The Kähler mean of Block-Toeplitz matrices with Toeplitz structured blocks

Ben Jeuris
Raf Vandebril

Report TW 660, May 2015

KU Leuven
Department of Computer Science
Celestijnenlaan 200A – B-3001 Heverlee (Belgium)

The Kähler mean of Block-Toeplitz matrices with Toeplitz structured blocks

Ben Jeuris
Raf Vandebril

Report TW 660, May 2015

Department of Computer Science, KU Leuven

Abstract

When computing 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. Also, 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.

Keywords : matrix mean, Toeplitz-Block Block-Toeplitz matrices, differential geometry, optimization on manifolds, Riemannian manifolds.
MSC : 15B05, 26E60, 53B21, 65K10.

THE KÄHLER MEAN OF BLOCK-TOEPLITZ MATRICES WITH TOEPLITZ STRUCTURED BLOCKS

B. JEURIS† AND R. VANDEBRIL†

Abstract. When computing 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. Also, 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.

Key words. matrix mean, Toeplitz-Block Block-Toeplitz matrices, differential geometry, optimization on manifolds, Riemannian manifolds

AMS subject classifications. 15B05, 26E60, 53B21, 65K10

1. Introduction. In radar theory and other signal processing applications [4, 6, 7, 31, 45], autocorrelation matrices are very popular to represent a window of some discrete or continuous signal.

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, the resulting autocorrelation matrix will be a positive definite (PD) Toeplitz matrix. A popular detection technique in radar theory consists of 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. [12], 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 vectorspace, 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 based on signal processing theory can be found [3, 5]. The transformed space can be endowed with a natural geometry 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 Block-Toeplitz (TBBT). We derive first order optimization techniques for the computation of these generalized Kähler means and analyze their properties.

This paper is organized in the following way. In Section 2, the transformation of PD Toeplitz matrices and its underlying interpretation are discussed. Afterwards, the natural geometry of

†Department of Computer Science, KU Leuven, Leuven, Belgium. {ben.jeuris,raf.vandebril}@kuleuven.be.

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 in Section 3. Moreover, we also discuss two different distance measures for the second generalized transformation. The generalized Kähler means for PD BT matrices and PD TBBT matrices are presented in Section 4 and 5, respectively. Finally, in Section 6 we compare the resulting algorithms in numerical experiments.

1.1. Definitions and notation. The set of PD matrices, denoted by $\mathcal{P}_n$, is defined as the set

$$\mathcal{P}_n = \left\{ A \in \mathbb{C}^{n\times n} \;\middle|\; x^H A x > 0, \forall x \in \mathbb{C}^n/\{0\} \right\}.$$

This characterization of PD matrices is equivalent to the condition that A is Hermitian and has positive eigenvalues [11], and is also denoted as $A > 0$. $\mathcal{P}_n$ is naturally endowed with the following distance measure and inner product

$$d(A,B) = \left\| \log\left(A^{-1/2} B A^{-1/2} \right) \right\|_F, \tag{1.1}$$

$$\langle E, F \rangle_A = \operatorname{trace}\left(A^{-1} E A^{-1} F \right), \tag{1.2}$$

where $A, B \in \mathcal{P}_n$, $E, F \in \mathcal{H}_n$, the set of $n \times n$ Hermitian matrices, and $\|.\|_F$ denotes the Frobenius norm.

The vectorspace of Toeplitz matrices consists of all matrices having identical elements along the diagonals,

$$\mathcal{T}_n = \left\{ \begin{bmatrix} t_0 & t_1 & \cdots & t_{n-1} \\ t_{-1} & t_0 & \ddots & t_{n-2} \\ \vdots & \ddots & \ddots & \vdots \\ t_{-n+1} & t_{-n+2} & \cdots & t_0 \end{bmatrix} \;\middle|\; t_{-n+1}, \ldots, t_{n-1} \in \mathbb{C} \right\}. \tag{1.3}$$

The intersection of this set of Toeplitz matrices with the Hermitian matrices $\mathcal{H}_n$ is given by the elements in (1.3) for which $t_{-i} = t_i^*, i = 0, \ldots, n-1$. 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 vectorspace 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 vectorspace 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 $A_1, \ldots, A_k$ 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:

$$\mathcal{B}_{\mathcal{S}}(A_1, \ldots, A_k) = \arg\min_{X \in \mathcal{S}} \frac{1}{2} \sum_{i=1}^{k} d_{\mathcal{S}}^2(X, A_i). \tag{1.4}$$

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 positive definite matrices with its natural distance measure (1.1), this barycenter is identical to the Karcher mean, the main instance of the geometric mean of positive definite matrices [11, 12, 13, 21, 27, 29, 33, 35]. The structured geometric mean, proposed by Bini et al. [12], 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.

Throughout the paper, 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

$$f(X) = g(A, B, C), \quad \left[\begin{array}{l} A = g_1(X), \\ B = g_2(X), \\ C = g_3(X), \end{array} \right.$$

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 counter-identity, the $n \times n$ matrix with ones on the anti-diagonal 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 $\overline{A}$ to represent the form JA^*J. Note that this operation corresponds to taking the conjugate transpose of A and reflecting the result over the anti-diagonal.

2. The Kähler mean for Toeplitz matrices.

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. [12]. Here, we will discuss a different geometry on $\mathcal{T}_n \cap \mathcal{P}_n$, along with its underlying interpretation and its properties.

2.1. The transformation.

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 signal $x(k)$, its autocorrelation coefficients $r_t = E\left[x(k+t)x(k)^*\right]$ can be obtained for differents lags t. If this autocorrelation is performed on the above autoregressive model, the following system is found:

$$R_n \widetilde{a}_n = -\widetilde{r}_n, \quad \widetilde{a}_n = [a_1^n, \ldots, a_n^n]^T, \quad \widetilde{r}_n = [r_1, \ldots, r_n]^T, \tag{2.1}$$

where R_n is the PD Toeplitz matrix of size n with elements $[R]_{i,j} = [R]_{j,i}^* = r_{i-j}$, $i, j = 0, \pm 1, \ldots, \pm(n-1)$. 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 [28, 43] can be used to find the solution to system (2.1) by solving the system for $n = 1$, and sequentially obtain the prediction coefficients $\widetilde{a}_n$ for increasing n. The Levinson recurrence relation for the prediction coefficients is

given by:

$$
\begin{aligned}
\widetilde{a}_1 &= a_1^1 = -\frac{r_1}{r_0}, \\
a_\ell^\ell &= -\frac{r_\ell + \sum_{j=1}^{\ell-1} r_{\ell-j} a_j^{\ell-1}}{r_0 + \sum_{j=1}^{\ell-1} r_j a_j^{\ell-1*}}, \\
\widetilde{a}_\ell &= \begin{bmatrix} a_1^\ell \\ \vdots \\ a_{\ell-1}^\ell \\ a_\ell^\ell \end{bmatrix} = \begin{bmatrix} a_1^{\ell-1} \\ \vdots \\ a_{\ell-1}^{\ell-1} \\ 0 \end{bmatrix} + a_\ell^\ell \begin{bmatrix} a_{\ell-1}^{\ell-1*} \\ \vdots \\ a_1^{\ell-1*} \\ 1 \end{bmatrix},
\end{aligned} \tag{2.2}
$$

with $\ell = 2, \ldots, n$. It can be shown that the factors a_ℓ^ℓ all lie within the complex unit disk $\mathbb{D}$, $|a_\ell^\ell| < 1, \forall \ell = 1, \ldots, n$.

Our main interest in the above is the one-to-one relation between the PD Toeplitz matrix R_n and the scalars $\left(r_0, a_1^1, \ldots, a_{n-1}^{n-1}\right)$. 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 (2.1), but not of R_n.

The transformation of the matrix R_n is the following:

$$
\begin{array}{rcl}
\mathcal{T}_n^+ & \rightarrow & \mathbb{R}^{++} \times \mathbb{D}^{n-1} \\
R_n & \mapsto & (p_0, \mu_1, \ldots, \mu_{n-1}),
\end{array} \tag{2.3}
$$

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.

2.2. The potential, the metric, and the cost function. In order to define the Kähler metric, the set of PD Toeplitz matrices is considered to be a Kähler manifold [3, 5]. 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 information geometry in general), the *Kähler potential* is often chosen to be the process entropy $\Phi(R_n)$ [8], defined as follows:

$$
\Phi(R_n) = \log\left(\det R_n^{-1}\right) - \log(\pi \mathrm{e}), \tag{2.4}
$$

where π 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 (2.3) of R_n, the process entropy $\Phi(R_n)$ can be rewritten as a function of the parameter space $\mathbb{R}^{++} \times \mathbb{D}^{n-1}$:

$$
\Phi(R_n) = -n \log\left(p_0\right) - \sum_{\ell=1}^{n-1} (n-\ell) \log\left(1 - |\mu_\ell|^2\right) - \log(\pi \mathrm{e}),
$$

where R_n is identified with its transformation $(p_0, \mu_1, \ldots, \mu_{n-1})$. This decomposition of the determinant of R_n is discussed in more detail for the block matrix case in Section 3.1.

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 $\xi^{(n)} = [p_0, \mu_1, \ldots, \mu_{n-1}]^T$, then

$$
[H]_{i,j} = \frac{\partial^2 \Phi}{\partial \xi_i^{(n)} \overline{\partial} \xi_j^{(n)}}.
$$

The desired metric can be found as

$$\begin{aligned} \mathrm{d}s^2 &= \mathrm{d}\xi^{(n)^H} H \, \mathrm{d}\xi^{(n)} \\ &= n\frac{\mathrm{d}p_0^2}{p_0^2} + \sum_{\ell=1}^{n-1}(n-\ell)\frac{|\,\mathrm{d}\mu_\ell|^2}{\left(1-|\mu_\ell|^2\right)^2}. \end{aligned} \tag{2.5}$$

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 (1.1) and (1.2) (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:

$$\begin{aligned} \forall a, b \in \mathbb{R}^{++}, \forall e, f \in \mathbb{R}: \quad & \langle e, f \rangle_a = n\frac{ef}{a^2}, \\ & d_{\mathbb{R}^{++}}(a,b) = \sqrt{n}\left|\log\frac{b}{a}\right|; \\ \forall \mu, \nu \in \mathbb{D}, \forall \varepsilon, \varsigma \in \mathbb{C}: \quad & \langle \varepsilon, \varsigma \rangle_\mu = \frac{n-\ell}{2}\frac{\varepsilon\varsigma^* + \varsigma\varepsilon^*}{\left(1-|\mu|^2\right)^2}, \\ & d_{\mathbb{D}}(\mu,\nu) = \frac{\sqrt{n-\ell}}{2}\log\left(\frac{1+\left|\frac{\mu-\nu}{1-\mu\nu^*}\right|}{1-\left|\frac{\mu-\nu}{1-\mu\nu^*}\right|}\right), \end{aligned} \tag{2.6}$$

where ℓ is chosen corresponding to the coordinate (μ_ℓ, $\ell = 1, \ldots, n-1$, from (2.3)) 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

$$\begin{aligned} d^2_{\mathcal{T}_n^+}(T_1,T_2) &= d^2_{\mathcal{T}_n^+}\Big((p_{0,1},\mu_{1,1},\ldots,\mu_{n-1,1}),(p_{0,2},\mu_{1,2},\ldots,\mu_{n-1,2})\Big) \\ &= n\log^2\left(\frac{p_{0,2}}{p_{0,1}}\right) + \sum_{\ell=1}^{n-1}\frac{n-\ell}{4}\log^2\left(\frac{1+\left|\frac{\mu_{\ell,1}-\mu_{\ell,2}}{1-\mu_{\ell,1}\mu_{\ell,2}^*}\right|}{1-\left|\frac{\mu_{\ell,1}-\mu_{\ell,2}}{1-\mu_{\ell,1}\mu_{\ell,2}^*}\right|}\right). \end{aligned} \tag{2.7}$$

By entering this distance measure into definition (1.4), the Kähler mean is obtained as the barycenter $\mathcal{B}_{\mathcal{T}_n^+}$. Endowing the manifold $\mathcal{T}_n^+$ with the Kähler metric (2.5) results in a complete, simply connected manifold with non-positive sectional curvature everywhere, or a Cartan–Hadamard manifold. Hence, existence and uniqueness are guaranteed for the barycenter with respect to this metric [15, 30].

2.3. The properties. 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 (2.3) of a matrix $T_i \in \mathcal{T}_n^+$ by $(p_{0,i}, \mu_{1,i}, \ldots, \mu_{n-1,i})$, then, for any $\alpha_i > 0$, the transformation of $\alpha_i T_i$ is $(\alpha_i p_{0,i}, \mu_{1,i}, \ldots, \mu_{n-1,i})$. Hence, from the explicit expression of the first coordinate $p_{0,\mathcal{B}} = (p_{0,1}\cdots p_{0,k})^{1/k}$ 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 $\widetilde{T}_1, T_1, T_2 \in \mathcal{T}_n^+$ with $\widetilde{T}_1 \geq T_1$, then in general $\mathcal{B}_{\mathcal{T}_n^+}(\widetilde{T}_1, T_2) \not\geq \mathcal{B}_{\mathcal{T}_n^+}(T_1, T_2)$.

When experimenting with the Kähler mean, results have shown that its averaging properties cooperate very well with the application from which it was derived [5, 7, 31, 45]. 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:

$$\begin{array}{ccccccc} T_1 & \mapsto & (& \boxed{p_{0,1},} & \boxed{\mu_{1,1},} & \cdots, & \boxed{\mu_{n-1,1}} &) \\ \vdots & & & \boxed{\vdots} & \boxed{\vdots} & & \boxed{\vdots} & \\ T_k & \mapsto & (& \boxed{p_{0,k},} & \boxed{\mu_{1,k},} & \cdots, & \boxed{\mu_{n-1,k}} &) \\ & & & \downarrow & \downarrow & & \downarrow & \\ T & \leftarrow & (& p_0, & \mu_1, & \cdots, & \mu_{n-1} &) \end{array}$$

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. [12].

3. Generalization of the Toeplitz structure. Our real interest goes out to the linear autoregressive model for multichannel signals [32], 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-1} square matrices. Taking the normal equations of the multichannel model, the so-called Yule-Walker equations are obtained:

$$\begin{aligned} \widetilde{A}_n \widetilde{R}_n &= -U_n \\ \widetilde{A}_n &= [A_1^n, \ldots, A_n^n], \\ U_n &= [R_1, \ldots, R_n], \\ \widetilde{R}_n &= \begin{bmatrix} R_0 & R_1 & \cdots & R_{n-1} \\ R_1^H & R_0 & \ddots & R_{n-2} \\ \vdots & \ddots & \ddots & \vdots \\ R_{n-1}^H & R_{n-2}^H & \cdots & R_0 \end{bmatrix}, \end{aligned} \tag{3.1}$$

where $\widetilde{R}_n \in \mathcal{B}_{n,N}^+$ 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 2D signal, when interpreted as a multichannel signal) result in a matrix $\widetilde{R}_n$ which is not only PD BT, but also has the Toeplitz structure in the individual blocks [26, 28, 39]. Hence it will become a PD TBBT matrix. In practice, these Toeplitz blocks will often be Hermitian themselves, $R_\ell = R_\ell^H, \ell = 0, \ldots, n-1$, but we will develop our theory for the more general case in which only the entire matrix $\widetilde{R}_n$ is Hermitian. The results remain valid in the more specified setting.

3.1. A first generalized transformation.

The transformation. With $\widetilde{R}_n$ now defined as a PD TBBT matrix, we would like to generalize the transformation (2.3) to $\mathcal{T}_{n,N}^+$. Similar to the link between the recursion (2.2) and the transformation (2.3), this generalization is obtained using a recursive computation of the prediction matrices in $\widetilde{A}_n$. This recursive computation goes as follows [28, 32, 42, 44],

$$A_1^1 = -R_1 R_0^{-1}, \tag{3.2}$$

$$A_\ell^\ell = -\Delta_\ell P_{\ell-1}^{-1}, \tag{3.3}$$

$$\begin{bmatrix} \Delta_\ell = R_\ell + \sum_{j=1}^{\ell-1} A_j^{\ell-1} R_{\ell-j}, \\ P_{\ell-1} = R_0 + \sum_{j=1}^{\ell-1} J A_j^{\ell-1*} J R_j = R_0 + \sum_{j=1}^{\ell-1} \overline{A_j^{\ell-1}} R_j, \end{bmatrix} \tag{3.4}$$

$$\widetilde{A}_\ell = \left[\widetilde{A}_{\ell-1}, 0\right] + A_\ell^\ell \left[\overline{A_{\ell-1}^{\ell-1}}, \ldots, \overline{A_1^{\ell-1}}, I\right], \tag{3.5}$$

with $\ell = 2, \ldots, n$. Similar to the prediction coefficients a_ℓ^ℓ from before, the factors A_ℓ^ℓ 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 $\widetilde{R}_n$ (3.1) are assumed to be Toeplitz matrices, we have $\overline{R_\ell} = R_\ell^H, \ell = 0, \ldots, n-1$, 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 Schur complement, for the inversion of block matrices,

$$\widetilde{R}_{\ell+1}^{-1} = \begin{bmatrix} \alpha_\ell & -\alpha_\ell U_\ell \widetilde{R}_\ell^{-1} \\ -\widetilde{R}_\ell^{-1} U_\ell^H \alpha_\ell & \widetilde{R}_\ell^{-1} + \widetilde{R}_\ell^{-1} U_\ell^H \alpha_\ell U_\ell \widetilde{R}_\ell^{-1} \end{bmatrix},$$

with $\alpha_\ell = \left(R_0 - U_\ell \widetilde{R}_\ell^{-1} U_\ell^H\right)^{-1}$. Note that α_ℓ is a principal submatrix of the PD matrix $\widetilde{R}_{\ell+1}^{-1}$ and is therefore also PD.

Now, the auxiliary matrix P_ℓ in the recursive computation (3.4) can be written as

$$\overline{P_\ell} = R_0 + \widetilde{A}_\ell U_\ell^H = R_0 - U_\ell \widetilde{R}_\ell^{-1} U_\ell^H = \alpha_\ell^{-1},$$

hence $\overline{P_\ell}$ (and P_ℓ) is also a PD matrix. Using the recursion expression (3.5), an updating rule can be found for $\overline{P_\ell}$ (and consequently for α_ℓ^{-1}),

$$\overline{P_\ell} = \overline{P_{\ell-1}} - \Delta_\ell P_{\ell-1}^{-1} \overline{\Delta_\ell} = \left(I - A_\ell^\ell \overline{A_\ell^\ell}\right) \overline{P_{\ell-1}}, \tag{3.6}$$

where $\overline{P_0} = R_0$.

Finally, we show that the matrices A_ℓ^ℓ belong to the set

$$\mathcal{D}_N = \left\{ \Gamma \in \mathbb{C}^{N \times N} \;\middle|\; I - \Gamma\overline{\Gamma} > 0 \right\}.$$

Note that for $N = 1$, this set reduces to the complex numbers γ for which $\gamma\overline{\gamma} = \gamma\gamma^* < 1$, which is exactly the complex unit disk $\mathbb{D}$. To prove that all matrix factors A_ℓ^ℓ belong to $\mathcal{D}_N$, we start from the positive definiteness of $\overline{P_\ell}$:

$$\begin{aligned} & \overline{P_\ell} = \overline{P_{\ell-1}} - \Delta_\ell P_{\ell-1}^{-1} \overline{\Delta_\ell} > 0, \\ \xrightarrow{congruence} \quad & I - \overline{P_{\ell-1}^{-1/2}} \Delta_\ell P_{\ell-1}^{-1} \overline{\Delta_\ell} \overline{P_{\ell-1}^{-1/2}} > 0, \\ \xrightarrow{similarity} \quad & I - \Delta_\ell P_{\ell-1}^{-1} \overline{\Delta_\ell} \overline{P_{\ell-1}^{-1}} = I - A_\ell^\ell \overline{A_\ell^\ell} > 0. \end{aligned}$$

The resulting transformation will be a mapping between the PD BT (not TBBT) matrices and the new parameter space, and it is defined as

$$\begin{array}{rcl} \mathcal{B}_{n,N}^+ & \to & \mathcal{P}_N \times \mathcal{D}_N^{n-1} \\ \widetilde{R}_n & \mapsto & (P_0, \Gamma_1, \ldots, \Gamma_{n-1}), \end{array} \tag{3.7}$$

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 $(P_0, \Gamma_1, \ldots, \Gamma_{n-1}) \in \mathcal{P}_N \times \mathcal{D}_N^{n-1}$ 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 $\widetilde{R}_n$ [34]:

$$\begin{aligned} \det\left(\widetilde{R}_n\right) &= \det\left(\widetilde{R}_{n-1}\right) \det\left(R_0 - U_{n-1} \widetilde{R}_{n-1}^{-1} U_{n-1}^H\right) \\ &= \det\left(\widetilde{R}_{n-1}\right) \det\left(\alpha_{n-1}^{-1}\right) \\ &= \det\left(\widetilde{R}_{n-1}\right) \det\left(I - A_{n-1}^{n-1} \overline{A_{n-1}^{n-1}}\right) \ldots \det\left(I - A_1^1 \overline{A_1^1}\right) \det\left(R_0\right) \\ &= \det\left(I - A_{n-1}^{n-1} \overline{A_{n-1}^{n-1}}\right) \ldots \det\left(I - A_1^1 \overline{A_1^1}\right)^{n-1} \det\left(R_0\right)^n, \end{aligned} \tag{3.8}$$

where the recursive updating rule (3.6) for α_ℓ^{-1} (and $\overline{P_\ell}$) is used. The resulting factorization of the Kähler potential (2.4) becomes (in parameter space $\mathcal{P}_N \times \mathcal{D}_N^{n-1}$):

$$\Phi\left(\widetilde{R}_n\right) = -n\log\left(\det P_0\right) - \sum_{\ell=1}^{n-1}(n-\ell)\log\left(\det\left(I - \Gamma_\ell\overline{\Gamma_\ell}\right)\right) - \log(\pi e),$$

where $\widetilde{R}_n$ is identified with $(P_0, \Gamma_1, \ldots, \Gamma_{n-1})$ under transformation (3.7).

As before, we use complex differentiation to determine the Hessian of the Kähler potential and obtain the *generalized metric*:

$$\begin{aligned}\mathrm{d}s^2 =& n\,\mathrm{trace}\left(P_0^{-1}\,\mathrm{d}P_0 P_0^{-1}\,\mathrm{d}P_0\right)\\ &+ \sum_{\ell=1}^{n-1}(n-\ell)\,\mathrm{trace}\left(\left(I-\Gamma_\ell\overline{\Gamma_\ell}\right)^{-1}\mathrm{d}\Gamma_\ell\left(I-\overline{\Gamma_\ell}\Gamma_\ell\right)^{-1}\overline{\mathrm{d}\Gamma_\ell}\right).\end{aligned}$$

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 (1.1) and (1.2). 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

$$\mathcal{SD}_N = \left\{\Omega \in \mathbb{C}^{N\times N} \;\middle|\; I - \Omega\Omega^H > 0\right\},$$

which is (almost) the Siegel disk [36] and which has been well-studied along with the Siegel upper halfplane. In the next section 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.

3.2. A second generalized transformation. In this section, we present a different generalized transformation, where the set $\mathcal{D}_N$ in transformation (3.7) 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 [40, 41] 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 [14, 18, 19, 25]. 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 Szegő's recursion [38], in which the Verblunsky coefficients arise. It turns out that these coefficients are equal to the prediction coefficients a_ℓ^ℓ (2.2) used in transformation (2.3) [9].

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 [18, 20]. On the other hand, constructing orthogonal matrix polynomials on the unit circle w. r. t. the matrix measure results in a generalization of the Szegő recursion, with corresponding generalized Verblunsky coefficients [16, 20, 37].

We use the proposed generalization of the Verblunsky coefficients [20] to define a new *transformation* of a PD BT matrix as follows,

$$\begin{array}{rcl}\mathcal{B}_{n,N}^{+} & \to & \mathcal{P}_N \times \mathcal{SD}_N^{n-1}\\ \widetilde{R}_n & \mapsto & (P_0, \Omega_1, \ldots, \Omega_{n-1}),\end{array} \tag{3.9}$$

where P_0 is still equal to R_0, but now

$$\Omega_\ell := L_{\ell-1}^{-\frac{1}{2}} \left(R_\ell - M_{\ell-1}\right) K_{\ell-1}^{-\frac{1}{2}}, \tag{3.10}$$
$$\left[\begin{array}{l} L_{\ell-1} = R_0 - \left[R_1, \ldots, R_{\ell-1}\right] \widetilde{R}_{\ell-1}^{-1} \left[R_1, \ldots, R_{\ell-1}\right]^H, \\ K_{\ell-1} = R_0 - \left[R_{\ell-1}^H, \ldots, R_1^H\right] \widetilde{R}_{\ell-1}^{-1} \left[R_{\ell-1}^H, \ldots, R_1^H\right]^H, \\ M_{\ell-1} = \left[R_1, \ldots, R_{\ell-1}\right] \widetilde{R}_{\ell-1}^{-1} \left[R_{\ell-1}^H, \ldots, R_1^H\right]^H, \end{array}\right.$$

for $\ell = 1, \ldots, n-1$. Comparing this transformation to the previous one, the following relations can be found for the auxiliary matrices P_ℓ and Δ_ℓ (3.4): $K_{\ell-1} = P_{\ell-1}$, $L_{\ell-1} = \overline{P_{\ell-1}}$, 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 (3.11).

It still remains to show that the coordinate matrices Ω_ℓ 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 [20] and Fritzsche and Kirstein [23]. 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 in Section 4.

Suppose we have $\widetilde{R}_\ell \in \mathcal{T}_{n,N}^+$, then by exploiting the Toeplitz structure of the blocks and $\overline{\widetilde{R}_\ell} = \widetilde{R}_\ell$, we can show that

$$\begin{aligned} \overline{\Delta_\ell} &= \overline{R_\ell} - \overline{M_{\ell-1}} \\ &= R_\ell^H - J_N \left[R_1, \ldots, R_{\ell-1}\right]^* \widetilde{R}_{\ell-1}^{-1^*} \left[R_{\ell-1}^H, \ldots, R_1^H\right]^{H^*} J_N \\ &= R_\ell^H - J_N \left[R_1, \ldots, R_{\ell-1}\right]^* J_{nN} \widetilde{R}_{\ell-1}^{-1} J_{nN} \left[R_{\ell-1}^H, \ldots, R_1^H\right]^{H^*} J_N \\ &= R_\ell^H - \left[R_{\ell-1}^H, \ldots, R_1^H\right] \widetilde{R}_{\ell-1}^{-1} \left[R_1, \ldots, R_{\ell-1}\right]^H \\ &= \Delta_\ell^H, \end{aligned}$$

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 (3.8) 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* are

$$\begin{aligned}\Phi\left(\widetilde{R}_n\right) =& -n\log\left(\det P_0\right) - \sum_{\ell=1}^{n-1}(n-\ell)\log\left(\det\left(I-\Omega_\ell\Omega_\ell^H\right)\right) - \log(\pi e),\\ \mathrm{d}s^2 =& n\,\mathrm{trace}\left(P_0^{-1}\,\mathrm{d}P_0 P_0^{-1}\,\mathrm{d}P_0\right)\\ &+\sum_{\ell=1}^{n-1}(n-\ell)\,\mathrm{trace}\left(\left(I-\Omega_\ell\Omega_\ell^H\right)^{-1}\mathrm{d}\Omega_\ell\left(I-\Omega_\ell^H\Omega_\ell\right)^{-1}\mathrm{d}\Omega_\ell^H\right).\end{aligned} \tag{3.11}$$

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 halfplane described by Siegel himself [36], using the link

$$\begin{aligned}\Omega &= (B-iI)(B+iI)^{-1},\\ B &= i(I+\Omega)(I-\Omega)^{-1},\end{aligned}$$

where B is an element of the Siegel upper halfplane ($\mathfrak{Im}(B) > 0$). 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 halfplane, 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

$$\begin{aligned}&\forall A,B\in\mathcal{P}_N, \forall E,F\in\mathcal{H}_N:\\ &\qquad \langle E,F\rangle_A = n\,\mathrm{trace}\left(A^{-1}EA^{-1}F\right),\\ &\qquad d_{\mathcal{P}_N}(A,B) = \sqrt{n}\left\|\log\left(A^{-1/2}BA^{-1/2}\right)\right\|_F;\end{aligned} \tag{3.12}$$

$$\begin{aligned}&\forall\Omega,\Psi\in\mathcal{SD}_N, \forall v,\omega\in\mathbb{C}^{N\times N}:\\ &\qquad \langle v,\omega\rangle_\Omega = \frac{n-\ell}{2}\,\mathrm{trace}\left(\left(I-\Omega\Omega^H\right)^{-1}v\left(I-\Omega^H\Omega\right)^{-1}\omega^H\right)\\ &\qquad\qquad + \frac{n-\ell}{2}\,\mathrm{trace}\left(\left(I-\Omega\Omega^H\right)^{-1}\omega\left(I-\Omega^H\Omega\right)^{-1}v^H\right),\\ &\qquad d^2_{\mathcal{SD}_N}(\Omega,\Psi) = \frac{n-\ell}{4}\,\mathrm{trace}\left(\log^2\left(\frac{I+C^{\frac{1}{2}}}{I-C^{\frac{1}{2}}}\right)\right),\\ &\qquad\qquad \left[C = (\Psi-\Omega)\left(I-\Omega^H\Psi\right)^{-1}\left(\Psi^H-\Omega^H\right)\left(I-\Omega\Psi^H\right)^{-1},\right.\end{aligned} \tag{3.13}$$

where ℓ is chosen corresponding to the coordinate matrix (Ω_ℓ, $\ell = 1\ldots,n-1$, from (3.9)) to which it relates. Note that both inner products and distance measures reduce to the scalar expressions (Section 2.2) 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 (4.1)) 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 $\widetilde{T}_1$ and $\widetilde{T}_2$, with transformations $(P_{0,1},\Omega_{1,1},\ldots,\Omega_{n-1,1})$ and $(P_{0,2},\Omega_{1,2},\ldots,\Omega_{n-1,2})$, is defined as

$$\begin{aligned}d^2_{BT}(\widetilde{T}_1,\widetilde{T}_2) &= d^2_{BT}\Big((P_{0,1},\Omega_{1,1},\ldots,\Omega_{n-1,1}),(P_{0,2},\Omega_{1,2},\ldots,\Omega_{n-1,2})\Big)\\ &= n\left\|\log\left(P_{0,1}^{-1/2}P_{0,2}P_{0,1}^{-1/2}\right)\right\|_F^2 + \sum_{\ell=1}^{n-1}\frac{n-\ell}{4}\,\mathrm{trace}\left(\log^2\left(\frac{I+C_\ell^{\frac{1}{2}}}{I-C_\ell^{\frac{1}{2}}}\right)\right),\\ &\left[C_\ell = (\Omega_{\ell,2}-\Omega_{\ell,1})\left(I-\Omega_{\ell,1}^H\Omega_{\ell,2}\right)^{-1}\left(\Omega_{\ell,2}^H-\Omega_{\ell,1}^H\right)\left(I-\Omega_{\ell,1}\Omega_{\ell,2}^H\right)^{-1}.\right.\end{aligned} \tag{3.14}$$

Using the definition of a barycenter (1.4), the generalized Kähler mean can now be found as $\mathcal{B}_{BT}$.

3.3. An alternative for the distance measure on $\mathcal{SD}_N$. The distance measure discussed in the previous section 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 [8, 10, 22]

$$d_K(\Omega, \Psi) = \frac{1}{2} \log \left(\frac{1 + \|\phi_\Omega(\Psi)\|_2}{1 - \|\phi_\Omega(\Psi)\|_2} \right),$$

$$\left[\phi_\Omega(\Psi) = \left(I - \Omega\Omega^H\right)^{-\frac{1}{2}} (\Psi - \Omega) \left(I - \Omega^H\Psi\right)^{-1} \left(I - \Omega^H\Omega\right)^{\frac{1}{2}}, \right. \tag{3.15}$$

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 to 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 (3.11), our differential metric on the Siegel disk becomes $\mathrm{d}s^2 = \mathrm{trace}(\mathrm{d}\Omega\ \mathrm{d}\Omega^H) = \|\mathrm{d}\Omega\|_F^2$. The differential metric corresponding to the Kobayashi distance measure at the zero matrix is given by $\mathrm{d}s^2 = \|\mathrm{d}\Omega\|_2^2$ [22, Theorem IV.1.8 and Lemma V.1.5], which is clearly not the same.

However, the main advantage of this distance measure lies in the transformation ϕ_Ω (3.15), which acts as an automorphism on the Siegel disk. The distance between two matrices and between their transformations under ϕ_Ω 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.

We note already that this translation to the origin is no longer practical once we enforce the Toeplitz structure on the individual blocks R_ℓ, $\ell = 0, \ldots, n-1$, i. e., when we go from PD BT matrices to PD TBBT matrices. As will be fully explained in the next section, once an iteration step ω at the translated origin is computed, the actual iteration point Ω_ℓ (with respect to the original matrices) should be updated to $\phi_{(-\Omega_\ell)}(\omega)$. Imposing the Toeplitz structure on the blocks R_ℓ now results in a very involved condition for the step ω. The process of exploiting the translation itself is further explained in Section 4.

4. The generalized mean for PD BT matrices. 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, and in the next section, 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 (3.9) 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 $\mathcal{B}_{\mathcal{P}_N}(P_{0,1}, \ldots, P_{0,k})$ of the involved PD matrices, as defined in Section 1.1.

For the other coordinates ($\Omega_{\ell,i} \in \mathcal{SD}_N$), the optimization at each level of ℓ ($= 1, \ldots, n-1$) can be formulated in the same way, hence we omit the dependence on ℓ in the definition of the barycenter

$$\mathcal{B}_{\mathcal{SD}_N}(\Omega_1, \ldots, \Omega_k) = \arg\min_{X \in \mathcal{SD}_N} \frac{1}{2} \sum_{i=1}^{k} \left\| \log \left(\frac{I + C_i^{\frac{1}{2}}}{I - C_i^{\frac{1}{2}}} \right) \right\|_F^2,$$

$$\left[C_i = I - (I - \Omega_i\Omega_i^H)^{\frac{1}{2}} (I - X\Omega_i^H)^{-1} (I - XX^H) (I - \Omega_i X^H)^{-1} (I - \Omega_i\Omega_i^H)^{\frac{1}{2}}, \right. \tag{4.1}$$

where the cost function has been rescaled and C_i is written in the Hermitian form which was mentioned in Section 3.2. The cost function in this optimization problem will be denoted as $f_{\mathcal{B}_{\mathcal{SPD}_N}}(X)$.

A first order optimization algorithm requires us to determine the (Riemannian) gradient of the cost function, defined for $\mathcal{SPD}_N$ as

$$\mathrm{D}f_{\mathcal{B}_{\mathcal{SPD}_N}}(X)[\omega_X] = \langle \mathrm{grad}\, f_{\mathcal{B}_{\mathcal{SPD}_N}}(X), \omega_X \rangle_X, \tag{4.2}$$

with the inner product (3.13). 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

$$\begin{aligned} &\mathrm{D}g(X)[\omega] = -X^{-1}\omega X^{-1}, && \text{inversion [17]},\\ &\mathrm{D}h(X)[\omega]\, X^{\frac{1}{2}} + X^{\frac{1}{2}}\, \mathrm{D}h(X)[\omega] = \omega, && \text{square root}, \end{aligned}$$

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. After some calculations, the emerging gradient is

$$\mathrm{grad} f_{\mathcal{B}_{\mathcal{SPD}_N}}(X) = \left(I - XX^H\right) \sum_{i=1}^{k} \left(V_i \left(X - \Omega_i\right) \left(I - X^H\Omega_i\right)^{-1} \right) \left(I - X^H X\right), \tag{4.3}$$

$$\left[\begin{array}{l} V_i = \left(I - \Omega_i X^H\right)^{-1} \left(I - \Omega_i\Omega_i^H\right)^{\frac{1}{2}} Z_i \left(I - \Omega_i\Omega_i^H\right)^{\frac{1}{2}} \left(I - X\Omega_i^H\right)^{-1}, \\ Z_i = \mathfrak{L}\left(C_i^{\frac{1}{2}}, (I - C_i)^{-1} \log\left(\frac{I + C_i^{\frac{1}{2}}}{I - C_i^{\frac{1}{2}}} \right) \right), \end{array} \right.$$

where C_i is defined as in (4.1) and $\mathfrak{L}(A, Q)$ stands for the solution X of the continuous Lyapunov equation $AX + XA^H = Q$. Note that the second argument in the Lyapunov operator $\mathfrak{L}$ is a Hermitian matrix, hence the continuous Lyapunov equation (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 ϕ (3.15), computations can be greatly simplified. Suppose the initial guess for the barycenter $\mathcal{B}_{\mathcal{SPD}_N}$ is given by a matrix X_0. The translation ϕ_{X_0} maps the matrix X_0 exactly onto the origin and by applying the same transformation to the original matrices Ω_i, the distances and hence the barycenter cost function do not change. 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 Ψ_1. We can translate this new point again to the origin using the next translation ϕ_{Ψ_1}. However, in order to keep track of the barycenter approximations with respect to the original matrices, we need to keep in mind that Ψ_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 Algorithm 1. Note that $\Omega_i^{(j+1)}$ can also be computed as $\phi_{\Psi_{j+1}}(\Omega_i^{(j)})$ [8]. 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 round-off errors.

Finally, we present the simplified form of the gradient at the origin,

$$\mathrm{grad} f_{\mathcal{B}_{\mathcal{SPD}_N}}(0; \Omega_1, \ldots, \Omega_k) = -\sum_{i=1}^{k} V_i\Omega_i, \tag{4.4}$$

$$\left[V_i = \mathfrak{L}\left(\left(\Omega_i\Omega_i^H\right)^{\frac{1}{2}}, \log\left(\frac{I + \left(\Omega_i\Omega_i^H\right)^{\frac{1}{2}}}{I - \left(\Omega_i\Omega_i^H\right)^{\frac{1}{2}}} \right) \right), \right.$$

where V_i is now obtained directly as the solution of a CLE.

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 Ψ_{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: $\mathcal{B}_{\mathcal{SD}_N}(\Omega_1, \ldots, \Omega_k)$

5. The generalized mean for PD TBBT matrices. 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 (3.9) of the matrices, with the $n-1$ coordinate matrices in the Siegel disk given by (3.10).

At first sight, imposing the Toeplitz structure requires the matrix R_ℓ in each Ω_ℓ to be Toeplitz. However, the matrices $L_{\ell-1}$, $K_{\ell-1}$, and $M_{\ell-1}$ depend on the matrices $R_0, \ldots, R_{\ell-1}$, which should also be Toeplitz matrices now. All these Toeplitz restrictions are translated in an involved way to the search space in which each Ω_ℓ 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 (3.9) ($P_0 \to \Omega_1 \to \ldots \to \Omega_{n-1}$).

5.1. Global version of the mean. Instead of translating the Toeplitz restriction towards involved conditions on the coordinate matrices $(P_0, \Omega_1, \ldots, \Omega_{n-1})$, we consider the barycenter cost function $f_{\mathcal{B}_{BT}}$, based on the total Kähler distance function d_{BT} (3.14), as a function of the blocks $R_0, \ldots, R_{n-1}$ of the matrix $\widetilde{R}_n$. 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 Ω_ℓ depends on the blocks $R_0, \ldots, R_\ell$, for $\ell = 1, \ldots, n-1$. Or reversely, R_0 will influence all coordinate matrices, and for each $\ell = 1, \ldots, n-1$, block R_ℓ is present in coordinate matrices $\Omega_\ell, \ldots, \Omega_{n-1}$.

The gradient. As shown in (4.2), 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

$$\begin{aligned}
&\mathrm{D} f_{\mathcal{B}_{BT}}\big((R_0, \ldots, R_{n-1})\big)\big[(E_0, \omega_1, \ldots, \omega_{n-1})\big] \\
&\quad = \big\langle \operatorname{grad} f_{\mathcal{B}_{BT}}\big((R_0, \ldots, R_{n-1})\big), (E_0, \omega_1, \ldots, \omega_{n-1})\big\rangle_{(R_0, \ldots, R_{n-1})} \\
&\quad := \langle \operatorname{grad} f_{\mathcal{B}_{BT}}\big((R_0, \ldots, R_{n-1})\big)_0, E_0\rangle_{P_0} \\
&\qquad + \sum_{\ell=1}^{n-1} \langle L_{\ell-1}^{-\frac{1}{2}} \operatorname{grad} f_{\mathcal{B}_{BT}}\big((R_0, \ldots, R_{n-1})\big)_\ell K_{\ell-1}^{-\frac{1}{2}}, L_{\ell-1}^{-\frac{1}{2}} \omega_\ell K_{\ell-1}^{-\frac{1}{2}}\rangle_{\Omega_\ell},
\end{aligned} \tag{5.1}$$

where $(P_0, \Omega_1, \ldots, \Omega_{n-1})$ is the image of $\widetilde{R}_n$ under transformation (3.9) 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 (3.12) and (3.13), respectively, and the $(\ell+1)$th component of the gradient is represented by $\operatorname{grad} f_{\mathcal{B}_{BT}}\big((R_0, \ldots, R_{n-1})\big)_\ell$. 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_ℓ versus the tangent space at Ω_ℓ.

To demonstrate the complexity of the relations, we present the gradient below. The point at which the gradient is computed is denoted by $\widetilde{R}_n$, with blocks $(R_0, \ldots, R_{n-1})$ and transformation $(P_0, \Omega_1, \ldots, \Omega_{n-1})$, while the PD TBBT matrices of which the barycenter is computed will be denoted by $\widetilde{R}_{n,i}$, with blocks $(R_{0,i}, \ldots, R_{n-1,i})$ and transformation $(P_{0,i}, \Omega_{1,i}, \ldots, \Omega_{n-1,i})$, $i = 1, \ldots, k$.

In the expressions, the matrices $A_j^{\ell-1}$ (3.2–3.5), associated with the creation of Δ_ℓ and $P_{\ell-1}$ (and therefore $L_{\ell-1}$, $K_{\ell-1}$, and $M_{\ell-1}$) in the transformation of $\widetilde{R}_n$, are used to increase readability and computational efficiency. The *first component of the gradient* becomes the following

$$\begin{aligned}
&\operatorname{grad} f_{\mathcal{B}_{BT}}\big((R_0, \ldots, R_{n-1})\big)_0 \\
&\qquad = P_0 \sum_{i=1}^{k} \left(P_0^{-1} \log\left(P_0 P_{0,i}^{-1}\right) + \sum_{\ell=1}^{n-1} \frac{n-\ell}{2n} G_{\ell,i} \right) P_0, \qquad (5.2)\\
&\left[\begin{array}{l}
G_{\ell,i} = -D_{\ell.i}^L - D_{\ell,i}^K + \sum_{j=1}^{\ell-1} \Big(- {A_j^{\ell-1}}^H D_{\ell,i}^L A_j^{\ell-1} - \overline{A_j^{\ell-1}}^H D_{\ell,i}^K \overline{A_j^{\ell-1}} \\
\qquad\qquad + {A_j^{\ell-1}}^H L_{\ell-1}^{-\frac{1}{2}} {V_{\ell,i}^{(1)}}^H K_{\ell-1}^{-\frac{1}{2}} \overline{A_{\ell-j}^{\ell-1}} + \overline{A_{\ell-j}^{\ell-1}}^H K_{\ell-1}^{-\frac{1}{2}} V_{\ell,i}^{(1)} L_{\ell-1}^{-\frac{1}{2}} A_j^{\ell-1} \Big), \\
D_{\ell,i}^L = \mathfrak{L}\left(L_{\ell-1}^{\frac{1}{2}}, \Omega_\ell V_{\ell,i}^{(1)} L_{\ell-1}^{-\frac{1}{2}} + L_{\ell-1}^{-\frac{1}{2}} {V_{\ell,i}^{(1)}}^H \Omega_\ell^H \right), \\
D_{\ell,i}^K = \mathfrak{L}\left(K_{\ell-1}^{\frac{1}{2}}, \Omega_\ell^H {V_{\ell,i}^{(1)}}^H K_{\ell-1}^{-\frac{1}{2}} + K_{\ell-1}^{-\frac{1}{2}} V_{\ell,i}^{(1)} \Omega_\ell \right), \\
V_{\ell,i}^{(1)} = (I - \Omega_{\ell,i}^H \Omega_\ell)^{-1} (\Omega_\ell^H - \Omega_{\ell,i}^H) V_{\ell,i}, \\
V_{\ell,i} = (I - \Omega_{\ell,i} \Omega_\ell^H)^{-1} (I - \Omega_{\ell,i} \Omega_{\ell,i}^H)^{\frac{1}{2}} Z_{\ell,i} (I - \Omega_{\ell,i} \Omega_{\ell,i}^H)^{\frac{1}{2}} (I - \Omega_\ell \Omega_{\ell,i}^H)^{-1}, \\
Z_{\ell,i} = \mathfrak{L}\left(C_{\ell,i}^{\frac{1}{2}}, (I - C_{\ell,i})^{-1} \log\left(\frac{I + C_{\ell,i}^{\frac{1}{2}}}{I - C_{\ell,i}^{\frac{1}{2}}} \right) \right), \\
C_{\ell,i} = I - \Big((I - \Omega_{\ell,i} \Omega_{\ell,i}^H)^{\frac{1}{2}} (I - \Omega_\ell \Omega_{\ell,i}^H)^{-1} (I - \Omega_\ell \Omega_\ell^H) \ldots \\
\qquad\qquad (I - \Omega_{\ell,i} \Omega_\ell^H)^{-1} (I - \Omega_{\ell,i} \Omega_{\ell,i}^H)^{\frac{1}{2}} \Big),
\end{array}\right.
\end{aligned}$$

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 $q = 1, \ldots, n-1$, given by

$$\begin{aligned}
\operatorname{grad} f_{\mathcal{B}_{BT}}\big((R_0, \ldots, R_{n-1})\big)_q &= L_{q-1}^{\frac{1}{2}} \left(I - \Omega_q \Omega_q^H\right) \sum_{i=1}^{k} {V_{q,i}^{(1)}}^H \left(I - \Omega_q^H \Omega_q\right) K_{q-1}^{\frac{1}{2}} \\
&+ L_{q-1}^{\frac{1}{2}} \left(I - \Omega_q \Omega_q^H\right) L_{q-1}^{\frac{1}{2}} \sum_{i=1}^{k} \left(\sum_{\ell=q+1}^{n-1} \frac{n-\ell}{n-q} W_{\ell,i}^{(q)} \right) K_{q-1}^{\frac{1}{2}} \left(I - \Omega_q^H \Omega_q\right) K_{q-1}^{\frac{1}{2}}, \qquad (5.3)
\end{aligned}$$

$$\left[\begin{array}{l}
W_{\ell,i}^{(q)} = -D_{\ell,i}^L A_q^{\ell-1} - \overline{A_q^{\ell-1}}^H D_{\ell,i}^K \\
\quad + {A_{\ell-q}^{\ell-1}}^H L_{\ell-1}^{-\frac{1}{2}} {V_{\ell,i}^{(1)}}^H K_{\ell-1}^{-\frac{1}{2}} + L_{\ell-1}^{-\frac{1}{2}} {V_{\ell,i}^{(1)}}^H K_{\ell-1}^{-\frac{1}{2}} \overline{A_{\ell-q}^{\ell-1}} \\
\quad + \sum_{j=q+1}^{\ell-1} \Big(- {A_{j-q}^{\ell-1}}^H D_{\ell,i}^L A_j^{\ell-1} - \overline{A_j^{\ell-1}}^H D_{\ell,i}^K \overline{A_{j-q}^{\ell-1}} \\
\qquad + {A_{j-q}^{\ell-1}}^H L_{\ell-1}^{-\frac{1}{2}} {V_{\ell,i}^{(1)}}^H K_{\ell-1}^{-\frac{1}{2}} \overline{A_{\ell-j}^{\ell-1}} + \overline{A_{\ell-j+q}^{\ell-1}}^H K_{\ell-1}^{-\frac{1}{2}} V_{\ell,i}^{(1)} L_{\ell-1}^{-\frac{1}{2}} A_j^{\ell-1} \Big),
\end{array}\right.$$

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 $(R_0, \ldots, R_{n-1})$ instead of the coordinate matrices $(P_0, \Omega_1, \ldots, \Omega_{n-1})$. 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 [1]. In our case, the embedding manifold is the set $\left(\mathbb{C}^{N\times N}\right)_+^n$ containing all tuples $(R_0, \ldots, R_{n-1})$ which represent the blocks of an element in $\mathcal{B}_{n,N}^+$. The submanifold is given by the set $(\mathcal{T}_N)_+^n$ which contains all tuples $(R_0, \ldots, R_{n-1})$ 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 $\left(\mathbb{C}^{N\times N}\right)_+^n$ since no additional structure was imposed on the blocks. Hence, we need an orthogonal projection of this gradient at any point $(R_0, \ldots, R_{n-1}) \in (\mathcal{T}_N)_+^n \subset \left(\mathbb{C}^{N\times N}\right)_+^n$ from $T_{(R_0,\ldots,R_{n-1})}\left(\mathbb{C}^{N\times N}\right)_+^n$ onto $T_{(R_0,\ldots,R_{n-1})}(\mathcal{T}_N)_+^n$. This *projection* should be orthogonal with respect to the inner product (5.1) and, for $(E_0, \omega_1, \ldots, \omega_{n-1}) \in T_{(R_0,\ldots,R_{n-1})}\left(\mathbb{C}^{N\times N}\right)_+^n$, is given by

$$E_0 \mapsto \operatorname{vec}^{-1}\left(U_{\mathrm{H}}\left(U_{\mathrm{H}}^H\left({P_0^{-1}}^T \otimes P_0^{-1}\right)U_{\mathrm{H}}\right)^{-1} U_{\mathrm{H}}^H \operatorname{vec}\left(P_0^{-1} E_0 P_0^{-1}\right)\right), \tag{5.4}$$

$$\omega_\ell \mapsto \operatorname{vec}^{-1}\left(U_{\mathrm{T}}\left(U_{\mathrm{T}}^H\left({S_\ell^K}^T \otimes S_\ell^L\right)U_{\mathrm{T}}\right)^{-1} U_{\mathrm{T}}^H \operatorname{vec}\left(S_\ell^L \omega_\ell S_\ell^K\right)\right), \tag{5.5}$$

$$\left[\begin{array}{l} S_\ell^L = L_{\ell-1}^{-\frac{1}{2}}\left(I - \Omega_\ell \Omega_\ell^H\right)^{-1} L_{\ell-1}^{-\frac{1}{2}}, \\ S_\ell^K = K_{\ell-1}^{-\frac{1}{2}}\left(I - \Omega_\ell^H \Omega_\ell\right)^{-1} K_{\ell-1}^{-\frac{1}{2}}, \end{array}\right.$$

for $\ell = 1, \ldots, n-1$, where $(P_0, \Omega_1, \ldots, \Omega_{n-1})$ is the transformation (3.9) of $\widetilde{R}_n$, the BT matrix containing blocks $(R_0, \ldots, R_{n-1})$, with associated matrices $L_{\ell-1}$ and $K_{\ell-1}$. The vec-operator is the columnwise vectorization of a matrix, and the matrices U_{H} and U_{T} are parametrization matrices for Hermitian Toeplitz and general Toeplitz matrices, respectively. Hence, e. g., we write $\operatorname{vec}(T_1) = U_{\mathrm{H}} t_1$, with $t_1 \in \mathbb{R}^{2N-1}$ the parametrization of $T_1 \in \mathcal{T}_N \cap \mathcal{H}_N$, and $\operatorname{vec}(T_2) = U_{\mathrm{T}} t_2$, 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 inner product (5.1) for both the Riemannian gradient and the orthogonal projection.

5.2. Greedy version of the mean. 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 from the previous section that the coordinate matrix P_0 only depends on the block R_0, coordinate matrix Ω_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 $(P_{0,1}, \ldots, P_{0,k})$ as described by Bini et al. [12] (Section 1.1).

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 Ω_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 Ω_ℓ, $\ell = 1, \ldots, n-1$ is performed using a combination of constructions which have already been derived. From Section 4, we remember the barycenter cost function $f_{\mathcal{B}_{\mathcal{SD}_N}}$ with associated gradient (4.3). Because of the Toeplitz restriction and the assumption that all previously optimized coordinate matrices are fixed, the tangent space at Ω_ℓ is given by

$$T_{\Omega_\ell}\left(L_{\ell-1}^{-1/2}\left(\mathcal{T}_N - M_{\ell-1}\right)K_{\ell-1}^{-1/2}\right) \simeq \left\{L_{\ell-1}^{-\frac{1}{2}} T K_{\ell-1}^{-\frac{1}{2}} \;\middle|\; T \in \mathcal{T}_N\right\}.$$

We are now working directly on the level of Ω_ℓ instead of R_ℓ, hence the projection of the gradient onto this tangent space slightly differs from the one presented in (5.5) as follows

$$\omega_\ell \mapsto \operatorname{vec}^{-1}\left(U_{\mathrm{T}}\left(U_{\mathrm{T}}^H\left(S_\ell^{K^T} \otimes S_\ell^L\right)U_{\mathrm{T}}\right)^{-1} U_{\mathrm{T}}^H \operatorname{vec}\left(S_\ell^L L_{\ell-1}^{\frac{1}{2}} \omega_\ell K_{\ell-1}^{\frac{1}{2}} S_\ell^K\right)\right),$$

where U_{T}, S_ℓ^K, and S_ℓ^L are the same as in (5.5).

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.

5.3. Properties of the generalized Kähler mean. When considering the properties of the generalized Kähler mean, an intuitive approach is to start from the properties of the Kähler mean for Toeplitz matrices (Section 2.3).

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* (3.9) when a PD (TB)BT matrix $\widetilde{R}_n$ is replaced with $\alpha\widetilde{R}_n$, for any real $\alpha > 0$. We denote the transformation of $\widetilde{R}_n$ by $(P_0, \Omega_1, \ldots, \Omega_{n-1})$, with corresponding prediction matrices A_j^ℓ and auxiliary matrices $P_{\ell-1}$ and Δ_ℓ, and that of $\alpha\widetilde{R}_n$ by $(P_0', \Omega_1', \Omega_{n-1}')$, now with corresponding prediction matrices $A_j^{\ell\prime}$ and auxiliary matrices $P_{\ell-1}'$ and Δ_ℓ'.

First, the change of the prediction matrices $A_j^{\ell\prime}$, and auxiliary matrices $P_{\ell-1}'$ and Δ_ℓ', $\ell = 1, \ldots, n-1$, $j = 1, \ldots, \ell$, can be found using induction. Considering (3.2)–(3.5), it is clear to see that $A_1^{1\prime} = A_1^1$, $P_0' = \alpha P_0$, and $\Delta_1' = \alpha\Delta_1$. Now assuming $\widetilde{A}_{\ell-1}' = \widetilde{A}_{\ell-1}$, we find $P_{\ell-1}' = \alpha P_{\ell-1}$, $\Delta_\ell' = \alpha\Delta_\ell$, and $A_\ell^{\ell\prime} = A_\ell^\ell$. As a consequence of (3.5), $\widetilde{A}_\ell' = \widetilde{A}_\ell$, which closes the induction.

By writing the coordinate matrices Ω_ℓ' in the form $\overline{P_{\ell-1}'^{-1/2}}\Delta_\ell' P_{\ell-1}'^{-1/2}$, we now find that $\Omega_\ell' = \Omega_\ell$, $\ell = 1, \ldots, n-1$. Summarized, the transformation of $\alpha\widetilde{R}_n$ is given by $(\alpha P_0, \Omega_1, \ldots, \Omega_{n-1})$, which is consistent with the Kähler transformation of PD Toeplitz matrices. Note that transformation (3.7) behaves in the same way for positive scaling.

Now, as for *joint homogeneity*, suppose we have PD (TB)BT matrices $\widetilde{T}_i$, $i = 1, \ldots, k$, with a corresponding transformation $(P_{0,i}, \Omega_{1,i}, \ldots, \Omega_{n-1,i})$, and k positive scalars α_i. The generalized Kähler mean for PD BT matrices (Section 4) 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, 27] is sufficient to prove the property in this case.

The global version of the Kähler mean for PD TBBT matrices (Section 5.1) 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 $\widetilde{R}_n$ with given matrices $\widetilde{T}_i$, $i = 1, \ldots, k$, it can be checked that the same happens for $(\alpha_1 \cdots \alpha_k)^{1/k}\widetilde{R}_n$ with given matrices $\alpha_i\widetilde{T}_i$, $i = 1, \ldots, k$.

Moreover, the greedy approximation (Section 5.2) also satisfies the property, which can be seen as follows. We will denote the transformation of the greedy Kähler mean of the unscaled $\widetilde{T}_1, \ldots, \widetilde{T}_k$ by $(P_0, \Omega_1, \ldots, \Omega_{n-1})$. The greedy Kähler mean of the scaled matrices $\alpha_1\widetilde{T}_1, \ldots, \alpha_k\widetilde{T}_k$ now starts by averaging the first coordinate matrices, resulting in $\mathcal{B}_{\mathcal{T}_N^+}(\alpha_1 P_{0,1}, \ldots, \alpha_k P_{0,1}) = (\alpha_1 \cdots \alpha_k)^{1/k} P_0$

because of the joint homogeneity of the structured geometric mean for linear structures [12]. 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 $(\Omega_{1,1}, \ldots, \Omega_{1,k})$ we still minimize the cost function $f_{\mathcal{B}_{\mathcal{SD}_N}}(X; \Omega_{1,1}, \ldots, \Omega_{1,k})$. However, the search space has changed from

$$\overline{P_0^{-1/2}}\mathcal{T}_N P_0^{-1/2} \cap \mathcal{SD}_N \qquad \text{to} \qquad (\alpha_1 \cdots \alpha_k)^{-1/k}\overline{P_0^{-1/2}}\mathcal{T}_N P_0^{-1/2} \cap \mathcal{SD}_N,$$

from which it can be seen that the resulting coordinate matrix Ω_1 remains the same as in the unscaled setting (since a scaling of vectorspace $\mathcal{T}_N$ does not change the space). The other coordinate matrices Ω_ℓ, $\ell = 2, \ldots, n-1$, similarly do not change. Finally, the greedy Kähler mean of the scaled matrices is obtained with coordinate matrices $((\alpha_1 \cdots \alpha_k)^{1/k} P_0, \Omega_1, \ldots, \Omega_{n-1})$, 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.

6. Numerical experiments. In this section, 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.

6.1. The Siegel and Kobayashi barycenter on $\mathcal{SD}_N$. In this paper, we have endowed the Siegel disk $\mathcal{SD}_N$ with the Siegel distance measure $d_{\mathcal{SD}_N}$ (Section 3.2) and with the Kobayashi distance measure d_K (Section 3.3). Since each distance measure 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.

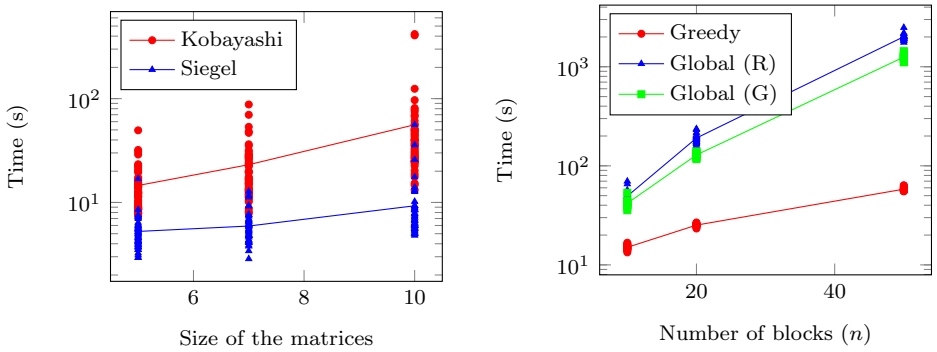
When investigating the distance between the barycenters, a relative distance of the order $\mathcal{O}(10^{-1})$ can be found consistently for varying matrix sizes. Note that the diameter of the Siegel disk becomes infinity for both distance measures.

As for computational time, we display some results of both barycenters for varying sizes of matrices in Figure 1a. The Siegel barycenter $\mathcal{B}_{\mathcal{SD}_N}$ requires less computational time, which also increases more slowly.

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 [24]. During the computation of the barycenter $\mathcal{B}_K$, it is possible that a matrix with 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.

6.2. The generalized Kähler mean. 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 1b. This was expected, since the gradient for the greedy optimization problem can be found in the gradient of the global optimization problem (5.2)–(5.3) by setting the factors $G_{\ell,i}$ (for the first component) and $W_{\ell,i}^{(q)}$ (for the other components) to zero.



(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.

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 1b, the computational time of the global version of the mean is displayed when using 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.

7. Conclusions. In this paper, we have focused on a geometry for positive definite Toeplitz matrices and a generalization thereof towards positive definite (Toeplitz-Block) Block-Toeplitz 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 [5, 7, 31, 45].

Afterwards, two possible generalizations of the Kähler transformation towards positive defi-

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 vs. Global	2.28e-04	1.36e-04	8.24e-05
Size global gradient at Greedy	2.34	2.44	3.14

nite (Toeplitz-Block) Block-Toeplitz 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 Toeplitz-Block Block-Toeplitz 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 of 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.

REFERENCES

[1] Pierre-Antoine Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, 2008.

[2] T. Ando, C.K. Li, and R. Mathias, *Geometric means*, Linear Algebra and its Applications, 385 (2004), pp. 305–334.

[3] F. Barbaresco, *Information intrinsic geometric flows*, American Institute of Physics Conference Series, 872 (2006), pp. 211–218.

[4] ———, *Innovative tools for radar signal processing based on Cartan's geometry of SPD matrices and information geometry*, in RADAR '08, IEEE International Radar Conference, Rome, 2008, pp. 1–6.

[5] ———, *Interactions between symmetric cone and information geometries: Bruhat-Tits and Siegel spaces models for high resolution autoregressive Doppler imagery*, Lecture Notes in Computer Science, 5416 (2009), pp. 124–163.

[6] ———, *Robust statistical radar processing in Fréchet metric space: OS-HDR-CFAR and OS-STAP processing in Siegel homogeneous bounded domains*, in IRS '11, International Radar Conference, Leipzig, 2011, pp. 639–644.

[7] ———, *Information Geometry of Covariance Matrix: Cartan-Siegel Homogeneous Bounded Domains, Mostow/Berger Fibration and Fréchet Median*, Springer, 2012, ch. 9, pp. 199–255.

[8] ———, *Information geometry manifold of Toeplitz Hermitian positive definite covariance matrices: Mostow/Berger fibration and Berezin quantization of Cartan-Siegel domains*, International Journal of Emerging Trends in Signal Processing, 1 (2013), pp. 1–11.

[9] ———, *Eidetic Reduction of Information Geometry Through Legendre Duality of Koszul Characteristic Function and Entropy: From MassieuDuhem Potentials to Geometric Souriau Temperature and Balian Quantum Fisher Metric*, Springer, 2014, ch. 7, pp. 141–217.

[10] Giovanni Bassanelli, *On horospheres and holomorphic endomorfisms of the Siegel disc*, Rendiconti del Seminario Matematico della Universit di Padova, 70 (1983), pp. 147–165.

[11] Rajendra Bhatia, *Positive Definite Matrices*, Princeton Series in Applied Mathematics, Princeton University Press, 2007.

[12] D.A. Bini, B. Iannazzo, B. Jeuris, and R. Vandebril, *Geometric means of structured matrices*, BIT Numerical Mathematics, 54 (2014), pp. 55–83.

[13] Dario A. Bini and Bruno Iannazzo, *Computing the Karcher mean of symmetric positive definite matrices*, Linear Algebra and its Applications, 438 (2011), pp. 1700–1710.

[14] Albrecht Böttcher and Sergei M. Grudsky, *Spectral Properties of Banded Toeplitz Matrices*, SIAM, 2005.

[15] E. Cartan, *Leçons sur la géométrie des espaces de Riemann*, Uspekhi Matematicheskikh Nauk, 3 (1948), pp. 218–222.

[16] David Damanik, Alexander Pushnitski, and Barry Simon, *The analytic theory of matrix orthogonal polynomials*, Surveys in Approximation Theory, 4 (2008), pp. 1–85.
[17] J. Dehaene, *Continuous-time matrix algorithms, systolic algorithms and adaptive neural networks*, PhD thesis, Department of Electrical Engineering, KU Leuven, Belgium, 1995.
[18] Philippe Delsarte, Yves V. Genin, and Yves G. Kamp, *Orthogonal polynomial matrices on the unit circle*, IEEE Transactions on Circuits and Systems, 25 (1978), pp. 149–160.
[19] Holger Dette and William J. Studden, *The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis*, Wiley and Sons, 1997.
[20] Holger Dette and Jens Wagener, *Matrix measures on the unit circle, moment spaces, orthogonal polynomials and the Geronimus relations*, Linear Algebra and its Applications, 432 (2010), pp. 1609–1626.
[21] 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, 2006.
[22] T. Franzoni and E. Vesentini, *Holomorphic Maps and Invariant Distances*, vol. 40 of Mathematics studies, North-Holland, 1980.
[23] B. Fritzsche and B. Kirstein, *An extension problem for non-negative Hermitian block Toeplitz matrices*, Mathematische Nachrichten, 131 (1987), pp. 287–297.
[24] Mike Giles, *An extended collection of matrix derivative results for forward and reverse mode algorithmic differentiation*, tech. report, Oxford University Computing Laboratory, Oxford, England, 2008.
[25] Matthias Humet and Marc Van Barel, *Algorithms for the Geronimus transformation for orthogonal polynomials on the unit circle*, Journal of Computational and Applied Mathematics, 267 (2014), pp. 195–217.
[26] Andreas Jakobsson, S. Lawrence Marple, and Petre Stoica, *Computationally efficient two-dimensional Capon spectrum analysis*, IEEE Transactions on Signal Processing, 48 (2000), pp. 2651–2661.
[27] 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 (2012), pp. 379–402.
[28] R. Kanhouche, *A modified Burg algorithm equivalent in results to Levinson algorithm*, 2003. http://arxiv.org/abs/math/0309384.
[29] H. Karcher, *Riemannian center of mass and mollifier smoothing*, Communications on Pure and Applied Mathematics, 30 (1977), pp. 509–541.
[30] W. Kendall, *Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence*, Proceedings London Mathematical Society, 61 (1990), pp. 371–406.
[31] J. Lapuyade-Lahorgue and F. Barbaresco, *Radar detection using Siegel distance between autoregressive processes, application to HF and X-band radar*, in RADAR '08, IEEE International Radar Conference, Rome, 2008.
[32] S. L. Marple, *Digital Spectral Analysis with Applications*, Prentice-Hall, Englewood Cliffs, 1980.
[33] Miklos Palfia, *The Riemannian barycenter computation and means of several matrices*, International Journal of Computational and Mathematical Sciences, 3 (2009), pp. 128–133.
[34] Philip D. Powell, *Calculating determinants of block matrices*, 2011. Available on arXiv - http://arxiv.org/abs/1112.4379v1.
[35] Quentin 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.
[36] Carl Ludwig Siegel, *Symplectic geometry*, American Journal of Mathematics, 65 (1943), pp. 1–86.
[37] Ann Sinap and Walter Van Assche, *Orthogonal matrix polynomials and applications*, Journal of Computational and Applied Mathematics, 66 (1996), pp. 27–52.
[38] Gabor Szegő, *Orthogonal Polynomials*, American Mathematical Society, 1939.
[39] C.W. Therrien, *Relations between 2-D and multichannel linear prediction*, IEEE Transactions on Acoustics, Speech and Signal Processing, 29 (1981), pp. 454–456.
[40] S. Verblunsky, *On positive harmonic functions: a contribution to the algebra of Fourier series*, Proceedings London Mathematical Society, 38 (1935), pp. 125–157.
[41] ———, *On positive harmonic functions (second paper)*, Proceedings London Mathematical Society, 40 (1936), pp. 290–320.
[42] P. Whittle, *On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix*, Biometrika, 50 (1963), pp. 129–134.
[43] N. Wiener, *The Wiener RMS (Root Mean Square) Error Criterion in Filter Design and Prediction*, MIT Press, 1 ed., 1949, pp. 129–148.
[44] Ralph A. Wiggins and Enders A. Robinson, *Recursive solution to the multichannel fitering problem*, Journal of Geophysical Research, 70 (2012), pp. 1885–1891.
[45] L. Yang, *Medians of probability measures in Riemannian manifolds and applications to radar target detection*, PhD thesis, Université de Poitiers, France, 2011.