A Riemannian steepest descent approach over the Inhomogeneous Symplectic Group: Application to the Averaging of Linear Optical Systems

Simone Fiori*

Abstract

The present manuscript describes a Riemannian-steepest-descent approach to compute the average out of a set of optical system transference matrices on the basis of a Lie-group averaging criterion function. The devised averaging algorithm is compared with the Harris' exponential-mean-logarithm averaging rule, especially developed in computational ophthalmology to compute the average character of a set of biological optical systems. Results of numerical experiments show that the iterative algorithm based on gradient steepest descent implemented by exponential-map stepping converges to solutions that are in good agreement with those obtained by the application of Harris' exponential-mean-logarithm averaging rule. Such results seem to confirm that Harris' exponential-mean-logarithm averaging rule is *numerically optimal* in a Lie-group averaging sense.

Keywords: Averaging on Lie groups; Optimization on Riemannian Lie groups; Symplectic matrices; Hamiltonian matrices.
AMS subject classifications: 35Q93, 37N40, 53C21, 65D99

1 Introduction

In computational optics, it is assumed that the first-order physical features of an optical system, such as a human eyeball, are described by a 5×5 real matrix termed *optical transference matrix* [10, 11, 15]. The transference matrix associated to an optical section is a function of its geometric features, such as its thickness, and of its light-distorting features, such as its refractive index. Formal descriptions based on transference matrices allow one to characterize coaxial spherical systems, astigmatic surfaces and optical devices containing prisms and decentered lenses. A compound optical system may then be described, through a law of composition, via the optical transference matrices of its constituent subsystems. This is the case, for example, of a human eyeball, that may be described as composed by the cornea, the anterior chamber, the anterior surface of the lens, the interior of the lens, the posterior surface of the lens and the gap between the lens and the retina. A fairly complete survey of such model was offered in the paper [6].

*Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche, Via Brecce Bianche, Ancona I-60131, Italy. eMail: s.fiori@univpm.it

A problem of computational optics is how to define and calculate a meaningful *average character* out of a collection of optical systems. An average character of a set of optical systems is computed by averaging their transference matrices through an algorithm that is able to preserve the underlying mathematical structure. In fact, optical transference matrices are *structured*, in the sense that their entries obey mutual non-linear symmetry laws. For example, *averaging eyeball transference matrices should result in a plausible model of an eyeball* [9, 11]. Averaging algorithms for optical systems possess potential technological and clinical applications, such as in the design of quantum computing systems [7], in laser ablation of the front surface of the cornea and in the design and fit of soft contact lenses [18].

Since, in general, a matrix to qualify as an optical transference matrix needs to obey nonlinear constraints, the set of optical transferences, that will be later denoted as $\mathbb{T}$, is a curved space. This implies that linear-algebra operations do not work on the space $\mathbb{T}$ which, in turn, implies that the common arithmetic averaging will fail on $\mathbb{T}$. Even if the optical transference space is endowed with matrix multiplication and inverse as a group structure, the common geometric averaging will fail because, in general, matrix multiplication is not commutative (for an explanatory numerical example, see Subsection 4.1).

The definition of average of a set of structured data is not unique and the development of numerical algorithms to compute the mean element out of a set of structured data has attracted a considerable interest in the mathematical literature, in applied sciences and in engineering. For example, the contribution [14] tackled the general problem of taking averages over Riemannian manifolds, the contribution [5] dealt with averaging over Lie groups, the contributions [8, 13] dealt with numerical methods to take averages over the Stiefel and the Grassmann manifolds, and the paper [3] dealt with taking averages over the special Euclidean group and the unipotent group of matrices. A few contributions from the scientific literature tackled the problem of averaging a set of optical transfer matrices. In the seminal paper [11], Harris proposed a plausible non-iterative scheme to compute the average out of a set of optical devices. The contribution [6] investigated on the application of a fixed-point-type averaging algorithm, introduced by the present author and a coworker in [5], to averaging optical systems. In such contribution, Harris' average transference matrix was chosen as an initial guess to start the fixed-point iteration and it was observed that, in a number of experiments, the iterative algorithm is able to improve over the initial guess only slightly, leaving the impression that Harris' average is already a satisfactory solution to the averaging problem.

Since Harris' rule was derived empirically and does not possess any optimization ground, but it seems a satisfactory answer to the averaging problem in linear optics, it is natural to ask whether Harris' mean transference matrix is optimal in any sense. In order to investigate such a question, in the present contribution the line of thinking below is pursued:

- The space of optical transference matrices is studies in details and it is shown that it coincides with the *inhomogeneous symplectic group* $\mathbb{IS}(4, \mathbb{R})$. It is recalled that, being an algebraic group and a smooth manifold, the inhomogeneous symplectic group is a *Lie group*.
- The problem of computing an average matrix of a set of optical system transference matrices is formulated as an optimization problem over the

inhomogeneous symplectic group.

- The above optimization problem over the inhomogeneous symplectic group is tackled by a Riemannian steepest descent optimization algorithm upon metrization of the inhomogeneous symplectic group.
- The results provided by the above optimization algorithm are compared with the result provided by Harris' averaging rule.

Numerical simulations show that the Riemannian gradient steepest descent optimization rule devised in the present contribution is able to improve over Harris' mean transference matrix only slightly, hence suggesting the conclusion that *Harris' averaging rule is numerically optimal* in a Riemannian gradient-steepest descent sense.

Gradient-based optimization on matrix Lie groups is based on the notion of Riemannian gradient which, in the context of Lie-group theory, may be brought back to the Lie algebra associated with the Lie group. While, typically, Lie-group matrices obey non-linear symmetry laws, Lie-algebra matrices are subjected to linear constraints that induce symmetry in their entries.

The notion of computing averages over structured spaces applied to data-sets of only *two* data, recalls immediately the notion of continuous interpolation between two structured objects. The useful concept of *continuous interpolation* between two given optical systems will be discussed very briefly within the present paper as well, as a natural extension of the main topic.

2 Matrix Lie groups and Lie-group averaging

The present Section recalls those notions from differential geometry and Lie group theory that are instrumental in the subsequent development. For a complete reference on differential geometry, readers might want to see, e.g., the textbook [2] and the collection of textbooks [19]. The present Section also recalls the notion of gradient-based steepest-descent optimization on a Lie group upon group metrization and sets out a definition of *Lie-group averaging.*

2.1 Lie groups, metrization and optimization by greadient-steepest-descent

A *Lie group* $\mathbb{G}$ is an *algebraic group* that also possesses the structure of a *smooth manifold.* Within the present manuscript, our interest lays upon real-valued matrix Lie groups.

The algebraic matrix group structure $(\mathbb{G}, \mu, \iota, I)$ is made of a set $\mathbb{G}$ endowed with a multiplication operation μ, an inverse operation ι and an identity element I, such that for any $X, Y \in \mathbb{G}$, it holds that $\mu(X, Y) \in \mathbb{G}$, $\mu(X, \iota(X)) = \mu(\iota(X), X) = I$, and $\mu(X, I) = \mu(I, X) = X$. Group identity and inverse need to be unique and the group multiplication needs to be associative, namely, $\mu(X, \mu(Y, Z)) = \mu(\mu(X, Y), Z)$ for every $X, Y, Z \in \mathbb{G}$. In general, group multiplication is not commutative, though.

In addition, the differentiable matrix manifold $\mathbb{G}$ is a continuous set of structured matrices. The tangent space of the manifold $\mathbb{G}$ at a point $X \in \mathbb{G}$ is denoted by $T_X\mathbb{G}$. A tangent space $T_X\mathbb{G}$ is a vector space spanned by vectors tangent to each possible smooth curve passing through the point X. Namely, denoting by $\Gamma_X : [-a,\ a] \to \mathbb{G}$, with $a > 0$, any smooth curve $\Gamma_X(t)$ such that $\Gamma_X(0) = X$, the tangent space $T_X\mathbb{G}$ is spanned by vectors $\dot{\Gamma}_X(0)$ for all possible Γ_X's, where the over-dot denotes differentiation with respect to the parameter t. Associated with the Lie group $\mathbb{G}$ is a Lie algebra $\mathfrak{g} \stackrel{\text{def}}{=} T_I\mathbb{G}$, namely, the tangent space at identity.

Group multiplication may be extended to compute the multiplication between an element of a Lie group and an element of any tangent space. In fact, take two points $X, Y \in \mathbb{G}$ and a curve Γ_Y. The quantity $\mu(X, \Gamma_Y(t))$ belongs to the Lie group $\mathbb{G}$ for every t and, in addition, it holds that:

$$\frac{d}{dt}\mu(X, \Gamma_Y(t)) = \mu(X, \dot{\Gamma}_Y(t)).$$

At $t = 0$, the right-hand side involves a multiplication between an element of the Lie group $\mathbb{G}$ (namely, X) and an element of the tangent space $T_Y\mathbb{G}$ (namely, $\dot{\Gamma}_Y(0)$).

Given a Lie group $\mathbb{G}$, for every $U \in \mathfrak{g}$, there exists a unique smooth homomorphism $\Phi_U : (\mathbb{R}, +) \to (\mathbb{G}, \mu, \iota, I)$ such that $\dot{\Phi}_U(0) = U$. The *exponential map* $\exp : \mathfrak{g} \to \mathbb{G}$ is defined as $\exp(U) \stackrel{\text{def}}{=} \Phi_U(1)$. The exponential map is a local diffeomorphism at $0 \in \mathfrak{g}$ and it holds that $\exp(0) = I$. In turn, the homomorphism Φ may be written in terms of the exponential map as $\Phi_U(t) = \exp(tU)$. The inverse of the exponential map is termed *logarithmic map* and is denoted by $\log : \mathbb{G} \to \mathfrak{g}$. Since the exponential map is a local diffeomorphism, the logarithmic map is defined only in a neighborhood of the identity. The exponential map gets shifted to the map $\exp_X(V)$, where $X \in \mathbb{G}$ and $V \in T_X\mathbb{G}$, by the following rule:

$$\exp_X(V) \stackrel{\text{def}}{=} \mu(X, \exp(\mu(\iota(X), V)), \tag{1}$$

so that $\exp_X(0) = X$. Likewise, the logarithmic map at X is computed by $\log_X(Y) = \mu(X, \log(\mu(\iota(X), Y)))$. It is worth recalling that, on the group $(\mathbb{R}^n, +, -, 0)$, endowed with the Euclidean inner product, it holds that $\exp_X(V) = X + V$ and $\log_X(Y) = Y - X$, hence, the exponential map may be regarded as a generalization of the notion of *vector addition*, while the logarithmic map may be regarded as a generalization of *vector subtraction*.

A (positive-definite) inner product $\langle\cdot,\cdot\rangle_X : T_X\mathbb{G} \times T_X\mathbb{G} \to \mathbb{R}$ turns a Lie group into a metric space. The inner product is a function of the point X. The inner product at the identity of the Lie group $\mathbb{G}$ is denoted by $\langle\cdot,\cdot\rangle_{\mathfrak{g}}$. By the Lie group theory, the inner product over the tangent space at any given point is related to the inner product on the Lie algebra by the relationship:

$$\langle W, V\rangle_X = \langle \mu(\iota(X), W), \mu(\iota(X), V)\rangle_{\mathfrak{g}},\ \forall W, V \in T_X\mathbb{G}. \tag{2}$$

Given a smooth function $\varphi : \mathbb{G} \to \mathbb{R}$ and a point $X \in \mathbb{G}$, the *Riemannian gradient* $\nabla_X\varphi \in T_X\mathbb{G}$ is the unique solution of the equation:

$$\langle V, \nabla_X\varphi\rangle_X = \left.\frac{d}{dt}\varphi(\Gamma_{X,V}(t))\right|_{t=0}, \tag{3}$$

where $\Gamma_{X,V} : [-a,\ a]$, with $a > 0$, is any smooth curve such that $\Gamma_{X,V}(0) = X$ and $\dot{\Gamma}_{X,V}(0) = V$, for any $V \in T_X\mathbb{G}$. Since V and Γ are arbitrary, the gradient

$\nabla_X\varphi$ depends only on X, φ and on the geometric/topological structure of the Lie group $\mathbb{G}$.

The notion of *pushforward map* is of prime importance in the calculus of differentiable manifolds and Lie groups. Take two spaces $\mathbb{M}$ and $\mathbb{N}$ and a smooth map $\psi : \mathbb{M} \to \mathbb{N}$. Let $X \in \mathbb{M}$, $\Gamma_{X,V}(t)$ denote a curve in $\mathbb{M}$ with $t \in [-a,\ a]$, $a > 0$, such that $\Gamma_{X,V}(0) = X \in \mathbb{M}$ and $\dot{\Gamma}_{X,V}(0) = V \in T_X\mathbb{M}$. The pushforward map $\psi'_X : T_X\mathbb{M} \to T_{\psi(X)}\mathbb{N}$ is defined as:

$$\psi'_X(V) \stackrel{\text{def}}{=} \frac{d}{dt}\psi(\Gamma_{X,V}(t))\bigg|_{t=0}. \tag{4}$$

The pushforward map is linear in the tangent direction $V \in T_X\mathbb{M}$. Assume that $\mathbb{M}$ and $\mathbb{N}$ are matrix spaces and that the function ψ is analytic about a point $X_0 \in \mathbb{M}$, namely, that it may be expressed as:

$$\psi(X) = a_0 I + \sum_{k=1}^{\infty} a_k (X - X_0)^k, \tag{5}$$

for a set of coefficients $a_k \in \mathbb{R}$. Below, the analytic expansion of some matrix-to-matrix maps are recalled, which will be of particular interest to complete the calculations involved in the present paper. In the following list, the symbol $\mathbb{G}$ denotes a matrix Lie group:

- **Matrix exponential**. $\psi(X) = \exp(X)$, $\mathbb{M} = \mathfrak{g}$, $\mathbb{N} = \mathbb{G}$: $X_0 = 0$, $a_k = \frac{1}{k!}$ (where symbol $k!$ denotes factorial),
- **Matrix principal logarithm**. $\psi(X) = \log(X)$, $\mathbb{M}$ coincides with a subset of $\mathbb{G}$ where the logarithmic function is defined, $\mathbb{N} = \mathfrak{g}$: $X_0 = I$, $a_0 = 0$, $a_k = -\frac{(-1)^k}{k}$ for $k \geq 1$,
- **Matrix inverse**. $\psi(X) = X^{-1}$, $\mathbb{M}$ and $\mathbb{N}$ coincide with the subset of $\mathbb{G}$ where the series converges: $X_0 = I$, $a_k = (-1)^k$.

The pushforward map $\psi'_X(V)$ may thus be expressed as:

$$\psi'_X(V) = \sum_{k=1}^{\infty} a_k \sum_{r=1}^{k} (X - X_0)^{r-1} V (X - X_0)^{k-r}. \tag{6}$$

The above expression of the pushforward map is obtained through a term-by-term derivation of the series (5). In particular, it holds that:

$$\exp'_{tV}(V) = \exp(tV)V, \tag{7}$$
$$\log'_I(V) = V. \tag{8}$$

Gradient steepest descent optimization on flat (Euclidean) spaces may be extended to (curved, in general) Lie groups. To this purpose, take the smooth criterion function $\varphi : \mathbb{G} \to \mathbb{R}$ and the differential equation on the manifold $\mathbb{G}$:

$$\dot{X}(t) = -\nabla_{X(t)}\varphi,\ X(0) = \bar{X} \in \mathbb{G},\ t \geq 0. \tag{9}$$

Any limiting solution of such differential equation coincides with the local minimum of the function φ in $\mathbb{G}$ that belongs to the basin of attraction of the initial point $\bar{X}$. In fact, by definition of Riemannian gradient, it holds that:

$$\frac{d}{dt}\varphi(X(t)) = \langle \dot{X}(t), \nabla_{X(t)}\varphi \rangle_{X(t)} = -\langle \nabla_{X(t)}\varphi, \nabla_{X(t)}\varphi \rangle_{X(t)} \leq 0, \tag{10}$$

for all $t \geq 0$ and the equality holds only when $\nabla_X \varphi = 0$.

In the case of the Lie group $(\mathbb{R}^n, +, -, 0)$ endowed with the standard Euclidean inner product, the following gradient steepest descent algorithm is used to find a local minimizer of a function $\varphi : \mathbb{R}^n \to \mathbb{R}$:

$$X_{a+1} = X_a - t_a \left.\frac{\partial \varphi}{\partial X}\right|_{X=X_a}, \ a = 0,\ 1,\ 2,\ \dots\ , \tag{11}$$

with $X_0 \in \mathbb{R}^n$ being an initial guess and $t_a \geq 0$ being an optimization step-size schedule. The rationale of the above iteration algorithm is that the next step X_{a+1} is located on a straight line that departs from the point X_a and that is directed toward the opposite direction of the Euclidean gradient $\left.\frac{\partial \varphi}{\partial X}\right|_{X=X_a}$.

For the problem at hand, the concept of *straight line* is replaced by its generalized version and, in the language of Lie group theory, such concept is captured by the exponential map. In other words, in order to integrate numerically the differential equation (9) on a Lie group $(\mathbb{G}, \mu, \iota, I)$, a discrete-time sequence is generated by moving from each point $X_a \in \mathbb{G}$ to the next point $X_{a+1} \in \mathbb{G}$ along a short arc in the opposite direction of the Riemannian gradient of the criterion function. Namely, upon setting

$$V_a \stackrel{\text{def}}{=} \nabla_{X_a} \varphi, \tag{12}$$

the numerical optimization algorithm may be expressed as:

$$X_{a+1} = \exp_{X_a}(-t_a V_a) = \mu(X_a, \exp(-\mu(\iota(X_a), V_a) t_a)), \tag{13}$$

where $t_a \geq 0$ denotes any suitable optimization step-size schedule that drives the iterative algorithm (12)-(13) to convergence and $a = 0, \dots, A$, with the number of iterations A being sufficiently large. The above stepping method may be traced back to [16] and has been extensively explored, see, e.g., [1, 4].

2.2 Lie-group averaging and measure of spread

In order to define a notion of *empirical average matrix* out of a collection of N samples Y_k in a real-valued matrix Lie group $\mathbb{G}$, the following restrictions are considered:

- An average matrix of a set of matrices in a Lie group $\mathbb{G}$ must belong to the same space $\mathbb{G}$.
- The notion of *spread* of a set of samples with respect to a given point on a Lie group must be defined in a way that accounts for the amount of dispersion of the objects about the reference point.
- A definition of average matrix of a set of matrices in a Lie group must embody the intuitive notion that the average matrix minimizes the spread measure with respect to all the available samples.

In the case of the group $\mathbb{R}^n$ endowed with the Euclidean inner product/norm, a measure of spread σ^2 of a collection of points $\{Y_k\}_{k=1}^N$ around a given point $X \in \mathbb{R}^n$ is given by:

$$\sigma^2(X) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{k=1}^{N} \|Y_k - X\|^2. \tag{14}$$

For a matrix Lie group $\mathbb{G}$, such dispersion may be directly recast as:

$$\sigma^2(X) \stackrel{\text{def}}{=} \frac{1}{N}\sum_{k=1}^{N}\langle \log_X(Y_k), \log_X(Y_k)\rangle_X. \tag{15}$$

The criterion function (15) appears as a specialization, to Lie groups, of the Karcher criterion [14], valid for metrizable manifolds. A fundamental result of the Karcher theory is that *if the data to average are sufficiently close one to another, the criterion function (15) possesses exactly one minimizer.* Such minimizer is, by definition, a Lie-group average matrix. The average, thus, depends on the differential geometric structure of the Lie group $\mathbb{G}$ and on the chosen metrization via the inner product. The Figure 1 displays the notion of Karcher-Lie-group average.

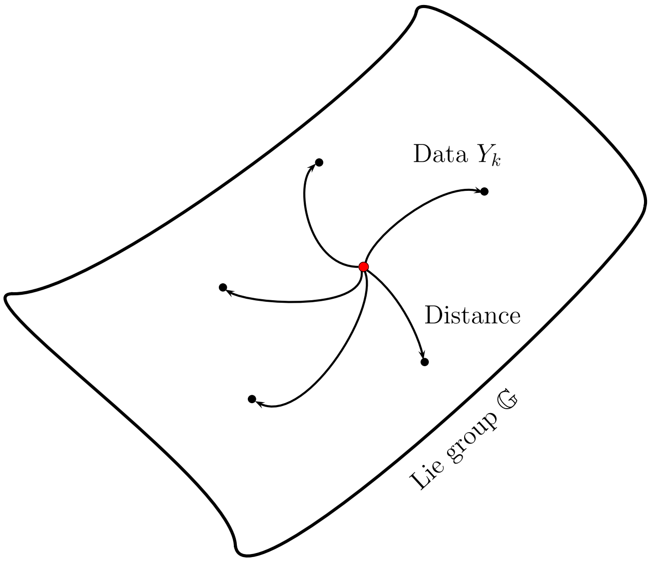

Figure 1: Notion of Karcher average over a Lie group $\mathbb{G}$. The black dots denote the data Y_k to average, the central (red) dot denotes the Karcher mean which shows equal Riemannian distance from alla data.

The minimization of the spread function (15) is achieved by the algorithm (12)-(13) upon setting $\varphi = \sigma^2$. The corresponding value of the function (15) is the spread of the samples about their empirical average matrix, namely, the *empirical variance of the data set.*

2.3 Continuous geometric interpolation

Consider the special case that the set of matrix-type Lie-group data to average counts only two elements $\{Y_1, Y_2\} \subset \mathbb{G}$. The mean matrix of the set $\{Y_1, Y_2\}$ may be thought of as a midpoint on a line connecting the endpoints Y_1 and Y_2. Other points on such a line might be of use in applications, for example:

- To simulate the effects of a slight misalignment or a defect in accommodation of an optical system,
- To fill-in missing data (or to augment the number of available samples) in on-site measurements.

Continuous geometric interpolation of structured matrices on a Lie group is the analogous of interpolation over flat spaces.

On the flat space $\mathbb{G} = \mathbb{R}^n$, the continuous interpolation $X(t)$, $t \in [0,\ 1]$, of two assigned points $Y_1, Y_2 \in \mathbb{R}^n$ may be defined via the combination:

$$X(t) \stackrel{\text{def}}{=} f_1(t)Y_1 + f_2(t)Y_2, \tag{16}$$

where $f_1, f_2 : [0,\ 1] \to \mathbb{R}$ are smooth weighting functions such that $f_1(0) = f_2(1) = 1$ and $f_1(1) = f_2(0) = 0$ and the variable $t \in [0,\ 1]$ provides a continuous parametrization for the interpolatory scheme. The simplest choice of weighting functions, which gives rise to a convex combination, is $f_1(t) \stackrel{\text{def}}{=} 1 - t$ and $f_2(t) \stackrel{\text{def}}{=} t$.

In a non-flat Lie group, though, the linear combination involved in the definition (16) is incompatible with its geometric/algebraic structure. A possible way to generalize such interpolatory scheme is via the exponential map, namely:

$$X(t) \stackrel{\text{def}}{=} \exp_{Y_1}(tV), \tag{17}$$

where $Y_1, Y_2 \in \mathbb{G}$ and the tangent vector $V \in T_{Y_1}\mathbb{G}$ is chosen in such a way that $\exp_{Y_1}(V) = Y_2$, namely, $V = \log_{Y_1}(Y_2)$. Care should be taken that, as the logarithmic map is a *local* diffeomorphism, two assigned points $Y_1, Y_2 \in \mathbb{G}$ are not necessarily connectable by a curve like (17). It is worth mentioning that the curve $X(t)$ might be extendable to values $t \notin [0,\ 1]$, in which case, an extrapolation of the two data Y_1, Y_2 within $\mathbb{G}$ is being invoked.

3 Computation of averages over the space of optical system transference matrices

The aim of the present Section is to recall the properties of the space of optical transfer matrices and develop an optimization principle that leads to the definition of average optical transference matrix. Simple proofs of the mathematical assertions below may be found, e.g., in [6, Subsection 3.1].

3.1 The symplectic group and the optical transference matrix group

Recall the notion of *real symplectic group*:

$$\mathbb{S}(2p, \mathbb{R}) \stackrel{\text{def}}{=} \{S \in \mathbb{R}^{2p \times 2p} \mid S^T J S = J\},\ J \stackrel{\text{def}}{=} \begin{bmatrix} 0 & I_p \\ -I_p & 0 \end{bmatrix}, \tag{18}$$

where the symbol I_p denotes the $p \times p$ identity matrix, symbol ^{T} denotes matrix transpose and the symbol J denotes the *skew-symmetric unit matrix*. The skew-symmetric unit matrix enjoys the following properties: $J^2 = -I$,

$J^{-1} = J^T = -J$. The symplectic group has fundamental applications in computational ophthalmology [10], in the study of Hamiltonian systems [20] as well as in quantum theory [21]. The real symplectic group may be regarded as a special instance of *quadratic* matrix Lie group [17].

The space $\mathbb{S}(2p, \mathbb{R})$ is a curved smooth manifold that may be endowed with smooth algebraic-group operations compatible with the manifold structure. The space $\mathbb{S}(2p, \mathbb{R})$ is an algebraic group under matrix multiplication and inversion:

- **Multiplication**: For every $S_1, S_2 \in \mathbb{S}(2p, \mathbb{R})$, the group multiplication $\mu(S_1, S_2)$ is defined as the matrix product $S_1 S_2 \in \mathbb{S}(2p, \mathbb{R})$. The **identity element** of the group $\mathbb{S}(2p, \mathbb{R})$ with respect to the standard matrix multiplication is the identity matrix I_p.

- **Inversion**: Denote by $\det(\cdot)$ the determinant of a matrix and take any symplectic matrix $S \in \mathbb{S}(2p, \mathbb{R})$. From the condition $S^T JS = J$, it follows that $\det(S)\det(J)\det(S) = \det(J)$, namely, any matrix $S \in \mathbb{S}(2p, \mathbb{R})$ is such that $\det^2(S) = 1$. An implication of the above property is that any symplectic matrix is invertible. In addition, if $S \in \mathbb{S}(2p, \mathbb{R})$, then $S^{-1} \in \mathbb{S}(2p, \mathbb{R})$. Therefore, the group inverse $\iota(S)$, defined as the matrix inversion S^{-1}, is well-posed.

The space $\mathbb{S}(2p, \mathbb{R})$ has the structure of a Lie group.

To characterize the structure of the tangent space to the Lie group $\mathbb{S}(2p, \mathbb{R})$ at every point $S \in \mathbb{S}(2p, \mathbb{R})$, denoted as $T_S\mathbb{S}(2p, \mathbb{R})$, one may take the route indicated in the Subsection 2.1. It is readily found that the tangent space $T_S\mathbb{S}(2p, \mathbb{R})$ has the structure:

$$T_S\mathbb{S}(2p, \mathbb{R}) = \{V \in \mathbb{R}^{2p\times 2p} | V^T JS + S^T JV = 0\}. \tag{19}$$

In particular, the tangent space at the identity of the Lie group, namely the Lie algebra $\mathfrak{h}(2p, \mathbb{R}) \stackrel{\text{def}}{=} T_I\mathbb{S}(2p, \mathbb{R})$, has the structure:

$$\mathfrak{h}(2p, \mathbb{R}) = \{H \in \mathbb{R}^{2p\times 2p} | H^T J + JH = 0\}, \tag{20}$$

namely, it coincides with the space of $2p \times 2p$ Hamiltonian matrices. It is worth noting that if $H \in \mathfrak{h}(2p, \mathbb{R})$, then $JH - (JH)^T = 0$, namely, a matrix H is Hamiltonian if and only if the product JH is symmetric; therefore, any Hamiltonian matrix may be parameterized by a symmetric matrix via multiplication by the skew-symmetric unit matrix.

The *inhomogeneous real symplectic group* $\mathbb{IS}(2p, \mathbb{R})$ is defined as:

$$\mathbb{IS}(2p, \mathbb{R}) \stackrel{\text{def}}{=} \left\{ \left[\begin{array}{cc} S & \delta \\ 0 & 1 \end{array} \right] \middle| S \in \mathbb{S}(2p, \mathbb{R}),\ \delta \in \mathbb{R}^{2p} \right\}. \tag{21}$$

The space of optical system transference matrices, as defined in [11], coincides with the inhomogeneous real symplectic group $\mathbb{IS}(4, \mathbb{R})$ and will be denoted by $\mathbb{T}$ for the sake of notation conciseness.

The space $\mathbb{T}$ may be endowed with the structure of an algebraic group by matrix multiplication and inversion, namely, given optical transference $T, T_1 \in \mathbb{T}$, the group multiplication is defined as $\mu(T, T_1) \stackrel{\text{def}}{=} TT_1$ and the group inverse is defined as $\iota(T) = T^{-1}$. The identity element in $\mathbb{T}$ with respect to matrix multiplication is the unit matrix I_5.

The space $\mathbb{T}$ has the structure of a Riemannian manifold as well. It is instrumental to characterize the tangent spaces $T_X\mathbb{T}$ for every transference $X \in \mathbb{T}$ and to define an inner product $\langle\cdot,\cdot\rangle_X : T_X\mathbb{T} \times T_X\mathbb{T} \to \mathbb{R}$. The tangent space at a point $X = \left[\begin{smallmatrix} S & \delta \\ 0 & 1 \end{smallmatrix}\right] \in \mathbb{T}$ has structure:

$$T_X\mathbb{T} \stackrel{\text{def}}{=} \left\{ \begin{bmatrix} V & r \\ 0 & 0 \end{bmatrix} \mid V \in T_S\mathbb{S}(4,\mathbb{R}),\ r \in \mathbb{R}^4 \right\}. \tag{22}$$

The tangent space at identity $T_I\mathbb{T}$ is denoted by the symbol $\mathfrak{t}$ and has the structure:

$$\mathfrak{t} \stackrel{\text{def}}{=} \left\{ \begin{bmatrix} H & r \\ 0 & 0 \end{bmatrix} \mid H \in \mathfrak{h}(4,\mathbb{R}),\ r \in \mathbb{R}^4 \right\}. \tag{23}$$

The Lie algebra $\mathfrak{t}$ is endowed with the standard Euclidean inner product of matrices:

$$\langle U_1, U_2\rangle_{\mathfrak{t}} \stackrel{\text{def}}{=} \operatorname{tr}(U_1^T U_2) = \operatorname{tr}(H_1^T H_2) + r_1^T r_2, \tag{24}$$

where symbol $\operatorname{tr}(\cdot)$ denotes the trace of the matrix argument and $U_1, U_2 \in \mathfrak{t}$.

For a matrix Lie group, the exponential map coincides with the matrix exponential, therefore, for the Lie group $(\mathbb{T}, \cdot, ^{-1}, I_5)$, the definition (1) particularizes to:

$$\exp_X(V) = X \exp(X^{-1}V), \tag{25}$$

where (with a slight notation clash) 'exp' denotes the matrix exponential, while its inverse, that is the logarithmic map, coincides with the principal matrix logarithm, namely:

$$\log_X(Y) = X \log(X^{-1}Y), \tag{26}$$

where (again with a slight notation clash) 'log', without any footer, denotes the matrix logarithm, and $X, Y, \in \mathbb{T}$, while $V \in T_X\mathbb{T}$.

3.2 Design of a Lie-group averaging algorithm over the Lie group $\mathbb{T}$

In order to compute a Lie-group average of a set of N transference matrices $Y_k \in \mathbb{T}$, it is necessary to minimize the spread function $\sigma^2(X)$ defined in (15). To such aim, it is necessary to compute the Riemannian gradient of the function $X \mapsto \langle\log_X(Y), \log_X(Y)\rangle_X$ for a given matrix $Y \in \mathbb{T}$. Given the chosen algebraic/geometric structure of the Lie group $\mathbb{T}$, such function may be defined as:

$$\varphi(X) \stackrel{\text{def}}{=} \langle\log(X^{-1}Y), \log(X^{-1}Y)\rangle_{\mathfrak{t}}. \tag{27}$$

According to the definition (3), the Riemannian gradient $\nabla_X\varphi$ is defined by the equation

$$\operatorname{tr}(V^T(XX^T)^{-1}\nabla_X\varphi) = \frac{d}{dt}\operatorname{tr}\left(\log^T\left(\Gamma_{X,V}^{-1}(t)Y\right)\log\left(\Gamma_{X,V}^{-1}(t)Y\right)\right)\Big|_{t=0}, \tag{28}$$

where $\Gamma_{X,V}(t) = X\exp(tX^{-1}V)$. Set $U \stackrel{\text{def}}{=} -X^{-1}V \in \mathfrak{t}$ and $\Theta \stackrel{\text{def}}{=} X^{-1}Y \in \mathbb{T}$. The inner product in the right-hand side of the equation (28) casts as:

$$F(t) \stackrel{\text{def}}{=} \operatorname{tr}(\log^T(\exp(tU)\Theta)\log(\exp(tU)\Theta)), \tag{29}$$

while the left-hand-side inner product casts as $-\mathrm{tr}((X^{-1}\nabla_X\varphi)^T U)$. It holds that:

$$\dot{F}(t) = 2\mathrm{tr}\left(\log^T(\exp(tU)\Theta)\left(\frac{d}{dt}\log(\exp(tU)\Theta)\right)\right). \tag{30}$$

Thanks to the notion of pushforward map, the derivative in the right-hand side of equation (30) casts as follows:

$$\frac{d}{dt}\log(\exp(tU)\Theta) = \log'_{\exp(tU)\Theta}\left(\exp'_{tU}(U)\Theta\right). \tag{31}$$

Recall from Subsection 2.1 that:

$$\log'_X(V) = -\sum_{k=1}^{\infty}\frac{(-1)^k}{k}\sum_{r=1}^{k}(X-I)^{r-1}V(X-I)^{k-r},$$
$$\exp'_{tU}(U) = \exp(tU)U.$$

As a consequence, it holds that: $\dot{F}(t) =$

$$-2\mathrm{tr}\left(\log^T(e^{tU}\Theta)\sum_{k=1}^{\infty}\frac{(-1)^k}{k}\sum_{r=1}^{k}(e^{tU}\Theta-I)^{r-1}e^{tU}U\Theta(e^{tU}\Theta-I)^{k-r}\right).$$

Since two analytic functions of the same matrix-variable commute, the above expression may be rewritten as:

$$\dot{F}(t) = -2\sum_{k=1}^{\infty}\frac{(-1)^k}{k}\sum_{r=1}^{k}\mathrm{tr}\left((e^{tU}\Theta-I)^{r-1}\log^T(e^{tU}\Theta)e^{tU}U\Theta(e^{tU}\Theta-I)^{k-r}\right).$$

By the properties of the trace operator, it follows that:

$$\begin{aligned}
\dot{F}(t) &= -2\sum_{k=1}^{\infty}\frac{(-1)^k}{k}\sum_{r=1}^{k}\mathrm{tr}\left(\log^T(\exp(tU)\Theta)\exp(tU)U\Theta(\exp(tU)\Theta-I)^{k-1}\right)\\
&= 2\sum_{k=1}^{\infty}(-1)^{k-1}\mathrm{tr}\left(\log^T(\exp(tU)\Theta)\exp(tU)U\Theta(\exp(tU)\Theta-I)^{k-1}\right)\\
&= 2\mathrm{tr}\left(\log^T(\exp(tU)\Theta)\exp(tU)U\Theta\sum_{k=1}^{\infty}(-1)^{k-1}(\exp(tU)\Theta-I)^{k-1}\right).
\end{aligned}$$

Recall from Subsection 2.1 that:

$$X^{-1} = \sum_{k=0}^{\infty}(-1)^k(X-I)^k,$$

therefore, the derivative $\dot{F}(t)$ may be recast as:

$$\begin{aligned}
\dot{F}(t) &= 2\mathrm{tr}\left(\log^T(\exp(tU)\Theta)\exp(tU)U\Theta(\exp(tU)\Theta)^{-1}\right)\\
&= 2\mathrm{tr}\left(\log^T(\exp(tU)\Theta)\exp(tU)U\Theta\Theta^{-1}\exp(-tU)\right)\\
&= 2\mathrm{tr}\left(\log^T(\exp(tU)\Theta)U\right).
\end{aligned} \tag{32}$$

Setting $t = 0$, it is readily obtained that $\dot{F}(0) = 2\mathrm{tr}\left(\log^T(\Theta)U\right)$. The equation (28) hence reads:

$$\mathrm{tr}((X^{-1}\nabla_X \varphi)^T X^{-1}V) = -2\mathrm{tr}\left(\log^T(X^{-1}Y)X^{-1}V\right). \tag{33}$$

Since the above equality must hold for every matrix $V \in T_X\mathbb{T}$, the sought Riemannian gradient reads:

$$\nabla_X \varphi = -2X\log(X^{-1}Y). \tag{34}$$

The above result is made use of in order to compute the gradient of the spread function (15), namely:

$$\nabla_X \sigma^2(X) = -\frac{2}{N}X\sum_{k=1}^{N}\log(X^{-1}Y_k). \tag{35}$$

The optimization algorithm (13) is customized as:

$$X_{a+1} = X_a \exp\left(\frac{2t_a}{N}\sum_{k=1}^{N}\log(X_a^{-1}Y_k)\right). \tag{36}$$

3.3 Optimal stepsize schedule for the averaging algorithm on $\mathbb{T}$

With the aim to compute a nearly-optimal stepsize schedule for the averaging algorithm (36), it is necessary to compute the first-order and second-order derivatives of the spread function $\sigma^2(X_{a+1})$, thought of as a function of the stepsize t_a, with respect to the parameter t_a at the point $t_a = 0$, in order to obtain the expansion:

$$\sigma^2(X_{a+1}) \approx \frac{1}{2}\phi_{2,a}t_a^2 + \phi_{1,a}t_a + \sigma^2(X_a), \tag{37}$$

where the $\phi_{i,a}$'s, $i = 1,\ 2$, are the coefficients of the expansion. In fact, the coefficients are computed by:

$$\phi_{2,a} = \left.\frac{d^2\sigma^2(X_{a+1})}{dt_a^2}\right|_{t_a=0},\ \phi_{1,a} = \left.\frac{d\sigma^2(X_{a+1})}{dt_a}\right|_{t_a=0}. \tag{38}$$

The nearly-optimal value of the stepsize that grants fastest descent is $t_a^\star \approx -\frac{\phi_{1,a}}{\phi_{2,a}}$, provided that $\phi_{2,a} > 0$.

In order to accomplish such calculations, it is convenient to rewrite the iterative algorithm (36) as follows:

$$X_{a+1} = X_a \exp\left(t_a U_a\right),\ U_a \stackrel{\mathrm{def}}{=} \frac{2}{N}\sum_{k=1}^{N}\log(X_a^{-1}Y_k). \tag{39}$$

The spread function $\sigma^2(X_{a+1})$ may thus be customized as:

$$\sigma^2(X_{a+1}) = \frac{1}{N}\sum_{k=1}^{N}\mathrm{tr}\left(\log^T(\exp(-t_aU_a)X_a^{-1}Y_k)\log(\exp(-t_aU_a)X_a^{-1}Y_k)\right). \tag{40}$$

Setting $U = -U_a$ and $\Theta = X_a^{-1} Y_k$ to ease the notation, the function within the summation is exactly the function $F(t)$ defined in (29), where the variable t plays the role of the stepsize t_a, while its first-order derivative is given by equation (32). Therefore, it holds that:

$$\phi_{1,a} = -\frac{2}{N} \sum_{k=1}^{N} \mathrm{tr}\left(\log^T(X_a^{-1} Y_k) U_a\right) = -\mathrm{tr}(U_a^T U_a). \tag{41}$$

On the basis of the first-order derivative (32), the second-order derivative $\ddot{F}(t)$ is readily found, in fact:

$$\ddot{F}(t) = 2\mathrm{tr}\left(\left(\log'_{\exp(tU)\Theta}(\exp'_{tU}(U)\Theta)\right)^T U\right).$$

Setting $t = 0$ yields:

$$\ddot{F}(0) = 2\mathrm{tr}\left(U^T \left(\log'_{\Theta}(U\Theta)\right)\right). \tag{42}$$

Near convergence, X_a is close to all samples Y_k's, hence, it may be assumed that $\Theta \approx I$. From Subection 2.1, it follows that:

$$\ddot{F}(0)|_{\Theta=I} = 2\mathrm{tr}\left(U^T \log'_I(U)\right) \approx 2\mathrm{tr}(U^T U).$$

As a consequence, the coefficient $\phi_{2,a}$ may be approximated as:

$$\phi_{2,a} \approx \frac{2}{N} \sum_{k=1}^{N} \mathrm{tr}(U_a^T U_a) = 2\mathrm{tr}(U_a^T U_a), \tag{43}$$

which is larger than zero (unless convergence is achieved, in which case it equals zero). Therefore, the obtained nearly-optimal stepsize schedule is constant:

$$t_a^\star \approx \frac{1}{2}. \tag{44}$$

The iteration algorithm (36) with stepsize (44) may be rewritten explicitly as:

$$X_{a+1} = X_a \exp\left(\frac{1}{N} \sum_{k=1}^{N} \log(X_a^{-1} Y_k)\right), \quad a = 0,\ 1,\ \ldots,\ A, \tag{45}$$

with $X_0 \in \mathbb{T}$ denoting an initial guess.

3.4 Harris 'exponential-mean-logarithm' transference

In the paper [11], Harris proposed an *exponential-mean-logarithm* formula to compute the average $X_\mathrm{H} \in \mathbb{T}$ of a set of optical system transference matrices $Y_k \in \mathbb{T}$. Such formula is expressed as:

$$X_\mathrm{H} \stackrel{\mathrm{def}}{=} \exp\left(\frac{1}{N} \sum_{k=1}^{N} \log Y_k\right). \tag{46}$$

In the subsequent paper [12], it was proved that such averaging expression is consistent with the geometry of the space $\mathbb{T}$.

Comparing the expression (46) with the algorithm (45), it is readily seen that the *exponential-mean-logarithm rule coincides with the first iteration of the optimization algorithm*, where the initial guess is chosen as the identity of the Lie group $\mathbb{T}$.

For objective evaluation purposes, it is convenient to define the spread of the transference matrices around Harris' average optical transference matrix as:

$$\sigma_{\mathrm{H}}^2 \stackrel{\text{def}}{=} \sigma^2(X_{\mathrm{H}}) = \frac{1}{N}\sum_{k=1}^{N} \operatorname{tr}\left(\log^T(X_{\mathrm{H}}^{-1}Y_k)\log(X_{\mathrm{H}}^{-1}Y_k)\right). \tag{47}$$

3.5 An interpolatory scheme on the space of optical transference matrices

Assume that two optical system transference matrices $Y_1, Y_2 \in \mathbb{T}$ are given and that their geometric interpolation is sought in $\mathbb{T}$. The solution to such problem was discussed in Subsection 2.3 and is given by the curve:

$$X(t) = Y_1 \exp(t \log(Y_1^{-1}Y_2)),\ 0 \le t \le 1. \tag{48}$$

The matrix logarithm $\log(Z)$ exists as long as its matrix-argument Z satisfies the inequality $\|Z - I\| < 1$. As a consequence, the interpolation rule (48) makes sense as long as the two inhomogeneous symplectic matrices Y_1 and Y_2 are "sufficiently close" one to another, namely, they are such that $\|Y_1^{-1}Y_2 - I\| < 1$. In addition, since the matrix exponential exists for every value taken by its argument, the *interpolation* parameter range $[0,\ 1]$ may be extended besides its margins affording *extrapolation* of the data.

4 Numerical experiments

In the present Section, the averaging algorithm (36) – based on Riemannian-gradient optimization – is used to compute averages over the inhomogeneous symplectic group $\mathbb{T}$ in order to answer to the original question whether Harris' averaging rule is optimal with respect to a Riemannian-gradient-steepest descent procedure. In order to compare different optical system transference matrices, the following *discrepancy* measure $\mathrm{D} : \mathbb{T} \times \mathbb{T} \to \mathbb{R}_{\ge 0}$ is defined:

$$\mathrm{D}(X, Y) \stackrel{\text{def}}{=} \operatorname{tr}\left(\log^T(X^{-1}Y)\log(X^{-1}Y)\right). \tag{49}$$

The discrepancy $\mathrm{D}(X, Y)$ coincides with the squared Riemannian distance between the inhomogeneous symplectic matrices $X \in \mathbb{T}$ and $Y \in \mathbb{T}$ with respect to the metrics (24).

4.1 Numerical experiments on averaging two inhomogeneous symplectic matrices

An experiment described in [11] is repeated in order to confirm numerically that standard (i.e., arithmetic and geometric) averaging is not suitable for data belonging to the space $\mathbb{T}$ and to compare the proposed optimziation-based averaging algorithm with Harris' rule. The average of the following two matrices is sought:

$$Y_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \; Y_2 = \begin{bmatrix} 1 & 0 & \frac{1}{10} & 0 & 0 \\ 0 & 1 & 0 & \frac{1}{10} & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{50}$$

Comparing the matrix Y_1 above with the general form of an optical transference matrix $\begin{smallmatrix} S & \delta \\ 0 & 1 \end{smallmatrix}$, it is easy to identify the astigmatic submatrix S_1 and the decentering vector $\delta_1 = 0$. Recalling the definition of the skew-symmetric unit matrix J given in the Subsection 3.1, it is readily verified that $S_1^T J S_1 = J$, hence that S_1 is a real symplectic matrix and, therefore, that Y_1 is an inhomogeneous symplectic (optical transference) matrix. The same may be concluded about the matrix Y_2.

A simple way to compute the midpoint between the above two matrices would be to invoke their *arithmetic average* $M_{\mathrm{a}} \stackrel{\mathrm{def}}{=} \frac{1}{2}(Y_1 + Y_2)$. The rationale of arithmetic averaging is that the mean value of two data is the midpoint of a segment connecting them. Since the set $\mathbb{T}$ of optical transference matrices is a curved space, it is very unlikely that a rigid segment will be entirely contained in $\mathbb{T}$, hence such midpoint will most probably lay outside $\mathbb{T}$. Computing the arithmetic average of the matrices (50) and extracting the astigmatic submatrix S_{a}, it is found that:

$$S_{\mathrm{a}}^T J S_{\mathrm{a}} = \left(1 + \frac{1}{40}\right) J.$$

Since the matrix S_{a} is *not* symplectic and hence $M_{\mathrm{a}} \notin \mathbb{T}$, the matrix M_{a} cannot be taken as an average of the transference matrices Y_1, Y_2 (it does not meet the first of the requirements stated in the Subsection 2.2).

Another known way of defining the average of two matrices is by the geometric mean $M_{\mathrm{g}}^{12} \stackrel{\mathrm{def}}{=} \sqrt{Y_1 Y_2}$ or $M_{\mathrm{g}}^{21} \stackrel{\mathrm{def}}{=} \sqrt{Y_2 Y_1}$. Since the matrix square root is computed by $\sqrt{Z} = \exp\left(\frac{1}{2}\log(Z)\right)$, there are two versions of the geometric mean:

$$M_{\mathrm{g}}^{12} = \exp\left(\tfrac{1}{2}\log(Y_1 Y_2)\right) \text{ and } M_{\mathrm{g}}^{21} = \exp\left(\tfrac{1}{2}\log(Y_2 Y_1)\right).$$

By the Lie-group property of the space $\mathbb{T} \ni Y_1, Y_2$, the matrix product $Y_1 Y_2$ belongs to $\mathbb{T}$, hence the matrix $\frac{1}{2}\log(Y_1 Y_2)$ belongs to the Lie algebra $\mathfrak{t}$ (provided that the logarithm exists), which implies that the matrix M_{g}^{12} is an optical transference pattern. The same may be concluded about the matrix M_{g}^{21}.

It is worth recalling that, in general, the matrix logarithm is not distributive with respect to the matrix product, namely, $\log(Y_1 Y_2) \neq \log(Y_1) + \log(Y_2)$, therefore, in general, $M_{\mathrm{g}}^{12} \neq M_{\mathrm{g}}^{21}$ (unless the matrices Y_1 and Y_2 commute, namely, $Y1_Y 2 - Y_2 Y1 = 0$). Moreover, such observation implies that none of the geometric means coincide with Harris' mean transference. Numerically, the discrepancy $\mathrm{D}(M_{\mathrm{g}}^{12}, M_{\mathrm{g}}^{21}) \approx 0.1064$. Consequently, none of them can be retained as *the* mean transference matrix and choosing one of them would be an unwarranted arbitrariness. Moreover, the above geometric means do not warrant equidistance from the data, in fact, for example, it is found that $\mathrm{D}(M_{\mathrm{g}}^{21}, Y_1) \approx 0.6536$, while $\mathrm{D}(M_{\mathrm{g}}^{21}, Y_2) \approx 0.7166$.

The average matrix computed by Harris' rule (46) is:

$$X_{\mathrm{H}} = \begin{bmatrix} 0.9875 & 0 & 0.0498 & 0 & 0 \\ 0 & 0.9875 & 0 & 0.0498 & 0 \\ -0.4979 & 0 & 0.9875 & 0 & 0 \\ 0 & -0.4979 & 0 & 0.9875 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Extracting the astigmatic submatrix S_{H}, it is readily verified that it is real symplectic, hence that $X_{\mathrm{H}} \in \mathbb{T}$.

An average matrix may be computed through the optimization algorithm (36) for comparison. The numerical behavior of the proposed averaging algorithm (36) is evaluated through the following figures:

- The value of the spread $\sigma^2(X_a)$ corresponding to each iteration, with $a = 0, 1, \ldots, A$. The spread function is expected to show a decreasing trend.
- Since the sequence of $\mathbb{T}$-matrices X_a is expected to converge to a matrix that locates amid the samples Y_k, the quantities $X_a^{-1}Y_k$ are expected to converge to the identity. Hence, the following *eigenvalues spread*:

$$\frac{1}{N}\sum_{k=1}^{N}\sum_{i=1}^{4} |\lambda_a^{(i,k)} - 1|^2, \tag{51}$$

where $\lambda_a^{(i,k)}$ denotes the i^{th} eigenvalue of the matrix $X_a^{-1}Y_k$, is expected to converge to a minimal value toward zero. (In general, it does not tend to zero, unless the optical transference sample-matrices Y_k are equal one to another.)

The Figure 2 displays the result of running the algorithm (36). The number of iterations was set to $A = 20$. The initial guess was set as $X_0 = Y_1$. From the top panel of Figure 2, it may be observed that the averaging algorithm converges steadily and that the spread σ^2_{H} of the distribution of the transference matrices around the mean transference matrix computed by the Harris' rule (46) is comparable with the spread $\sigma^2(X_A)$ corresponding to the Lie-group average. The bottom panel of Figure 2 also confirms that the eigenvalue spread tends to a low value, in such experiment.

The average matrix X_A computed by the algorithm (36) coincides with Harris' average matrix X_{H} up to the fourth digit. The discrepancy $\mathrm{D}(X_A, X_{\mathrm{H}}) \approx 5 \times 10^{-5}$, which looks low compared to the discrepancy $\mathrm{D}(Y_1, Y_2) \approx 1.4015$. It is also interesting to compare the distance $\sqrt{\mathrm{D}(X_A, Y_1)} \approx 0.8336$ with the distance $\sqrt{\mathrm{D}(X_A, Y_2)} \approx 0.8405$, which confirms that the computed average matrix is truly a midpoint (i.e., it is located at approximately equal distances from the two data).

4.2 Numerical experiments on averaging several inhomogeneous symplectic matrices

In order to test numerically the averaging algorithm (36), N random matrices $Y_k \in \mathbb{T}$ are generated, around a given center, by exploiting the geodesic curve

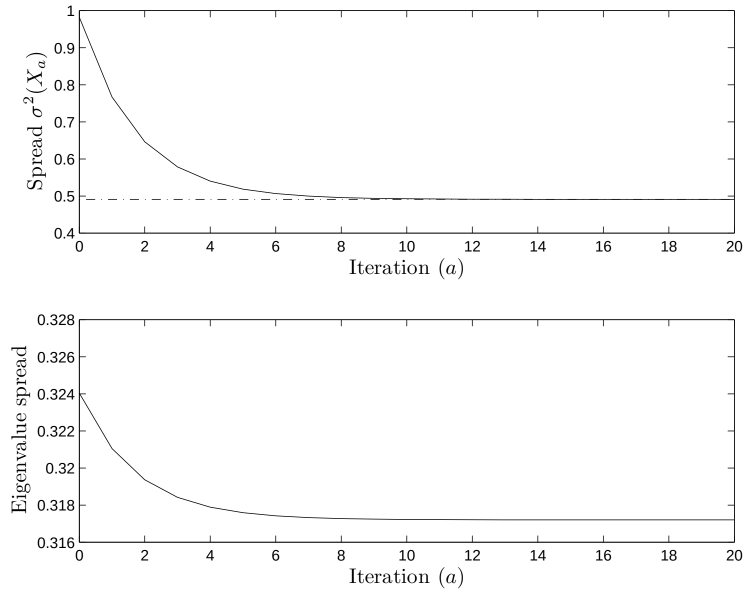

Figure 2: Experiment on averaging two optical transference matrices. Top panel: Discrepancy $\sigma^2(X_a)$ (solid line) versus iteration index compared with the discrepancy σ^2_{H} (dot-dashed line). Bottom panel: Eigenvalue spread during iteration.

$\Gamma_{X,V}(t) = X\exp(tX^{-1}V)$, with $X \in \mathbb{T}$ and $V \in T_X\mathbb{T}$. The central optical transference matrix may be generated by a curve departing from the identity:

$$X_\star = \exp(U), \tag{52}$$

with U randomly generated in $\mathfrak{t}$. The matrices in the Lie-algebra $\mathfrak{t}$ have the general shape (23). Random Hamiltonian matrices may be generated by the rule $H = \frac{1}{2}J(B^T + B)$, with $B \in \mathbb{R}^{4\times 4}$ being any unstructured random matrix. The sample-set of matrices Y_k may now be generated via the rule:

$$Y_k = X_\star \exp(U_k), \tag{53}$$

where U_k is a randomly generated matrix in $\mathfrak{t}$.

The numerical behavior of the averaging algorithm (36) is evaluated through the following figures of performance:

- The value of the spread $\sigma^2(X_a)$ corresponding to each iteration.
- The discrepancy $\mathrm{D}(X_A, X_\star)$ evaluated at the end of the iterations, compared with the discrepancy $\mathrm{D}(Y_k, X_\star)$ between each sample Y_k and the actual center of the distribution $X_\star$. The discrepancy between the computed average X_A and the actual center of the distribution is expected to locate in an intermediate level between the discrepancies between each of the samples and their actual center $X_\star$.
- The eigenvalue spread (51.

Figure 3 displays the result of a numerical experiment obtained with $N = 50$ transference matrices. The number of iterations of the algorithm was set to $A = 20$. The initial guess was set to $X_0 = I$. The top panel of Figure 3 con-

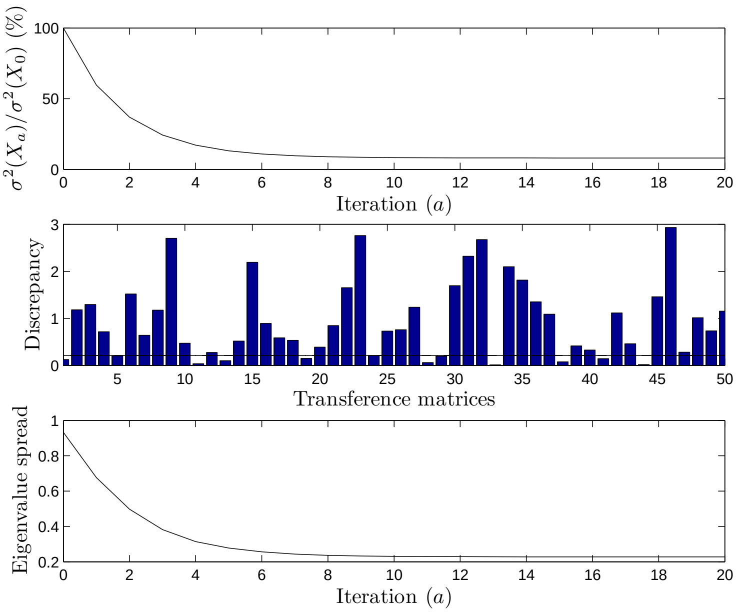

Figure 3: Experiment on averaging 50 random optical transference matrices. Top panel: Relative spread σ_a^2/σ_0^2, in percentage, versus iteration index. Middle panel: Discrepancy between the optical transference matrices Y_k and the central transference $X_\star$ (bars) compared with the discrepancy between the computed average matrix X_A and the central transference (horizontal solid line) and with the discrepancy between Harris' average transference matrix X_{H} with the central transference (horizontal dot-dashed line). Bottom panel: Eigenvalue spread (51) during iteration.

firms that the averaging algorithm (36) converges steadily. The middle panel of Figure 3 shows that the discrepancy $\mathrm{D}(X_A, X_\star)$ is an average value if compared with the discrepancies $\mathrm{D}(Y_k, X_\star)$. The same panel shows that the spread σ_{H}^2 of the distribution of the transference matrices around the average transference matrix computed by the Harris' rule (46) coincides with the spread $\sigma^2(X_A)$ corresponding to the computed Lie-group average. The bottom panel of Figure 3 also confirms that the eigenvalue spread tends to a minimum value.

4.3 Numerical experiments on averaging 'eyes'

An experiment described in [9, Table 2] is repeated. It is interesting to note that, in such experiment, the matrices Y_k are quite ill-conditioned and that they differ considerably from the identity matrix. However, these facts do not influence the numerical behavior of the averaging algorithm that only depends on the products $X_a^{-1}Y_k$ which keep close to the identity matrix, provided that X_0 be chosen sufficiently close to all samples. An effective strategy to ensure that the initial guess X_0 is located amidst the data cloud is to choose X_0 as one of the inhomogeneous symplectic matrices to average.

Figure 4 displays the result of running the algorithm (36). The number of iterations of the algorithm was set to $A = 20$. The initial guess was set as one of the matrices to average. The averaging algorithm converges steadily and reaches

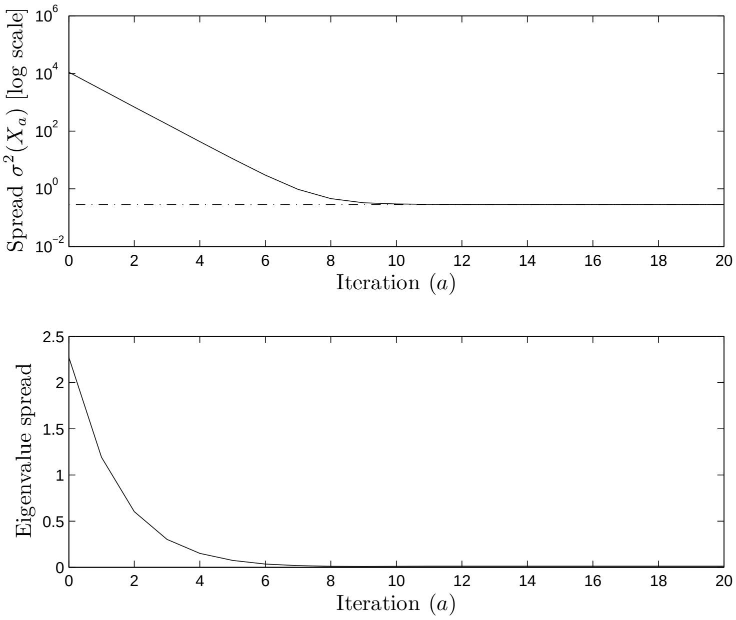

Figure 4: Experiment on averaging eyes. Top panel: Discrepancy $\sigma^2(X_a)$ (solid line) versus iteration index a compared with the discrepancy σ^2_{H} (dot-dashed line). Bottom panel: Eigenvalue spread (51) during iteration.

a discrepancy value $\sigma^2(X_A)$ comparable with the spread σ^2_{H}. The bottom panel of Figure 4 confirms that the eigenvalue spread reaches low values.

5 Conclusion

The present manuscript aims at describing a Riemannian-steepest-descent based solution to the problem of computing the average out of a set of optical system transference matrices. It was shown that such transference matrices form a Lie group that coincides with the inhomogeneous real symplectic group $\mathbb{IS}(4, \mathbb{R})$ and, as such, the averaging problem may be tackled in the context of *Lie-group averaging* theory.

The devised averaging algorithm was compared with the Harris' exponential-mean-logarithm averaging rule, especially developed by Harris to compute the average character out of a set of biological optical systems. Harris' rule was developed empirically and was not known to be optimal under any optimality principle.

The results of numerical experiments show that the iterative algorithm based on gradient steepest descent implemented by exponential-map stepping converges to solutions that are in good agreement with the value obtainable by applying Harris' exponential-mean-logarithm averaging rule. Such results confirm that Harris' exponential-mean-logarithm averaging rule is optimal in the sense that it minimizes the Lie-group spread function (15) corresponding to the Euclidean inner product.

6 Acknowledgments

I had the great pleasure of meeting Prof. William Harris during a Workshop on matrix analysis in Manchester in April 2013. I would like to thank Prof. Harris for sharing his papers on optical system transference theory and for his encouragement on pursuing the present research endeavor.

References

[1] E. Celledoni and S. Fiori, *Descent methods for optimization on homogeneous manifolds*, Mathematics and Computers in Simulation, Vol. 79, No. 4, pp. 1298 – 1323, December 2008

[2] G.S. Chirikjian, *Stochastic Models, Information Theory, and Lie Groups, Volume 1: Classical Results and Geometric Methods*, 2009 Edition, Birkhäuser, September 2009

[3] X. Duan, H. Sun and L. Peng, *Riemannian means on special Euclidean group and unipotent matrices group*, The Scientific World Journal, Vol. 2013, Article ID: 292787, 9 pages, DOI: 10.1155/2013/292787, 2013

[4] S. Fiori, *Quasi-geodesic neural learning algorithms over the orthogonal group: A tutorial*, Journal of Machine Learning Research, Vol. 6, pp. 743 – 781, May 2005

[5] S. Fiori and T. Tanaka, *An algorithm to compute averages on matrix Lie groups*, IEEE Transactions on Signal Processing, Vol. 57, No. 12, pp. 4734 – 4743, December 2009

[6] S. Fiori, *Averaging over the Lie group of optical systems transference matrices*, Frontiers of Electrical and Electronic Engineering in China (Springer), Special issue of the Sino foreign-interchange Workshop on Intelligence Science and Intelligent Data Engineering 2010 - Part A, Vol. 6, No. 1, pp. 137 – 145, March 2011

[7] S. Fiori, *Solving minimal-distance problems over the manifold of real symplectic matrices*, SIAM Journal on Matrix Analysis and Applications, Vol. 32, No. 3, pp. 938 – 968, 2011

[8] S. Fiori, T. Kaneko and T. Tanaka, *Tangent-bundle maps on the Grassmann Manifold: Application to empirical arithmetic averaging*, IEEE Transactions on Signal Processing, Vol. 63, N. 1, pp. 155 – 168, January 2015

[9] R.D. van Gool and W.F. Harris, *The concept of the average eye*, The South African Optometrist, Vol. 64, No. 2, pp. 38 – 43, 2005

[10] W.F. Harris, *Paraxial ray tracing through noncoaxial astigmatic optical systems, and a* 5×5 *augmented system matrix*, Optometry and Vision Science, Vol. 71, No. 4, pp. 282 – 285, 1994

[11] W.F. Harris, *The average eye*, Ophthalmic and Physiological Optics, Vol. 24, No. 6, pp. 580 – 585, 2004

[12] W.F. Harris and J.R. Cardoso, *The exponential-mean-log-transference as a possible representation of the optical character of an average eye*, Ophthalmic and Physiological Optics, Vol. 26, No. 4, pp. 380 – 383, 2006

[13] T. Kaneko, S. Fiori and T. Tanaka, *Empirical arithmetic averaging over the compact Stiefel manifold*, IEEE Transactions on Signal Processing, Vol. 61, No. 4, pp. 883 – 894, February 2013

[14] H. Karcher, *Riemannian center of mass and mollifier smoothing*, Communications on Pure and Applied Mathematics, Vol. 30, pp. 509 – 541, 1977

[15] M.P. Keating, *A matrix formulation of spectacle magnification*, Ophthalmic and Physiological Optics, Vol. 2, No. 2, 145 – 158, 1982

[16] D.G. Luenberger, *The gradient projection methods along geodesics*, Management Science, Vol. 18, No. 11, pp. 620 – 631, 1972

[17] L.M. Machado and F.S. Leite, *Optimization on quadratic matrix Lie groups*, Pré-Publicações do Departamento de Matemática, No. 02-21, University of Coimbra, 2002

[18] A. Martínez-Finkelshtein, D. Ramos López, G.M. Castro and J.L. Alió, *Adaptive cornea modeling from keratometric data*, Investigative Ophthalmology & Visual Science, Vol. 52, No. 8, pp. 4963 – 4970, July 2011

[19] M. Spivak, *A Comprehensive Introduction to Differential Geometry*, Volume 1, 2^{nd} Edition, Berkeley, CA: Publish or Perish Press, 1979

[20] R.M. Tudoran, *On a class of three-dimensional quadratic Hamiltonian systems*, Applied Mathematics Letters, Vol. 25, No. 9, pp. 1214 – 1216, September 2012

[21] R.-B. Wu, R. Chakrabarti and H. Rabitz, *Critical landscape topology for optimization on the symplectic group*, Journal of Optimization Theory and Applications, Vol. 145, No. 2, pp. 387 – 406, May 2010