., $M$ is a complete geodesic space satisfying the following property: for any $x, y \in M$, there exists a point $m \in M$ such that

$$d(z, m)^2 \leq \frac{1}{2} d(z, x)^2 + \frac{1}{2} d(z, y)^2 - \frac{1}{4} d(x, y)^2 \text{ for any } z \in M. \tag{88}$$

Note that any Hadamard space is a 2-uniformly convex metric space with parameter $c_M = 1$, and that the metric function $x \mapsto d(z,x)$ is convex (indeed, $(x,y) \mapsto d(x,y)$ is doubly convex) (see [231]).

Let $(\Omega, \mathcal{F}, P)$ be a probability space. A function $X: \Omega \to M$ is called an $M$-valued random variable (or simply, random variable) if $X$ is a Borel measurable function. For a Borel subset $B$ of $M$, we put $P_X(B) := P(X^{-1}(B))$, which is called the distribution (or law) of $X$. A sequence $\{X_n\}$ of random variables is said to be independent if for any finite subset $I$ of $\mathbb{N}$,

$$P\left(\bigcap_{i\in I} X_i^{-1}(B_i)\right) = \prod_{i\in I} P\left(X_i^{-1}(B_i)\right),$$

where $\{B_i\}_{i\in I}$ is any finite sequence of Borel subsets of $M$, and is identically distributed if $P_{X_i}(B) = P_{X_j}(B)$ for any Borel subset $B$ of $M$ and $i,j \in \mathbb{N}$.

For $1 \le p < \infty$, let $L^p(\Omega, M)$ be the set of all random variables $X$ such that

$$\int_\Omega [d(z, X(\omega))]^p dP(\omega) = \int_M [d(z,x)]^p dP_X(x) < \infty$$

for some $z \in M$, and $L^\infty(\Omega, M)$ be the set of all random variables $X$ such that

$$d(z, X(\omega)) \le R \text{ a.s.}$$

for some $z \in M$ and $R \ge 0$.

For a given real-valued function $\phi$ on $M$, if there exists a point $x \in M$ such that $\phi(x) = \inf_{z\in M} \phi(z)$, then $x$ is called a minimizer and denoted by $x := \operatorname{argmin}_{z\in M} \phi(z)$. We now define the expectation (or barycenter) of $X$ in $L^1(\Omega, M)$[231]: for each fixed $y \in M$,

$$\mathbf{E}[X] := \operatorname*{argmin}_{z\in M} E[d(z,X)^2 - d(y,X)^2] := \operatorname*{argmin}_{z\in M} \int_M [d(z,x)^2 - d(y,x)^2] dP_X(x).$$

For a random variable $X \in L^1(\Omega, M)$, we define the variance of $X$ by

$$\mathbf{V}(X) := \inf_{z\in M} E[d(z,X)^2].$$

**Remark (6.3.5)[345]:** If we restrict to $X \in L^2(\Omega, M)$, then $\mathbf{E}[X]$ is the unique minimizer of $z \mapsto d(z,X)^2$ (see [231]), i.e.,

$$\mathbf{E}[X] = \operatorname*{argmin}_{\gamma\subset M} E[d(z,X)^2] = \operatorname*{argmin}_{z\in M} \int_M d(z,x)^2 dP_X(x). \qquad (89)$$

Hence, $\mathbf{V}(X) = E[d(\mathbf{E}[X], X)^2] < \infty$.

**Proposition (6.3.6)[345]:** Let $\mathbf{X} = \{X_i\}_{i=1}^n (n \in \mathbb{N})$ and $\mathbf{Y} = \{Y_i\}_{i=1}^n$ be two sequences of random variables valued in $(M,d)$ and $w := \{w_{mi}\}$ be a positive weighted sequence. Then we have

(i) For all $z \in M, d(z, S_n(w;\mathbf{X}))^2 \le \sum_{i=1}^n w_{mi} d(z, X_i)^2$.

(ii) $d(H_n(w;\mathbf{X}), H_n(w;\mathbf{Y})) \le \sum_{i=1}^n w_{w^2} d(X_i, Y_i)$.

(iii) $d(H_{n-1}(\hat{\boldsymbol{w}}_{n-1}; X_1, \cdots X_{n-1}), H_n(\boldsymbol{w};\mathbf{X})) \le \frac{w_{nn}}{1-w_{nn}} \sum_{i=1}^{n-1} w_n d(X_i, X_n)$, where $\widehat{w}_{n-1} := \left\{\frac{w_m}{1-w_{i=}}\right\}_{i=1}^{n-1}$.

(iv) For $n \ge 2$, it hold that for all $z \in M$

$$d(z, H_n(w;\mathbf{X}))^2 \le \sum_{i=1}^n w_m d(z, X_i)^2 - w_{mn} \sum_{i=1}^{n-1} w_m d(X_i, X_n)^2. \quad (90)$$

**Proof.** For the proof of (i) and (ii), we refer to [347]. By using (ii) and the convexity of $x \mapsto d(x,z)$, we have

$$d(H_{n-1}\widehat{\boldsymbol{w}}_{n-1}; X_1, \cdots X_{n-1}), H_n(\boldsymbol{w}; \mathbf{X}))$$
$$= d(H_{n-1}\widehat{\boldsymbol{w}}_{n-1}; X_1, \cdots X_{n-1}), H_{n-1}\big(\widehat{\boldsymbol{w}}_{n-1}; X_1 \#_{w_{nn}} X_n, \cdots X_{n-1} \#_{w_{mn}} X_n\big)$$
$$\le \frac{1}{1-w_{nn}} \sum_{i=1}^{n-1} w_{ni} d\big(X_i, X_i \#_{w_{nn}} X_n\big) \le \frac{w_{nn}}{1-w_{nn}} \sum_{i=1}^{n-1} w_{ni} d(X_i, X_n),$$

where for the second inequality, we used (ii) and for the third inequality, we used the convexity of $x \mapsto d(x,z)$. Thus, the proof of (iii) is completed. We use the mathematical induction to prove (90). Indeed, it is clear for $n = 2$. Suppose that (90) is hold for $n-1$. Then, by the convexity of $x \mapsto d(x,z)^2$, we obtain that

$$d\quad (z, H_n(\boldsymbol{w}; \mathbf{X}))^2 = d\left(z, H_{n-1}\big(\widehat{\boldsymbol{w}}_{n-1}; X_1 \#_{w_{nn}} X_n, \cdots, X_{n-1} \#_{w_{nn}} X_n\big)\right)^2$$
$$\le \quad \frac{1}{1-w_{nn}} \left( \sum_{i=1}^{n-1} w_i d\big(z, X_i \#_{w_{nn}} X_n\big)^2 \right)$$
$$- \frac{w_{n(n-1)}}{(1-w_{nn})^2} \left( \sum_{i=1}^{n-2} w_{ni} d\big(X_i \#_{w_{nn}} X_n, X_{n-1} \#_{w_{nn}} X_n\big)^2 \right)$$
$$\le \quad \frac{1}{1-w_{nn}} \sum_{i=1}^{n-1} w_{ni} \{(1-w_{nn}) d(z, X_i)^2 + w_{nn} d(z, X_n)^2 - (1-w_{nn}) w_{nn} d(X_i, X_n)^2\}$$
$$= \quad \sum_{i=1}^{n-1} w_{ni} d(z, X_i)^2 + \frac{w_{nn}}{1-w_{nn}} (1-w_{nn}) d(z, X_n)^2 - w_{nn} \sum_{i=1}^{n-1} w_{ni} d(X_i, X_n)^2$$
$$= \quad \sum_{i=1}^{n} w_{ni} d(z, X_i)^2 - w_{nn} \sum_{i=1}^{n-1} w_{ni} d(X_i, X_n)^2.$$

The proof is completed.

**Proposition (6.3.7)[345]:** Let $\mathbf{X} = \{X_i\}_{i=1}^n$ be a sequence of random variable valued in $(M, d)$. For all $n \ge 2$, it hold that

$$d(S_n(\boldsymbol{w}; \mathbf{X}), H_n(\boldsymbol{w}; \mathbf{X}))^2 \le \sum_{i \ne j}^{n} w_{ni} w_{nj} d\big(X_i, X_j\big)^2 - w_{nn} \sum_{i=1}^{n-1} w_{ni} d(X_i, X_n)^2. \quad (91)$$

**Proof.** By using (i) and (iv) in the Proposition (6.3.6), we have

$$d(S_n(\boldsymbol{w}; \mathbf{X}), H_n(\boldsymbol{w}; \mathbf{X}))^2 \quad \le \sum_{j=1}^{n} w_{nj} d\big(S_n(\boldsymbol{w}; \mathbf{X}), X_j\big)^2 - w_{nn} \sum_{i=1}^{n-1} w_{ni} d(X_i, X_n)^2$$
$$\le \sum_{i \ne j}^{n} w_{ni} w_{nj} d\big(X_i, X_j\big)^2 - w_{nn} \sum_{i=1}^{n-1} w_{ni} d(X_i, X_n)^2,$$

where for the first inequality, we use (iv) and for the second inequality, we use (i). The proof is completed.

We recall the strong law of large numbers for weighted inductive means in a Hadamard space.

**Theorem (6.3.8)[345]:** ([335]). Let $(M, d)$ be a Hadamard space. Let $\mathrm{X} = \{X_i\}_{i=1}^n$ be a sequence of independent identically distributed random variables in $L^\infty(\Omega, M)$. If $\boldsymbol{w} := \{w_{ni}\}$ is a positive weighted sequence satisfying the following condition:

(C) There exists a constant $C \geq 1$ such that $\max_{1\leq i\leq n} w_{ni} \leq C\min_{1\leq i\leq n} w_{ni}$ then

$$S_n(\boldsymbol{w};\mathbf{X}) \to \mathbf{E}[X_1] \text{ a.s.} \tag{92}$$

as $n \to \infty$, where $S_n$ is given as in (82).
[347], constructed some examples that the Hansen's inductive mean does not converge to same limit for the inductive mean $S_n$ (see Examples 5.1 and 5.3 in [347]). The following theorem give an asymptotic property for the Hansen's inductive geometric mean.
**Theorem (6.3.9)[345]:** Under the same assumption in Theorem (6.3.8), we have for some $R > 0$,

$$\lim_{n\to\infty} d(\mathbf{E}[X_1], H_n(\boldsymbol{w};\mathbf{X})) \leq 2R\sqrt{2\left(1-\sum_{i=1}^{\infty} w_{ni}^2\right)} \tag{93}$$

as $n \to \infty$, where $H_n$ is given as in (83).
**Proof.** Since $\{X_i\} \subseteq L^\infty(\Omega, M)$, there exists a constant $R > 0$ such that for all $i, d(z, X_i(\omega)) < R$ a.s. for some $z \in M$. Thus by Proposition (6.3.7), we have

$$\begin{aligned} d(\mathbf{E}[X_1], H_n(\boldsymbol{w};\mathbf{X})) &\leq d(\mathbf{E}[X_1], S_n(\boldsymbol{w};\mathbf{X})) + d(S_n(\boldsymbol{w};\mathbf{X}), H_w(\boldsymbol{w};\mathbf{X})) \\ &\leq d(\mathbf{E}[X_1], S_n(\boldsymbol{w};\mathbf{X})) + 2R\sqrt{\sum_{i+j}^{n} w_n w_{nj}} \\ &= d(\mathbf{E}[X_1], S_n(\boldsymbol{w};\mathbf{X})) + 2R\sqrt{\left(1-\sum_{i=1}^{n} w_m^2\right)} \end{aligned}$$

Therefore, by Theorem (6.3.8), we obtain that

$$\lim_{n\to\infty} d(\mathbf{E}[X_1], H_n(w;\mathbf{X})) \leq 2R\sqrt{\left(1-\sum_{i=1}^{\infty} w_{\mu i}^2\right)}.$$

The proof is completed.
(see Problem 1 in [347]). The following theorem is a partial answer to the open problem in [347] (see Problem 1 in [347]). For the proof, we use the Toeplitz lemma (Theorem (6.3.3)).
**Theorem (6.3.10)[345]:** Under the same assumption in Theorem (6.3.8), we have

$$H_n(\boldsymbol{w}; S_1, \cdots, S_n) \to \mathbf{E}[X_1] \text{ a.s.} \tag{94}$$

as $n \to \infty$, where $S_n$ is given as in (82) and $H_n$ is given as in (83).
**Proof.** Since $S_n(\boldsymbol{w};\mathbf{X}) \to \mathbf{E}[X_1]$ a.s. as $n \to \infty$, by Theorem (6.3.3), the proof is completed.

We consider two interesting examples of metric spaces satisfying the geometric version of the Toeplitz lemma.

Let $\mathbb{H} := \mathbb{H}_n$ be the spaces of $n \times n$ Hermitian matrices (i.e., $A \in \mathbb{H}$ if and only if $A^* = A$ ) and $\mathbb{P} := \mathbb{P}_n$ the convex cone of positive definite element in $\mathbb{H}$ (i.e., $A \in \mathbb{P}$ if and only if $\langle Ax, x\rangle > 0$, for all $x \in \mathbb{R}^n$, where $\langle\cdot,\cdot\rangle$ is the usual inner product on $\mathbb{R}^n$ ). We now recall the Riemannian trace metric on $\mathbb{P}$ as following:

$$\delta(A, B) := \|\log A^{-1/2} B A^{-1/2}\|_2,$$

where $\| A \|_2 := (\operatorname{tr} A^* A)^{1/2}$. Then it is well-known fact that $(\mathbb{P}, \delta)$ is a complete geodesic metric space having a unique geodesic $A \#_t B := A^{1/2}\left(A^{-1/2} B A^{-1/2}\right)^t A^{1/2}$. Also, for fixed $Z \in \mathbb{P}, A \mapsto \delta(Z, A)^2$ is a convex function (see [3]). Then we have the following result.

**Theorem (6.3.11)[345]:** (Toeplitz Lemma for matrix geometric means). Let $A := \{A_i\}_{i=1}^n$ be a sequence of elements in $(\mathbb{P}, \delta)$ and $w := \{w_{ni}\}$ be a positive weighted sequence. Suppose that there exists an element $A$ in $\mathbb{P}$ such that $A_n \to A$, i.e., $\lim_{n\to\infty} \delta(A_n, A) = 0$ and $\lim_{n\to\infty} w_{ni} = 0$ for any fixed $i \geq 1$. Then we have

$$S_n(\boldsymbol{w}; \boldsymbol{A}) \to A \text{ and } H_n(\boldsymbol{w}; \boldsymbol{A}) \to A, \tag{95}$$

as $n \to \infty$, where $S_n$ is given as in (82) and $H_n$ is given as in (83). In particular, if $b_n := \sum_i^n a_i$, $a_i > 0$ with $b_n \to \infty$ as $n \to \infty$, then $A_n \to$ A implies that

$$S_n(\boldsymbol{w}; \boldsymbol{A}) \to A \text{ and } H_n(\boldsymbol{w}; \boldsymbol{A}) \to A,$$

where $\boldsymbol{w} := \left\{w_{ni} = \frac{a_i}{b_n}\right\}$ for $n \in \mathbb{N}$ and $i = 1,2,\cdots,n$

**Proof.** Since $(\mathbb{P}, \delta)$ is a 2 -uniformly convex metric space with parameter $c_{\mathbb{P}} = 1$( see [3]), the assertions are immediate from Theorem (6.3.3).

Let $H$ be a Hilbert space with inner product $\langle\cdot,\cdot\rangle_H$ and $\mathcal{P}$ be the set of all positive invertible elements in $\mathcal{B}(H)$, where $\mathcal{B}(H)$ is the Banach space of all bounded linear operators on $H$ equipped with the operator norm. The Thompson metric on $\mathcal{P}$ is defined by

$$d(S,T) := \text{m } \{\log M(S \setminus T), \log M(T \setminus S)\},\ S,T \in \mathcal{P},$$

where $M(S \setminus T) = \inf\{\lambda \mid S \leq \lambda T\}$ (see [349]). It is one of important fact that for fixed $U \in \mathcal{P}$, the map $\mathcal{P} \ni S \mapsto d(U,S)$ is a convex function for the geodesic $S \#_t T := S^{1/2}\left(S^{-1/2} T S^{-1/2}\right)^t S^{1/2}$ (see [346], [348], [150]).

**Theorem (6.3.12)[345]:** (Toeplitz Lemma for positive operators). Let $T := \{T_i\}_{i=1}^n$ be a sequence of elements in $(\mathcal{P}, d)$ and $\boldsymbol{w} := \{w_{ni}\}$ be a positive weighted sequence. Suppose that there exists an element $T$ in $\mathcal{P}$ such that $T_n \to T$, i.e., $\lim_{n\to\infty} d(T_n, T) = 0$ and $\lim_{n\to\infty} w_{ni} = 0$ for any fixed $i \geq 1$. Then we have

$$S_n(\boldsymbol{w}; \boldsymbol{T}) \to T \text{ and } H_n(\boldsymbol{w}; \boldsymbol{T}) \to T, \tag{96}$$

as $n \to \infty$, where $S_n$ is given as in (82) and $H_n$ is given as in (83). In particular, if $b_n := \sum_i^n a_i$, $a_i > 0$ with $b_n \to \infty$ as $n \to \infty$, then $T_n \to T$ implies that

$$S_n(\boldsymbol{w}; \boldsymbol{T}) \to T \text{ and } H_n(\boldsymbol{w}; \boldsymbol{T}) \to T.$$

where $\boldsymbol{w} := \left\{w_{ni} = \frac{a_i}{b_n}\right\}$ for $n \in \mathbb{N}$ and $i = 1,2,\cdots,n$

**Proof.** Since the metric function $d(U,\cdot)$ is convex, and using the proof of Theorem (6.3.3), the proof is straightforward.

**Corollary (6.3.13)[350]:** Let $(M,d)$ be a $(1+\epsilon)$-uniformly convex metric space. Let $\boldsymbol{x}^m := \{x_i^m\}_{i=1}^n$ be a sequence of elements in $(M,d)$ and $\boldsymbol{w}^m := \{w_{ni}^m\}$ be a positive weighted sequence. Then we have for any $z^m$ in $M$,

$$d(S_n^m(\boldsymbol{w}^m; \boldsymbol{x}^m), z^m)^{2+\epsilon} \leq \sum_{k=1}^n \sum_m w_{nk}^m d(x_k^m, z^m)^{2+\epsilon}, \tag{97}$$

$$d(H_n(\boldsymbol{w}^m; \boldsymbol{x}^m), z^m)^{2+\epsilon} \leq \sum_{k=1}^n \sum_m w_{nk}^m d(x_k^m, z^m)^{2+\epsilon}. \tag{98}$$

**Proof.** Using the convexity of $x^m \mapsto d(x^m, z^m)^{2+\epsilon}$ (it is clear by the definition of a $(2+\epsilon)$-uniformly convex metric space) and the mathematical induction, we can prove the desired assertion.

**Corollary (6.3.14)[350]:** (Toeplitz Lemma for inductive means). Let $(M,d)$ be a p-uniformly convex metric space. Let $\boldsymbol{x}^m := \{x_i^m\}_{i=1}^n$ be a sequence of elements in $(M,d)$ and $\boldsymbol{w}^m := \{w_{ni}^m\}$ be a positive weighted sequence. Suppose that there exists a element $z^m$ in $M$ such that $x_n^m \to z^m$, i.e., $\lim_{n\to\infty} d(x_n^m, z^m) = 0$ and $\lim_{n\to\infty} w_{ni}^m = 0$ for any fixed $i \geq 1$. Then we have

$$S_n^m(\boldsymbol{w}^m;\boldsymbol{x}^m) \to z^m \text{ and } H_n(\boldsymbol{w}^m;\boldsymbol{x}^m) \to z^m, \tag{99}$$

as $n \to \infty$. In particular, if $b_n := \sum_i^n a_i^m, a_i^m > 0$ with $b_n \to \infty$ as $n \to \infty$, then $x_n^m \to z^m$ implies that

$$S_n^m(\boldsymbol{w}^m;\boldsymbol{x}^m) \to z^m \text{ and } H_n(\boldsymbol{w}^m;\boldsymbol{x}^m) \to z^m,$$

where $\boldsymbol{w}^m := \left\{w_{ni}^m = \frac{a_i^m}{b_n}\right\}$ for $n \in \mathbb{N}$ and $i = 1,2,\cdots,n$.

**Proof.** By Corollary (6.3.13), we have

$$d(S_n^m(\boldsymbol{w}^m,\boldsymbol{x}^m),z^m)^{1+\epsilon} \leq \sum_{i=1}^n \sum_m w_{ni}^m d(x_i^m,z^m)^{1+\epsilon},\ d(H_n(\boldsymbol{w}^m;\boldsymbol{x}^m),z^m)^{1+\epsilon}$$
$$\leq \sum_{i=1}^n \sum_m w_{ni}^m d(x_i^m,z^m)^{1+\epsilon}.$$

Therefore, by applying Theorem (6.3.2) to the right-hand side in above inequalities, the proof is completed. For the second part, by putting $w_{ni}^m = \frac{a_i^m}{b_n}$ with $a_i^m > 0$ for $i = 1,2,\cdots,n$ in (99), the proof is immediate.

**Corollary (6.3.15)[350]:** (See [345]). Let $\mathbf{X} = \{X_i\}_{i=1}^n (n \in \mathbb{N})$ and $\mathbf{Y} = \{Y_i\}_{i=1}^n$ be two sequences of random variables valued in $(M,d)$ and $w^m := \{w_{m_i}^m\}$ be a positive weighted sequence. Then we have

(i) For all $z^m \in M, d(z^m, S_n^m(w^m;\mathbf{X}))^2 \leq \sum_{i=1}^n w_{m_i}^m d(z^m,X_i)^2$.

(ii) $d(H_n(w^m;\mathbf{X}),H_n(w^m;\mathbf{Y})) \leq \sum_{i=1}^n \sum_m \ w_{w^{2m}}^m d(X_i,Y_i)$.

(iii) $d(H_{n-1}(\hat{\boldsymbol{w}}_{n-1}^m;X_1,\cdots X_{n-1}),H_n(\boldsymbol{w}^m;\mathbf{X})) \leq \frac{w_{nn}^m}{1-w_{nn}^m}\sum_{i=1}^{n-1}\sum_m \ w_n^m d(X_i,X_n)$, where $\widehat{w}_{n-1}^m := \left\{\frac{w_{m_0}^m}{1-w_i^m}\right\}_{i=1}^{n-1}$.

(iv) For $n \geq 2$, it hold that for all $z^m \in M$

$$d(z^m,H_n(w^m;\mathbf{X}))^2 \leq \sum_{i=1}^n \sum_m \ w_{m_0}^m d(z^m,X_i)^2 - w_{m_0 n}^m \sum_{i=1}^{n-1}\sum_m \ w_{m_0}^m d(X_i,X_n)^2. \tag{100}$$

**Proof.** For the proof of (i) and (ii), we refer to [347]. By using (ii) and the convexity of $x^m \mapsto d(x^m,z^m)$, we have

$$d(H_{n-1}\widehat{\boldsymbol{w}}_{n-1}^m;X_1,\cdots X_{n-1}),H_n(\boldsymbol{w}^m;\mathbf{X}))$$
$$= d(H_{n-1}\widehat{\boldsymbol{w}}_{n-1}^m;X_1,\cdots X_{n-1}),H_{n-1}\left(\widehat{\boldsymbol{w}}_{n-1}^m;X_1\#_{w_{nn}^m}X_n,\cdots X_{n-1}\#_{w_{m_0 n}^m}X_n\right)$$
$$\leq \sum_m \frac{1}{1-w_{nn}^m}\sum_{i=1}^{n-1} w_{ni}^m d\left(X_i,X_i\#_{w_{nn}^m}X_n\right) \leq \sum_m \frac{w_{nn}^m}{1-w_{nn}^m}\sum_{i=1}^{n-1} w_{ni}^m d(X_i,X_n),$$

where for the second inequality, we used (ii) and for the third inequality, we used the convexity of $x^m \mapsto d(x^m,z^m)$. Thus, the proof of (iii) is completed. We use the mathematical induction to prove (100). Indeed, it is clear for $n = 2$. Suppose that (100) is hold for $n-1$. Then, by the convexity of $x^m \mapsto d(x^m,z^m)^2$, we obtain that

$$d(z^m, H_n(\boldsymbol{w}^m; \mathbf{X}))^2 = d\left(z^m, H_{n-1}\left(\widehat{\boldsymbol{w}}_{n-1}^m; X_1 \#_{w_{nn}^m} X_n, \cdots, X_{n-1} \#_{w_{nn}^m} X_n\right)\right)^2$$

$$\leq \sum_m \frac{1}{1 - w_{nn}^m}\left(\sum_{i=1}^{n-1} w_i^m d\left(z^m, X_i \#_{w_{nn}^m} X_n\right)^2\right)$$

$$- \sum_m \frac{w_{n(n-1)}^m}{(1 - w_{nn}^m)^2}\left(\sum_{i=1}^{n-2} w_{ni}^m d\left(X_i \#_{w_{nn}^m} X_n, X_{n-1} \#_{w_{nn}^m} X_n\right)^2\right)$$

$$\leq \sum_m \frac{1}{1 - w_{nn}^m} \sum_{i=1}^{n-1} w_{ni}^m \{(1 - w_{nn}^m) d(z^m, X_i)^2 + w_{nn}^m d(z^m, X_n)^2$$

$$- (1 - w_{nn}^m) w_{nn}^m d(X_i, X_n)^2\}$$

$$= \sum_{i=1}^{n-1} \sum_m w_{ni}^m d(z^m, X_i)^2 + \sum_m \frac{w_{nn}^m}{1 - w_{nn}^m}(1 - w_{nn}^m) d(z^m, X_n)^2$$

$$- \sum_m w_{nn}^m \sum_{i=1}^{n-1} w_{ni}^m d(X_i, X_n)^2$$

$$= \sum_{i=1}^{n} \sum_m w_{ni}^m d(z^m, X_i)^2 - \sum_m w_{nn}^m \sum_{i=1}^{n-1} w_{ni}^m d(X_i, X_n)^2.$$

The proof is completed.

**Corollary (6.3.16)[350]:** (See [21]). Let $\mathbf{X} = \{X_i\}_{i=1}^n$ be a sequence of random variable valued in $(M, d)$. For all $n \geq 2$, it hold that

$$d(S_n^m(\boldsymbol{w}^m; \mathbf{X}), H_n(\boldsymbol{w}^m; \mathbf{X}))^2$$
$$\leq \sum_{i \neq j}^{n} \sum_m w_{ni}^m w_{nj}^m d\left(X_i, X_j\right)^2 - \sum_m w_{nn}^m \sum_{i=1}^{n-1} w_{ni}^m d(X_i, X_n)^2. \qquad (101)$$

**Proof.** By using (i) and (iv) in the Corollary (6.3.15), we have

$$d(S_n^m(\boldsymbol{w}^m; \mathbf{X}), H_n(\boldsymbol{w}^m; \mathbf{X}))^2 \leq \sum_{j=1}^{n} \sum_m w_{nj}^m d\left(S_n^m(\boldsymbol{w}^m; \mathbf{X}), X_j\right)^2 - \sum_m w_{nn}^m \sum_{i=1}^{n-1} w_{ni}^m d(X_i, X_n)^2$$
$$\leq \sum_{i \neq j}^{n} \sum_m w_{ni}^m w_{nj}^m d\left(X_i, X_j\right)^2 - \sum_m w_{nn}^m \sum_{i=1}^{n-1} w_{ni}^m d(X_i, X_n)^2,$$

where for the first inequality, we use (iv) and for the second inequality, we use (i). The proof is completed.

**Corollary (6.3.17)[350]:** Under the same assumption in Theorem (6.3.8), we have for some $\epsilon \geq 0$,

$$\lim_{n \to \infty} d(\mathbf{E}[X_1], H_n(\boldsymbol{w}^m; \mathbf{X})) \leq 2(1 + \epsilon)\sqrt{2\left(1 - \sum_{i=1}^{\infty} w_{ni}^{2m}\right)} \qquad (102)$$

as $n \to \infty$, where $H_n$ is given as in (83).

**Proof.** Since $\{X_i\} \subseteq L^\infty(\Omega, M)$, there exists a constant $\epsilon \geq 0$ such that for all $i, d(z^m, X_i(\omega)) < 1 + \epsilon$ a.s. for some $z^m \in M$. Thus by Corollary (6.3.16), we have

$$
\begin{aligned}
d(\mathbf{E}[X_1], H_n(\boldsymbol{w}^m;\mathbf{X})) &\leq d(\mathbf{E}[X_1], S_n^m(\boldsymbol{w}^m;\mathbf{X})) + d(S_n^m(\boldsymbol{w}^m;\mathbf{X}), H_{w^m}(\boldsymbol{w}^m;\mathbf{X})) \\
&\leq d\big(\mathbf{E}[X_1], S_n^m(\boldsymbol{w}^m;\mathbf{X})\big) + 2(1+\epsilon)\sqrt{\sum_{i+j}^{n}\sum_{m} w_n^m w_{nj}^m} \\
&= d(\mathbf{E}[X_1], S_n^m(\boldsymbol{w}^m;\mathbf{X})) + 2(1+\epsilon)\sqrt{\left(1-\sum_{i=1}^{n}\sum_{m} w_{m_0}^{2\mathrm{m}}\right)}
\end{aligned}
$$

Therefore, by Theorem (6.3.8), we obtain that

$$
\lim_{n\to\infty} d(\mathbf{E}[X_1], H_n(w^m;\mathbf{X})) \leq 2(1+\epsilon)\sqrt{\left(1-\sum_{i=1}^{\infty}\sum_{m} w_{\mu i}^{2m}\right)}.
$$

The proof is completed.

**Corollary (6.3.18)[350]:** Under the same assumption in Theorem (6.3.8), we have

$$
H_n(\boldsymbol{w}^m; S_1^m, \cdots, S_n^m) \to \mathbf{E}[X_1] \text{ a.s.} \tag{103}
$$

as $n \to \infty$, where $S_n^m$ is given as in (82) and $H_n$ is given as in (83).

**Proof.** Since $S_n^m(\boldsymbol{w}^m;\mathbf{X}) \to \mathbf{E}[X_1]$ a.s. as $n \to \infty$, by Corollary (6.3.14), the proof is completed.

**Corollary (6.3.19)[350]:** (Toeplitz Lemma for matrix geometric means). Let $A := \{A_i\}_{i=1}^n$ be a sequence of elements in $(\mathbb{P}, \delta)$ and $\boldsymbol{w}^m := \{w_{ni}^m\}$ be a positive weighted sequence. Suppose that there exists an element $A$ in $\mathbb{P}$ such that $A_n \to A$, i.e., $\lim_{n\to\infty}\delta(A_n, A) = 0$ and $\lim_{n\to\infty} w_{ni}^m = 0$ for any fixed $i \geq 1$. Then we have

$$
S_n^m(\boldsymbol{w}^m;\boldsymbol{A}) \to A \text{ and } H_n(\boldsymbol{w}^m;\boldsymbol{A}) \to A, \tag{104}
$$

as $n \to \infty$, where $S_n^m$ is given as in (82) and $H_n$ is given as in (83). In particular, if $b_n := \sum_i^n a_i^m$, $a_i^m > 0$ with $b_n \to \infty$ as $n \to \infty$, then $A_n \to$ A implies that

$$
S_n^m(\boldsymbol{w}^m;\boldsymbol{A}) \to A \text{ and } H_n(\boldsymbol{w}^m;\boldsymbol{A}) \to A,
$$

where $\boldsymbol{w}^m := \left\{w_{ni}^m = \frac{a_i^m}{b_n}\right\}$ for $n \in \mathbb{N}$ and $i = 1,2,\cdots,n$

**Proof.** Since $(\mathbb{P}, \delta)$ is a 2 -uniformly convex metric space with parameter $c_{\mathbb{P}} = 1$( see [3]), the assertions are immediate from Corollary (6.3.14).

**Corollary (6.3.20).** (See [345]) (Toeplitz Lemma for positive operators). Let $T := \{T_i\}_{i=1}^n$ be a sequence of elements in $(\mathcal{P}, d)$ and $\boldsymbol{w}^m := \{w_{ni}^m\}$ be a positive weighted sequence. Suppose that there exists an element $T$ in $\mathcal{P}$ such that $T_n \to T$, i.e., $\lim_{n\to\infty} d(T_n, T) = 0$ and $\lim_{n\to\infty} w_{ni}^m = 0$ for any fixed $i \geq 1$. Then we have

$$
S_n^m(\boldsymbol{w}^m;\boldsymbol{T}) \to T \text{ and } H_n(\boldsymbol{w}^m;\boldsymbol{T}) \to T, \tag{105}
$$

as $n \to \infty$, where $S_n^m$ is given as in (82) and $H_n$ is given as in (83). In particular, if $b_n := \sum_i^n a_i^m$, $a_i^m > 0$ with $b_n \to \infty$ as $n \to \infty$, then $T_n \to T$ implies that

$$
S_n^m(\boldsymbol{w}^m;\boldsymbol{T}) \to T \text{ and } H_n(\boldsymbol{w}^m;\boldsymbol{T}) \to T.
$$

where $\boldsymbol{w}^m := \left\{w_{ni}^m = \frac{a_i^m}{b_n}\right\}$ for $n \in \mathbb{N}$ and $i = 1,2,\cdots,n$

**Proof.** Since the metric function $d(U,\cdot)$ is convex, and using the proof of Corollary (6.3.14), the proof is straightforward.

**List of Symbols**

| Symbol | | Page |
|---|---|---|
| ALM: | Ando – Li – Mathias | 1 |
| det: | determinant | 3 |
| max: | maximum | 3 |
| diag: | diagonal | 13 |
| tr: | trace | 15 |
| $\otimes$: | tensor product (Kronecker product) | 20 |
| $mod$: | modulo | 20 |
| $BMP$: | Beurling – Malliavin problem | 24 |
| $AHG$: | Arithmetic-Harmonic-Geometric | 26 |
| $Vec$: | vector | 29 |
| min: | minimum | 32 |
| grad: | gradient | 40 |
| $ops$: | operations | 42 |
| inf: | infimum | 49 |
| $ker$: | kernel | 49 |
| $eff$: | effecting | 49 |
| dist: | distance | 49 |
| $BM$: | distance | 49 |
| $L^2$: | Hilbert space | 49 |
| $L^\infty$: | essential Lebesgue space | 50 |
| $H^2$: | Hardy space | 50 |
| $\ominus$: | Direct divergence | 50 |
| $\oplus$: | Direct sum | 50 |
| $PW_a$: | Paley - Wiener | 50 |
| loc: | local | 51 |
| $L^1$: | Lebesgue space integral on the real line | 51 |
| $H^p$: | Hardy space | 51 |
| $L^p$: | Lebesgue space | 51 |
| arg: | argument | 51 |
| $SL$: | Sturm - Liouville | 51 |
| $H^\infty$: | essential Hardy space | 53 |
| supp: | support | 54 |
| mult: | multiplicity | 58 |
| cont: | constant | 67 |
| Res: | Residue | 68 |
| $dim$: | dimension | 69 |
| spec: | spectrum | 75 |
| $a.s$: | Almost sure | 123 |
| $\mathcal{L}_p$: | absolute $pth$ power integrable random variables is a Banach space | 127 |
| $i.i.d$: | Independent and identically distributecl | 141 |
| $Var$: | Variance | 142 |
| Re: | Real | 157 |

| | | |
|---|---|---|
| $SGM$: | Structured geometric mean | 161 |
| $KMM$: | Kähler metric mean | 161 |
| $PD$: | Posit hue definite | 162 |
| $TBBT$ | Toeplitz – block – Block - Toeplitz | 163 |
| $CLE$: | Continuous Lyapunov equation | 174 |
| $ess$: | essential | 184 |
| $inn$: | inner | 192 |
| $out$ | outer | 192 |
| $Ind$ | Index | 193 |
| $Ran$: | Range | 195 |
| $Lip$: | Lipschitz | 201 |
| $Im$: | Imaginary | 201 |
| $Hol$: | Holmerphic | 206 |
| $UP$; | Uncertainty principle | 215 |
| $MIF$; | Meromorphic inner functions | 217 |
| $HB$: | Hermite – Biehler | 220 |
| $TE$: | Toeplitz equivalence | 221 |
| $To$: | Toeplitz order | 221 |
| $Dom$: | Domain | 241 |

## References

[1] Dario A. Bini, Beatrice Meini, And Federico Poloni, An Effective Matrix Geometric Mean Satisfying the Ando–Li–Mathias Properties, Volume 79, Number 269, January 2010, Pages 437–452.

[2] T. Ando, Chi-Kwong Li, and Roy Mathias, Geometric means, Linear Algebra Appl. 385 (2004), 305–334. MR2063358 (2005f:47049)

[3] Rajendra Bhatia, Positive definite matrices, Princeton Series in Applied Mathematics, Princeton University Press, Princeton, NJ, 2007. MR2284176 (2007k:15005)

[4] Rajendra Bhatia and John Holbrook, Noncommutative geometric means, Math. Intelligencer 28 (2006), no. 1, 32–39. MR2202893 (2007g:47023)

[5] , Riemannian geometry and matrix geometric means, Linear Algebra Appl. 413 (2006), no. 2-3, 594–618. MR2198952 (2007c:15030)

[6] R. F. S. Hearmon, The elastic constants of piezoelectric crystals, J. Appl. Phys. 3 (1952), 120–123.

[7] Nicholas J. Higham, The Matrix Computation Toolbox, http://www.ma.man.ac.uk/~higham/ mctoolbox.

[8] Bruno Iannazzo and Beatrice Meini, Palindromic matrix polynomials and their relationships with certain functions of matrices, Tech. Report, Dipartimento di Matematica, Universit`a di Pisa, 2009.

[9] Yongdo Lim, On Ando-Li-Mathias geometric mean equations, Linear Algebra Appl. 428 (2008), no. 8-9, 1767–1777. MR2398117

[10] Maher Moakher, On the averaging of symmetric positive-definite tensors, J. Elasticity 82 (2006), no. 3, 273–296. MR2231065 (2007a:74007)

[11] D.A. Bini, B. Iannazzo, A note on computing matrix geometric means, Adv. Comput. Math. 35 (2011) 175–192.

[12] D. Bini and B. Iannazzo. On the numerical solution of the matrix equation Pk i=1 log(XA−1 i ) = 0. 16-th ILAS Conference, Pisa, June 21–25, 2010, www.dm.unipi.it/eilas2010/abstracts-im4.pdf.

[13] D. A. Bini and B. Iannazzo. The Matrix Means Toolbox. http://bezout.dm.unipi.it/software/mmtoolbox/. Retrieved on May 7, 2010.

[14] N. J. Higham. Functions of Matrices: Theory and Computation. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2008.

[15] J. D. Lawson and Y. Lim. The geometric mean, matrices, metrics, and more. Amer. Math. Monthly, 108(9):797–812, 2001.

[16] J. H. Manton. A globally convergent numerical algorithm for computing the centre of mass on compact lie groups. In Eighth International Conference on Control, Automation, Robotics and Vision, 2004. ICARCV 2004 8th, Kunming, China, December 2004, 2004.

[17] M. Moakher. A differential geometric approach to the geometric mean of symmetric positivedefinite matrices. SIAM J. Matrix Anal. Appl., 26(3):735–747, 2005.

[18] F. Poloni. Constructing matrix geometric means, 2009. arXiv:0906.3132v1.

[19] G. Pusz and S. L. Woronowicz. Functional calculus for sesquilinear forms and the purification map. Rep. Math. Phys., 8:159–170, 1975.

[20] D. Bini, B. Iannazzo, B. Jeuris, and R. Vandebril, Geometric means of structured matrices, BIT, 54 (2014), pp. 55–83.

[21] P.-A. Absil, R. Mahony, and R. Sepulchre. Optimization Algorithms on Matrix Manifolds. Princeton University Press, 2008.
[22] F. Barbaresco. Information Intrinsic Geometric Flows. In A. Mohammad-Djafari, editor, Bayesian Inference and Maximum Entropy Methods In Science and Engineering, volume 872 of American Institute of Physics Conference Series, pages 211–218, Nov. 2006.
[23] F. Barbaresco. Interactions between symmetric cone and information geometries: Bruhat-tits and siegel spaces models for high resolution autoregressive doppler imagery. In F. Nielsen, editor, Emerging Trends in Visual Computing, volume 5416 of Lecture Notes in Computer Science, pages 124–163. Springer Berlin / Heidelberg, 2009.
[24] P. G. Batchelor, M. Moakher, D. Atkinson, F. Calamante, and A. Connelly. A Rigorous Framework for Diffusion Tensor Calculus. Magnetic Resonance in Medicine, 53:221–225, 2005.
[25] D. Bini and V. Pan. Polynomial and Matrix Computations. Birkh¨auser, 1994.
[26] D. A. Bini and B. Iannazzo. Computing the Karcher mean of symmetric positive definite matrices. Linear Algebra and its Applications, 2011. In Press, doi 10.1016/j.laa.2011.08.052.
[27] D. A. Bini and B. Iannazzo. Computing the Karcher mean of symmetric positive definite matrices. Linear Algebra Appl., 2012. To appear.
[28] S. Bonnabel and R. Sepulchre. Geometric distance and mean for positive semi-definite matrices of fixed rank. SIAM Journal on Matrix Analysis and Applications, 31(3):1055–1070, 2009.
[29] M. Bridson and A. Haefliger. Metric Spaces of Non-Positive Curvature. Springer-Verlag, 1999.
[30] E. Cartan. Le¸cons sur la g´eom´etrie des espaces de Riemann. Uspekhi Matematicheskikh Nauk, 3(3):218–222, 1948.
[31] M. T. Chu and G. H. Golub. Structured inverse eigenvalue problems. Acta Numerica, 11:1–71, 2002.
[32] Y. Eidelman and I. C. Gohberg. On a new class of structured matrices. Integral Equations and Operator Theory, 34:293–324, 1999.
[33] S. Fiori. Learning the Fr´echet mean over the Manifold of Symmetric Positive-Definite Matrices. Springer Science, Business Media, 2009.
[34] R. A. Horn and C. R. Johnson. Topics in Matrix Analysis. Cambridge University Press, Cambridge, 1991.
[35] B. Iannazzo. The geometric mean of two matrices from a computational viewpoint. Technical report, Dipartimento di Matematica e Informatica, Universit`a di Perugia, Perugia, Italy, 2011. arXiv:1201.0101v1 [math.NA].
[36] B. Jeuris, R. Vandebril, and B. Vandereycken. A survey and comparison of contemporary algorithms for computing the matrix geometric mean. Electronic Transactions on Numerical Analysis, 39:379–402, 2012.
[37] H. Karcher. Riemannian center of mass and mollifier smoothing. Comm. Pure Appl. Math., 30(5):509–541, 1977.
[38] W. Kendall. Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence. Proceedings London Mathematical Society, 61(2):371–406, 1990.

[39] S. Lang. Fundamentals of differential geometry. Graduate texts in mathematics. Springer, New York, 1999.

[40] J. Lapuyade-Lahorgue and F. Barbaresco. Radar detection using Siegel distance between autoregressive processes, application to HF and X-band radar. In Radar Conference, 2008. RADAR ’08, IEEE, Rome, May 2008.

[41] J. Lawson and Y. Lim. Monotonic properties of the least squares mean. Mathematische Annalen, 351(2):267–279, 2011.

[42] Y. Lim. Riemannian and Finsler structures of symmetric cones. Trends in Mathematics, 4(2):111–118, 2001.

[43] K. Nakamura. Geometric means of positive operators. Kyunpook Mathematical Journal, 49:167–181, 2009.

[44] M. P´alfia. The Riemann barycenter computation and means of several matrices. International Journal of Computational and Mathematical Sciences, 3(3):128–133, 2009.

[45] C. Qi. Numerical optimization methods on Riemannian manifolds. PhD thesis, Florida State University, College of arts and sciences, 2011.

[46] Q. Rentmeester and P.-A. Absil. Algorithm comparison for Karcher mean computation of rotation matrices and diffusion tensors. In European Signal Processing, 2011, EUSIPCO. 19th conference on, 2011.

[47] R. Vandebril, M. Van Barel, and N. Mastronardi. Matrix Computations and Semiseparable Matrices. Vol. 1, Linear Systems. The Johns Hopkins University Press, 2008.

[48] L. Yang, M. Arnaudon, and F. Barbaresco. Geometry of covariance matrices and computation of median. In AIP Conference Proceedings, 30th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, volume 1305, pages 479–486, 2011.

[49] Makarov, N., Poltoratski, A. Beurling-Malliavin theory for Toeplitz kernels, preprint, arXiv:math/0702497

[50] Beurling, A., Malliavin, P. On Fourier transforms of measures with compact support, Acta Math. 107 (1962), 291–302.

[51] Beurling, A., Malliavin, P. On the closure of characters and the zeros of entire functions, Acta Math. 118 (1967), 79-93.

[52] Bruna J., Olevskii, A., Ulanovskii, A. Completeness in L1 (R) of discrete translates and related questions for quasi-analytic classes, Rev. Mat. Iberoamericana 22 (2005), 1-16.

[53] Calderon, A., Zygmund, A. ´ On the existence of certain singular integrals, Acta Math. 88 (1952), 85–139.

[54] Coifman, R., Weiss, G. Extensions of Hardy spaces and their use in analysis, Bull. AMS 83 (1977), 569–645.

[55] De Branges, L. Hilbert spaces of entire functions. Prentice-Hall, Englewood Cliffs, NJ, 1968.

[56] Garnett, J. Bounded analytic functions. Academic Press, New York, 1981

[57] Gohberg, I., Krein, M. Theory and applications of Volterra operators in Hilbert space. AMS, Providence, RI, 1970

[58] Havin, V., Joricke, B. ¨ The uncertainty principle in harmonic analysis. Springer-Verlag, Berlin, 1994.

[59] Havin, V., Mashreghi, J. Admissible majorants for model subspaces of H2 ; I. Slow winding of the generating inner function, II. Fast winding of the generating inner function, Canad. J. Math. 55 (2003), 1231–1263, 1264–1301.
[60] Higgins, J. Completeness and basis properties of sets of special functions. Cambridge Univ. Press, Cambridge, 1977
[61] Hruschev S., Nikolskii, N., Pavlov, B. Unconditional bases of exponentials and of reproducing kernels, Lecture Notes in Math., Vol. 864, 214–335
[62] Hunt R., Muckenhoupt B., Wheeden R. Weighted norm inequalities for the conjugate functions and Hilbert transform, Trans. AMS 176 (1973), 227-251
[63] Kahane, J.-P. Travaux de Beurling et Malliavin, Seminaire Bourbaki. Expos´es 223 ´a 228, 1962
[64] Khabibullin, B. Completeness of exponential systems and uniqueness sets. Bashkir State Univ. Press, Ufa, 2006
[65] Koosis, P. Introduction to Hp spaces. Cambridge Univ. Press, Cambridge, 1980
[66] Koosis, P. The logarithmic integral, Vol. I & II. Cambridge Univ. Press, Cambridge, 1988
[67] Koosis, P. Lecons sur le Theorem de Beurling et Malliavin. Les Publications CRM, Montreal, 1996
[68] Koosis, P. Kargaev's proof of Beurling's lemma (unpublished manuscript)
[69] Levin, B. Distribution of zeros of entire functions. AMS, Providence, RI, 1980
[70] Levinson, N. Gap and density theorems. AMS, New York, 1940
[71] Lyubarskii, Ju., Seip, K. Complete interpolating sequences for Paley-Wiener spaces and Muckenhoupt's (Ap) condition, Rev. Mat. Iberoamericana 13 (1997), 361–376
[72] Makarov, N., Poltoratski, A. Meromorphic inner functions, Toeplitz kernels, and the uncertainty principle, in Perspectives in Analysis, Springer Verlag, Berlin, 2005, 185–252
[73] Mashreghi, J., Nazarov, F., Havin, V. Beurling-Malliavin multiplier theorem: the seventh proof, St. Petersburg Math. J. 17 (2006), 699–744
[74] Nazarov, F. The Beurling lemma via the Bellman function (unpublished manuscript)
[75] Nikolskii, N.K. Bases of exponentials and the values of reproducing kernels, Dokl. Acad. Nauk SSSR 252, (1980), 1316–1320
[76] Nikolskii, N. Operators, functions, and systems: an easy reading, Vol. I & II. AMS, Providence, RI, 2002
[77] Ortega-Cedra, J., Seip, K. ´ Fourier frames, Annals of Math. 155 (2002), 789–806
[78] Paley, R., Wiener, N. Fourier transform in the complex domains. AMS, New York, 1934
[79] Pavlov, B. The basis property of a system of exponentials and the condition of Muckenhoupt, Dokl. Acad. Nauk SSSR 247, (1979), 37-40
[80] Redheffer, R. Completeness of sets of complex exponentials, Advances in Math. 24, 1977, 1-62
[81] Schwartz, L. Etudes des sommes d'exponentielles r´eelles. ´ Hermann, Paris, 1943
[82] Seip, K. Interpolation and sampling in spaces of analytic functions, AMS, 2004
[83] Treil S., Volberg, A. Embedding theorems for invariant subspaces of inverse shift, Proceedings of LOMI Seminars, 149 (1986), 38–51
[84] A. POLTORATSK, SPECTRAL GAPS FOR SETS AND MEASURES, arXiv:0908.2079v2 [math.CV]20 Aug 2009.

[85] Alexandrov, A. Isometric embeddings of coinvariant subspaces of the shift operator, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) 232 (1996), Issled. po Linein Oper. i Teor. Funktsii. 24, 5–15, 213; translation in J. Math. Sci. (New York) 92 (1998), no. 1, 3543–3549
[86] Benedicks, M. The support of functions and distributions with a spectral gap, Math. Scand., 55 (1984), 285–309
[87] Beurling, A. On quasianalyticity and general distributions, Mimeographed lecture notes, Summer institute, Stanford University (1961)
[88] Borichev, A. and Sodin, M.Private communications
[89] Cornu, F. and Jancovici, B. On the two-dimensional Coulomb gas, Journal of statistical physics, 49:1-21-2, 33-56, Springer, 1987.
[90] Dym, H. On the span of trigonometric sums in weighted L 2 spaces, Linear and Complex Analysis Problem Book 3, Part II, Lecture Notes in Math., Springer, 1994, 87 – 88
[91] Dym H, McKean H.P. Gaussian processes, function theory and the inverse spectral problem Academic Press, New York, 1976
[92] Kerov, S. V. Equilibrium and orthogonal polynomials, Algebra i Analiz, 12:6 (2000), 224-237
[93] Krein, M. G. On an interpolation problem of A. N. Kolmogorov, Dokl. Akad. Nauk SSSR 46 (1945), 306–309 (Russian).
[94] Krein, M. G. On a basic approximation problem of the theory of extrapolation and filtration of stationary random processes, Doklady Akad. Nauk SSSR (N.S.) 94, (1954), 13–16 (Russian).
[95] Levin, B. Lectures on entire functions AMS, Providence, RI, 1996
[96] Mitkovski, M. and Poltoratski, A. Polya sequences, Toeplitz kernels and gap theorems, preprint, arXiv:0903.4499
[97] Nienhuis, B. Coulomb gas formulation of two-dimensional phase transitions, Phase transitions and critical phenomena, vol. 11, C. Domb and J.L. Lebowitz, eds. (Academic, 1987.)
[98] Nikolskii, N. K. Treatise on the shift operator, Springer-Verlaag, Berlin (1986)
[99] Poltoratski, A. Kreˇın's spectral shift and perturbations of spectra of rank one, Algebra i Analiz, 10 (1998), no. 5, 143–183; translation in St. Petersburg Math. J. 10 (1999), no. 5, 833–859
[100] Poltoratski, A. and Sarason, D. Aleksandrov-Clark measures, Recent advances in operator-related function theory, 1–14, Contemp. Math., 393, Amer. Math. Soc., Providence, RI, 2006
[101] Samaj L. The statistical mechanics of the classical two-dimensional Coulomb gas is exactly solved, J. Phys. A36:5913-5920, 2003
[102] Simon, B. Spectral analysis of rank one perturbations and applications, Mathematical quantum theory. II. Schroedinger operators (Vancouver, BC, 1993), 109–149, CRM Proc. Lecture Notes, 8, Amer. Math. Soc., Providence, RI, 1995
[103] MICHAEL T. LACEY, ERIC T. SAWYER, CHUN-YEN SHEN, AND IGNACIO URIARTE-TUERO, TWO WEIGHT INEQUALITY FOR THE HILBERT TRANSFORM: A REAL VARIABLE CHARACTERIZATION, I, arXiv:1201.4319v11 [math.CA]1 Mar 2014

[104] Guy David and Jean-Lin Journé, A boundedness criterion for generalized Calderón-Zygmund operators, Ann. of Math. (2) 120 (1984), no. 2, 371–397. MR763911 (85k:42041)
[105] V. P. Havin and N. K. Nikolski (eds.), Linear and complex analysis. Problem book 3. Part I, Lecture Notes in Mathematics, vol. 1573, Springer-Verlag, Berlin, 1994. MR1334345 (96c:00001a)
[106] Tuomas P. Hytönen, The sharp weighted bound for general Calderón-Zygmund operators, Ann. of Math. (2) 175 (2012), no. 3, 1473–1506.
[107] Tuomas P. Hytönen, Michael T. Lacey, Henri Martikainen, Tuomas Orponen, Maria Carmen Reguera, Eric T. Sawyer, and Ignacio Uriarte-Tuero, Weak and strong type estimates for maximal truncations of CalderónZygmund operators on Ap weighted spaces, J. Anal. Math. 118 (2012), 177–220.
[108] Michael T. Lacey, Two Weight Inequality for the Hilbert Transform: A Real Variable Characterization, II, Submitted (2013), available at http://www.arxiv.org/abs/1301.4663.
[109] Michael T. Lacey, Stefanie Petermichl, and Maria Carmen Reguera, Sharp A2 inequality for Haar shift operators, Math. Ann. 348 (2010), no. 1, 127–141. MR2657437
[110] Michael Lacey, Eric T. Sawyer, and Ignacio Uriarte-Tuero, A characterization of two weight norm inequalities for maximal singular integrals with one doubling measure, Anal. PDE 5 (2012), no. 1, 1–60.
[111] Michael T. Lacey, Eric T. Sawyer, and Ignacio Uriarte-Tuero, A Two Weight Inequality for the Hilbert transform Assuming an Energy Hypothesis, J Funct Anal 263 (2012), 305–363.
[112] Michael T. Lacey, Eric T. Sawyer, Chun-Yun Shen, and Ignacio Uriarte-Tuero, The two weight inequality for the Hilbert transform, coronas, and energy conditions (2011), available at http://www.arXiv.org/abs/1108.2319.
[113] Benjamin Muckenhoupt, Weighted norm inequalities for the Hardy maximal function, Trans. Amer. Math. Soc. 165 (1972), 207–226. MR0293384 (45 #2461)
[114] , Hardy's inequality with weights, Studia Math. 44 (1972), 31–38. Collection of articles honoring the completion by Antoni Zygmund of 50 years of scientific activity, I. MR0311856 (47 #418)
[115] F. Nazarov, A counterexample to Sarason's conjecture, Preprint, MSU (1997), available at http://www.math.msu.edu/~fedja/prepr.html.
[116] F. Nazarov, S. Treil, and A. Volberg, Cauchy integral and Calderón-Zygmund operators on nonhomogeneous spaces, Internat. Math. Res. Notices 15 (1997), 703–726.
[117] , Accretive system Tb-theorems on nonhomogeneous spaces, Duke Math. J. 113 (2002), no. 2, 259– 312.
[118] , The Tb-theorem on non-homogeneous spaces, Acta Math. 190 (2003), no. 2, 151–239.
[119] F. Nazarov, S. Treil, A. Volberg, Two weight estimate for the Hilbert transform and Corona decomposition for non-doubling measures (2004), available at http://arxiv.org/abs/1003.1596.
[120] , Two weight inequalities for individual Haar multipliers and other well localized operators, Math. Res. Lett. 15 (2008), no. 3, 583–597. MR2407233 (2009e:42031)

[121] F. Nazarov and A. Volberg, The Bellman function, the two-weight Hilbert transform, and embeddings of the model spaces Kθ, J. Anal. Math. 87 (2002), 385–414. Dedicated to the memory of Thomas H. Wolff.
[122] Nikolai Nikolski and S. Treil, Linear resolvent growth of rank one perturbation of a unitary operator does not imply its similarity to a normal operator, J. Anal. Math. 87 (2002), 415–431. Dedicated to the memory of Thomas H. Wolff. MR1945291 (2004e:47004)
[123] Carlos Pérez, S. Treil, and A. Volberg, On A2 conjecture and Corona decomposition of weights (2010), available at http://arxiv.org/abs/1006.2630.
[124] Sharp weighted estimates for dyadic shifts and the A2 conjecture (2010), available at http://arxiv.org/abs/1010.0755.
[125] Maria Carmen Reguera and James Scurry, On joint estimates for maximal functions and singular integrals on weighted spaces, Proc. Amer. Math. Soc. 141 (2013), no. 5, 1705–1717.
[126] Donald Sarason, Exposed points in H1 . II, Topics in operator theory: Ernst D. Hellinger memorial volume, Oper. Theory Adv. Appl., vol. 48, Birkhäuser, Basel, 1990, pp. 333–347. MR1207406 (94a:46031)
[127] Eric T. Sawyer, A characterization of a two-weight norm inequality for maximal operators, Studia Math. 75 (1982), no. 1, 1–11. MR676801 (84i:42032)
[128] A characterization of two weight norm inequalities for fractional and Poisson integrals, Trans. Amer. Math. Soc. 308 (1988), no. 2, 533–545. MR930072 (89d:26009)
[129] E. Sawyer and R. L. Wheeden, Weighted inequalities for fractional integrals on Euclidean and homogeneous spaces, Amer. J. Math. 114 (1992), no. 4, 813–874. MR1175693 (94i:42024)
[130] A. Volberg, Calderón-Zygmund capacities and operators on nonhomogeneous spaces, CBMS Regional Conference Series in Mathematics, vol. 100, Published for the Conference Board of the Mathematical Sciences, Washington, DC, 2003.
[131] ANTONIO LINERO AND ANDREW ROSALSKY, On the Toeplitz Lemma, Convergence in Probability, and Mean Convergence, Stochastic Analysis and Applications, 31: 684–694, 2013.
[132] von Bahr, B., and Esseen, C.-G. 1965. Inequalities for the rth absolute moment of a sum of random variables, $1 \leq r \leq 2$. Ann. Math. Stat. 36:229–303.
[133] Bartle, R.G. 1976. The Elements of Real Analysis. 2nd ed. John Wiley, New York.
[134] Chow, Y.S., and Teicher, H. 1997. Probability Theory: Independence, Interchangeability, Martingales. 3rd ed. Springer-Verlag, New York.
[135] Curtiss, J.H. 1942. A note on the theory of moment generating functions. Ann. Math. Stat. 13:430–433.
[136] Dugué, D. 1957. Sur la convergence stochastique au sens de Cesàro et sur des diffèrences importantes entre la convergence presque certaine et les convergences en probabilité et presque completés. Sankhya¯ 18:127–138.
[137] Jain, N.C., and Marcus, M.B. 1975. Integrability of infinite sums of independent vectorvalued random variables. Trans. Amer. Math. Soc. 212:1–36.
[138] Loève, M. 1977. Probability Theory I. 4th ed. Springer-Verlag, New York.
[139] Meyer, P.A. 1966. Probability and Potentials. Blaisdell Publishing Company; Ginn and Company, Waltham, MA.

[140] Rosenthal, H.P. 1970. On the subspaces of Lp p > 2spanned by sequences of independent random variables. Israel J. Math. 8:273–303.
[141] Taylor, R.L. 1978. Stochastic Convergence of Weighted Sums of Random Elements in Linear Spaces. Lecture Notes in Mathematics, No. 672. Springer-Verlag, Berlin.
[142] Frank Hansen, Regular operator mappings and multivariate geometric means, arXiv:1403.3781v4 [math.FA]22 Jul 2014
[143] C. Davis. A Schwarz inequality for convex operator functions. Proc. Amer. Math. Soc., 8:42–44, 1957.
[144] E. Effros and F. Hansen. Non-commutative perspectives. Annals of Functional Analysis, 5(2):74–79, 2014.
[145] E.G. Effros. A matrix convexity approach to some celebrated quantum inequalities. Proc. Natl. Acad. Sci. USA, 106:1006–1008, 2009.
[146] F. Hansen. Means and concave products of positive semi-definite matrices. Math. Ann., 264:119–128, 1983.
[147] F. Hansen and G.K. Pedersen Jensen's operator inequality. Bull. London Math. Soc., 35:553–564, 2003.
[148] H. Kosaki. Geometric mean of several operators. 1984.
[149] F. Kubo and T. Ando. Means of positive linear operators. Math. Ann., 246:205–224, 1980.
[150] J. Lawson and Y. Lim. Karcher means and Karcher equations of positive definite operators. Trans. Amer. Math. Soc., Series B, 1:1–22, 2014.
[151] Y. Lim and M. Pálfia. Matrix power means and the Karcher mean. Journal of Functional Analysis, 262:1498–1514, 2012.
[152] M. Sagae and K. Tanabe. Upper and lower bounds for the arithmeticgeometric-harmonic means of positive definite matrices. Linear and Multilinear Algebra, 37:279–282, 1994.
[153] Jiyanglin Li and Ze-Chun Hu, Toeplitz Lemma, Complete Convergence and Complete Moment Convergence, arXiv:1406.0768v3 [math.PR]23 Jan 2015
[154] Baum, L. E., Katz, M. (1965). Convergence rates in the law of large numbers. Trans. Amer. Math. Soc. 120:108-123.
[155] Chen, P. (2006). Complete moment convergence for sequences of independent random elements in Banach spaces. Stoch. Anal. Appl. 24:999-1010.
[156] Chen, Y.-Y., Zhang, L.-X. (2010). Second moment convergence rates for uniform empirical processes. J. Inequal. Appl. 2010:972324, 9 pages (doi: 10.1155/2010/972324).
[157] Chen, P., Hu. T.-C., Volodin, A. (2007). Limiting behavior of moving average processes under negative association. Teor. Imovir. ta Matem. Statyst. 7:154-166.
[158] Chen, P., Sung S. H. (2014). A Baum-Katz theorem for i.i.d. random variables with higher order moments. Statis. Probab. Lett. 94:63-68.
[159] Chow, Y. (1988). On the rate of moment convergence of sample sums and extremes. Bull. Inst. Math. Acad. Sin. 16:177-201.
[160] Erd¨os, P. (1949). On a theorem of Hsu and Robbins. Ann. Math. Statist. 20:286-291.
[161] Erd¨os, P. (1950). Remark on my paper "On a theorem of Hsu and Robbins". Ann. Math. Statist. 21:138.
[162] Fu, K.-A., Zhang, L.-X. (2008). On the moment convergence rates of LIL in Hilbert space. Math. Comput. Modelling 47:153-167.

[163] Guo M. (2013). On complete moment convergence of weighted sums for arrays of row-wise negatively associated random variables. Stochastics An International Journal of Probability and Stochastic Processes DOI: 10.1080/17442508.2013.801971.
[164] Guo, M.-L., Xu, J. (2006). Moment complete convergence for arrays of B-valued random elements (in Chinese). Appl. Math. J. Chinese Univ. Ser. A 21:432-438.
[165] Gut, A. (1978). Marcinkiwicz laws and convergence rates in the law of large numbers for random variables with multidimensionial indices. Ann. Probab. 6:469-482.
[166] Gut, A. (1980). Convergence rates for probabilities of moderate deviations for sums of random variables with multidimensional indices. Ann. Probab. 8:298-313.
[167] Gut, A., Stadtm¨uller, U. (2011). An intermediate Baum-Katz theorem. Statis. Probab. Lett. 81:1486-1492.
[168] Hsu, P., Robbins, H. (1947). Complete convergence and the law of large numbers. Proc. Natl. Acad. Sci. USA 33:25-31.
[169] Jiang, Y., Zhang, L.-X. (2006). Precise asymptotics in complete moment convergence of i.i.d. random variables (in Chinese). Acta. Math. Sci. Ser. A Chin. Ed. 26:917-925.
[170] Kim T.-S., Ko M.-H., Choi Y.-K. (2008). Complete moment convergence of moving average processes with dependent innovations. J. Korean Math. Soc. 45:355-365.
[171] Lanzinger, H. (1998). A Baum-Katz theorem for random variables under exponential moment conditions. Statis. Probab. Lett. 39:89-95.
[172] Li, D., Rao, M.B., Jiang, T., Wang X. (1995). Complete convergence and almost sure convergence of weighted sums of random variables. J. Theoret. Probab. 8:49-76.
[173] Li Y.-X. (2006). Precise asymptotics in complete moment convergence of moving-average processes. Statis. Probab. Lett. 76:1305-1315.
[174] Li, Y.-X., Zhang, L.-X. (2004). Complete moment convergence of moving-average processes under dependence assumptions. Statis. Probab. Lett. 70:191-197.
[175] Liang, H. Y., Li, D. L., Rosalsky, A. (2010). Complete moment and integral convergence for sums of negatively associated random variables. Acta Math. Sin. Engl. Ser. 26:419-432.
[176] Liu, W., Lin Z. (2006). Precise asymptotics for a new kind of complete moment convergence. Statis. Probab. Lett. 76:1787-1799.
[177] Qiu, D., Chen P. (2014). Complete moment convergence for i.i.d. random variables. Statis. Probab. Lett. 91:76-82.
[178] Qiu, D., Urmeneta, H., Volodini, A. (2014). Complete moment convergence for weighted sums of sequences of independent random elements in Banach spaces. Collect. Math. 65:155-167.
[179] Rosalsky, A., Thanh, L. V., Volodin, A. (2006). On complete convergence in mean of normed sums of independent random elements in Banach spaces. Stoch. Anal. Appl. 24:23-35.
[180] Sung, S. H. (2007). Complete convergence for weighted sums of random variables. Statis. Probab. Lett. 77:303-311.
[181] Sung, S.H., Volodin, A. (2006). On the rate of complete convergence for weighted sums of arrays of random elements. J. Korean Math. Soc. 43:815-828.
[182] Wang, D., Su, C. (2004). Moment complete convergence for sequences of B-valued iid random elements (in Chinese). Acta Math. Appl. Sin. 27:440-448.

[183] Wang, Y., Zhong, S. M., Wang, D. C. (2005). Moment complete convergence for sums of i.i.d. random elements in Banach space (in Chinese). Dianzi Keji Daxue Xuebao 34:410- 412.
[184] Wang, D., Zhao, W. (2006). Moment complete convergence for sums of a sequence of NA random variables (in Chinese). Appl. Math. J. Chinese Univ. Ser. A 21:445-450.
[185] Ye, F., Zhu, D. J. (2007). Moment complete convergence for B-valued random elements (in Chinese). J. Hefei Univ. Technol. Nat. Sci. 30:525-528.
[186] Ye, J., Zhang L.-X., Pang, T.-X. (2007). Precise rates in the law of logarithm for the moment convergence of i.i.d. random variables. J. Math. Anal. Appl. 327:695-714.
[187] Zhou X. (2010). Complete moment convergence of moving average processes under '- mixing assumptions. Statis. Probab. Lett. 80:285-292.
[188] Zhao Y., Tao J. (2008). Precise asymptotics in complete moment convergence for selfnormalized sums. Comptut. Math. Appl. 56:1779-1786.
[189] Elham Nobari, A monotone geometric mean for a class of Toeplitz matrices, Linear Algebra and its Applications 511 (2016) 1–18
[190] D.A. Bini, B. Iannazzo, Computing the Karcher mean of symmetric positive definite matrices, Linear Algebra Appl. 438(4) (2013) 1700–1710.
[191] D.A. Bini, B. Iannazzo, B. Jeuris, R. Vandebril, Geometric mean of structured matrices, BIT 54(1) (2014) 55–83.
[192] A. Böttcher, B. Silbermann, Analysis of Toeplitz Operator, second edition, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2006, prepared jointly with Alexie Karlovich.
[193] J.B. Conway, A Course in Functional Analysis, second edition, Grad. Texts in Math., vol.96, Springer-Verlag, New York, 1990.
[194] U. Grenander, G. Szegö, Toeplitz Forms and Their Applications, Univ. of California Press, Berkeley, 1985.
[195] Y. Rathi, O. Michailovich, A. Tannenbaum, Segmenting images on the tensor manifold, in: Pro-ceedings of Computer Vision and Pattern Recognition, 2007, pp.1–8.
[196] B. JEURIS AND R. VANDEBRIL, THE K¨AHLER MEAN OF BLOCK-TOEPLITZ MATRICES WITH TOEPLITZ STRUCTURED BLOCKS, SIAM J. MATRIX ANAL. APPL. Vol. 37, No. 3, pp. 1151–1175
[197] M. Arnaudon, F. Barbaresco, and Y. Le, Medians and means in Riemannian geometry: Existence, uniqueness and computation, in Matrix Information Geometry, Springer, Berlin, 2012, pp. 169–197.
[198] M. Arnaudon, F. Barbaresco, and Y. Le, Riemannian medians and means with applications to radar signal processing, IEEE J. Sel. Top. Signal Process., 7 (2013), pp. 595–604.
[199] F. Barbaresco, Super resolution spectrum analysis regularization: Burg, Capon and agoantagonistic algorithms, in EUSIPCO-96, Trieste, Italy, Lint, Trieste, 1996, pp. 2005–2008.
[200] F. Barbaresco, Innovative tools for radar signal processing based on Cartan’s geometry of SPD matrices and information geometry, in RADAR ’08, IEEE, Piscataway, NJ, 2008, pp. 1–6.

[201] F. Barbaresco, Robust statistical radar processing in Fr´echet metric space: OS-HDR-CFAR and OS-STAP processing in Siegel homogeneous bounded domains, in IRS '11, Leipzig, 2011, IEEE, Piscataway, NJ, pp. 639–644.
[202] F. Barbaresco, Information geometry of covariance matrix: Cartan-Siegel homogeneous bounded domains, Mostow/Berger fibration and Fr´echet median, in Matrix Information Geometry, Springer, Berlin, 2012, pp. 199–255.
[203] F. Barbaresco, Information geometry manifold of Toeplitz Hermitian positive definite covariance matrices: Mostow/Berger fibration and Berezin quantization of Cartan-Siegel domains, Int. J. Emerg. Trends Signal Process., 1 (2013), pp. 1–11.
[204] F. Barbaresco, Eidetic reduction of information geometry through Legendre duality of Koszul characteristic function and entropy: From Massieu–Duhem potentials to geometric Souriau temperature and Balian quantum Fisher metric, in Geometric Theory of Information, Cham, Switzerland, Springer, 2014, pp. 141–217.
[205] F. Barbaresco, Koszul information geometry and Souriau geometric temperature/capacity of Lie group thermodynamics, Entropy, 16 (2014), pp. 4521–4565.
[206] G. Bassanelli, On horospheres and holomorphic endomorfisms of the Siegel disc, Rend. Semin. Mat. Univ. Padova, 70 (1983), pp. 147–165.
[207] M. Baˇc´ak, Computing medians and means in Hadamard spaces, SIAM J. Optim., 24 (2014), pp. 1542–1566.
[208] A. B¨ottcher and S. Grudsky, Spectral Properties of Banded Toeplitz Matrices, SIAM, Philadelphia, 2005.
[209] J. Burg, Maximum Entropy Spectral Analysis, IEEE Press, New York, 1978, pp. 34-41.
[210] D. Damanik, A. Pushnitski, and B. Simon, The analytic theory of matrix orthogonal polynomials, Surv. Approx. Theory, 4 (2008), pp. 1–85.
[211] J. Dehaene, Continuous-Time Matrix Algorithms, Systolic Algorithms and Adaptive Neural Networks, Ph.D. thesis, Department of Electrical Engineering, KU Leuven, Leuven, Belgium, 1995.
[212] P. Delsarte, Y. Genin, and Y. Kamp, Orthogonal polynomial matrices on the unit circle, IEEE Trans. Circuits Syst., 25 (1978), pp. 149–160.
[213] Ph. Delsarte, Y. Genin, and Y. Kamp, Parametrization of positive definite block-Toeplitz systems, SIAM J. Appl. Math., 36 (1979), pp. 34–46.
[214] H. Dette andW. Studden, The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis, Wiley, New York, 1997.
[215] H. Dette and J. Wagener, Matrix measures on the unit circle, moment spaces, orthogonal polynomials and the Geronimus relations, Linear Algebra Appl., 432 (2010), pp. 1609-1626.
[216] R. Ferreira, J. Xavier, J. Costeira, and V. Barroso, Newton method for Riemannian centroid computation in naturally reductive homogeneous spaces, in IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, Piscataway, NJ, 2006.
[217] T. Franzoni and E. Vesentini, Holomorphic Maps and Invariant Distances, Math. Stud. 40, North-Holland, 1980.
[218] B. Fritzsche and B. Kirstein, An extension problem for non-negative Hermitian block Toeplitz matrices, Math. Nachr., 131 (1987), pp. 287–297.

[219] M. Giles, An Extended Collection of Matrix Derivative Results for Forward and Reverse Mode Algorithmic Differentiation, Technical report, Oxford University Computing Laboratory, Oxford, England, 2008.
[220] M. Humet and M. Van Barel, Algorithms for the Geronimus transformation for orthogonal polynomials on the unit circle, J. Comput. Appl. Math., 267 (2014), pp. 195–217.
[221] A. Jakobsson, S. Marple, and P. Stoica, Computationally efficient two-dimensional Capon spectrum analysis, IEEE Trans. Signal Process., 48 (2000), pp. 2651–2661.
[222] R. Kanhouche, A Modified Burg Algorithm Equivalent in Results to Levinson Algorithm, preprint, arXiv:0309384, 2003.
[223] R. Kanhouche, Mathematical Methods in Signal Processing for Spectral Estimation, Ph.D. thesis, Centre de Math´ematiques et de Leurs Applications (CMLA), ´Ecole normale sup´erieure de Cachan, Cachan, France, 2006.
[224] M. Kre˘in and A. Nudel'man, The Markov Moment Problem and Extremal Problems, Transl. Math. Monogr. 50, AMS, Providence, RI, 1977.
[225] Y. Lim and M. P´alfia, Weighted deterministic walks for the least squares mean on Hadamard spaces, Bull. Lond. Math. Soc., 46 (2014), pp. 561–570.
[226] S. L. Marple, Digital Spectral Analysis with Applications, Prentice-Hall, Englewood Cliffs, 1980.
[227] M. P´alfia, The Riemannian barycenter computation and means of several matrices, Int. J. Comput. Math. Sci., 3 (2009), pp. 128–133.
[228] C. Siegel, Symplectic geometry, Amer. J. Math., 65 (1943), pp. 1–86.
[229] J. R. Silvester, Determinants of block matrices, Math. Gazette, 84 (2000), pp 460–467.
[230] A. Sinap and W. Van Assche, Orthogonal matrix polynomials and applications, J. Comput. Appl. Math., 66 (1996), pp. 27–52.
[231] K.-T. Sturm, Probability measures on metric spaces of nonpositive curvature, in Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces, P. A. et al., ed., American Mathematical Society, Contemp. Math. 338, 2003, pp. 357–390.
[232] G. Szeg˝o, Orthogonal Polynomials, AMS, Providence, RI, 1939.
[233] C. Therrien, Relations between 2-D and multichannel linear prediction, IEEE Trans. Acoust. Speech Signal Process., 29 (1981), pp. 454–456.
[234] W. Trench, An algorithm for the inversion of finite Toeplitz matrices, J. Soc. Ind. Appl. Math., 12 (1964), pp. 515–522.
[235] S. Verblunsky, On positive harmonic functions: A contribution to the algebra of Fourier series, Proc. Lond. Math. Soc. (2), 38 (1935), pp. 125–157.
[236] S. Verblunsky, On positive harmonic functions (second paper), Proc. Lond. Math. Soc. (2), 40 (1936), pp. 290–320.
[237] P. Whittle, On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix, Biometrika, 50 (1963), pp. 129–134.
[238] N. Wiener, The Wiener RMS (Root Mean Square) Error Criterion in Filter Design and Prediction, 1st ed., MIT Press, 1949, Cambridge, MA, pp. 129–148.
[239] R. Wiggins and E. Robinson, Recursive solution to the multichannel fitering problem, J. Geophys. Res., 70 (2012), pp. 1885–1891.

[240] L. Yang, Medians of Probability Measures in Riemannian Manifolds and Applications to Radar Target Detection, Ph.D. thesis, Universit´e de Poitiers, Poitiers, France, 2011.
[241] Elham Nobari, Bijan Ahmadi Kakavandi, A geometric mean for Toeplitz and Toeplitz-block block-Toeplitz matrices, Linear Algebra and its Applications 548 (2018) 189–202.
[242] G.B. Folland, Real Analysis. Modern Techniques and Their Applications, second edition, Pure and Applied Mathematics (New York), A Wiley-Interscience Publication, John Wiley & Sons, Inc., New York, 1999.
[243] P. Fletcher, S. Joshi, Riemannian geometry for the statistical analysis of diffusion tensor data, Signal Process. 87 (2007) 250–262.
[244] F. Nielsen, R. Bhatia (Eds.), Matrix Information Geometry, Springer, Heidelberg, 2013.
[245] E. Andruchow, E. Chiumiento and G. Larotonda, Geometric significance of Toeplitz kernels, arXiv:1608.05737v1 [math.FA]19 Aug 2016
[246] E. Andruchow. Operators which are the difference of two projections. J. Math. Anal. Appl. 420 (2014), no. 2, 1634–1653
[247] E. Andruchow. The Grassmann manifold of a Hilbert space. Proceedings of the XIIth Dr. Antonio A. R. Monteiro Congress, 41–55, Univ. Nac. Sur Dep. Mat. Inst. Mat., Bah´ıa Blanca, 2014.
[248] E. Andruchow, E. Chiumiento, M. E. Di Iorio y Lucero. Essentially commuting projections. J. Funct. Anal. 268 (2015), no. 2, 336–362.
[249] E. Andruchow, G. Larotonda. Hopf-Rinow theorem in the Sato Grassmannian. J. Funct. Anal. 255 (2008), no. 7, 1692-1712.
[250] E. Andruchow, G. Larotonda. Lagrangian Grassmanian in infinite dimension. J. Geom. Phys. 59 (2009), no. 3, 306-320.
[251] J. Antezana, G. Larotonda, A. Varela. Optimal paths for symmetric actions in the unitary group. Comm. Math. Phys. 328 (2014), no. 2, 481–497.
[252] D. Beltit¸˘a, T. S. Ratiu, A. B. Tumpach. The restricted Grassmannian, Banach Lie-Poisson spaces, and coadjoint orbits. J. Funct. Anal. 247 (2007), no. 1, 138–168.
[253] A. B¨ottcher, A. Karlovich, B. Silbermann. Generalized Krein algebras and asymptotics of Toeplitz determinants. Methods Funct. Anal. Topology 13 (2007), no. 3, 236-261.
[254] A. B¨ottcher, B. Silbermann. Analysis of Toeplitz operators. Springer-Verlag, Berlin, 1990.
[255] J. Bourgain, G. Kozma. One cannot hear the winding number. J. Eur. Math. Soc. (JEMS) 9 (2007), no. 4, 637–658.
[256] A. L. Carey. Some homogeneous spaces and representations of the Hilbert Lie group U(H)2. Rev. Roumaine Math. Pures Appl. 30 (1985), no. 7, 505–520.
[257] D. N. Clark, On the point spectrum of a Toeplitz operator. Trans. Amer. Math. Soc. 126 (1967), 251–266.
[258] G. Corach, H. Porta, L. Recht. The geometry of spaces of projections in C∗-algebras. Adv. Math. 101 (1993), no. 1, 59–77.
[259] C. Davis. Separation of two linear subspaces. Acta Sci. Math. Szeged 19 (1958), 172–187.
[260] J. Dixmier. Position relative de deux veri´et´es lin´eaires ferm´ees dans un espace de Hilbert. (French) Revue Sci. 86 (1948), 387–399.

[261] R. G. Douglas. Banach algebra techniques in operator theory. Second edition. Graduate Texts in Mathematics, 179. Springer-Verlag, New York, 1998.
[262] T. Goli´nski, A. Odzijewicz. Hierarchy of Hamilton equations on Banach Lie-Poisson spaces related to restricted Grassmannian, J. Funct. Anal. 258 (2010), no. 10, 3266-3294.
[263] P.R. Halmos. Two subspaces. Trans. Amer. Math. Soc. 144 1969 381–389.
[264] A. Hartmann, M. Mitkovski. Kernels of Toeplitz operators. Contemp. Math., to appear.
[265] S. V. Hruˇsˇchev, N. K. Nikol'skiˇı, B. S. Pavlov. Unconditional basis of exponentials and of reproducing kernels. Lecture Notes in Math. 864 (1981), 214–335.
[266] Z. V. Kovarik. Manifolds of linear involutions. Linear Algebra Appl. 24 (1979), 271–287.
[267] M. Lee, D. Sarason. The spectra of some Toeplitz operators. J. Math. Anal. Appl. 33 (1971), 529–543.
[268] M. Pavlovi´c. Introduction to function spaces on the disk. Posebna Izdanja [Special Editions], 20. Matematiki Institut SANU, Belgrade, 2004.
[269] H. Porta, L. Recht. Minimality of geodesics in Grassmann manifolds, Proc. Amer. Math. Soc. 100 (1987), 464–466.
[270] A. Pressley, G. Segal. Loop groups. Oxford Mathematical Monographs. Oxford Science Publications. The Clarendon Press, Oxford University Press, New York, 1986.
[271] D. Sarason. Algebras of functions on the unit circle. Bull. Amer. Math. Soc. 79 (1973), no. 2, 286–299.
[272] W. Seidel. On the distribution of values of bounded analytic functions. Trans. Amer. Math. Soc. 36 (1934), 201–226.
[273] G. Segal, G. Wilson. Loop groups and equations of KdV type. Inst. Hautes tudes Sci. Publ. Math. No. 61 (1985), 5–65.
[274] G. Segal, G. Wilson. Loop groups and equations of KdV type. Surveys in differential geometry: integral systems [integrable systems], 403–466, Surv. Differ. Geom., IV, Int. Press, Boston, MA, 1998.
[275] A. B. Tumpach. Hyperkh¨aler structures and infinite-dimensional Grassmannians. J. Funct. Anal. 243 (2007), no. 1, 158–206.
[276] A. Zygmund. Trigonometric series. Vol. I, II. Third edition. With a foreword by Robert A. Fefferman. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2002.
[277] Emmanuel Fricain, Andreas Hartmann, William T. Ross. Multipliers between model spaces. Studia Mathematica, Instytut Matematyczny - Polska Akademii Nauk, 2018, 240, pp.177-191.
[278] P. R. Ahern and D. N. Clark. Radial limits and invariant subspaces. Amer. J. Math., 92:332–342, 1970.
[279] A. D. Baranov. Differentiation in the Branges spaces and embedding theorems. J. Math. Sci. (New York), 101(2):2881–2913, 2000. Nonlinear equations and mathematical analysis.
[280] M. C. Cˆamara, M. T. Malheiro, and J. R. Partington. Model spaces and Toeplitz kernels in reflexive Hardy space. Oper. Matrices, 10(1):127–148, 2016.

[281] J. Cima, A. Matheson, and W. Ross. The Cauchy transform, volume 125 of Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 2006.
[282] Bill Cohn. Carleson measures for functions orthogonal to invariant subspaces. Pacific J. Math., 103(2):347–364, 1982.
[283] William S. Cohn. Carleson measures and operators on star-invariant subspaces. J. Operator Theory, 15(1):181–202, 1986.
[284] William S. Cohn. A maximum principle for star invariant subspaces. Houston J. Math., 14(1):23–37, 1988.
[285] R. Bruce Crofoot. Multipliers between invariant subspaces of the backward shift. Pacific J. Math., 166(2):225–246, 1994.
[286] P. L. Duren. Theory of Hp spaces. Academic Press, New York, 1970.
[287] K. M. D′ yakonov.Moduli and arguments of analytic functions from subspaces in Hp that are invariant under the backward shift operator. Sibirsk. Mat. Zh., 31(6):64–79, 1990.
[288] Stephan Ramon Garcia, Robert T. W. Martin, and William T. Ross. Partial order on partial isometries. J. Operator Theory, 75(2):409–442, 2016.
[289] Stephan Ramon Garcia, Javad Mashreghi, and William T. Ross. Introduction to model spaces and their operators, volume 148 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2016.
[290] J. Garnett. Bounded analytic functions, volume 236 of Graduate Texts in Mathematics. Springer, New York, first edition, 2007.
[291] B. Ya. Levin. Lectures on entire functions, volume 150 of Translations of Mathematical Monographs. American Mathematical Society, Providence, RI, 1996. In collaboration with and with a preface by Yu. Lyubarskii, M. Sodin and V. Tkachenko, Translated from the Russian manuscript by Tkachenko.
[292] B. A. Lotto. Toeplitz operators on weighted Hardy spaces. In Function spaces (Edwardsville, IL, 1990), volume 136 of Lecture Notes in Pure and Appl. Math., pages 295–300. Dekker, New York, 1992.
[293] Yurii I. Lyubarskii and Kristian Seip.Weighted Paley-Wiener spaces. J. Amer. Math. Soc., 15(4):979–1006 (electronic), 2002.
[294] A. G. Poltoratski˘ı. Boundary behavior of pseudocontinuable functions. Algebra I Analiz, 5(2):189–210, 1993.
[295] W. Rudin. Functional analysis. McGraw-Hill Inc., New York, second edition, 1991.
[296] D. Sarason. Sub-Hardy Hilbert spaces in the unit disk. University of Arkansas Lecture Notes in the Mathematical Sciences, 10. John Wiley & Sons Inc., New York, 1994. A Wiley-Interscience Publication.
[297] Dan Timotin. On a preorder relation for contractions. Acta Sci. Math. (Szeged), 82(3-4):629–640, 2016.
[298] A. A. Yukhimenko. Canonical products generated by perturbations of an integer sequence and their asymptotic estimates. Izv. Ross. Akad. Nauk Ser. Mat., 74(5):205–224, 2010.
[299] A. Poltoratski, Toeplitz order, Journal of Functional Analysis 275 (2018) 660–697
[300] A. Baranov, D. Sarason, Quotient representations of inner functions, in: Recent Trends in Analysis, Proceedings of the Conference in Honor of Nikolai Nikolski, Theta Foundation, Bucharest, 2013, pp.35–46.

[301] Yu. Belov, Complementability of exponential systems, C. R. Math. 353(3) (March 2015) 215–218.
[302] D. Clark, One dimensional perturbations of restricted shifts, J. Anal. Math. 25 (1972) 169–191.
[303] K. Dyakonov, Zero sets and multiplier theorems for star-invariant subspaces, J. Anal. Math. 86 (2002) 247–269.
[304] M.G. Krein, On an extrapolation problem of A.N. Kolmogorov, Dokl. Akad. Nauk SSSR 46 (1945) 306–309 (Russian).
[305] M.T. Lacey, The two weight inequality for the Hilbert transform: a primer, Preprint.
[306] N. Makarov, A. Poltoratski, Two spectra theorem with uncertainty, Preprint.
[307] N. Makarov, A. Poltoratski, Etudes in spectral problems, Preprint.
[308] N. Makarov, A. Poltoratski, M. Sodin, Lectures on linear complex analysis, in preparation.
[309] A. Poltoratski, On the boundary behavior of pseudocontinuable functions, St. Petersburg Math. J. 5 (1994) 389–406.
[310] A. Poltoratski, Problem on completeness of exponentials, Ann. Math. 178 (2013) 983–1016.
[311] A. Poltoratski, Toeplitz Approach to Problems of the UncertaintyPrinciple, CBMS Series, AMS/NSF, 2015.
[312] M. Sodin, P. Yuditskii, Another approach to de Branges' theorem on weighted polynomial approx-imation, in: Proceedings of the Ashkelon Workshop on Complex Function Theory (1996), in: Israel Math. Conf. Proc., vol.11, Bar-Ilan Univ., Ramat Gan, 1997, pp.221–227.
[313] N. Wiener, Extrapolation, Interpolation, and Smoothing of Stationary Time Series, M.I.T. Press, 1949.
[314] L. Accardi, U.C. Ji, K. Saitˆo, Higher order multi-dimensional extensions of Ces`aro theorem, Infin. Dimens. Anal. Quantum Probab. Relat. Top. 18 (2015) 1550030, 14 pp.
[315] L. Accardi, A. Barhoumi and H. Ouerdiane: A Quantum approach to Laplace operators, Infin. Dimens. Anal. Quantum Probab. Relat. Top. 9 (2006), 215–248.
[316] L. Accardi, P. Gibilisco and I. V. Volovich: Yang-Mills gauge fields as harmonic functions for the L_evy Laplacian, Russian J. Math. Phys. 2 (1994), 235–250.
[317] L. Accardi, U. C. Ji and K. Saitˆo: Exotic Laplacians and associated stochastic processes, Infin. Dimens. Anal. Quantum Probab. Relat. Top. 12 (2009), 1–19.
[318] L. Accardi, U. C. Ji, K. Saitˆo: Exotic Laplacians and derivatives of white noise, Infin. Dimens. Anal. Quantum Probab. Relat. Top. 14 (2011), 1–14.
[319] L. Accardi, U. C. Ji and K. Saitˆo: The exotic (higher order L_evy) Laplacians generate the Markov processes given by distribution derivatives of white noise, Infin. Dimens. Anal. Quantum Probab. Relat. Top. 16 (2013), no. 3, 1350020.
[320] L. Accardi, H. Ouerdiane and O. G. Smolyanov: L_evy Laplacian acting on operators, Russian J. Math. Phys. 10 (2003), 359–380.
[321] L. Accardi, P. Roselli and O. G. Smolyanov: Brownian motion generated by the L_evy Laplacian, Math. Notes 54 (1993), 1174–1177.
[322] L. Accardi and O. G. Smolyanov: On Laplacians and traces, Confer. Sem. Mat. Univ. Bari No. 250 (1993).
[323] L. Accardi and O. G. Smolyanov: Classical and nonclassical L_evy Laplacians, Dokl. Math. 76 (2007), no. 3, 801–805.

[324] L. Accardi and O. G. Smolyanov: Generalized Levy Laplacians and Ces_aro means, Doklady Akademii Nauk 424 (2009), 583-587.
[325] D. M. Chung, U. C. Ji and K. Saitˆo: Cauchy problems associated with the L_evy Laplacian in white noise analysis, Infin. Dimens. Anal. Quantum Probab. Relat. Top. 2 (1999), 131–153.
[326] M. N. Feller: In_nite-dimensional elliptic equations and operators of L_evy type, Russian Math. Surveys 41 (1986), 119–170.
[327] G. M. Fikhtengol'ts: "A Course in Differential and Integral Calculus," Vol. 2 [in Russian], Moscow, 1959.
[328] U. C. Ji, N. Obata and H. Ouerdiane: Quantum L_evy Laplacian and associated heat equation, J. Funct. Anal. 249 (2007), 31–54.
[329] U. C. Ji and K. Saitˆo: A similarity between the Gross Laplacian and the L_evy Laplacian, Infin. Dimens. Anal. Quantum Probab. Relat. Top. 10 (2007), 261–276.
[330] H.-H. Kuo, N. Obata and K. Saitˆo: L_evy Laplacian of generalized functions on a nuclear space, J. Funct. Anal. 94 (1990), 74–92.
[331] P. L´evy: "Le, cons d'Analyse Fonctionnelle," Gauthier-Villars, Paris, 1922.
[332] P. L´evy: "Problemes Concrets d'Analyse Fonctionnelles," Gauthier-Villars, Paris, 1951.
[333] N. Obata: Quadratic quantum white noises and L_evy Laplacian, Nonlinear Anal. 47 (2001), 2437–2448.
[334] K. Saitˆo: A stochastic process generated by the L_evy Laplacian, Acta Appl. Math. 63 (2000), 363–373.
[335] B.J. Choi, J. Heo, U.C. Ji, Law of large numbers for weighted inductive means in a Hadamard space, Math. Nachr. 290 (2017) 2775–2787.
[336] A. Bozorgnia and M. B. Rao, Limit theorems for weighted sums of random elements in separable Banach spaces, J. Multivariate Anal. 9 (1979), 428–433.
[337] B. D. Choi and S. H. Sung, Almost sure convergence theorems of weighted sums of random variables, Stochastic Anal. Appl. 5 (1987), 365–377.
[338] C. Conde, Nonpositive curvature in $p$-Schatten class, J. Math. Anal. Appl. 356 (2009), 664–673.
[339] J. Cuzick, A strong law for weighted sums of i.i.d. random variables, J. Theoret. Probab. 8 (1995), 625–641.
[340] B. Y. Jing and H. Y. Liang, Strong limit theorems for weighted sums of negatively associated random variables, J. Theoret. Probab. 21 (2008), 890–909.
[341] A. McCarthy Charles, $cp$, Israel J. Math. 5 (1967), 249–271.
[342] M. Ordóñez Cabrera, Limit theorems for randomly weighted sums of random elements in normed linear spaces, J. Multivariate Anal. 25 (1988), 139–145.
[343] V. K. Rohatgi, Convergence of weighted sums of independent random variables, Proc. Cambridge Philos. Soc. 69 (1971), 305–307.
[344] R. Thrum, A remark on almost sure convergence of weighted sums, Probab. Theory Related Fields 75 (1987), 425–430.
[345] B. Jin Choi, U.C. Ji, Toeplitz Lemma in Geodesic Metric Space and Convergence of Inductive Means, J. Math. Anal. Appl. (2018), https://doi.org/10.1016/j.jmaa.2018.05.004
[346] G. Corach, H. Porta, L. Recht, Convexity of the geodesic distance on spaces of positive operators, Illinois J. Math. 38 (1994) 87–94.

[347] S. Kim, H. Lee, Y. Lim, A fixed point mean approximation to the Cartan barycenter of positive definite matrices, Linear Algebra Appl. 496 (2016) 420–437.
[348] J. Lawson, Y. Lim, Metric convexity of symmetric cones, Osaka J. Math. 44 (2007) 795–816.
[349] A.C. Thompson, On certain contraction mappings in a partially ordered vector space, Proc. Amer. Math. Soc. 14 (1963) 438–443.
[350] Shawgy Hussein and Saad Mohammed, Geometric Mean for Toeplitz-Block and Toeplitz Lemma in Geodesic Metric Space, Ph.D. Thesis Sudan University of Science and Technology, Sudan (2022).