Least-Squares on the Real Symplectic Group

Simone Fiori
Dipartimento di Ingegneria Biomedica, Elettronica
e Telecomunicazioni (DiBET)
Facoltà di Ingegneria, Università Politecnica delle Marche
Via Brecce Bianche, Ancona I-60131, Italy
E-mail: s.fiori@univpm.it

January 6, 2010

Abstract

The present paper discusses the problem of least-squares over the real symplectic group of matrices $\text{Sp}(2n, \mathbb{R})$. The least-squares problem may be extended from flat spaces to curved spaces by the notion of geodesic distance. The resulting non-linear minimization problem on manifold may be tackled by means of a gradient-descent algorithm tailored to the geometry of the space at hand. In turn, gradient steepest descent on manifold may be implemented through a geodesic-based stepping method. As the space $\text{Sp}(2n, \mathbb{R})$ is a non-compact Lie group, it is convenient to endow it with a pseudo-Riemannian geometry. Indeed, a pseudo-Riemannian metric allows the computation of geodesic arcs and geodesic distances in closed form on $\text{Sp}(2n, \mathbb{R})$.

Keywords: Real symplectic group of matrices. Pseudo-Riemannian geometry. Geodesic least-squares. Geodesic stepping method.

1 Introduction

Least-squares problems over compact Lie groups have been extensively studied due to their broad application range. A feature of these problems is that the domain of the objective function to optimize, being a compact Lie group, may be endowed with the structure of a differential manifold with a bi-invariant Riemannian metric. The formulation of least-squares problems on compact Riemannian Lie groups relies on closed forms of geodesic curves and geodesic distances.

On the Euclidean matrix space $\mathbb{R}^{n \times m}$, a (weighted) least-squares problem may be formulated via the criterion function $f : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}_0^+$:

$$f(x) \overset{\text{def}}{=} \sum_k \alpha_k \|x - \tau_k\|^2, \quad (1)$$

where $\tau_k \in \mathbb{R}^{n \times m}$ denote optimization target matrices, $\alpha_k > 0$ denote weights and symbol $\| \cdot \|$ denotes the Frobenius matrix norm. On a Riemannian curved
manifold $M$, the above least-squares problem may be reformulated via the generalized criterion function $f : M \rightarrow \mathbb{R}^+_0$:

$$f(x) \overset{\text{def}}{=} \sum_k \alpha_k d^2(x, \tau_k), \quad (2)$$

where $\tau_k \in M$ and the function $d(\cdot, \cdot)$ denotes a geodesic distance on the manifold $M$ corresponding to the metric that the manifold is endowed with.

On non-compact Riemannian Lie groups, the problem of formulating a least-squares criterion and of its optimization is substantially more involved, because it might be hard to compute geodesic distances in closed form. However, a non-compact Lie group may be treated as a pseudo-Riemannian manifold with a bi-invariant pseudo-Riemannian metric [20].

One of such non-compact Lie groups is the real symplectic group.

Real symplectic matrices form an algebraic group under standard matrix multiplication and inversion, denoted as $\text{Sp}(2n, \mathbb{R})$. The least-squares problem over the set of real symplectic matrices plays an important role in applied fields. Some applications are:

- **Quantum computing** [3, 16]: An important application of quantum mechanics is quantum computing, as quantum mechanics allows for information processing that cannot be performed classically. In particular, it may be possible to design algorithms on a quantum computer that are more efficient than on a classical computer.

- **Control of beam systems in particle accelerators** [8]: Lie-group tools may be applied to the characterization of beam dynamics in charged-particle optical systems. These methods are applicable to accelerator design, charge-particle beam transport and electron microscopes. Lie-group methods potentially provide a way to analyze and control non-linear behavior with the same completeness of linear behavior.

- **Computational ophthalmology** [17, 18]: In the study of optical systems in ophthalmology, it is assumed that the optical nature of a centered optical system is completely described by a real symplectic matrix.

- **Vibration analysis** [23]: Transfer matrices are widely used for the dynamic analysis of engineering structures as well as for static analysis, and are particularly useful in the treatment of repetitive structures. Transfer matrices are real symplectic.

- **Control theory** [10]: Real symplectic matrices find applications in linear control theory for discrete-time systems.

Other applications may be found in coding theory [5] and in time-series prediction [2].

Although some noticeable results are available about the real symplectic group [9, 10], optimization on the real symplectic group appears to be far less studied than for other Lie groups. In the present manuscript, we discuss the problem of least-squares on the real symplectic group. After a review of results known from literature about optimization on the real symplectic manifold, an optimization method based on endowing it with a pseudo-Riemannian geometry will be discussed.
2 Criterion optimization on Riemannian and on pseudo-Riemannian manifolds

For a reference on differential geometry, see [24].

Let \( M \) be a smooth manifold. The tangent space at \( x \in M \) to the manifold is denoted by \( T_x M \). A metric on \( M \) is a non-degenerate, smooth, symmetric, bilinear map which assigns a real number to pairs of tangent vectors at each tangent space of the manifold \( M \). Let us denote the metric by \( \langle \cdot, \cdot \rangle_x : T_x M \times T_x M \to \mathbb{R} \).

A Riemannian manifold is a smooth manifold endowed with a positive definite metric, namely:

\[
\langle v, v \rangle_x \geq 0, \forall v \in T_x M.
\]

The metric \( \langle \cdot, \cdot \rangle_x \) also defines the norm \( \| v \|_x = \sqrt{\langle v, v \rangle_x} \) for \( v \in T_x M \).

The geodesic curve connecting two points \( x_1, x_2 \in M \) is the curve \( G(t) \in M \), parameterized by \( t \in [0, 1] \), that minimizes the energy integral:

\[
\int_0^1 \langle \dot{x}, \dot{x} \rangle_x dt,
\]

under the normalization condition that \( \langle \dot{G}, \dot{G} \rangle_G \) is constant. By the calculus of variation on manifold, the geodesic equation may be written in normal form as:

\[
\ddot{x} + \Gamma_x (\dot{x}, \dot{x}) = 0.
\]

In the above expressions, the over-dot and the double over-dot denote first-order and second-order derivation with respect to parameter \( t \), respectively, while symbol \( \Gamma (\cdot, \cdot) \) denotes the Christoffel operator. The solution of the geodesic equation may be written in terms of two known quantities that serve as boundary conditions for the second-order geodesic differential equation. The values \( x = x(0) \in M \) and \( v = \dot{x}(0) \in T_x M \) might be specified, in which case the solution of the geodesic equation \( (5) \) will be denoted as \( G_{x,v}(t) \).

The squared geodesic distance between the geodesic’s endpoints is defined as:

\[
d^2(x_1, x_2) \overset{\text{def}}{=} \left( \int_0^1 \langle \dot{G}, \dot{G} \rangle_G dt \right)^2 = \langle \dot{G}, \dot{G} \rangle_G \bigg|_{t=0}.
\]

Note that if the geodesic curve is expressed as \( G_{x,v}(t) \), then the squared geodesic distance equals \( \| v \|_x^2 \).

The gradient \( \nabla_x f \in T_x M \) of a regular criterion function \( f : M \to \mathbb{R} \) in a point \( x \in M \) may be defined as:

\[
\nabla_x f = (df)^{\sharp},
\]

where symbol \( \sharp \) denotes the ‘sharp’ isomorphism and symbol \( df \) denotes differential. On a Riemannian manifold, the Riemannian gradient may be computed by the metric compatibility condition:

\[
\langle \partial_x f, v \rangle^E = \langle \partial_x f, v \rangle_x, \forall v \in T_x M,
\]

where symbol \( \langle \cdot, \cdot \rangle^E \) denotes Euclidean metric.
When a Riemannian manifold of interest $\mathcal{M}$ and a regular criterion function $f: \mathcal{M} \to \mathbb{R}$ are specified, a known optimization rule is ‘gradient steepest descent’, that may be expressed as the differential equation on manifold:

$$\dot{x} = -\nabla_x f.$$ 

(9)

The gradient flow $x(t)$ associated to this system tends toward a local minimum of the criterion function $f$, in fact:

$$\dot{f} = \langle \nabla_x f, \dot{x} \rangle_x = -\|\nabla_x f\|_x^2 \leq 0,$$

with equality holding if and only if $\nabla_x f = 0$, namely, when the flow $x(t)$ approaches a stationary point of the criterion $f$. The optimization system (9) is based on the knowledge that the Riemannian gradient $\nabla_x f \in T_x \mathcal{M}$ points toward the direction of the maximum growth of the function $f$ around the point $x \in \mathcal{M}$.

The basic idea to implement the optimization scheme represented by equation (9) is to replace the continuous-time state-variable $x(t) \in \mathcal{M}$ with a discrete-time state-variable $x_k \in \mathcal{M}$. This operation requires a numerical scheme of integration of the equation (9).

The differential equation (9) may be solved numerically by the help of the notion of geodesic curve, namely, by the numerical optimization algorithm [6, 12, 13]:

$$x_{k+1} = G_{x_k, -\nabla_x f}(\eta),$$

(10)

where $\eta \in [0, 1]$ denotes an integration step-size and $G_{x,v}(t)$ denotes a geodesic arc departing from the point $x \in \mathcal{M}$ with initial direction $v \in T_x \mathcal{M}$ and parameter $t \in [0, 1]$.

A pseudo-Riemannian manifold is a manifold endowed with a metric that is not positive definite. (On a pseudo-Riemannian manifold $\mathcal{M}$, the quantity $\|v\|_x^2$ may be positive, negative or null even for $0 \neq v \in T_x \mathcal{M}$.)

In order to extend the gradient-based optimization algorithm (10) to the case of pseudo-Riemannian manifold $\mathcal{M}$, it is necessary to compute geodesic arcs and gradients on $\mathcal{M}$.

The basic idea to cope with pseudo-Riemannian manifolds is to decompose each tangent space $T_x \mathcal{M}$ as follows:

$$\begin{align*}
T^+_x \mathcal{M} & \overset{\text{def}}{=} \{v \in T_x \mathcal{M} \text{ such that } \|v\|_x^2 > 0\}, \\
T^0_x \mathcal{M} & \overset{\text{def}}{=} \{v \in T_x \mathcal{M} \text{ such that } \|v\|_x^2 = 0\}, \\
T^-_x \mathcal{M} & \overset{\text{def}}{=} \{v \in T_x \mathcal{M} \text{ such that } \|v\|_x^2 < 0\}.
\end{align*}$$

(11)

The notion of geodesics extends to pseudo-Riemannian manifolds by the calculus of variation on the energy integral (4). A geodesic arc $G_{x,v}(t)$ will be the solution of the differential equation (5). On a pseudo-Riemannian manifold $\mathcal{M}$, however, it holds $\|v\|_x^2 < 0$ for $v \in T^-_x \mathcal{M}$, therefore the notion of squared pseudo-Riemannian geodesic distance may be defined as:

$$d^2(x_1, x_2) = \|v\|_x^2.$$

(12)

Likewise, pseudo-Riemannian gradient defines as $\nabla_x f = (df)^2$ again. It is, however, worth remarking that the pseudo-Riemannian gradient $\nabla_x f$ does not
point to the direction of the maximum growth of the function $f$. Therefore, the optimization equation (9) needs to turn into:

$$\dot{x} = \begin{cases} -\nabla_x f & \text{if } \nabla_x f \in T_x^+ M \cup T_x^0 M, \\ \nabla_x f & \text{if } \nabla_x f \in T_x^- M, \end{cases} \tag{13}$$

whose solution is a minimization flow. In fact, it induces the dynamics:

$$f = \begin{cases} -\|\nabla_x f\|^2_x & \text{if } \nabla_x f \in T_x^+ M \cup T_x^0 M \\ \|\nabla_x f\|^2_x & \text{if } \nabla_x f \in T_x^- M \end{cases} \leq 0. \tag{14}$$

The equation (13) may be solved numerically by the optimization algorithm:

$$x_{k+1} = \begin{cases} G_{x_k, -\nabla x_k f}(\eta) & \text{if } \nabla_x f \in T_x^+ M \cup T_x^0 M, \\ G_{x_k, \nabla x_k f}(\eta) & \text{if } \nabla_x f \in T_x^- M, \end{cases} \tag{15}$$

where symbol $G_{x, v}(t)$ denotes a pseudo-Riemannian geodesic arc departing from the point $x \in M$ with initial direction $v \in T_x M$ and parameter $t \in [0, 1]$ and the succession $x_k \in M$ denotes an approximation of the actual solution of the differential equation (13).

### 3 The real symplectic group

The present section aims at recalling the definition of the real symplectic group and its properties, along with some recent results about optimization on it.

#### 3.1 Definitions and properties

The real symplectic group is defined as follows:

$$\text{Sp}(2n, \mathbb{R}) \overset{\text{def}}{=} \{ x \in \mathbb{R}^{2n \times 2n} \mid x^T q_{2n} x = q_{2n} \}, \tag{16}$$

$$q_{2n} \overset{\text{def}}{=} \begin{bmatrix} 0_n & e_n \\ -e_n & 0_n \end{bmatrix}, \tag{17}$$

where symbol $e_n$ denotes the $n \times n$ identity matrix, while symbol $0_n$ denotes a whole-zero $n \times n$ matrix. The skew-symmetric matrix $q_{2n}$ enjoys the following properties:

$$q_{2n}^2 = -e_{2n}, \tag{18}$$

$$q_{2n}^{-1} = q_{2n}^T = -q_{2n}. \tag{19}$$

From now on, the subscript $2n$ on the symbol $q_{2n}$ will drop for a tidier notation.

The space $\text{Sp}(2n, \mathbb{R})$ is a curved smooth manifold that may also be endowed with an algebraic-group structure (namely, group multiplication and group inverse and possesses a identity element) in a manner that is compatible with the manifold structure. Therefore, the space $\text{Sp}(2n, \mathbb{R})$ has the structure of a Lie group (of dimension $n(2n + 1)$).

In particular, standard matrix multiplication and inversion work as algebraic-group operations. For the matrix multiplication, the property may be proven by noting that, for every $x, y \in \text{Sp}(2n, \mathbb{R})$, the product $xy \in \text{Sp}(2n, \mathbb{R})$, in fact:

$$(xy)^T q(xy) = y^T (x^T q x) y = y^T q y = q.$$
Any symplectic matrix $x \in \text{Sp}(2n, \mathbb{R})$ is such that $\det^2(x) = 1$, where symbol $\det(\cdot)$ denotes determinant. Therefore, symplectic matrices are invertible. It may be proven that if matrix $x \in \text{Sp}(2n, \mathbb{R})$, then also matrix $x^{-1} \in \text{Sp}(2n, \mathbb{R})$ by negation, namely, by trying to prove that $(x^{-1})^T qx^{-1} \neq q$. It holds:

$$(x^{-1})^T qx^{-1} \neq q \Rightarrow -(x^T)^{-1} qx^{-1} \neq -q \Rightarrow q \neq x^T qx,$$

that would contradict the hypothesis that $x \in \text{Sp}(2n, \mathbb{R})$. The identity element of the group $\text{Sp}(2n, \mathbb{R})$ is clearly the matrix $e_{2n}$.

In addition, the following identities hold:

$$\det(x) = 1, \quad (20)$$
$$x^T = -qx^{-1}q, \quad (21)$$
$$x^{-T} = -qxq. \quad (22)$$

From the identity $(20)$, it follows that the group $\text{Sp}(2n, \mathbb{R})$ is a subgroup of $\text{Si}(2n, \mathbb{R})$ as well as of $\text{Gl}(2n, \mathbb{R})$. The proof of identity $(20)$ is far from trivial.

The tangent space $T_x \text{Sp}(2n, \mathbb{R})$ has structure:

$$T_x \text{Sp}(2n, \mathbb{R}) = \{ v \in \mathbb{R}^{2n \times 2n} | v^T qx + x^T qv = 0_{2n} \}. \quad (23)$$

The tangent space at the identity of the Lie group, namely the Lie algebra $\text{sp}(2n, \mathbb{R})$, has structure:

$$\text{sp}(2n, \mathbb{R}) = \{ h \in \mathbb{R}^{2n \times 2n} | h^T q + qh = 0 \} \quad (24)$$

and it coincides with the space of $2n \times 2n$ Hamiltonian matrices, in fact. By the embedding of the manifold $\text{Sp}(2n, \mathbb{R})$ into the Euclidean space $\mathbb{R}^{2n \times 2n}$, the embedded real symplectic group may be endowed with normal spaces as well, as:

$$N_x \text{Sp}(2n, \mathbb{R}) \equiv \{ n \in \mathbb{R}^{2n \times 2n} | \text{tr}(n^T v) = 0, \forall v \in T_x \text{Sp}(2n, \mathbb{R}) \}, \quad (25)$$

where symbol $\text{tr}(\cdot)$ denotes the trace operator. The tangent space, the Lie algebra and the normal space associated to the real symplectic group may be characterized as follows:

$$T_x \text{Sp}(2n, \mathbb{R}) = \{ xqs | q \in \mathbb{R}^{2n \times 2n}, s \in \mathbb{R}^{2n \times 2n}, s^T = s \},$$
$$\text{sp}(2n, \mathbb{R}) = \{ qsq | q, s \in \mathbb{R}^{2n \times 2n}, s^T = s \},$$
$$N_x \text{Sp}(2n, \mathbb{R}) = \{ qro | q, r, o \in \mathbb{R}^{2n \times 2n}, o^T = -o \}.$$

A noteworthy property of symplectic matrices is as follows. Let $x \in \text{Sp}(2n, \mathbb{R})$, $v \in T_x \text{Sp}(2n, \mathbb{R})$ and $y \in \mathbb{R}^{2n \times 2n}$. The following identity holds:

$$\text{tr}(x^{-1}qyqx^{-1}v) = \text{tr}(y^T v). \quad (26)$$

To prove such identity, use first the parametrization $v = xqs$, with $s = s^T \in \mathbb{R}^{2n \times 2n}$. Then it holds:

$$\text{tr}(x^{-1}qyqx^{-1}v) = \text{tr}(x^{-1}qyqx^{-1}(xqs)) = -\text{tr}(x^{-1}qys) = -\text{tr}(sx^{-1}qy) = -\text{tr}(y^T q x^{-T} s^T) = \text{tr}(y^T q (-qxq)s) = \text{tr}(y^T qxq) = \text{tr}(y^T v).$$
Note that the identity (26) holds for any real-valued matrix $y$ of appropriate size.

A result concerning optimization on the manifold $\text{Sp}(2n, \mathbb{R})$ is adapted from [4]. Let $\sigma : \mathfrak{sp}(2n, \mathbb{R}) \to \mathfrak{sp}(2n, \mathbb{R})$ be a symmetric positive-definite operator with respect to the Euclidean inner product $\langle \cdot, \cdot \rangle^E$ on the space $\mathfrak{sp}(2n, \mathbb{R})$ given by $\text{tr}(h_1^T h_2)$ for every $h_1, h_2 \in \mathfrak{sp}(2n, \mathbb{R})$. The minimizing curve of the integral:

$$\int_{t_1}^{t_2} \langle h, \sigma(h) \rangle^E dt$$

over all curves $x(t) \in \text{Sp}(2n, \mathbb{R})$ with $t \in [t_1, t_2]$ and with fixed endpoints $x(t_1) = x_1 \in \text{Sp}(2n, \mathbb{R})$ and $x(t_2) = x_2 \in \text{Sp}(2n, \mathbb{R})$, and where $h$ is defined by $\dot{x} = xh$, so that $h \in \mathfrak{sp}(2n, \mathbb{R})$, is the solution of the system:

$$\begin{aligned}
\dot{h} &= xh, \\
\dot{m} &= \sigma^T(h)m - m\sigma^T(h), \\
\dot{m} &= \sigma^{-1}(m),
\end{aligned}$$

(27)

where symbol $\sigma^{-1}$ denotes the inverse of the operator $\sigma$.

The simplest choice for the symmetric positive-definite operator $\sigma$ is $\sigma(h) = h$, which corresponds to a Riemannian metric for the real symplectic group $\text{Sp}(2n, \mathbb{R})$ given by:

$$\langle u, v \rangle_x = \text{tr}((x^{-1}u)^T(x^{-1}v)), \quad \forall u, v \in T_x\text{Sp}(2n, \mathbb{R}).$$

(28)

Such a metric leads to the ‘natural gradient’ on the space of real invertible matrices $\text{Gl}(n, \mathbb{R})$ studied in [1]. The above choice for the operator $\sigma$ implies that $m = h$ and that the corresponding curve on the real symplectic group satisfies the equations:

$$\begin{aligned}
\dot{h} &= h^T h - hh^T, \\
\dot{h} &= x^{-1}\dot{x},
\end{aligned}$$

(29)

or, in normal form:

$$\ddot{x} = \dot{x}x^{-1}\dot{x} + xx^T qxx^{-1}\dot{x} - \dot{x}x^T qxq = 0.$$  

(30)

The above equations describe geodesic arcs on the real symplectic group corresponding to the metric (28). Closed form solutions of the above equations are unknown to the authors of [4] and to the present author.

The Riemannian gradient $\nabla_x f$ of a regular function $f : \text{Sp}(2n, \mathbb{R}) \to \mathbb{R}$ corresponding to the metric (28) may be calculated as the unique vector in $T_x\text{Sp}(2n, \mathbb{R})$ that satisfies the following compatibility condition:

$$\text{tr}(v^T \partial_x f) = \text{tr}((x^{-1}v)^T(x^{-1}\nabla_x f)), \quad \forall v \in T_x\text{Sp}(2n, \mathbb{R}).$$

(31)

By recalling the structures of the tangent and normal spaces to the real symplectic group, the solution of the above equation is found as:

$$\nabla_x f = xq(\omega - x^{-1}q\partial_x f),$$

$$\omega = \frac{1}{2}x^{-1}q\partial_x f + \frac{1}{2}(\partial_x f)^T xq,$$

from which the expression of the Riemannian gradient associated to the metric (28) follows:

$$\nabla_x f = \frac{1}{2}xq((\partial_x f)^T xq - qx^T \partial_x f).$$

(32)
3.2 Symplectic group as a pseudo-Riemannian manifold

Let us consider the following pseudo-Riemannian metric on the general linear group of matrices Gl(n) [20]:

$$\langle u, v \rangle_x \overset{\text{def}}{=} \text{tr}(x^{-1}ux^{-1}v), \quad \forall u, v \in T_x\text{Gl}(n).$$  \hfill (33)

Under the above pseudo-Riemannian metric, it is indeed possible to solve the geodesic equation in closed form. The energy integral in this case reads:

$$\int_0^1 \text{tr}((x^{-1}x)^2)dt,$$ \hfill (34)

and the corresponding geodesic curve and the squared geodesic distance have the following expressions:

$$G_{x,v}(t) = x \exp(tx^{-1}v),$$ \hfill (35)

$$d^2(x_1, x_2) = \text{tr}(|\log^2(x_1^{-1}x_2)|).$$ \hfill (36)

On the general linear group Gl(n), matrix logarithm and exponential may be defined by the series:

$$\log y \overset{\text{def}}{=} -\sum_{k=1}^{\infty} \frac{(e_n - y)^k}{k}, \quad y \in \text{Gl}(n),$$ \hfill (37)

$$\exp y \overset{\text{def}}{=} \sum_{k=0}^{\infty} \frac{y^k}{k!}, \quad y \in \text{gl}(n).$$ \hfill (38)

The pseudo-Riemannian gradient of a regular function $f : \text{Sp}(2n, \mathbb{R}) \to \mathbb{R}$ associated to the metric (33) is the solution of the compatibility condition:

$$\text{tr}(\partial_x^Tfv) = \text{tr}(x^{-1}\nabla_xfx^{-1}v), \quad \forall v \in T_x\text{Sp}(2n, \mathbb{R}).$$ \hfill (39)

that reads:

$$\nabla_x f = \frac{1}{2} xq \left( x^T \partial_x f q - q(\partial_x f)^T x \right).$$ \hfill (40)

It is straightforward to show that the above vector $\nabla_x f$ belongs to $T_x\text{Sp}(2n, \mathbb{R})$. It may be shown that it satisfies the condition (39) by the help of the identity (26).

Having chosen a pseudo-Riemannian metric for the real symplectic group, the expression of the geodesic arc as well as the expression of the geodesic distance may be computed in closed form. The least-squares problem:

$$f(x) \overset{\text{def}}{=} \sum_k d^2(x, \tau_k),$$ \hfill (41)

may be thus set up, where target matrices $\tau_k$ belong to Sp(2n, \mathbb{R}). Moreover, the numerical scheme (15) may be effectively implemented.

The Euclidean gradient of the criterion function (41) has expression:

$$\partial_x \sum_k d^2(x, \tau_k) = 2 \sum_k (\log(\tau_k^{-1}x)x^{-1})^T \text{sign} \left( \text{tr}(\log^2(x^{-1}\tau_k)) \right).$$ \hfill (42)
4 Numerical tests

The numerical behavior of the developed least-squares algorithm will be examined via two different tests. A numerical test concerns the computation of the intrinsic mean of a set of given real symplectic matrices. A further numerical test concerns the interpolation of two given real symplectic matrices.

4.1 Computing the intrinsic mean of a set of real symplectic matrices

As a numerical test about least squares on the real symplectic group, consider the computation of the mean of a set of given real symplectic matrices. (An application is described in [18]. Some general remarks on averaging over Lie groups may be found in [14].)

On a metrizable manifold $M$, the ‘intrinsic mean’ may be defined as [15, 19]:

$$
\mu \overset{\text{def}}{=} \arg \min_{x \in M} \sum_k d^2(x, \tau_k),
$$

(43)

where the matrices $\tau_k \in M$ are distributed around a center-of-mass to be estimated by $\mu \in M$. The minimum value of the criterion function in (43) is the ‘intrinsic variance’ of the distribution, namely:

$$
\sigma^2 \overset{\text{def}}{=} \frac{1}{N} \sum_k d^2(\mu, \tau_k).
$$

(44)

Setting $M = \text{Sp}(2n, \mathbb{R})$, the problem (43) is a least-squares problem on the real symplectic group $\text{Sp}(2n, \mathbb{R})$.

The Figure 1 shows a result obtained with the iterative algorithm (15) for $n = 5$. The picture shows the value of the criterion function $\frac{1}{N} \sum_k d^2(x, \tau_k)$ as well as the value of the squared norm of its pseudo-Riemannian gradient during iteration. In the shown example, the squared norm assumes negative as well as positive values during optimization. The Figure also shows the squared distances $d^2(x, \tau_k)$ before iteration (with initial guess chosen as $x = e^{10}$) and after the iteration. The Figure shows that the algorithm converges steadily toward the minimal variance (in fact, the distances from the found center of mass are much smaller than the distances from the initial guess).

The group $\text{Sp}(10, \mathbb{R})$ is a 55-dimensional manifold whose elements are represented by $10 \times 10$ real-valued matrices. It is impossible to render graphically the target matrices $\tau_k$, their center of mass and the computed mean matrix. An illustration of the distribution of the target matrices around the actual center and of the computed mean may, however, be gotten through the application of an appropriate dimensionality reduction technique. We chose the ‘multidimensional scaling’ (MDS, see Appendix A) as dimensionality reduction method onto the real plane $\mathbb{R}^2$. The result is depicted in Figure 2. From this Figure, it is possible to appreciate how close the computed (empirical) intrinsic mean lies to the actual center of mass.

A close-up of the numerical behavior of the least-squares optimization algorithm (15) comes from the examination of the case $n = 1$. The group $\text{Sp}(2, \mathbb{R})$ is a 3-dimensional manifold (in fact, $\text{Sp}(2, \mathbb{R}) = \text{Sl}(2, \mathbb{R})$), therefore the following
Figure 1: Optimization over the real symplectic group $\text{Sp}(10, \mathbb{R})$. Variance during iteration, value of the squared norm of criterion pseudo-Riemannian gradient during iteration, the squared distances $d^2(x, \tau_k)$ before iteration and after iteration.

The parametrization may be taken advantage of:

$$\mathbb{R}^3 \ni (a, b, c) \rightarrow \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \text{Sp}(2, \mathbb{R}).$$  \hspace{1cm} (45)

(In fact, the fourth parameter $d$ is constrained to the three free parameters $a$, $b$ and $c$ by the symplecticity condition $ad - bc = 1$.) Hence, the elements of the group $\text{Sp}(2, \mathbb{R})$ may be rendered on a 3-dimensional figure.

The Figure 3 shows a result obtained with the iterative algorithm (15) for $n = 1$. The Figure shows the location of the target matrices $\tau_k$ (circles), the location of the center-of-mass (cross), the trajectory of the optimization algorithm over the search space (solid-dotted line) and the location of the final point computed by the algorithm (diamond). The Figure shows that the algorithm is convergent toward the center of mass. The Figure 4 shows the variance during iteration, the value of the squared norm of criterion pseudo-Riemannian gradient during iteration, the squared distances $d^2(x, \tau_k)$ before iteration and after the iteration.
Figure 2: Optimization over the real symplectic group $\text{Sp}(10, \mathbb{R})$: Visualization via MDS of the cloud of points representing target matrices (open circles), the center of mass (cross) and the computed mean (open diamond).

### 4.2 Interpolating over the real symplectic group

As a further numerical test, consider the interpolation of two given real symplectic matrices. (An application is described in [22].)

On a metrizable manifold $M$, the continuous interpolate between two given matrices $\tau_1, \tau_2 \in M$ may be defined as the curve $x : [0, 1] \rightarrow M$ such that:

$$
  x(t) \overset{\text{def}}{=} \arg \min_{z(t) \in C} [(1 - t)d^2(z(t), \tau_1) + td^2(z(t), \tau_2)],
$$

where $C$ denotes the set of smooth curves that perform the mapping $[0, 1] \rightarrow M$. Setting $M = \text{Sp}(2n, \mathbb{R})$, the problem (46) is a least-squares problem on the real symplectic group $\text{Sp}(2n, \mathbb{R})$.

The solution of the above least-squares problem may be given in closed form. It coincides to a geodesic arc $G_{x,v}(t)$ parameterized so that $G_{x,v}(0) = \tau_1$ and $G_{x,v}(1) = \tau_2$, namely:

$$
  G(t) = \tau_1 \exp(t \log(\tau_1^{-1} \tau_2)).
$$

The Figure 5 shows a result of continuous interpolation for $n = 1$ (the parametrization (45) has been used again). The Figure shows the location of the endpoint matrices $\tau_1, \tau_2 \in \text{Sp}(2, \mathbb{R})$ (dotted-circles) and of the interpolates (dots).
5 Generalizations and applications

The present manuscript focuses on least-squares problems on the real symplectic group. The developed least-squares theory applies, however, to general non-compact manifolds.

Least-squares problems arise in a series of applications in signal processing, machine learning, pattern recognition and computational statistics. Here follows a non-exhaustive list of possible applications of least-squares methods on non-compact manifolds:

- **Computation of the empirical mean value on manifold**: Given a ‘cloud’ of points \( \tau_k \in M \) on a manifold \( M \) endowed with a distance function \( d(\cdot,\cdot) \), its center of mass \( \mu \) is defined as:

\[
\mu = \arg\min_{x \in M} \sum_k d^2(x, \tau_k).
\]

The empirical mean value \( \mu \) of a distribution of points on a manifold is instrumental in several applications. The mean value \( \mu \) is, by definition, close to all points in the distribution, therefore, the tangent space \( T_{\mu}M \) may serve as reference tangent space in the development of algorithms on
Figure 4: Optimization over the real symplectic group \( \text{Sp}(2, \mathbb{R}) \). Variance during iteration, value of the squared norm of criterion pseudo-Riemannian gradient during iteration, the squared distances \( d^2(x, \tau_k) \) before iteration and after iteration.

the manifold \( M \) (likewise the Lie algebra associated to a Lie group serves as reference tangent space in Lie-group theory).

- **Computation of the empirical high-order moments on manifold:** Given a cloud of \( N \) points \( \tau_k \in M \) on a manifold \( M \) endowed with a distance function \( d(\cdot, \cdot) \), its scalar \( m^{\text{th}} \)-order (centered) moment is defined as:

\[
\mu_m = \frac{1}{N} \sum_k d^m(\mu, \tau_k), \quad m \geq 2,
\]

(49)

where \( \mu \in M \) denotes the empirical mean of the cloud. The moment \( \mu_2 \) denotes the variance \( \sigma^2 \), which measures the width of the cloud around its center.

- **Applications based on the statistical distributions of data-points:** Some statistical techniques, such as maximum likelihood estimation, are based on assumptions about the statistical distributions of the data-points. In the present context, it could be worth considering the distribution of the distances of the points from their center, namely, by defining the following exemplary distributions: Gaussian \( p(\tau) \sim \exp\left( -\frac{d^2(\tau, \mu)}{2\sigma^2} \right) \), Lapla-
Figure 5: Interpolation over the real symplectic group Sp(2, ℚ). Endpoint matrices τ₁, τ₂ are denoted by dotted circles. Computed interpolates are denoted by dots.

cian \( p(τ) \sim \exp \left(-\lambda d(τ, μ)\right) \) and Rayleigh \( p(τ) \sim d(τ, μ) \exp \left(-\frac{d^2(τ, μ)}{2\sigma^2}\right) \), where \( τ \in M \) denotes the manifold-valued variable of interest, \( μ \in M \) denotes the center of mass of the distribution, \( σ^2 \) denotes the variance and \( λ > 0 \) denotes a dispersion parameter.

- Projection of a point on a curve: Given a point \( τ \in M \) and a smooth curve \( c_{x,v} : [-a, a] \to M \) such that \( c_{x,v}(0) = x \), the ‘foot’ of the projection of the point \( τ \) on the curve \( c_{x,v} \) is defined as the point \( c_{x,v}(φ) \), with:

\[
φ \overset{\text{def}}{=} \arg \min_{t \in [-a, a]} d^2(τ, c_{x,v}(t)). \quad (50)
\]

It is assumed that the projection is well-defined, namely, that the foot-parameter value \( φ \) is unique. The (scalar) projection of the point \( τ \) on the curve \( c_{x,v} \) is defined as:

\[
π^τ_{c_{x,v}} \overset{\text{def}}{=} \int_0^φ \langle \dot{c}_{x,v}(θ), c_{x,v}(θ) \rangle_\cdot \frac{1}{c_{x,v}(θ)} dθ. \quad (51)
\]

The projection computes as the (signed) distance (measured along the curve \( c_{x,v} \)) between the foot of the projection and the origin of the curve.
the analogous of the orthogonal projection on Euclidean spaces. In the special case that the curve \( c_{x,v} \) is a geodesic arc on the manifold \( M \) endowed with the metric \( \langle \cdot, \cdot \rangle \), then it holds:

\[
\pi^\tau_{c_{x,v}} = \phi \sqrt{\langle v, v \rangle}.
\]

Moreover, the quantity \( d(\tau, c_{x,v}(\phi)) \) denotes the distance of the point \( \tau \in M \) from the curve \( c_{x,v} \).

- **Interpolation and modeling:** Given a set of points \( \tau_k \in M \) on a manifold \( M \) endowed with a distance function \( d(\cdot, \cdot) \) and a curve \( c_{x,v} : [-a, a] \to \mathbb{R} \), interpolation/modeling is about finding parameters:

\[
(x, v) = \arg \min_{x \in M, v \in T_x M} \sum_k d^2(\tau_k, c_{x,v}(\phi_k)),
\]

with projection-feet defined as:

\[
\phi_k \overset{\text{def}}{=} \arg \min_{t \in [-a, a]} d^2(\tau_k, c_{x,v}(t)).
\]

The curve \( c_{x,v}(t) \) is an interpolator with optimal parameters.

- **Principal geodesic analysis:** First principal component analysis and one-unit independent component analysis may be extended to manifolds as follows. Let it be given a cloud of points \( \tau_k \in M \) on a manifold \( M \) endowed with a geodesic distance function \( d(\cdot, \cdot) \) associated to a geodesic curve \( G_{x,v} \). Let \( \mu \in M \) denote the empirical mean of the cloud. The principal geodesic arc that captures the largest data-power may be defined as \( G_{\mu,v} \) with:

\[
\bar{v} = \arg \max_{v \in T_\mu M} \frac{1}{N} \sum_k d^2(\mu, G_{\mu,v}(\phi_k)), \text{under } \langle \bar{v}, \bar{v} \rangle_\mu = 1,
\]

where each \( \phi_k \) denotes the foot-parameter associated to the projection of \( \tau_k \) over the geodesic arc. (The normalization condition plays the same role as in linear principal component analysis). As the geodesic distance are computed over geodesic arcs, the above expression simplifies as:

\[
\bar{v} = \arg \max_{v \in T_\mu M} \sum_k \phi_k^2.
\]

Likewise, the principal geodesic arc that corresponds to the largest \( m \)th moment finds by:

\[
\bar{v} = \arg \max_{v \in T_\mu M} \sum_k \phi_k^m.
\]

The principal geodesic arc that maximizes the projection variance may be used to ‘skeletonize’ data on manifolds and for data compression. In fact, a lossy representation of a data \( \tau_k \) is given by its foot-parameter \( \phi_k \). A principal geodesic arc that maximizes a high-order moment of the projection may be used in binary discriminant analysis to discern between data that possess or not possess certain high-order statistical features.
Fisher discriminant analysis on manifold: Let data-points \( \tau_k \in M \) belonging to two classes \( C_1 \) and \( C_2 \) of cardinality \( N_1 \) and \( N_2 \), respectively, be given. Let \( \mu_1 \) be the intrinsic mean of data-points in class \( C_1 \) and \( \mu_2 \) be the intrinsic mean of data-points in class \( C_2 \). Let \( G_{x,v} \) be a geodesic curve on the manifold \( M \) and \( d(\cdot, \cdot) \) be the associated geodesic distance function on \( M \). Let \( \Phi \) denote the foot-parameters associated to the projection of \( \mu_1 \) over the geodesic arc and of \( \mu_2 \) over the geodesic arc, respectively, and let \( \phi_k \) denote the foot-parameter associated to the projection of \( \tau_k \) over the geodesic arc. The Fisher ratio associated to the 2-class maximum discrimination problem is extended from the Euclidean case as:

\[
F(x, v) \overset{\text{def}}{=} \frac{d^2(G_{x,v}(\Phi_1), G_{x,v}(\Phi_2))}{\frac{1}{N_1} \sum_{k \in C_1} d^2(G_{x,v}(\phi_k), G_{x,v}(\Phi_1)) + \frac{1}{N_2} \sum_{k \in C_2} d^2(G_{x,v}(\phi_k), G_{x,v}(\Phi_2))}.
\]

The numerator of the above expression represents the between-class variance which amounts to the squared difference between the means projected on a properly oriented curve. The denominator represents the within-class variance which amounts at the variance of the projected elements of the first class and of the second class. Note that all geodesic distances are measured on a geodesic arc, therefore, the above expression simplifies considerably because, for instance, \( d^2(G_{x,v}(\Phi_1), G_{x,v}(\Phi_2)) = (\Phi_1 - \Phi_2)^T v_x (\Phi_1 - \Phi_2) \).

Hence:

\[
F = \frac{(\Phi_1 - \Phi_2)^T}{\frac{1}{N_1} \sum_{k \in C_1} (\phi_k - \phi_1)^2 + \frac{1}{N_2} \sum_{k \in C_2} (\phi_k - \phi_2)^2}.
\]

Kalman filtering on manifold: The state of a dynamical system evolving on a manifold may be estimated by setting up an appropriate least-squares problem. The extension from the case of Kalman filtering on Euclidean spaces is straightforward conceptually. An example of system model on a manifold \( M \) is:

\[
\begin{cases}
\tau_{n+1} = G_{\tau_n, v_n}(1), \\
v_n = S(\tau_n) + r_n,
\end{cases}
\]

where \( n \) denotes discrete time, \( \tau_n \in M \) denotes system state, \( G_{\cdot, \cdot} \) denotes a geodesic curve on the manifold \( M \), \( S(\cdot) \) denotes a system operator which maps \( \tau \in M \mapsto S(\tau) \in T_M \) and \( r_n \in T_{\tau_n} \) is a random noise term. (A map that sends a point from a manifold to its tangent space and that satisfies certain regularity conditions is termed lifting map.) More generally, the problem of Bayesian filtering on manifold is connected to the problem of extending standard state-space models from Euclidean spaces to manifolds \([7]\).

Least squares problems on manifolds may generalize the quadratic assignment problem \([11]\) that is encountered in the allocation of facilities while optimizing the distance between locations in combination with the flow between the facilities. Also, the traveling salesman problem and the plant location problem are special cases of the quadratic allocation problem \([11]\).
6 Conclusion

The present paper discusses the problem of least squares over the real symplectic group of matrices. The present research takes its moves from the following observations:

- The least-squares problem may be extended from flat spaces to curved smooth metrizable spaces by the help of the notion of ‘geodesic distance’.
- The resulting sum-of-squared-distance minimization problem on manifold may be tackled via a gradient-based descent algorithm tailored to the geometry of the symplectic group through a geodesic-based stepping method.

As the real symplectic group is a non-compact manifold, it might be hard to compute closed-forms quantities in a Riemannian context. Indeed, known results from scientific literature show that it is the case.

The key point of the present paper is to regard the real symplectic group as a pseudo-Riemannian manifold and chose a metric that allows for the computation of closed-forms of geodesic arcs and hence of geodesic distance. On the basis of these findings, the geodesic least-squared problem may be properly set up and the geodesic-based numerical stepping method may be properly implemented.

Numerical tests have been performed with reference to the computation of the intrinsic mean of a collection of symplectic matrices as well as to the interpolation of two symplectic matrices. Numerical results show that the pseudo-Riemannian-gradient-based algorithm, along with a pseudo-geodesic-based stepping method, is suitable to the numerical solution of a least-squares problem formulated in terms of pseudo-geodesic distance.

A Appendix: Metric Multidimensional Scaling

One of the purposes of multidimensional scaling (MDS) [21] is to provide a visual representation of the pattern of proximities among a set of high-dimensional objects. In this instance, MDS finds a set of vectors in the two-dimensional or three-dimensional Euclidean space such that the matrix of Euclidean distances among them corresponds – as closely as possible – to some function of the objects’ proximity matrix according to a criterion function termed stress.

The following description provides a short introduction to MDS, as reworded in terms of low-dimensional representation of pseudo-Riemannian-manifold-valued elements.

Let $M$ be a high-dimensional pseudo-Riemannian manifold with distance function $d(\cdot, \cdot)$ and let $\{\tau_k\}_k$ be a given collection of elements of $M$. The aim of MDS is to determine a collection of vectors $z_k \in \mathbb{R}^p$ (with $p = 2$ or $p = 3$) that replicate the pattern of proximities among the elements $\tau_k$. This may be achieved by minimizing the Kruskal stress function:

$$\phi(\{\tau_k\}_k) \triangleq \sum_i \sum_{k \neq i} (\|z_i - z_k\| - d(\tau_i, \tau_k))^2,$$

where symbol $\| \cdot \|$ denotes Euclidean norm. The above stress function gives rise to ‘metric MDS’. More elaborated stress functions are available in the specific literature.
It is worth noting that the correspondence $\tau_k \leftrightarrow z_k$ induced by MDS is not unique. In fact, if $\{z_k\}_k$ is a minimizer of the criterion $[21]$, $c \in \mathbb{R}^p$ and $R \in O(p)$ (namely, a $p$-dimensional rotation/reflection), also $\{Rz_k + c\}_k$ is a minimizer.

MDS may be used as a proximity/similarity visualization tool for high-dimensional data as it computes two-dimensional or three-dimensional vectors $z_k$, corresponding to the original elements $\tau_k$, that captures the fundamental information about mutual distances.

The axes corresponding to the coordinates of the vectors $z_k$, possess, in themselves, no particular meaning. Also, the orientation and scaling of the obtained visualization are arbitrary. All that matters in an MDS map is which point is close to which others.

On a MDS visualization corresponding to a non-zero stress, the distances among objects are imperfect representations of the relationships among original data: The greater the stress, the greater the distortion. In general, however, the larger distances are represented more accurately, because the Kruskal stress function accentuates discrepancies in the larger distances.

References

[1] S.-i. Amari, Natural gradient works efficiently in learning, Neural Computation, Vol. 10, 251 – 276, 1998

[2] M. Aoki, Two complementary representations of multiple time series in state-space innovation forms, Journal of Forecasting (Special Issue on “Times Series Advances in Economic Forecasting”), Vol. 13, No. 2, pp. 69 – 90, 2006

[3] S. D. Bartlett, B. Sanders, S. Braunstein and K. Nemoto, Efficient classical simulation of continuous variable quantum information processes, Physical Review Letters, Vol. 88, 097904/1-4, 2002

[4] A.M. Bloch, P.E. Crouch, J.E. Marsden and A.K. Sayal, Optimal control and geodesics on quadratic matrix Lie groups, Foundations of Computational Mathematics, Vol. 8, pp. 469 – 500, 2008

[5] Y. Borissov, Minimal/nonminimal codewords in the second order binary Reed-Muller codes: revisited, Proceedings of the Eleventh International Workshop on Algebraic and Combinatorial Coding Theory, pp. 29 – 34, June 16-22, 2008, Pamporovo, Bulgaria

[6] E. Celledoni and S. Fiori, Descent methods for optimization on homogeneous manifolds, Journal of Mathematics and Computers in Simulation (Special issue on “Structural Dynamical Systems: Computational Aspects”, Guest Editors: N. Del Buono, L. Lopez and T. Politi), Vol. 79, No. 4, pp. 1298 – 1323, December 2008

[7] Y. Chikuse, State space models on special manifolds, Journal of Multivariate Analysis, Vol. 97, No. 6, pp. 1284 – 1294, July 2006

[8] A.J. Draft, F. Neri, G. Rangarajan, D.R. Douglas, L.M. Healy and R.D. Ryne, Lie algebraic treatment of linear and nonlinear beam
dynamics, Annual Review of Nuclear and Particle Science, Vol. 38, pp. 455 – 496, December 1988

[9] L. DIECI and L. LOPEZ, Smooth singular value decomposition on the symplectic group and Lyapunov exponents approximation, CALCOLO, Vol. 43, No 1, pp. 1 – 15, March 2006

[10] F.M. DOPICO and C.R. JOHNSON, Complementary bases in symplectic matrices and a proof that their determinant is one, Linear Algebra and its Applications, Vol. 419, No.s 2-3, pp. 772 – 778, December 2006

[11] J. Fan and P. Nie, Quadratic problems over the Stiefel manifold, Operations Research Letters, Vol. 34, pp. 135 – 141, 2006

[12] 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

[13] S. FIORI, Geodesic-based and projection-based neural blind deconvolution algorithms, Signal Processing, Vol. 88, No. 3, pp. 521 – 538, March 2008

[14] S. FIORI and T. TANAKA, An algorithm to compute averages on matrix Lie groups, IEEE Transactions on Signal Processing. Accepted for publication

[15] M. FRÉCHET, Les éléments aléatoires de nature quelconque dans un espace distancié, Annales de l’Institut Henri Poincaré, Vol. 10, pp 215 – 310, 1948

[16] V. GUILLEMIN and S. STERNBERG, Symplectic Techniques in Physics, Cambridge University Press, 1984

[17] 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

[18] W.F. HARRIS, The average eye, Ophthalmic and Physiological Optics, Vol. 24, pp. 580 – 585, 2004

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

[20] A. KHVEDELIDZE and D. MLADENOV, Generalized Calogero-Moser-Sutherland models from geodesic motion on GL+(n, R) group manifold, Physics Letters A, Vol. 299, No.s 5-6, pp. 522 – 530, July 2002

[21] J.B. KRUSKAL and M. WISH, Multidimensional Scaling, Sage University Paper series on Quantitative Application in the Social Sciences, 07-011. Beverly Hills and London: Sage Publications, 1978

[22] W.W. MACKAY and A.U. LUCCIO, Symplectic interpolation, Proceedings of the Tenth European Particle Accelerator Conference (EPAC 2006), 26 - 30 June, 2006 Edinburgh, Scotland

[23] N.G. STEPHEN, Transfer matrix analysis of the elastostatics of one-dimensional repetitive structures, Proceedings of the Royal Society A, Vol. 462, No. 2072, pp. 2245 – 2270, August 2006
[24] M. Spivak. *A Comprehensive Introduction to Differential Geometry*, 2nd Edition, Berkeley, CA: Publish or Perish Press, 1979