
Optimal Data Fitting on Lie Groups: a Coset Approach

C. Lageman¹ and R. Sepulchre²

¹ Institut für Mathematik

Universität Würzburg

Am Hubland

97074 Würzburg, Germany

`christian.lageman@mathematik.uni-wuerzburg.de`

² Department of Electrical Engineering and Computer Science, B28

Université de Liège

B-4000 Liège Sart-Tilman, Belgium

`r.sepulchre@ulg.ac.be`

Summary. This work considers the problem of fitting data on a Lie group by a coset of a compact subgroup. This problem can be seen as an extension of the problem of fitting affine subspaces in \mathbb{R}^n to data which can be solved using principal component analysis. We show how the fitting problem can be reduced for biinvariant distances to a generalized mean calculation on an homogeneous space. For biinvariant Riemannian distances we provide an algorithm based on the Karcher mean gradient algorithm. We illustrate our approach by some examples on $SO(n)$.

1 Introduction

In this paper we consider the problem of fitting a submanifold to data points on a Lie group. Such fitting problems are relevant for dimension reduction and statistical analysis of data on Lie groups. In Euclidean space it is well-known that the best fitting k -dimensional linear subspace can be computed via principal component analysis (PCA) and this tool is widely used in applications in natural sciences, statistics and engineering.

However, in some applications the data naturally arises as points on an embedded or abstract manifold, e.g. points on spheres [2] or manifolds of shape representations [4, 5]. This raises the question of extending subspace fitting and dimension reduction methods like PCA to nonlinear spaces like Riemannian manifolds and Lie groups. In the recent years some approaches have been proposed to construct local extensions of PCA [4, 5] on Riemannian manifolds or to consider fitting by single geodesics and interpolation problems on manifolds [7, 8]. Here, we focus on compact Lie groups and propose the different

approach to fit a coset to the data. Our approach overcomes some limitations of the local approaches and leads to potentially efficient computational algorithms.

In Section 2 we recall basic facts on PCA. Section 3 discusses principal geodesic analysis from [4, 5]. Section 4 introduces our fitting of cosets approach and shows how it leads to a reduced optimization problem on a homogeneous space. For Riemannian distances we derive an algorithm based on known Karcher mean algorithms. Section 5 provides examples for fitting on $SO(n)$.

Notation

In this paper G will always denote a compact, connected Lie group. For more background on differential geometry, Lie groups etc. we refer to [1]. Recall that given a closed subgroup $H \subset G$ the quotient space G/H carries naturally a manifold structure. A Riemannian metric on G is called left- resp. right-invariant if it is invariant under the action of G on itself by left- resp. right-multiplication, i.e. for all $p, q \in G$, $v, w \in T_p G$ we have $\langle T_p L_q v, T_p L_q w \rangle = \langle v, w \rangle$ with L_q the left multiplication map $L_q(p) = qp$, analogously for the right-invariant case. A Riemannian metric is called biinvariant if it is left- and right-invariant. It can be shown that on any compact Lie group a biinvariant Riemannian metric exists. This is not the case for non-compact groups.

Furthermore, we recall the definition of a Karcher mean on a Riemannian manifold. Let M be a Riemannian manifold with Riemannian distance dist_R . The *Karcher mean* of points q_1, \dots, q_k on M is defined [10] as a minimum of the function $f(x) = \sum_{i=1}^k \text{dist}_R(q_i, x)^2$. Note that a Karcher mean does not have to be unique.

2 Principal Component Analysis

In Euclidean spaces the most common method for dimension reduction of data is principal component analysis (PCA). We recall some basic facts on PCA, for a detailed account see the numerous literature on this topic, e.g. [3].

Given k data points $q_1, \dots, q_k \in \mathbb{R}^n$, the problem is to determine an affine subspace $p + V$ of dimension m such that the sum of squared Euclidean distances

$$\sum_{i=1}^k \min_{v \in p+V} \|q_i - v\|^2 = \sum_{i=1}^k \text{dist}_E(q_i, p + V)^2 \quad (1)$$

is minimized, dist_E denoting the Euclidean distance to a closed subset.

This problem can be solved by computing an eigenvalue decomposition UDU^T , $D = \text{diag}(\lambda_1, \dots, \lambda_n)$, $\lambda_1 \geq \dots \geq \lambda_n$ of the symmetric, positive semidefinite matrix $\sum_{i=1}^k (q_i - \bar{q})(q_i - \bar{q})^T$ with $\bar{q} = \frac{1}{k} \sum_{i=1}^k q_i$ the mean of the data points. The best fitting affine subspace is given by $(p + V)_{\text{opt}} = \bar{q} + \text{span}\{u_1, \dots, u_m\}$ with u_1, \dots, u_m denoting the first m columns of U . The u_i are called the *principal components* of the q_i .

The orthogonal projection of the data points onto $(p + V)_{\text{opt}}$ in the basis u_1, \dots, u_m of $(p + V)_{\text{opt}}$ is given by $(u_1 \dots u_m)^T (q_i - \bar{q})$. This reduces the n -dimensional data points to m -dimensional data points.

In this paper we concentrate on generalizing the fitting of a subspace to the data (1) to Lie groups. This is justified by the statistical information hold by the $(p + V)_{\text{opt}}$ in the Euclidean case, cf. [3].

3 Principal geodesic analysis

Fletcher *et al.* propose *principal geodesic analysis (PGA)* — a local approach which lifts the data to a tangent space and performs PCA there — as a generalization of PCA to manifolds, [4, 5]. They consider data points q_1, \dots, q_k on a Riemannian manifold M and a Karcher mean \bar{q} . Let $\exp_{\bar{q}}$ denote the Riemannian exponential map. They define *principal geodesic submanifolds* recursively as submanifolds $N_1 := \exp_{\bar{q}}(V_1), \dots, N_n := \exp_{\bar{q}}(V_{n-1})$, $V_1 = \text{span}\{v_1\}, \dots, V_{n-1} = \text{span}\{v_1, \dots, v_{n-1}\}$ minimizing the squared distance to the data; we refer to [5] for details. To calculate the submanifolds the data points are first lifted to $T_{\bar{q}}M$ by computing $p_i = \exp_{\bar{q}}^{-1}(q_i)$. Since $T_{\bar{q}}M$ is a finite dimensional Hilbert space with the scalar product given by the Riemannian metric, one can choose an orthonormal basis of $T_{\bar{q}}M$ and perform PCA on the p_i as points in an Euclidean space. The principal components $u_i \in T_{\bar{q}}M$ yield an approximation $\tilde{V}_m = \text{span}\{u_1, \dots, u_m\} \subset T_{\bar{q}}M$ of the V_m and therefore an approximation of the fitting problem

$$\text{Minimize } \sum_{i=1}^k \text{dist}_R(q_i, \exp_{\bar{q}}(V))^2 \tag{2}$$

over the set of m -dimensional subspaces V of $T_{\bar{q}}M$ with dist_R the Riemannian distance. Note that for $M = \mathbb{R}^n$ with the Euclidean metric this yields precisely $(p + V)_{\text{opt}}$ of (1) since $(p + V)_{\text{opt}} = \exp_{\bar{q}}(\tilde{V}) = \bar{q} + \tilde{V}$.

For a sufficiently small neighborhood U of 0 in $T_{\bar{q}}M$ the set $\exp_{\bar{q}}(\tilde{V} \cap U)$ is an embedded submanifold and it is ‘close’ to the optimal $\exp_{\bar{q}}(\tilde{V})$ of (2). Therefore PGA is suitable if the data are clustered around a unique Karcher mean. However, if the data are not clustered around a point, one has to take into account that the Karcher mean is not unique, that $\exp_{\bar{q}}(\tilde{V})$ is not necessarily an embedded manifold, and that $\exp_{\bar{q}}(\tilde{V})$ is not an exact solution of the fitting problem (2). In such cases PGA is not well-suited to compute a best fitting submanifold and a global approach might be more desirable as a generalization of (1).

4 Fitting cosets

We propose here a global approach to generalize (1) to compact Lie groups. It is based on an alternative interpretation of the Euclidean fitting problem.

Recall that the special Euclidean group $SE(n) = \{(R, p) \mid R \in SO(n), p \in \mathbb{R}^n\}$ acts on \mathbb{R}^n transitively by $\phi: (x, (R, p)) \mapsto Rx + p$. Thus \mathbb{R}^n can be thought of as the homogeneous space $\mathbb{R}^n \cong SE(n)/SO(n)$. For the Euclidean distance dist_E we have $\text{dist}_E(x, y) = \text{dist}_E(\phi(x, (R, p)), \phi(y, (R, p)))$ for all (R, p) , i.e. dist_E is invariant under the action of $SE(n)$ on \mathbb{R}^n . In general, a distance dist on a homogeneous space M is called G -invariant if for all $x, y \in M$, $s \in G$ $\text{dist}(s \cdot x, s \cdot y) = \text{dist}(x, y)$, $s \cdot x$ denoting the action of G on M . Note further that given a fixed subspace $\tilde{V} \subset \mathbb{R}^n$, $\dim \tilde{V} = m$, any m -dimensional affine subspace can be written as $R\tilde{V} + p$ with $(R, p) \in SE(n)$. Thus minimizing (1) over the set of affine subspaces is equivalent to $\min_{(R, p) \in SE(n)} \sum_{i=1}^k \text{dist}_E(q_i, R\tilde{V} + p)^2$.

This motivates to consider the following fitting problem for invariant distances on homogeneous spaces as a generalization of (1).

Problem 1. Let M a homogeneous space with Lie group \tilde{G} acting transitively on M via $\phi: \tilde{G} \times M \rightarrow M$, N a submanifold on M and dist an invariant distance. Solve the optimization problem

$$\min_{g \in \tilde{G}} \sum_{i=1}^k \text{dist}(q_i, \phi(g, N))^2. \tag{3}$$

We have seen that (1) is a special case of (3) for $M = \mathbb{R}^n$, $\tilde{G} = SE(n)$, $N = \tilde{V} \cong \mathbb{R}^m$, $\text{dist} = \text{dist}_E$ and $\phi(x, (R, p)) = Rx + p$.

To use (3) for data on the Lie group G , we have to turn G into an homogeneous space, i.e. find another Lie group acting transitively on G . A naïve choice would be G with its action on itself by left- and right-multiplication. However, if e.g. N is a subgroup this would turn G into a fiber bundle, providing not enough degrees of freedom for a sensible fitting of the data by submanifolds diffeomorphic to N . The action ψ of $\tilde{G} = G \times G$ on G with $\psi: (x, (p, q)) \mapsto pxq^{-1}$ will be more suitable for our task: it will generate for subgroups N a larger class of submanifolds in G . The distances dist on G , invariant under the action ψ , are called *biinvariant* since for all $q, p, s \in G$ one has $\text{dist}(sq, sp) = \text{dist}(q, p) = \text{dist}(qs, ps)$.

Examples of biinvariant distances include the following:

(a) Let $\langle \cdot, \cdot \rangle$ be a biinvariant Riemannian metric on G . Then the Riemannian distance on G is biinvariant.

(b) Let $\rho: G \rightarrow \mathbb{C}^{m \times m}$ be a faithful, unitary representation of G , i.e. a homomorphism onto the group of unitary transformations of a finite dimensional Hilbert space with $\ker \rho = \{e\}$. Then $\text{dist}(q, p) = \|\rho(q) - \rho(p)\|_F$, $\|A\|_F = \text{tr}(A^\dagger A)^{1/2}$ the Frobenius norm, A^\dagger the Hermitian conjugate, is a biinvariant distance on G . In particular, for the special orthogonal and the unitary group, the Frobenius norm of the difference of two matrices $\|Q - P\|_F$ yields a biinvariant distance.

We have to choose the class of submanifolds which we use for fitting the data. For PCA in Euclidean space the fitting submanifolds are affine subspaces, i.e. totally geodesic submanifolds of \mathbb{R}^n . This suggests the use of totally

geodesic submanifolds at least for biinvariant Riemannian distances/metrics, too. However, since we want to exploit the group structure to obtain a reduced optimization problem, we restrict ourselves to closed, i.e. in this case compact, subgroups of G . Indeed subgroups of G are totally geodesic for any biinvariant metric.

Considering G as a homogeneous space with $G \times G$ acting on it by ψ , the fitting problem (3) for N a compact subgroup $H \subset G$ has the form

$$\min_{(p,q) \in G \times G} \sum_{i=1}^k \text{dist}(q_i, \psi((p, q), H))^2 = \sum_{i=1}^k \text{dist}(q_i, pHq^{-1})^2$$

with dist a ψ -invariant, i.e. biinvariant, distance on G . This gives the following fitting problem as a special case of (3) and a global generalization of (1) to Lie groups.

Problem 2. Let $H \subset G$ be a fixed, compact subgroup, $\text{dist}: G \times G \rightarrow \mathbb{R}$ a biinvariant distance function and $q_1, \dots, q_k \in G$ data points. Solve the optimization problem

$$\min_{p,q \in G} \sum_{i=1}^k \text{dist}(q_i, pHq^{-1})^2. \tag{4}$$

Any of the pHq^{-1} can be written as $\tilde{p}qHq^{-1}$, i.e. it is a coset of a subgroup of G conjugate to H . Therefore our approach consists of *optimally fitting to the data a coset of a subgroup conjugate to H* .

4.1 Reduction to a homogeneous space

Note that $G \times G$ is, especially for large subgroups H , a vast overparameterization of the family of submanifolds pHq^{-1} . Fortunately, this problem can be reduced to an optimization problem on the homogeneous space $G/H \times G/H$. The key insight is that the biinvariant distance on G induces a G -invariant distance on G/H .

Proposition 1. Let dist_G be a biinvariant distance on G and $H \subset G$ a compact subgroup. Then dist_G induces a G -invariant distance $\text{dist}_{G/H}$ on G/H , such that $\text{dist}_{G/H}(qH, pH) = \text{dist}_G(q, pH)$.

Proof. Since dist_G is right-invariant we have for all $k \in H$

$$\text{dist}_G(q, pH) = \min_{h \in H} \text{dist}_G(q, ph) = \min_{h \in H} \text{dist}_G(qk, ph) = \text{dist}_G(qk, pH).$$

Thus $\text{dist}_G(q, pH)$ induces a distance $\text{dist}_{G/H}$ on G/H . The G -invariance of $\text{dist}_{G/H}$ follows directly from the left-invariance of dist_G .

Induced distances on G/H include the following examples:

- (a) Let $\langle \cdot, \cdot \rangle_G$ be a biinvariant Riemannian metric on G . We can define on G the distribution $N(p) := (T_p pH)^\perp$, W^\perp the orthogonal complement with respect

to the Riemannian metric. Let $\pi: G \rightarrow G/H$ be the canonical projection $\pi(p) := pH$. Then, the formula

$$\langle v, w \rangle_{G/H} := \langle v^N, w^N \rangle \text{ for } v, w \in T_p H G/H$$

defines an G -invariant Riemannian metric on G/H with v^N, w^N uniquely defined by $v^N, w^N \in N(p), T_p \pi v^N = v, T_p \pi w^N = w$. This Riemannian metric is called the *normal metric* [9]. The distance on G/H induced by the Riemannian metric on G is the Riemannian distance of the normal metric.

(b) Let ρ be again a faithful, finite dimensional, unitary representation of G and $H = \text{stab}(v) = \{p \in G \mid \rho(p)v = v\}$ for a $v \in \mathbb{C}^m$. We can identify the orbit $O(v) = \{\rho(p)v \mid p \in G\}$ with G/H via $pH \mapsto \rho(p)v$. Then the distance $\text{dist}(p, q) = \|\rho(p) - \rho(q)\|_F$ induces the the Euclidean distance $\text{dist}(p, q) = \|\rho(p)(v) - \rho(q)(v)\|$ on $O(v) = G/H$.

Problem (4) thus leads to the following reduced optimization problem on $G/H \times G/H$.

Proposition 2. *Assume that dist is a biinvariant distance on G . Then $(p, q) \in G \times G$ is a solution of Problem (2) if and only if (qH, pH) is a minimum of $g: G/H \times G/H \rightarrow \mathbb{R}$,*

$$g(x, y) = \sum_{i=1}^k \text{dist}_{G/H}(q_i \cdot x, y)^2 \tag{5}$$

with $q \cdot x$ denoting the canonical action of G on G/H .

Proof. By the invariance of dist and Proposition 1 we have

$$\sum_{i=1}^k \text{dist}(q_i, pHq^{-1})^2 = \sum_{i=1}^k \text{dist}(q_i q, pH)^2 = \sum_{i=1}^k \text{dist}_{G/H}(q_i qH, pH)^2$$

Thus (p, q) solves (4) if and only if (qH, pH) is a minimum of g .

4.2 An algorithm for Riemannian fitting

If the distance on G is the Riemannian distance of a biinvariant Riemannian metric, we can derive a general gradient algorithm to find a minimum of (5). As discussed in the examples above the induced distance on G/H from the biinvariant metric on G is the Riemannian distance with respect to the normal metric on G/H . Thus we assume that G/H carries this normal metric in the remainder of this section. Note that

$$g(x, y) = \sum_{i=1}^k \text{dist}_{G/H}(q_i \cdot x, y)^2 = \sum_{i=1}^k \text{dist}_{G/H}(x, q_i^{-1} \cdot y)^2 \tag{6}$$

is in each variable the Karcher mean cost function for points $q_i \cdot x$ resp. $q_i^{-1} \cdot y$ on G/H . It is well-known that the gradient of the Karcher mean cost $c(x) = \sum_{i=1}^k \text{dist}(x, x_i)^2$ is given by $\text{grad } c(x) = \frac{1}{k} \sum_{i=1}^k \exp_x^{-1}(x_i)$, see [4, 7, 11]. Thus the gradient of g with respect to the product metric on $G/H \times G/H$ is

$$\text{grad } g(x, y) = \left(\frac{1}{k} \sum_{i=1}^k \exp_x^{-1}(q_i^{-1} \cdot y), \frac{1}{k} \sum_{i=1}^k \exp_y^{-1}(q_i \cdot x) \right).$$

The form (6) of the cost suggests the following gradient descent algorithm to minimize g as an adaption of the Karcher mean algorithm [4, 7, 11].

Riemannian fitting algorithm

1. Initialize $x_0, y_0 \in G/H$ and choose a $\varepsilon > 0$
2. $x_{j+1} = \exp_{x_j} \left(\frac{1}{k} \sum_{i=1}^k \exp_x^{-1}(q_i^{-1} \cdot y_j) \right)$
3. $y_{j+1} = \exp_{y_j} \left(\frac{1}{k} \sum_{i=1}^k \exp_y^{-1}(q_i \cdot x_j) \right)$
4. go to step 2 until $\text{dist}(x_j, x_{j+1}) < \varepsilon$ and $\text{dist}(y_j, y_{j+1}) < \varepsilon$
5. Let $x_j = qH$, $y_j = rH$.
6. Output: (r, q) as an approximation of the minimum of f

This algorithm requires that the $q_i^{-1} \cdot y_j$ resp. $q_i \cdot x_j$ are in the domain of $\exp_{x_j}^{-1}$ resp. $\exp_{y_j}^{-1}$ and is not necessarily globally defined. However, since these are exponential maps on G/H the algorithm will work for data clustered near a coset pHq^{-1} even if there is a continuum of Karcher means on G . An alternative would be to globalize the algorithm using non-smooth optimization methods, but this is beyond the scope of the present paper.

5 Example: Fitting on $SO(n)$

We illustrate the proposed approach on the special orthogonal group $SO(n)$.

The distances discussed in the examples (a), (b) above yield two choices for distances on $SO(n)$: (a) the Riemannian distance of a biinvariant metric and (b) the Frobenius norm distance on the matrix representation of $SO(n)$.

(a) In the Riemannian case the induced distance on $SO(n)/H$ is the normal Riemannian metric and the algorithm from Section 4.2 can be applied to compute the optimal coset on $SO(n)$. As a special case consider the problem of fitting data with a coset of a conjugate of a subgroup $\mathcal{H} \cong SO(n-1)$. The quotient space $SO(n)/\mathcal{H}$ can be identified with S^{n-1} via the diffeomorphism $Q\mathcal{H} \mapsto Qv$ for $v \in S^{n-1}$ such that $\text{stab}(v) = \mathcal{H}$. Any biinvariant Riemannian metric on $SO(n)$ has the form $\langle X\Omega, X\Theta \rangle = C \text{tr}(\Omega^T\Theta)$ with $C > 0$; w.l.o.g. assume $C = \frac{1}{2}$. Then the normal metric on S^{n-1} coincides with the Riemannian metric on S^{n-1} . Thus the exponential map on the sphere is given by $\exp_x(v) := \cos(\|v\|)x + \frac{\sin(\|v\|)}{\|v\|}v$ and its inverse by $\exp_x^{-1}(y) := \frac{s}{\sin(s)}(y - \cos(s)x)$ with $s = \arccos(y^T x)$. Using this information, it is straightforward to implement the algorithm from Section 4.2.

(b) As an example for the Frobenius norm distance on a matrix representation of $SO(n)$, consider the representation $\rho(U): \mathbb{C}^{n \times p} \rightarrow \mathbb{C}^{n \times p}$ with $\rho(U)(A) = UA$. We treat $\mathbb{C}^{n \times p}$ as the vector space \mathbb{C}^{np} . Then $\rho(U) = (I_p \otimes U)$ and the Frobenius norm distance is given by

$$\text{dist}_F(U, V) = \|\rho(U) - \rho(V)\|_F = \|(I_p \otimes U) - (I_p \otimes V)\|_F = p\|U - V\|_F.$$

Let $A = (I_p \ 0)^T \in \mathbb{C}^{n \times p}$. Assume that we want to fit a coset of a subgroup conjugate to $\mathcal{H} = \text{stab}(A) \cong SO(n - p)$ to the data. The orbit $O(A)$ is the compact Stiefel manifold $\text{St}(n, p)$ and we can identify $SO(n)/\mathcal{H}$ with $\text{St}(n, p)$ by $U\mathcal{H} \mapsto \rho(U)A$. By Section 4.1, Example (b), the induced distance on $SO(n)/\mathcal{H}$ is the Euclidean distance on $\text{St}(n, p)$, i.e.

$$\text{dist}_{SO(n)/\mathcal{H}}(UA, VA) = \|UA - VA\|_F.$$

Thus to find the best fitting coset $P\mathcal{H}Q^{-1}$, $P, Q \in SO(n)$, to data points Q_1, \dots, Q_k in $SO(n)$ one must minimize the cost

$$g(X, Y) = \sum_{i=1}^k \|X - Q_i Y\|_F^2$$

on $\text{St}(n, p) \times \text{St}(n, p)$. Here, we use the gradient descent with retractions from [6] on the product of the Stiefel manifold. To compute a gradient we use the Riemannian metric on the Stiefel manifold induced by the Euclidean one on $\mathbb{R}^{n \times p}$ and equip $\text{St}(n, p) \times \text{St}(n, p)$ with the product metric. The gradient with respect to this induced Riemannian metric is given by the orthogonal projection of the Euclidean gradient of an extension of g to $\mathbb{R}^{n \times p} \times \mathbb{R}^{n \times p}$ onto the tangent space $T_{(X, Y)}(\text{St}(n, p) \times \text{St}(n, p))$. Since the Euclidean gradient of g is given by $\text{grad}_E g(X, Y) = \left(\sum_{i=1}^k (X - Q_i^T Y), \sum_{i=1}^k (Y - Q_i X) \right)$ and the projection $\pi_X: \mathbb{R}^{n \times p} \rightarrow T_X \text{St}(n, p)$ by $\pi_X(V) = V - \frac{1}{2}X(X^T V + V^T X)$, cf. [6], we obtain $\text{grad} g(X, Y) = \left(\left(\frac{1}{2}X X^T - I_n \right) \sum_{i=1}^k Q_i^T Y + \frac{1}{2}X Y^T \sum_{i=1}^k Q_i X, \left(\frac{1}{2}Y Y^T - I_n \right) \sum_{i=1}^k Q_i X + \frac{1}{2}Y X^T \sum_{i=1}^k Q_i^T Y \right)$. A descent algorithm on a manifold needs suitable local charts R_X which map lines in the tangent space onto curves in the manifold. Here, we choose for the Stiefel manifold the polar decomposition retractions from [6], i.e. $R_X: T_X \text{St}(n, p) \rightarrow \text{St}(n, p)$, $R_X(V) = (X + V)(I_p + V^T V)^{-1/2}$. Since we have to optimize over the product of two Stiefel manifolds, we use this retraction in each component. The step length of the gradient descent is determined by an Armijo line search. This yields the following algorithm:

1. Initialize $X_0, Y_0 \in \text{St}(n, p)$ and choose a $\varepsilon > 0$, $\sigma \in (0, 1)$
2. Calculate $S_{X,j} = \sum_{i=1}^k Q_i X_j$ and $S_{Y,j} = \sum_{i=1}^k Q_i^T Y_j$.
3. Set $\eta_j := \left(\frac{1}{2}X_j X_j^T - I_n \right) S_{Y,j} + \frac{1}{2}X_j Y_j^T S_{X,j}$, $\zeta_j := \left(\frac{1}{2}Y_j Y_j^T - I_n \right) S_{X,j} + \frac{1}{2}Y_j X_j^T S_{Y,j}$
4. Choose the smallest $\alpha \in \mathbb{N}$ such that

$$g(X_j, Y_j) - g(R_{X_j}(-2^{-\alpha}\eta_j), R_{Y_j}(-2^{-\alpha}\zeta_j)) \geq \sigma 2^{-\alpha} (\|\eta_j\|_F^2 + \|\zeta_j\|_F^2)$$

5. Set $X_{j+1} := (X_j - 2^{-\alpha}\eta_j) (I_p + 2^{-2\alpha}\eta_j^T \eta_j)^{-1/2}$,
 $Y_{j+1} := (Y_j - 2^{-\alpha}\zeta_j) (I_p + 2^{-2\alpha}\zeta_j^T \zeta_j)^{-1/2}$

6. If $\|\eta_j\| > \varepsilon$ or $\|\zeta_j\| > \varepsilon$ then $j := j+1$ and go to step 2, otherwise go to step 7.
7. Find $Q, R \in SO(n)$ such that $X_j = QA, Y_j = RA$ and output (R, Q) as an approximation of the minimum of f .

Figure 1 shows the behavior of the algorithm for the Riemannian distance and the $\mathcal{H} \cong SO(n-1)$ with 30 data points in $SO(10)$. The data points for the left graph are constructed by choosing random points on a coset $\cong SO(9)$, while for the right graph randomly chosen data points on the coset were perturbed by multiplication with i.d.d. random rotations $R = \exp(N)$ with N the skew-symmetric parts of i.d.d. random matrices $M \sim N(0, \sqrt{0.1})$. For the unperturbed case the algorithm shows linear convergence as it is to be expected for a gradient method. In the perturbed case the algorithm converges quickly to a cost function value larger than 0 since an exact fitting is not possible anymore.

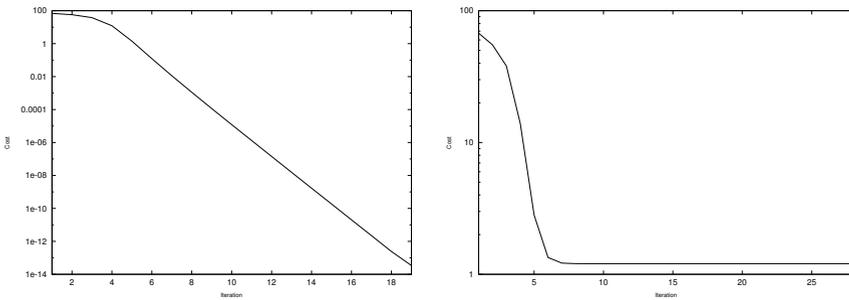


Fig. 1. Evolution of the cost for the first example in Section 5 with $n = 10$ and $k = 30$. The left figure shows the unperturbed case while the right the case of data points perturbed by random rotations.

Figure 2 illustrates the behavior of the algorithm for the Frobenius norm distance and the $\mathcal{H} = \text{stab}((I_p 0)^T) \cong SO(n-p)$ with $n = 10, p = 8$ and $k = 30$. The left graph shows the case of data points randomly chosen on a fixed coset, while the right graph shows the case of random points on the coset perturbed by a random rotations $R = \exp(N)$ with N the skew-symmetric part of random $M \sim N(0, \sqrt{0.1})$.

Acknowledgments

This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its authors.

The research was initiated during the postdoctoral stay of the first author at the University of Liège.

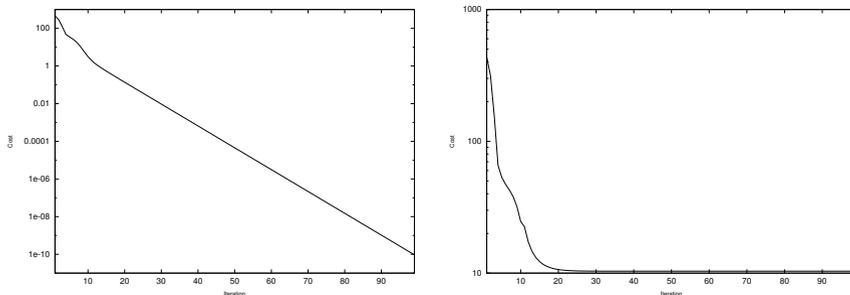


Fig. 2. Evolution of the cost for the second example in Section 5 with $n = 10$, $p = 8$ and $k = 30$. The left figure shows the case of unperturbed data on a coset while in the right one the data points have been perturbed by random rotations.

References

1. S. Helgason. (1994). Geometric analysis on symmetric spaces. American Math. Soc., Providence, RI.
2. K. V. Mardia, P. E. Jupp. (2000). Directional Statistics. Wiley, Chichester.
3. I. T. Jolliffe. (1986). Principal Component Analysis. Springer-Verlag, New York.
4. P.T. Fletcher, C. Lu, S. Joshi. (2003). Statistics of Shape via Principal Geodesic Analysis on Lie Groups. In: Proc. 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR03) p. I-95 – I-101
5. P.T. Fletcher, C. Lu, S.M. Pizer, S. Joshi. (2004). Principal Geodesic Analysis for the Study of Nonlinear Statistics of Shape. IEEE Trans. Medical Imaging 23(8):995–1005
6. P.-A. Absil, R. Mahony, R. Sepulchre. (2008). Optimization Algorithms on Matrix Manifolds. Princeton University Press, Princeton
7. L. Machado (2006) Least Squares Problems on Riemannian Manifolds. Ph.D. Thesis, University of Coimbra, Coimbra
8. L. Machado, F. Silva Leite (2006). Fitting Smooth Paths on Riemannian Manifolds. Int. J. Appl. Math. Stat. 4(J06):25–53
9. J. Cheeger, D. G. Ebin (1975). Comparison theorems in Riemannian geometry. North-Holland, Amsterdam
10. H. Karcher (1977). Riemannian center of mass and mollifier smoothing. Comm. Pure Appl. Math. 30:509–541
11. J. H. Manton. (2004). A Globally Convergent Numerical Algorithm for Computing the Centre of Mass on Compact Lie Groups. Eighth Internat. Conf. on Control, Automation, Robotics and Vision, December, Kunming, China. p. 2211–2216
12. M. Moakher. (2002). Means and averaging in the group of rotations. SIAM Journal on Matrix Analysis and Applications 24(1):1–16
13. J. H. Manton. (2006). A centroid (Karcher mean) approach to the joint approximate diagonalisation problem: The real symmetric case. Digital Signal Processing 16:468–478