Elsevier Editorial System(tm) for Journal of Multivariate Analysis
Manuscript Draft

Manuscript Number:

Title: Scale and Dimension Reduction in the Space of Positive Definite Matrices

Article Type: Regular Article

Keywords: dimension reduction, principal geodesic analysis (PGA), manifold-valued statistics, data scaling, diffusion tensors

Corresponding Author: Mr. Drew Lazar,

Corresponding Author's Institution: Central Michigan University

First Author: Drew Lazar

Order of Authors: Drew Lazar

Abstract: Using Riemannian manifolds to model data with non-linear constraints has been of increased interest recently. For this purpose, different methods of principal component analysis have been generalized to Riemannian manifolds. These methods, equivalent in the linear setting, produce different results when there is curvature. This paper considers and compares different ways to generalize principal component analysis on the space of positive definite matrices and introduces \textit{recursive principal geodesic analysis} as such a generalization. In order to explore the effect of curvature in its relation to the scale of data, a scaling parameter is introduced and expansions are obtained in this parameter. Also, numerical simulations are run to demonstrate each expansion and to compare the use of each procedure.

CMU CENTRAL MICHIGAN UNIVERSITY | College of Science & Technology
Department of Mathematics

Drew M. Lazar
214 Pearce Hall
1200 S. Franklin Street
Mount Pleasant, MI 48859

October 24, 2014

Dear Dr. Jan de Leeuw,

I wish to submit a new manuscript entitled "Scale and Dimension Reduction in the Space of Positive Definite Matrices" for consideration by the *Journal of Multivariate Analysis*.

I confirm that this work is original and has not been published nor is it currently under consideration for publication elsewhere.

In this paper, I report on an investigation of principal geodesic analysis (PGA) on the space of n-by-n positive definite matrices in relation to the scale and dispersion of data and a new, recursive method of dimension reduction on this space. This paper should be of interest to readers in the areas of dimension reduction, regression analysis, non-linear statistics and diffusion tensor analysis.

Dimension reduction on data with non-linear constraints should respect these constraints to produce meaningful results. Principal geodesic analysis (PGA) has been introduced for this purpose and is a non-linear, multivariate technique of generalizing principal component analysis.

In the tangent space at the intrinsic mean of data, where functions to be optimized in PGA have their domains, I introduce a scaling parameter in order to investigate the effect of the relationship between the scale of data and the curvature of the manifold on solutions to PGA. I also formulate and consider another such generalization of principal component analysis, which is entirely recursive and simplifies the projection of data to submanifolds in PGA to projection to geodesics.

In extending the widely used and straightforward techniques of principal component analysis these methods involve the powerful techniques of differential geometry to deal with the complications of curvature. While it addresses a multivariate statistical problem, I believe this manuscript would be of unique yet relevant interest to your readership. In 2009, you included a paper addressing this topic in the context of shape analysis, "Principal Component Geodesics for Planar Shape Spaces". I hope that my submission can make a similar contribution to your readers' understanding of this area of burgeoning interest to statisticians and applied mathematicians alike.

Thank you for your consideration.

Sincerely,

Drew M. Lazar

DEPARTMENT OF MATHEMATICS • PHONE (989) 774-3596 • FAX (989) 774-2414
CENTRAL MICHIGAN UNIVERSITY • MOUNT PLEASANT, MICHIGAN 48859
math@cmich.edu • www.cst.cmich.edu/units/mth

SCALE AND DIMENSION REDUCTION IN THE SPACE OF POSITIVE DEFINITE MATRICES

DREW M. LAZAR

Abstract. Using Riemannian manifolds to model data with non-linear constraints has been of increased interest recently. For this purpose, different methods of principal component analysis have been generalized to Riemannian manifolds. These methods, equivalent in the linear setting, produce different results when there is curvature. This paper considers and compares different ways to generalize principal component analysis on the space of positive definite matrices and introduces *recursive principal geodesic analysis* as such a generalization. In order to explore the effect of curvature in its relation to the scale of data, a scaling parameter is introduced and expansions are obtained in this parameter. Also, numerical simulations are run to demonstrate each expansion and to compare the use of each procedure.

Keywords. dimension reduction, principal geodesic analysis (PGA), manifold-valued statistics, data scaling, diffusion tensors

1. Introduction

In order to reduce dimensionality and explore the variance structure of manifold-valued data, one can identify explanatory directions in the tangent space at a mean of the data. In this paper we examine procedures for identifying such directions on the symmetric space of $n \times n$ positive-definite matrices.

We denote the space of $n \times n$ positive-definite matrices as $P(n)$ and the tangent space to $P(n)$ at a point $p \in P(n)$ as $\mathrm{T}_p P(n)$. Also, we let μ_D be the *intrinsic mean* of given data D in $P(n)$, that is, the point in $P(n)$ that minimizes the sum of squared geodesic distances in $P(n)$ to the data in D.

One procedure for identifying explanatory directions, *principal geodesic analysis* (PGA), as formulated in [1], successively identifies orthogonal explanatory directions in $\mathrm{T}_{\mu_D} P(n)$ and then exponentiates the span of the results to form submanifolds geodesic at the mean. In addition, in this paper, an alternative method of dimension reduction is formulated, which we call *recursive principal geodesic analysis* (RPGA). PGA and RPGA each generalize one of several methods of principal component analysis (PCA), which in the linear setting produce equivalent results.

In section 2, we review PCA and in section 3 we consider the geometry of $P(n)$ necessary to generalize PCA. In section 4, the formulation of PGA as in [1] is given and in section 5 RPGA is formulated.

RPGA removes variability in explanatory directions before finding the next direction. In this process, RPGA adjusts for the movement of the mean explained by variability in geodesic directions and the curvature of the manifold. Like PGA, RPGA locates explanatory submanifolds geodesic at the mean, but only involves projection to geodesics rather than to submanifolds of increasing dimension.

In section 6, we explain how we numerically compute explanatory directions as formulated in previous sections. Such computations will then be used in simulations in the last two sections.

By employing an isometry on $P(n)$, dimension reduction procedures can be carried out in the tangent space at the identity. We denote the identity by I and in section 7, we project data to $\mathrm{T_I}P(n)$ by the *Riemmanian log map* and then scale data in $\mathrm{T_I}P(n)$ by a scaling parameter ϵ. By expansion in ϵ we can then see how linear approximations in the tangent space interact with ϵ and higher order curvature terms to form parts of the dimension reduction procedures.

In section 8, data in $P(3)$ is generated by simulation and first through fourth explanatory directions for different methods are computed for different scales in the tangent space. Sum of squared residuals and computation times are reported and these results, in which RPGA is shown to be a useful method of computing explanatory directions and submanifolds, are discussed.

2. Principal Component Analysis

In an inner product space, principal component analysis (PCA) finds a sequence of linear, orthogonal principal component directions such that projection of variables to the first direction has the highest amount of variance, the projection to the next direction has the next highest amount of variance, *et cetera*. In this way a new set of uncorrelated variables is created which have lower dimension but which still account for much of the variability. These can then be used as inputs in a regression model while uncovering the variance structure of the initial data set.

Formally, assume X is an $\mathbb{R}^n$-valued random variable, or, more generally, a random variable in an inner product space of dimension n. Then let $A_m = \{S_m; S_m \text{ is an affine subspace of } \mathbb{R}^n \text{ with } \dim(S_m) = m\}$, $\pi_{S_m}(X)$ be the orthogonal projection of X to S_m for $S_m \in A_m$ and E be the expected value operator. PCA then locates $S_{X,m} \in A_m$ such that

S1. $\mathrm{var}(\pi_{S_{X,m}}(X)) \geq \mathrm{var}(\pi_{S_m}(X)) \ \forall S_m \in A_m$
or equivalently such that,

S2. $E(\mathrm{d}(X, \pi_{S_{X,m}}(X))^2) \leq E(\mathrm{d}(X, \pi_{S_m}(X))^2) \ \forall S_m \in A_m$.

The equivalence above follows from the Pythagorean theorem in an inner product space. Further, these subspaces are nested for increasing m. In particular, they all contain $S_{X,0} = E(X)$.

Assume $E(X) = 0$, let $X_1 = X$ and the v_i's below be unit vectors in $\mathbb{R}^n$ originating at the origin. Then $S_{X,m}$ can be found in either one of the following ways:

PCA 1. Constrained minimization

I. Find v_1 s.t. $\text{span}(v_1) = S_{X,1}$

II. For $i = 2$ to m find v_i s.t

i. $v_j \perp v_i \ \forall j < i$

ii. $S_{X,i} = \text{span}(\{v_1, \cdots, v_i\})$.

PCA 2. Recursively removing variability in principal directions

Set i=1.

I. Find v_i s.t. $\text{span}(v_i) = S_{X,1}$

II. Set $X' = X - \pi_{S_{X,1}}(X_i)$

III. If $i < m$ set $i = i + 1, X = X'$ and return to I, else stop.

At the i-th step PCA 1 locates v_i, orthogonal to all previous directions, that minimizes the expected squared distance of X to its projection to $\text{span}(v_i)$. As the projection of X to $\text{span}(v_i)$ is the projection of X with directions orthogonal to v_i removed to $\text{span}(v_i)$, PCA 1 and PCA 2 produce the same vectors at each step. These are known as *principal component directions* and they are also the first m normalized eigenvectors of the covariance matrix of the entries of X. If $E(X) = \mu$ then X can be demeaned, each v_i located as above and then the principal component directions of X are given by $v_i + \mu$ for $i = 1, \ldots, m$.

Assuming X is a random variable whose value has non-linear constraints then PCA may lead to projection outside of the those constraints. For example, the $n \times n$ positive definite matrices are in $\mathbb{R}^{n^2}$ but are not closed under scalar multiplication and thus not a linear space. Thus generalizations of PCA, such a *principal geodesic analysis* in [1] and *geodesic principal component analysis* in [4], that afford benefits of PCA while respecting the geometry of the constraints have been introduced.

3. Background on $P(n)$

Consider the action on $P(n)$:

$$\begin{aligned} &\varphi : GL(n) \times P(n) \to P(n), \\ &\varphi(g, p) = \varphi_g(p) = gpg^\top \end{aligned} \tag{1}$$

where $GL(n)$ is the general linear group. As $P(n)$ is an open set in the vector space of $n \times n$ symmetric matrices, we have the identification

$$\mathrm{T}_p P(n) \equiv n \times n \text{ symmetric matrices}$$

for any $p \in P(n)$ where $\mathrm{T}_p P(n)$ is the tangent space at p.

Further, φ is transitive as any $p \in P(n)$ can be decomposed as $p = gg^\top$ where $g \in GL(n)$. Then we have the Riemannian metric for which ϕ_g is an isometry

$$\langle X, Y \rangle_p = tr(g^{-1} X p^{-1} Y (g^{-1})^\top). \tag{2}$$

for $X, Y \in \mathrm{T}_p P(n)$ and where tr denotes the matrix trace.

For any smooth mapping between manifolds $f : M_1 \to M_2$ and $x \in M_1$ we denote the derivative map of f at x by $d_x f$ where $d_x f : \mathrm{T}_x M_1 \to \mathrm{T}_{f(x)} M_2$.

Let $\mathrm{I} \in P(n)$ be the identity matrix. With (2), any geodesic through $\mathrm{I} \in P(n)$, given by $t \to \mathrm{Exp}_{\mathrm{I}}(tX)$ for some $X \in \mathrm{T}_{\mathrm{I}} P(n)$, has $\mathrm{Exp}_{\mathrm{I}}$ as the matrix exponential. Then for any $p = gg^\top \in P(n)$ where $g \in GL(n)$ and $Y \in \mathrm{T}_p P(n)$ we have the geodesic $t \to \mathrm{Exp}_p(tY)$ given by

$$t \to \mathrm{Exp}_p(tY) = \phi_g\left(\mathrm{Exp}_{\mathrm{I}}\big(d_p \phi_{g^{-1}}(tY)\big)\right) = g\,\mathrm{Exp}_{\mathrm{I}}\Big(tg^{-1}Y(g^{-1})^\top\Big)g^\top.$$

The *Riemannian exponential map* at p of Y is then defined as $\mathrm{Exp}_p(Y)$. As $P(n)$ is of non-positive curvature Exp_p is injective. Thus its inverse, the *Riemannian log map* at p, denoted here by Log_p, is well-defined.

Using (2), for any $p, q \in P(n)$, the geodesic distance between p and q is $\mathrm{d}(p, q) = \|\mathrm{Log}_p(q)\|$. Then given $D = \{p_1, \dots, p_N\} \subset P(n)$ the *intrinsic mean* of D is defined

$$\mu_D = \operatorname*{argmin}_{x \in P(n)} \frac{1}{N} \sum \mathrm{d}(x, p_i)^2. \tag{3}$$

As $P(n)$ is of non-positive curvature, μ_D is guaranteed to exist and to be unique and considering the gradient of the cost function minimized above, μ_D is the point in $P(n)$ such that

$$\sum \mathrm{Log}_{\mu_D}(p_i) = 0. \tag{4}$$

Further, for any $p, q \in P(n)$, $V \subset \mathrm{T}_q P(n)$ and $S(V) = \mathrm{Exp}_q(\mathrm{span}(V))$ we define the *projection operator* as

$$\pi_{S(V)}(p) = \operatorname*{argmin}_{x \in S(V)} \mathrm{d}(x, p). \tag{5}$$

As $P(n)$ is of non-positive curvature, $\pi_{S(V)}(p)$ is also guaranteed to exist and to be unique.

4. Principal geodesic analysis

Principal geodesic analysis (PGA), as introduced in [1], generalizes PCA 1 in section 2 by projecting to submanifolds geodesic at the mean which are not necessarily linear subspaces.

Given $\mathrm{D} = \{p_1, \dots, p_N\} \subset P(n)$ with $\mu_D = \mathrm{I}$ and $m \le M = (n+1)n/2 = \dim(\mathrm{T}_{\mathrm{I}} P(n))$, PGA locates $\{v_1, \dots, v_m\} \subset \mathrm{T}_{\mathrm{I}} P(n)$ such that

$$\begin{aligned} v_1 &= \operatorname*{argmax}_{\|v\|=1} \frac{1}{N} \sum \mathrm{d}(\mathrm{I}, \pi_{S_1(v)}(p_i))^2 \\ &\vdots \qquad\qquad \vdots \\ v_m &= \operatorname*{argmax}_{\|v\|=1,\, v \in V_m^\perp} \frac{1}{N} \sum \mathrm{d}(\mathrm{I}, \pi_{S_m(v)}(p_i))^2 \end{aligned} \tag{6}$$

where $V_i = \mathrm{span}(\{v_1, \ldots, v_{i-1}\})$ and $S_i(v) = \mathrm{Exp}_{\mathrm{I}}(\mathrm{span}(\{v_1, \ldots, v_{i-1}, v\}))$ for $i = 2, \ldots, m$.

Alternatively, generalizing PCA 1 by finding submanifolds geodesic at I which minimize sum of squared residuals as in S2 in section 2, PGA is formulated as

$$\begin{aligned} v_1 = &\ \underset{\|v\|=1}{\operatorname{argmin}} \ \frac{1}{N} \sum \mathrm{d}(p_i, \pi_{S_1(v)}(p_i))^2 \\ \vdots &\qquad \vdots \\ v_m = &\ \underset{\|v\|=1,\, v \in V_m^{\perp}}{\operatorname{argmin}} \frac{1}{N} \sum \mathrm{d}(p_i, \pi_{S_m(v)}(p_i))^2 \end{aligned} \tag{7}$$

where $V_i = \mathrm{span}(\{v_1, \ldots, v_{i-1}\})$ and $S_i(v) = \mathrm{Exp}_{\mathrm{I}}(\mathrm{span}(\{v_1, \ldots, v_{i-1}, v\}))$ for $i = 2, \ldots, m$.

The $v_i's$ in either (6) or (7) are known as *principal geodesic directions.* A data set with mean $\mu = gg^{\top} \in P(n)$ can be demeaned by applying $\phi_{g^{-1}}$ to its data. Then solutions can be located as in (6) or (7) and principal geodesic directions are given by $d_{\mathrm{I}}\phi_g(v_i)$ for $i = 1, \ldots, m$.

As $P(n)$ in non-linear, (6) and (7) do not produce equivalent solutions as in S1 and S2 in PCA. In agreement with [9] I found (6) to be less computational stable than (7) and in this paper the ladder formulation is used.

4.1. **Approximate Principal Geodesic Analysis.** Solutions to PGA do not have closed form and can be computationally difficult to locate. Thus, a first-order approximation was suggested in [1] and is called *linear principal geodesic analysis* (LPGA) here.

Given $p \in P(n)$, $\{v_1, \ldots, v_k\} \subset \mathrm{T}_{\mathrm{I}}P(n)$ and $S_k = \mathrm{Exp}_{\mathrm{I}}(\mathrm{span}(\{v_1, \ldots, v_k\}))$ an approximation of the projection operator in (5) is given as

$$\begin{aligned} \mathrm{Log}_{\mathrm{I}}(\pi_{S_k}(p)) &\approx \sum tr\,(v_j \mathrm{Log}_{\mathrm{I}}(p))\, v_j \\ \pi_{S_k}(p) &\approx \mathrm{Exp}_{\mathrm{I}}\Big(\sum tr\,(v_j \mathrm{Log}_{\mathrm{I}}(p))\, v_j\Big) \end{aligned} \tag{8}$$

Now using (8) in (7) (or in (6)) we have

$$\begin{aligned} v_1 \approx \tilde{v}_1 = &\ \underset{\|v\|=1}{\operatorname{argmin}} \ -\frac{1}{N} \sum tr\,(v \mathrm{Log}_{\mathrm{I}}(p_i))^2 \\ \vdots &\qquad \vdots \\ v_m \approx \tilde{v}_m = &\ \underset{\|v\|=1,\, v \in \tilde{V}_m^{\perp}}{\operatorname{argmin}} -\frac{1}{N} \sum tr\,(v \mathrm{Log}_{\mathrm{I}}(p_i))^2 \end{aligned} \tag{9}$$

where $\tilde{V}_i = \mathrm{span}(\{\tilde{v}_1, \ldots, \tilde{v}_{i-1}\})$ for $i = 2, \ldots, m$ and $\{\tilde{v}_1, \ldots, \tilde{v}_m\}$ are approximate to principal geodesic directions in PGA. This is exactly PCA with the logs of the data in the tangent space at the identity.

Thus, letting $q^i = \mathrm{Log}_{\mathrm{I}}(p_i)$ for $i = 1, \ldots, N$, $\{\tilde{v}_1, \ldots, \tilde{v}_m\}$ are first m normalized eigenvectors of the covariance operator formed by $\{q^1, \ldots, q^N\}$.

That is, if

$$L(w) = \frac{1}{N} \sum tr\left(q^i w\right) q^i \quad \text{for } w \in \mathrm{T_I} P(n) \tag{10}$$

then $\{\tilde{v}_1, \ldots, \tilde{v}_m\}$ are the first m normalized eigenvectors of L.

Let $\{e_a\}_a$ and $\{E_{b,c}\}_{c \geqslant b}$ be the $M = n(n+1)/2$ standard basis vectors of $\mathbb{R}^M$ and $\mathrm{T_I}P(n)$, respectively. That is, for $1 \leq a \leq M$, where e_a is a column vector, let

$$(e_a)_k = \delta_{a,k}$$

for $1 \leq k \leq M$ and for $1 \leq b \leq c \leq n$, let

$$(E_{b,c})_{j,l} = \begin{cases} \delta_{b,j}\delta_{c,l} + \delta_{b,l}\delta_{c,j} & \text{if } c > b \\ \delta_{b,j}\delta_{b,l} & \text{if } b = c \end{cases}$$

for $1 \leq j, l \leq n$.

Then for $1 \leq \iota \leq \kappa \leq n$ let

$$s_{\iota,\kappa} = \begin{cases} \left(1/\sqrt{2}\right) E_{\iota,\kappa} & \text{if } \kappa > \iota \\ E_{\iota,\iota} & \text{if } \iota = \kappa \end{cases} \tag{11}$$

Define the linear isometry $\rho : P(n) \to \mathbb{R}^M$ by

$$\rho(s_{\iota,\kappa}) = e_{r(\iota,\kappa)} \quad \text{for } \kappa \geq \iota \tag{12}$$

where $r(\iota, \kappa) = \iota + \kappa(\kappa - 1)/2$. Then with

$$Q = \sum \rho(q^i)\rho(q^i)^\top$$

we have $\{\tilde{v}_1, \ldots, \tilde{v}_m\} = \{\rho^{-1}(w_1), \ldots, \rho^{-1}(w_m)\}$ where $w_1, \ldots, w_m$ are the first m normalized eigenvectors of Q.

5. Recursive Principal Component Analysis

Recursive principal geodesic analysis (RPGA), introduced here, generalizes PCA 2 in section 2 by recursively removing variability in first principal geodesic directions.

In PCA 2, as the space is linear, after removing variability in a principal component direction the mean remains at the identity. In $P(n)$ however, where there is curvature, removing variability in an explanatory direction results in data with a mean different than I. However, by demeaning the data after each such step this effect can be undone and the next direction can found in a recursive fashion. As a result, a sequence of explanatory direction in $\mathrm{T_I}P(n)$ is identified and data can be projected to the exponentiated span of these directions.

Note this is an almost analogous procedure to the location of *principal geodesic curves* in Lie groups given in [3], with group multiplications replaced by a group action and with an extra step to demean data after variability is removed in an explanatory direction.

Operators Γ and Φ_v, introduced below, demean data and remove variability in direction v from data with mean I, respectively. In both cases, decomposition is done so that when data is acted upon by the inverse of decompositions unnecessary rotational effects are avoided.

Let $B = \{b_1, \ldots, b_N\} \subset P(n)$ with $\mu_B = \mathrm{Exp_I}(x)$ for $x \in \mathrm{T_I}P(n)$ and $g = \mathrm{Exp_I}(x/2)$. Define

$$\Gamma(B) = \{\varphi_{g^{-1}}(b_1), \ldots, \varphi_{g^{-1}}(b_N)\}. \tag{13}$$

Let $\mathrm{C} = \{c_1, \ldots, c_N\} \subset P(n)$ with $\mu_C = \mathrm{I}$, $v \in \mathrm{T_I}P(n)$ and $S(v) = \mathrm{Exp_I}(\mathrm{span}(v))$. Further, let $\pi_{S(v)}(c_j) = h_j h_j$ with $h_j = \mathrm{Exp_I}((1/2)t_j v)$ and $t_j \in \mathbb{R}$ for $j = 1, \ldots, N$. Define

$$\Phi_v(C) = \{\varphi_{h_1^{-1}}(c_1), \ldots, \varphi_{h_n^{-1}}(c_N)\} \tag{14}$$

Then given $\mathrm{D} = \{p_1, \ldots, p_N\} \subset P(n)$ with $\mu_D = \mathrm{I}$ and $m \leqslant M = n(n+1)/2 = \dim(\mathrm{T_I}P(n))$, RPGA locates $\{v_1, \ldots, v_m\} \subset \mathrm{T_I}P(n)$ as follows:

RPGA. Recursive PGA

Set $i = 1$.

I. Find v_i s.t. $v_i = \underset{\|v\|=1}{\mathrm{argmin}} \sum \mathrm{d}(\pi_{S_v}(p_i), p_i)^2$.

II. Set $D' = \Gamma(\Phi_{v_i}(D))$

III. If $i < m$ set $i = i + 1, D = D'$ and return to I, else stop.

6. Numerical Computations

GNU Octave is used to compute the parts of the the procedures outlined above in simulations in section 7 and in section 8 . To compute the intrinsic mean of data the gradient damping algorithm detailed in [2] is implemented. The computations of the projection operator and principal geodesic directions are done using Octave minimization routines. A user-supplied gradient for the projection operator as formulated in [10] is used in these routines and specified for computations in $P(n)$ below.

The PGA directions are computed using spherical coordinates to parameterize unit spheres in subspaces of $\mathrm{T_I}P(n)$. This method leaves candidate directions on these unit spheres rather than normalizing them after each step and was found to be less noisy, particulary for data scaled close to the tangent space, in plots to test expansions in section 7.

6.1. **Computation of Projection Operator.**

Let $V = \{v_1, \ldots, v_m\} \subset \mathrm{T_I}P(n)$ and $S(V)$ be the submanifold of $P(n)$ given by $S(V) = \mathrm{Exp_I}(\mathrm{span}(V))$.

Further, define the functions γ, φ and f as

$$\gamma : \mathbb{R}^m \to P,\ \gamma(c_1, \ldots, c_m) = \mathrm{Exp_I}(c_1 v_1 + \ldots + c_m v_m),$$
$$\phi : P \to \mathbb{R},\ \varphi(y) = \left\|\mathrm{Log}_y(x)\right\|^2,\ \text{and}$$
$$f : \mathbb{R}^m \to \mathbb{R},\ f = \phi \circ \gamma.$$

Then

$$\pi_{S(V)}(p) = \gamma\left(\underset{c_1,\ldots,c_m}{\mathrm{argmin}}\ f(c_1, \ldots, c_m)\right). \tag{15}$$

As in [5] we have

$$d_y\phi(z) = -2\langle \mathrm{Log}_y(x), z\rangle_y$$

for $y \in P(n)$ and $z \in \mathrm{T}_I P$.

Thus, using the chain-rule we have

$$\frac{\partial f}{\partial c_i}(c_1^o, \dots, c_m^0) = -2\langle \mathrm{Log}_{\mathrm{Exp_I}(w)}(x), d_w \mathrm{Exp_I}(v_i)\rangle_{\mathrm{Exp_I}(w)} \tag{16}$$

for $c_1^0, \dots, c_m^0 \in \mathbb{R}^m$ where $w = c_1^0 v_1 + \ldots + c_m^0 v_m$. To compute $d_w\mathrm{Exp_I}(v_i)$ we use [7, Theorem 4.5] and then (16) can be used to compute the gradient of the objective function in (15).

In figure 1 below we take $V = \{v_1, v_2\}$ and $p = \mathrm{Log_I}\big(\pi_{S(V)}(x)\big)$ where the gradient evaluates to zero.

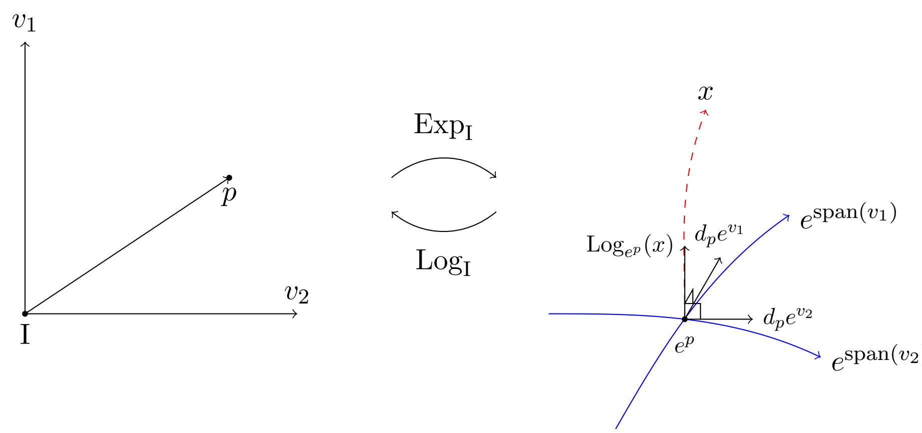


FIGURE 1. Gradient of the projection operator

As an initial guess, the first order approximation as in (8) can be used.

6.2. **Computation of PGA directions.** Let $D = \{p_1, \dots, p_N\} \subset P(n)$ be such that $\mu_D = \mathrm{I}$ and assume $V_i = \{v_1, \dots, v_{i-1}\} \subset \mathrm{T_I}P(n)$ makes up the set of the first $i-1$ PGA directions.

The i-th PGA direction lies in $U_{V_i} = \{w \in \mathrm{span}(V_i)^\perp; \|w\| = 1\}$, the unit sphere in the orthogonal complement of $\mathrm{span}(V_i)$. To find the i-th PGA direction, first an orthonormal basis for $\mathrm{span}(V_i)^\perp$, $B_{V_i^\perp} = \{b_1, \dots, b_k\}$, is located. Then U_{V_i} can be parameterized using spherical coordinates

$$\begin{aligned}(\phi_1, \dots, \phi_{k-1}) \xrightarrow{\beta} \cos\phi_1 b_1 + \sin\phi_1\cos\phi_2 b_2 + \ldots +&\\ \sin\phi_1 \cdots \sin\phi_{k-2}\cos\phi_{k-1}b_{k-1} +&\\ \sin\phi_1 \cdots \sin\phi_{k-2}\sin\phi_{k-1}b_k&\end{aligned}$$

where $0 \le \phi_1 < 2\pi$ and $0 \le \phi_j \le \pi$ for $j = 2, \dots, k-1$.

For $l = 1, \ldots, N$ let

$$\gamma_l : [0, 2\pi] \times [0, \pi) \times \cdots \times [0, \pi) \to P(n),$$
$$\gamma_l(\phi_1, \ldots, \phi_{k-1}) = \pi_{S(V_i,\beta)}(p_l)$$

for $0 \le \phi_1 < 2\pi$ and $0 \le \phi_j \le \pi$ for $j = 2, \ldots, k-1$ and where $S(V_i, \beta) = \mathrm{Exp}_{\mathrm{I}}(\mathrm{span}(V_i \cup \{\beta(\phi_1, \ldots, \phi_{k-1})\}))$.

Then let

$$(\phi_1^0, \ldots, \phi_{k-1}^0) = \underset{\phi_1,\ldots,\phi_{k-1}}{\mathrm{argmin}} \sum_{i=1}^{N} \mathrm{d}(\gamma(\phi_1, \ldots, \phi_{k-1}), p_i)^2.$$

The i-th PGA direction is given by $v_i = \beta^{-1}(\phi_1^0, \ldots, \phi_{k-1}^0)$.

For an initial guess, first project the data to $\mathrm{T_I}P(n)$ to form

$$\{\mathrm{Log}_{\mathrm{I}}(p_1), \ldots, \mathrm{Log}_{\mathrm{I}}(p_N)\}.$$

Then project each element of this set to $\mathrm{span}(V_i)^\perp$, do PCA with the resulting data in $\mathrm{span}(V_i)^\perp$ and take the initial guess as the first PCA direction. Follow the procedure in section 4.1, replacing the basis of $\mathrm{T_I}P(n)$ given in (11) with the basis of $\mathrm{span}(V_i)^\perp$ given in $B_{V_i^\perp}$ and the standard basis of $\mathbb{R}^M$, where $M = \dim(\mathrm{T_I}P(n))$, with the standard basis of $\mathbb{R}^k$ in the isometry given in (12).

7. Expansions in a scaling parameter

In dimension reduction procedures, data in $P(n)$ can scaled in the tangent space by a parameter $\epsilon \in \mathbb{R}$, $\epsilon \neq 0$ and expansions of solutions in ϵ can be derived. In this section expansions of parts of RPGA are obtained including projection to a geodesic, the first principal geodesic direction and the mean after removing a geodesic direction. First order terms will be solutions of corresponding procedures in the tangent space. Next higher order terms will include information about the effect of curvature of $P(n)$, in its relationship to ϵ, on full solutions.

7.1. **Computer Algebra.** All of the expansions in this section were found using the Maxima computer algebra system. A matrix Taylor function was coded to compute expansions involving $\mathrm{Log}_{\mathrm{I}}$ and $\mathrm{Exp}_{\mathrm{I}}$ in ϵ. These functions incorporate truncation rules to allow efficient computation. Also, a function was coded with the distributive and cyclic properties of the trace to simplify inner products. Numerical computations, as detailed in section 6, were used to test expansions by simulation and by comparing actual to predicted values according to each expansion.

7.2. **Projection coefficient.** Let $q, v \in \mathrm{T_I}P(n)$ with $\|v\| = 1$ and $p_\epsilon = \mathrm{Exp}_{\mathrm{I}}(\epsilon q)$ for $\epsilon \neq 0$. Then to project p_ϵ to $S(v) = \mathrm{Exp}_{\mathrm{I}}(\mathrm{span}(v))$ find

$$t(\epsilon) = \underset{s\in\mathbb{R}}{\mathrm{argmin}}\ \mathrm{d}(\mathrm{Exp}_{\mathrm{I}}(sv), \mathrm{Exp}_{\mathrm{I}}(\epsilon q))^2 \tag{17}$$

which has a unique solution as $P(n)$ is a space of non-positive curvature. Define $O_{\epsilon,s}(l)$ as

$$h(s,\epsilon) \text{ is } O_{\epsilon,s}(l) \Longleftrightarrow h(s,\epsilon) \leq \sum_{k=0}^{l} A_k \epsilon^k s^{l-k}$$

for some $A_1,\ldots,A_l \in \mathbb{R}$. Then let $g(s,\epsilon) = \mathrm{Exp}_{\mathrm{I}}(-sv/2)\mathrm{Exp}_{\mathrm{I}}(\epsilon q)\mathrm{Exp}_{\mathrm{I}}(-sv/2)$ and setting $f(s,\epsilon)$ as the cost function in (17) and expanding

$$\begin{aligned} f(s,\epsilon) &= \mathrm{d}(\mathrm{Exp}_{\mathrm{I}}(sv),\mathrm{Exp}_{\mathrm{I}}(\epsilon q))^2 \\ &= tr\left(\mathrm{Log}_{\mathrm{I}}(g(s,\epsilon))\mathrm{Log}_{\mathrm{I}}(g(s,\epsilon))\right) \\ &= \epsilon^2 - 2\epsilon s\, tr\,(qv) + \left(tr\left(q^2v^2\right) - tr\left((qv)^2\right)\right)\epsilon^2 s^2/6 + O_{\epsilon,s}(6). \end{aligned} \tag{18}$$

As $f(s,\epsilon)$ is real analytic in s and ϵ, letting $f_s = \frac{\partial f}{\partial s}$, $f_s(s,\epsilon)$ is real analytic in s and ϵ. Also,

$$f_s(t(\epsilon),\epsilon) = 0 \text{ and } \frac{\partial f_s}{\partial s}(t(\epsilon),\epsilon) < 0 \text{ for all } \epsilon.$$

Thus, by the implicit function theorem $t(\epsilon)$ is real analytic in ϵ. Further, as $P(n)$ is a symmetric space, the mapping

$$\iota : P(n) \circlearrowleft,\ \iota(p) = \mathrm{Exp}_{\mathrm{I}}(-\mathrm{Log}_{\mathrm{I}}(p)) \quad \text{for } p \in P(n)$$

is an isometry. Thus, $t(\epsilon)$ is odd in ϵ and we have

$$t(\epsilon) = t_1\epsilon + t_3\epsilon^3 + O(\epsilon^5)$$

for some $t_1, t_3 \in \mathbb{R}$.

Solving for t_1 and t_3 in $f_s(t(\epsilon)) = 0$ gives

$$t(\epsilon) = tr\,(qv)\,\epsilon + (1/6)tr\,(qv)\left(tr\,(qvqv) - tr\left(q^2v^2\right)\right)\epsilon^3 + O(\epsilon^5). \tag{19}$$

As $P(n)$ is a symmetric space, by [8],

$$\begin{aligned} \mathrm{R}_{q,v}\, v &= [v,[q,v]] \\ &= 2vqv - v^2q - qv^2 \end{aligned} \tag{20}$$

where R is the Riemannian curvature tensor at I. Further, letting $K(q,v)$ be the sectional curvature of span($\{q,v\}$), we have

$$\begin{aligned} K(q,v) &= \frac{tr\,((\mathrm{R}_{q,v}\, v)q)}{tr\,(qq)\, tr\,(vv) - tr\,(qv)^2} \\ &= \frac{2\, tr\left(vqvq - v^2q^2\right)}{tr\,(qq) - tr\,(qv)^2}. \end{aligned} \tag{21}$$

Using (21) in (19) we have

$$t(\epsilon) = \cos\theta\, \|q\|\,\epsilon + \frac{1}{12}\cos\theta\sin^2\theta K(v,q)\,\|q\|^3\,\epsilon^3 + O(\epsilon^5) \tag{22}$$

where θ is the angle formed by q and v.

As in figure 2 below, taking $\|q\| = 1$ so that $\epsilon = \mathrm{d}(p_\epsilon,\mathrm{I})$, as ϵ goes to zero the tangent vectors become more like their exponents and projection in the tangent space becomes more like projection in the manifold. The third order term accounts for the difference in the Euclidean triangle in the tangent space and the approximate geodesic triangle in the manifold. Further, as $P(n)$ is of non-positive sectional curvature, $\frac{1}{12}\cos\theta\sin^2\theta K(v,q)$ will be non-positive for acute angle θ.

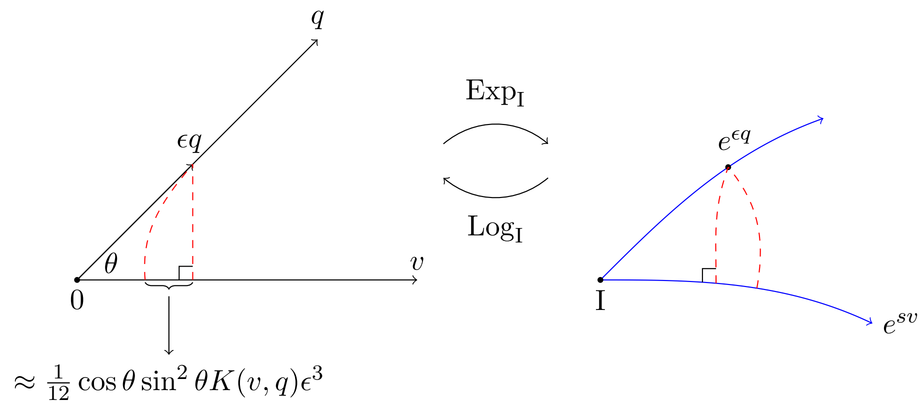


FIGURE 2. Approximation of projection coefficient

By considering the relative sizes of coefficients of orders of ϵ in $t(\epsilon)$ we get an idea of the relative contributions of the terms to the approximation.

First, as $t(\epsilon)$ is real analytic in ϵ, $t_i \to 0$ as $i \to \infty$. Then letting q and v vary set

$$r_3(q,v) = t3(q,v)/t1(q,v) = (1/6)\left(tr\,(qvqv) - tr\left(q^2v^2\right)\right)$$

the ratio of $t_3(q,v)$ to $t_1(q,v)$ in (19). For any $v, q \in \mathrm{T_I}P(n)$ take the spectral decomposition $v = U\Lambda U^\top$ and letting $\hat{q} = UqU^\top$ we have $r_3(q,v) = r_3(\hat{q},\Lambda)$. Letting $\lambda_1,\ldots,\lambda_n$ be the eigenvalues of v we have,

$$r_3(\hat{q},\Lambda) = -\frac{1}{6}\left(\Sigma_{i>j}(\lambda_i-\lambda_j)^2\hat{q}_{i,j}^2\right).$$

Optimizing with respect to $\lambda_1,\ldots,\lambda_n$ and the entries of $\hat{q}$ with constraints $\Sigma_i\lambda_i^2 = tr\,(\hat{q}\hat{q}) = 1$ gives $-1/6 \le t_3/t_1 \le 0$.

Let v, q be random i.i.d. vectors with uniform distributions on $U = \{w \in \mathrm{T_I}P(n); tr\,(ww) = 1\}$, the unit sphere in $\mathrm{T_I}P(n)$. Then v and q can be any two tangent vectors in U with equal probability and typical or expected values of the ratios of the coefficients of $t(\epsilon)$ can be computed.

As in [6], we can sample such v and q by first sampling two independent, standard normal vectors in $\mathbb{R}^M$ where $M = n(n+1)/2 = \dim(\mathrm{T_I}P(n))$ and then by applying the isometry in (12) to both and normalizing the results. That is, let $\hat{q}$ and $\hat{v}$ be i.i.d. with $\hat{q},\hat{v} \sim N(\mathbf{0}, I_M)$ and let $\tilde{q} = \rho(\hat{q})$ and $\tilde{p} = \rho(\hat{p})$ where ρ is as in (12). Then set $q = \tilde{q}/\|\tilde{q}\|$ and $v = \tilde{v}/\|\tilde{v}\|$.

We have

$$\mathrm{E}\left(r_3(\tilde{q},\tilde{v})\right) = (1/6)tr\left(\mathrm{E}\left(\tilde{q}\tilde{v}\tilde{q}\tilde{v}\right) - \mathrm{E}\left(\tilde{q}^2\tilde{v}^2\right)\right).$$

Fixing $\tilde{v}$, expanding in entries of $\tilde{q}$ and $\tilde{v}$ and computing conditional expected values gives
$$\mathrm{E}\left(\tilde{q}\tilde{v}\tilde{q}\tilde{v}|\tilde{v}\right) = (\tilde{v}^2 + tr\left(\tilde{v}\right)\tilde{v})/2.$$
Also,
$$\mathrm{E}\left(\tilde{v}^2\tilde{q}^2|\tilde{v}\right) = \tilde{v}^2\mathrm{E}\left(\tilde{q}^2\right) = \tilde{v}^2(n+1)/2.$$
So we have
$$\mathrm{E}\left(r_3(\tilde{v},\tilde{q})|\tilde{v}\right) = (1/12)\left(tr\left(\tilde{v}\right)^2 - n\,tr\left(\tilde{v}^2\right)\right).$$
Then as $\tilde{v}$ is independent of $\tilde{q}$, the expected value of $r_3(\tilde{q},\tilde{v})$ is given by
$$\begin{aligned}\mathrm{E}\left(r_3(\tilde{v},\tilde{q})\right) &= \mathrm{E}\left(\mathrm{E}\left(r_3(\tilde{v},\tilde{q})|\tilde{v}\right)\right)\\ &= (1/12)\mathrm{E}\left(tr\left(\tilde{v}\right)^2\right) - n\,tr\left(\mathrm{E}\left(\tilde{v}^2\right)\right)\\ &= (1/12)(n - n(n+1)n/2)\\ &= -(n/24)(n+2)(n-1).\end{aligned}$$
As $\hat{q}$ and $\hat{v}$ are i.i.d and $\hat{q},\hat{v} \sim N(\mathbf{0}, I_M)$ we have $tr\left(\tilde{q}\tilde{q}\right) \sim \mathcal{X}^2(M)$ and independent of q and $tr\left(\tilde{v}\tilde{v}\right) \sim \mathcal{X}^2(M)$ and independent of v. Further, as
$$r_3(\tilde{q},\tilde{v}) = r_3(q,v)\,tr\left(\tilde{q}\tilde{q}\right)tr\left(\tilde{v}\tilde{v}\right)$$
we have
$$\mathrm{E}\left(r_3(q,v)\right) = \mathrm{E}\left(r_3(\tilde{q},\tilde{v})\right)/(\mathrm{E}\left(tr\left(\tilde{q}\tilde{q}\right)\right)\mathrm{E}\left(tr\left(\tilde{v}\tilde{v}\right)\right))$$
so that
$$\mathrm{E}\left(r_3(q,v)\right) = -\frac{(n/24)(n+2)(n-1)}{((n+1)(n/2))^2} = -\frac{1}{6}\frac{(n+2)(n-1)}{n(n+1)^2}. \tag{23}$$

For $n = 3$, $\mathrm{E}\left(r_3(q,v)\right) = \mathrm{E}\left(t3(q,v)/t1(q,v)\right) = -5/144$. Further, expanding in the entries of $\tilde{q}$ and $\tilde{v}$ and though direct computation in Maxima, for $n = 3$, $\mathrm{E}\left(t_5(q,v)/t1(1,v)\right) = 7/3072$.

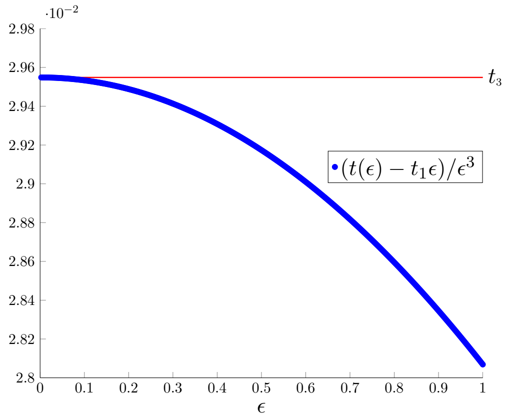
In table 1 below, these expected values are tested in Octave using the Central Limit Theorem.

Thus, in general, and specifically if v and q are uniformly distributed in $U = \{w \in \mathrm{T_I}P(n); tr\left(ww\right) = 1\}$, it is expected that the plot of $\log(t(\epsilon) - t_1\epsilon)$ against $\log(\epsilon)$ should be approximately a line with slope 3 and intercept of $\log(t_3)$.

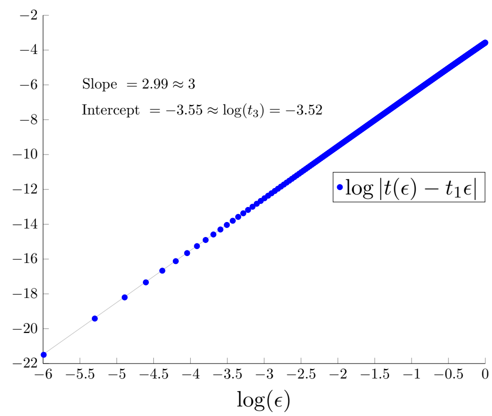
In figure 3 below, v and q are sampled from the uniform distribution on the unit sphere in $\mathrm{T_I}P(n)$ and plots are generated in Octave to test the expansions.

7.3. **First principal geodesic direction.** Let $q_i \in \mathrm{T_I}P(n)$ and $p_{i,\epsilon} = \mathrm{Exp_I}(\epsilon q_i)$ for $i = 1,\ldots,N$. Also, letting $D_\epsilon = \{p_{i,\epsilon}\}_i$ assume $\mu_{D_\epsilon} = \mathrm{I}$ for all ϵ, i.e., as in (4), assume $\sum q_i = 0$. Further, let $t_i(\epsilon, v)$ be the projection coefficient when projecting $p_{i,\epsilon}$ to $\mathrm{Exp_I}(\mathrm{span}(v))$ as in (17) for $i = 1,\ldots,N$.

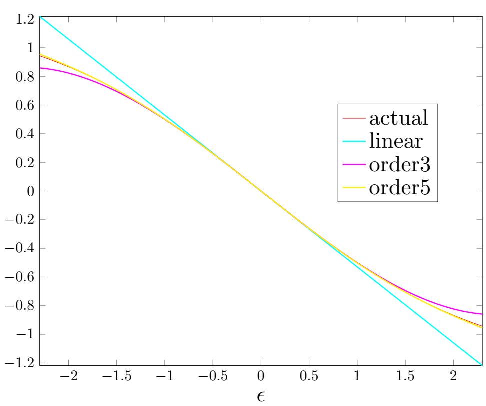
Set
$$f(v,\epsilon) = \frac{1}{N}\sum \mathrm{d}(\mathrm{Exp_I}(t_i(\epsilon,v)v), \mathrm{Exp_I}(\epsilon q_i))^2 \tag{24}$$



(A) "Solving" for t_3



(B) Log-log plot



(C) Approximations

FIGURE 3. Tests of expansion of $t(\epsilon)$

$n=3$

Samp. Size	Mean
10^2	-0.0310
10^3	-0.0333
10^4	-0.0348
10^5	-0.0347

$\mathrm{E}\left(\frac{t3(q,v)}{t1(q,v)}\right) \approx -0.0347$

$n=3$

Samp. Size	Mean
10^3	0.00236
10^4	0.00223
10^5	0.00229
10^6	0.00228

$\mathrm{E}\left(\frac{t5(q,v)}{t1(q,v)}\right) \approx 0.00228$

$n=4$

Samp. Size	Mean
10^2	-0.0317
10^3	-0.0298
10^4	-0.0301
10^5	-0.0300

$\mathrm{E}\left(\frac{t3(q,v)}{t1(q,v)}\right) = -0.0300$

TABLE 1. Tests of expected values of $t3/t1$ and $t5/t1$

for $v \in \mathrm{T_I}P(n)$, $\|v\| = 1$. Then using (19) in (24) and expanding in ϵ, $f(v,\epsilon) = f_2(v)\epsilon^2 + f_4(v)\epsilon^4 + O(\epsilon^6)$ where

$$f_2(v) = \frac{1}{N}\sum tr\,(q_i q_i) - tr\,(q_i v)^2 \text{ and}$$
$$f_4(v) = \frac{1}{N}\sum (1/6) tr\,(q_i v)^2\,(tr\left(q_i^2 v^2\right) - tr\,(q_i v q_i v)).$$

As in (7), the first principal geodesic direction is given by

$$v_1(\epsilon) = \operatorname*{argmin}_{\|v\|=1} f(v,\epsilon) \tag{25}$$

The expansion in ϵ of the gradient of $f(v,\epsilon)$ can be computed and using a Lagrange multiplier the expansion of $v_1(\epsilon)$ can be obtained. Specifically, we have,

$$\nabla_v f_2(v) = \frac{1}{N}\sum -2\,tr\,(q_i v)\,q_i$$
$$\nabla_v f_4(v) = \frac{1}{N}\sum (1/6)[2\,tr\,(q_i v)\,q_i(tr\left(q_i^2 v^2\right) - tr\,(q_i v q_i v)) +$$
$$tr\,(q_i v)^2\,(q_i^2 v + v q_i^2 - 2 q_i v q_i)]$$

so that

$$\nabla_v f(v,\epsilon) = \frac{1}{N}\sum -2\, tr\,(q_i v)\, q_i \epsilon^2 + (1/6)[2\, tr\,(q_i v)\, q_i (tr\left(q_i^2 w_2\right) - tr\,(q_i v q_i v)) + tr\,(q_i v)^2\,(q_i^2 v + v q_i^2 - 2 q_i v q_i)]\epsilon^4 + O(\epsilon^6).$$

Let λ be the multiplier and set

$$\nabla_v f(v,\epsilon) = \lambda \nabla_v tr\,(vv) = \lambda 2v \tag{26}$$

with constraint $tr\,(vv) = 1$. Then letting

$$v = v(\epsilon) = w_0 + w_2\epsilon^2 + 0(\epsilon^4) \tag{27}$$

we have

$$\begin{aligned} \nabla_v f(v(\epsilon),\epsilon) = \frac{1}{N}\sum & -2tr\,(q_i w_0)\, q_i \epsilon^2 + \\ & [(1/6)\big(2tr\,(q_i w_0)\, q_i (tr\left(q_i^2 w_0^2\right) - tr\,(q_i w_0 q_i w_0)) \\ & + tr\,(q_i w_0)^2\,(q_i^2 w_0 + w_0 q_i^2 - 2 q_i w_0 q_i)\big) \\ & - 2tr\,(q_i w_2)\, q_i]\epsilon^4 + O(\epsilon^6). \end{aligned} \tag{28}$$

Setting

$$\lambda = \lambda_{-2} - \lambda_0\epsilon^2 - \lambda_2\epsilon^4 + 0(\epsilon^6). \tag{29}$$

and using the expansions in (27), (28) and (29) in (26) and equating coefficients of orders of epsilon gives

$$\lambda_{-2} w_0 = 0 \ (\implies \lambda_{-2} = 0) \tag{30}$$

$$\lambda_0 w_0 = \frac{1}{N}\sum tr\,(q_i w_0)\, q_i \tag{31}$$

$$\begin{aligned} \lambda_0 w_2 + \lambda_2 w_0 = \frac{1}{N}\sum & -(1/12)[2tr\,(q_i w_0)\, q_i (tr\left(q_i^2 w_0^2\right) \\ & - tr\,(q_i w_0 q_i w_0)) + tr\,(q_i w_0)^2\,(q_i^2 w_0 + w_0 q_i^2 - 2 q_i w_0 q_i)] \\ & + tr\,(q_i w_2)\, q_i. \end{aligned} \tag{32}$$

The constraint in (26) then gives $tr\,(v(\epsilon)v(\epsilon)) = 1$

$$\begin{aligned} &\implies tr\,(w_0 w_0) + 2tr\,(w_0 w_2)\,\epsilon^2 + O(\epsilon^4) = 1 \\ &\implies tr\,(w_0 w_0) = 1 \text{ and } tr\,(w_0 w_2) = 0. \end{aligned}$$

Then from (31) and $tr\,(w_0 w_0) = 1$ it follows that $\lambda_0 = \frac{1}{N}\sum tr\,(q_i w_0)^2$.

Thus, if L is the covariance operator formed by $\{q_i\}_i$ as in (10), then w_0 is a normalized eigenvector of L with eigenvalue $\lambda_0 = \sum tr\,(q_i w_0)^2$.

Assuming $f(v(\epsilon),\epsilon) \leq f(w,\epsilon)$ $\forall \epsilon$ and $\forall w$ s.t. $tr\,(ww) = 1$

$$\begin{aligned}
&\Longrightarrow \sum (tr\,(q_i q_i) - tr\,(q_i w_0)^2)\epsilon^2 + 0(\epsilon^4) \\
&\qquad \leq \sum (tr\,(q_i q_i) - tr\,(q_i w)^2)\epsilon^2 \\
&\qquad\quad + 0(\epsilon^4)\ \forall \epsilon \text{ and } \forall w \text{ s.t. } tr\,(ww) = 1 \qquad (33)\\
&\Longrightarrow \frac{1}{N}\sum tr\,(q_i w_0)^2 \geq \frac{1}{N}\sum tr\,(q_i w)^2\ \ \forall w \text{ s.t. } tr\,(ww) = 1 \\
&\Longrightarrow w_o \text{ is the dominant, normalized eigenvector of } L.
\end{aligned}$$

Let $\{w_0, u_2, \ldots, u_M\}$ be the normalized eigenvectors of L with respective eigenvalues $\{\lambda_0, \beta_2, \ldots, \beta_M\}$ in descending order. Also, consider $w_2 = \sum_{i=2}^{M} c_i u_i$, the respective orthonormal expansion of w_2. Then let

$$\alpha = \frac{1}{N}\sum (1/12)[2tr\,(q_i w_0)\, tr\left(q_i^2 w_0^2 - q_i w_0 q_i w_0\right) q_i + tr\,(q_i w_0)^2\left(q_i^2 w_0 + w_0 q_i^2 - 2 q_i w_0 q_i\right)]$$

Using (20) and (21) we have

$$\alpha = -\frac{1}{N}\sum (1/12)(\cos\theta_i \sin^2\theta_i K(q_i, w_0)\tilde{q}_i + \cos^2\theta_i \mathrm{R}_{w_0,\tilde{q}_i}\,\tilde{q}_i)\,\|q_i\|^4$$

where θ_i is the angle formed by q_i and w_0 and $\tilde{q}_i = q_i / \|q_i\|$. From (32) we have

$$\begin{aligned}
&(L - \lambda_0) w_2 = \alpha + \lambda_2 w_0 \\
\Longrightarrow\ &(L - \lambda_0)\sum_{i=2}^{M} c_i u_i = \alpha + \lambda_2 w_0 \\
\Longrightarrow\ &c_j = tr\,(\alpha u_j)/(\beta_j - \lambda_0) \text{ for } j = 2, \ldots, M.
\end{aligned}$$

Thus, we have

$$w_2 = \sum_{i=2}^{M}\big(tr\,(\alpha u_i)/(\beta_i - \lambda_0)\big) u_i \qquad (34)$$

a weighted projection of α to the orthogonal complement of the dominant eigenvector w_0. Finally, we have the expansion of the first principal geodesic direction in ϵ of D_ϵ as

$$v_1(\epsilon) = w_0 + w_2\epsilon^2 + O(\epsilon^4). \qquad (35)$$

with w_0 and w_2 as in (33) and (34), respectively.

As eigenvalues of L, $\lambda_0, \beta_2, \ldots,$ and β_M give the relative proportions of variability accounted for by $w_0, u_2, \ldots,$ and u_M, respectively, of the total variability of D_ϵ projected to $\mathrm{T_I}P(n)$. This is reflected in the differences in the denominators in (34) which shows w_0 becoming a better approximation the greater the share of variability in the tangent space it accounts for. The computation of α in (34) and ϵ reflect the role of curvature and dispersion of data in the suitability of the linear approximation of the first PGA direction.

Also, evaluating (24) at our expansion of v_1 in ϵ gives

$$\begin{aligned}f(v_1(\epsilon),\epsilon) &= \sum_{i=1}^{N}(tr\,(q_iq_i) - tr\,(q_iw_0)^2)\epsilon^2 + [(tr\,(q_iw_0)^2\,/6)tr\,(q_iq_iw_0w_0 - q_iw_0q_iw_0) - \\ &\quad 2tr\,(q_iw_0)\,tr\,(q_iw_2)]\epsilon^4 + 0(\epsilon^6) \\ &= \sum_{i=1}^{N}\sin^2\theta_i \left\|q_i\right\|^2\epsilon^2 - [(1/12)\cos^2\theta_i\sin^2\theta_i K(q_i,w_0)\left\|q_i\right\|^2 \\ &\quad + 2\cos\theta_i\cos\hat{\theta}_i]\left\|q_i\right\|^2\epsilon^4 + O(\epsilon^6)\end{aligned}$$

where θ_i, $\hat{\theta}_i$ are the angles formed by q_i and w_0 and by q_i and w_2, respectively. The leading term is the solution in the tangent space. The next order term indicates how curvature and variability accounted for by w_0 in the tangent space contribute to the sum of squared distances of the data to $\mathrm{Exp}_{\mathrm{I}}(\mathrm{span}(v_1))$.

In figure 4 below, the expansion in (35) is tested. 100 matrices are sampled from $\mathrm{T_I}P(3)$ with entries having independent, normal distributions with mean 0. The variances vary by entry so that first principal geodesic directions can be identified. The sample is then demeaned to form $\{q_i\}_i \subset \mathrm{T_I}P(n)$ and the data, $D_\epsilon = \{p_{i,\epsilon}\}_i = \{\mathrm{Exp}_{\mathrm{I}}(\epsilon q_i)\}_i$ is as above.

7.4. **Mean after removing geodesic direction.** Let $v \in \mathrm{T_I}P(n)$ be s.t. $\|v\| = 1$. Also, let $\{q_i\}_i \subset \mathrm{T_I}P(n), D_\epsilon = \{p_{i,\epsilon}\}_i$ and $\{t_i(\epsilon,v)\}_i$ be as in section 7.3. Then let

$$x(\epsilon) = \mathrm{Log}_{\mathrm{I}}\big(\mu_{\Phi_v(D_\epsilon)}\big)$$

that is, the log at I of the mean of the data after the variability in the direction of v is removed.

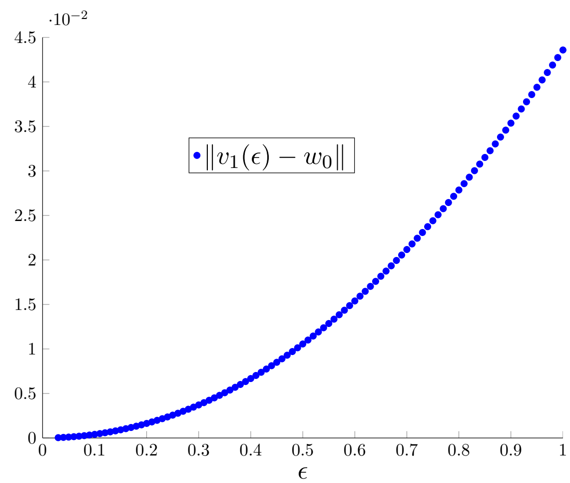
Using the gradient condition in (4), $x(\epsilon)$ is s.t.

$$\sum \mathrm{Log}_{\mathrm{I}}(\mathrm{Exp}_{\mathrm{I}}(-x(\epsilon)/2)\mathrm{Exp}_{\mathrm{I}}(\epsilon r_i)\mathrm{Exp}_{\mathrm{I}}(-x(\epsilon)/2)) = 0 \tag{36}$$

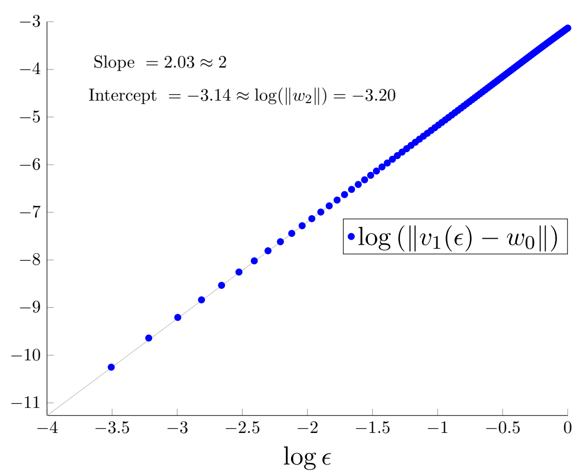
with $\mathrm{Exp}_{\mathrm{I}}(\epsilon r_i) = \mathrm{Exp}_{\mathrm{I}}(-t_i(\epsilon,v)v/2)\mathrm{Exp}_{\mathrm{I}}(\epsilon q_i)\mathrm{Exp}_{\mathrm{I}}(-t_i(\epsilon,v)v/2)$ and $r_i \in \mathrm{T_I}P(n)$ for $i = 1,\ldots,N$.

Letting $x(\epsilon) = x_0 + x_1\epsilon + x_2\epsilon^2 + x_3\epsilon^3 + x_4\epsilon^4 + O(\epsilon^5)$, substituting into (36) and solving for x_0, x_1, x_2, x_3 and x_4 gives the expansion $x(\epsilon) = x_3\epsilon^3 + O(\epsilon^5)$ where

$$\begin{aligned}x_3 &= \sum[(tr\,(q_iv)^2\,/24)(2vq_iv - q_ivv - vvq_i) + \\ &\quad (tr\,(q_iv)\,/12)(2q_ivq_i - q_iq_iv - vq_iq_i) \\ &\quad + (tr\,(q_iv)\,/6)\big(tr\,(q_iq_ivv) - tr\,(q_ivq_iv)\big)v] \\ &= \sum[(1/24)\cos^2\theta_i \mathrm{R}_{\tilde{q}_i,v}\,v + (1/12)\cos\theta_i \mathrm{R}_{v,\tilde{q}_i}\,\tilde{q}_i \\ &\quad - (1/12)\cos\theta_i\sin^2\theta_i K(q_i,v)v]\left\|q_i\right\|^3\end{aligned} \tag{37}$$



(A) "Solving" for w_0



(B) Log-log plot

FIGURE 4. Tests of expansion of $v_1(\epsilon)$

where the second equality follows from (20) and (21) where θ_i is the angle formed by q_i and v and $\tilde{q}_i = q_i / \|q_i\|$. Further,

$$\begin{aligned} \|x(\epsilon)\| &= \sqrt{tr\left((x_3\epsilon^3 + O(\epsilon^5))(x_3\epsilon^3 + O(\epsilon^5))\right)} \\ &= tr\left(x_3x_3\right)^2 \epsilon^3 \sqrt{1 + O(\epsilon^2)} \\ &= tr\left(x_3x_3\right)^2 \epsilon^3 (1 + O(\epsilon^2)) \\ &= O(\epsilon^3). \end{aligned} \tag{38}$$

We have $tr\,(vx_3) = 0$ and with (38),

$$tr\left(\frac{x(\epsilon)}{\|x(\epsilon)\|}v\right) = tr\left(\frac{(x_3 + O(\epsilon^5))}{\|x(\epsilon)\|}v\right) = \frac{O(\epsilon^2)}{\sqrt{1 + O(\epsilon^2)}} = O(\epsilon^2). \tag{39}$$

In (37), the leading term in the movement of the mean from the identity as a result of curvature in $P(n)$ is seen. In (38), the size of the movement in $\mathrm{T_I}P(n)$ is $O(\epsilon^2)$. In (39), the orthogonality of the movement with respect to v, the direction in which the variability was removed, is $O(\epsilon^2)$.

In RPGA, this movement, after removing variability in a principal geodesic direction, is accounted for by "shifting" the data by applying the isometry $\phi_{g^{-1}}$ where $g = \mathrm{Exp_I}(x(\epsilon)/2)$.

In figure 5 below, random $v \in \mathrm{T_I}P(n)$ and $D_\epsilon = \{p_{i,\epsilon}\}_i$ with distribution as in the tests in section 7.3 are generated and (37), (38), (39) are tested.

8. Simulation and comparison

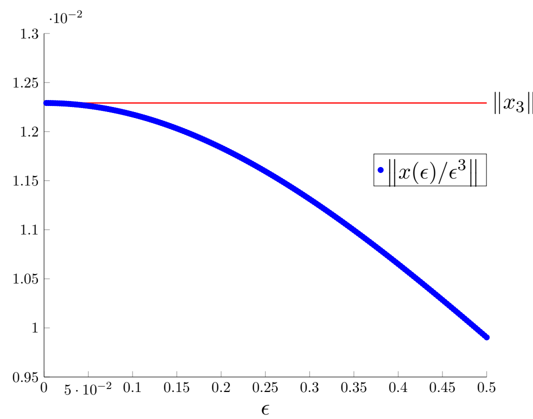
In the simulations that follow data $D = \{p_1, \ldots, p_{30}\} \subset P(3)$ is generated. Symmetric matrices $\{q_1, \ldots, q_{30}\}$ are sampled in $\mathrm{T_I}P(3)$, with entries having standard normal distributions but varying standard deviations, the largest having a standard deviation of 2 and the smallest having a standard deviation of $(0.25)^3$. This sample is then demeaned in $\mathrm{T_I}P(3)$ by subtracting the average of respective entries from each entry and then exponentiated to form D.

In each of 50 runs data is scaled in $\mathrm{T_I}P(n)$ by ϵ and first through fourth explanatory directions are computed using *PGA*, *LPGA*, and *RPGA*. The means and standard deviations of both the sum of squared residuals (SSR) and computation times (CT) for the 50 runs are computed for each method and reported in table 2 below. Standard deviations are given in parenthesis next to each mean. Also, as eigenvectors are computed all together in Octave, in LPGA only one computation time is reported.

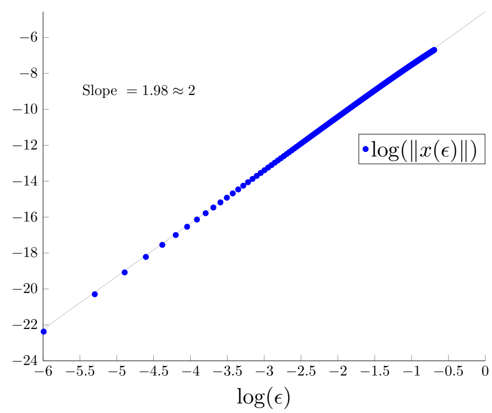
In locating explanatory directions, projection of data to submanifolds was computed in parallel with four-core processors. Intrinsic means and removal of variation in explanatory directions were computed in parallel as well.

While PGA computes explanatory directions in subspaces of $\mathrm{T_I}P(n)$ of decreasing dimension, as RPGA only projects to geodesics, in RPGA the projection operator always only minimizes a function of a single variable. Also, RPGA locates explanatory directions on data of increasingly less variability from the mean. All together, in locating the first three explanatory directions, in these simulations, RPGA has better computation times than PGA while PGA has better computations in locating the first four.

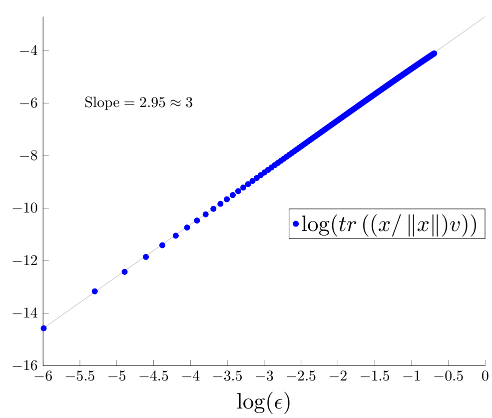
RPGA has comparable explanatory power of the located geodesic submanifolds in terms of sum of squared residuals to PGA. Across the table, the linear method has the highest mean SSR with the SSR of all four methods converging as the data is scaled closer to the tangent space at the mean by ϵ.



(A) "Solving" for x_3



(B) Log-log plot for $\|x(\epsilon)\|$



(C) Log-log plot for $tr\,((x/\,\|x\|)v)$

FIGURE 5. Tests of expansion of $x(\epsilon)$

Method	Dir.	1	2	3	4
PGA	SSR	1.0123 (.07)	0.5619 (0.05)	0.3116 (0.04)	0.0796 (0.01)
	CT	38.239 (11.5)	67.032 (13.2)	84.878 (16.7)	94.972 (17.4)
RPGA	SSR	1.0123 (0.07)	0.5621 (0.05)	0.3135 (0.04)	0.0799 (0.01)
	CT	38.245 (11.5)	51.705 (10.6)	79.596 (14.6)	104.97 (23.1)
LPGA	SSR	1.0239 (0.08)	0.5670 (0.05)	0.3202 (0.04)	0.0814 (0.003)
	CT	0.350 (0.20)	0.350 (0.20)	0.350 (0.20)	0.350 (0.20)

$\epsilon = 1$

Method	Dir.	1	2	3	4
PGA	SSR	0.7506 (0.27)	0.40726(0.16)	0.2208 (0.10)	0.0556 (0.005)
	CT	41.771 (12.0)	70.250 (13.6)	86.501 (16.8)	94.996 (17.4)
RPGA	SSR	0.7506 (0.27)	0.4074 (0.16)	0.2208 (0.10)	0.0557 (0.03)
	CT	43.363 (12.6)	60.053 (11.7)	80.961 (15.9)	126.19 (23.2)
LPGA	SSR	0.7530 (0.27)	0.4080 (0.17)	0.2221 (0.03)	0.0559 (0.005)
	CT	0.386 (0.20)	0.386 (0.20)	0.386 (0.20)	0.386 (0.20)

$\epsilon = .75$

Method	Dir.	1	2	3	4
PGA	SSR	0.4967 (0.52)	0.2652 (0.30)	0.1403 (0.17)	0.0353 (0.05)
	CT	43.024 (12.5)	68.592 (13.3)	80.720 (17.2)	86.097 (19.5)
RPGA	SSR	0.4967 (0.52)	.2652 (0.30)	0.1402 (0.18)	.0353 (0.05)
	CT	42.854 (12.4)	61.838 (11.6)	79.082 (16.8)	106.55 (24.5)
LPGA	SSR	0.4969 (0.52)	0.2655 (0.30)	0.1406 (0.18)	0.0354 (0.05)
	CT	0.0309 (0.20)	0.309 (0.20)	0.309 (0.20)	0.309 (0.20)

$\epsilon = .5$

Method	Dir.	1	2	3	4
PGA	SSR	0.2473 (0.77)	0.1308 (0.43)	0.0681 (0.25)	0.0172 (0.06)
	CT	35.676 (11.8)	48.356 (22.9)	57.545 (32.0)	61.622 (37.6)
RPGA	SSR	0.2473 (0.77)	0.1308 (0.43)	0.0681 (84.3)	0.0172 (0.013)
	CT	35.981 (11.7)	42.038 (19.7)	53.230 (30.8)	64.185 (44.3)
LPGA	SSR	0.2474 (0.77)	0.1308 (0.44)	0.0682 (0.25)	0.0172 (0.06)
	CT	0.270 (0.21)	0.270 (0.21)	0.270 (0.21)	0.270 (0.21)

$\epsilon = .25$

SSR = sum of squared residuals
CT = computation time

TABLE 2. Comparisons of PGA methods in $P(3)$

9. CONCLUSION AND OUTLOOK

In this paper, generalizing a method of principal component analysis, a new, recursive method of dimension reduction in the symmetric space of positive definite matrices was formulated. Computational methods for parts of PGA in $P(n)$ were specified and a formulation of PGA directions in terms of spherical coordinates was introduced.

Through the introduction of a scaling parameter the relationship of parts of RPGA to the scale of data and curvature was examined. These are also parts of PGA and the expansions obtained can help give an indication of how well first order-solutions approximate PGA.

These expansions were only obtained for projections to geodesics and first PGA directions. Further work can obtain such expansions for projection to submanifolds geodesic at the mean of any dimension and for i-th PGA directions. Also, additional interpretations of higher order terms involving curvature terms might be ascertained.

Given knowledge of the distribution of a type of data in $P(n)$, expected values and distributions of the these higher order terms might obtained, leading to conclusions about the appropriate method of dimension reduction to employ. In the basic case of projection to a geodesic, and with uniform distributions in the tangent space, in section 7.2 such expected values were derived. These results might be extended accordingly.

Simulation showed the usefulness of RPGA as an alternative method in finding explanatory directions and corresponding submanifolds of comparable explanatory power to PGA. Further, limiting projection to geodesics as in RPGA might increase computational stability when there is substantial noise and contamination in data and this can be an area of additional investigation.

References

[1] P. Thomas Fletcher and Sarang Joshi, *Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors*, Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis (Milan Sonka, IoannisA. Kakadiaris, and Jan Kybic, eds.), Lecture Notes in Computer Science, vol. 3117, Springer Berlin Heidelberg, 2004, pp. 87–98 (English).

[2] ———, *Riemannian geometry for the statistical analysis of diffusion tensor data*, Signal Processing **87** (2007), no. 2, 250 – 262, Tensor Signal Processing.

[3] P. Thomas Fletcher, Conglin Lu, and Sarang Joshi, *Statistics of shape via principal geodesic analysis on lie groups*, Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Washington, DC, USA), CVPR'03, IEEE Computer Society, 2003, pp. 95–101.

[4] Stephan Huckemann and Herbert Ziezold, *Principal component analysis for Riemannian manifolds, with an application to triangular shape spaces*, Adv. in Appl. Probab. **38** (2006), no. 2, 299–319. MR 2264946 (2007k:62102)

[5] Herman Karcher, *Riemannian center of mass and mollifier smoothing*, Comm. Pure Appl. Math. **30** (1977), no. 5, 509–541. MR 0442975 (56 #1350)

[6] Donald E. Knuth, *The art of computer programming: Seminumerical algorithms*, Addison-Wesley series in computer science and information processing, p. 242, Addison-Wesley, 1981.

[7] Igor Najfeld and Timothy F. Havel, *Derivatives of the matrix exponential and their computation*, Adv. in Appl. Math. **16** (1995), no. 3, 321–375. MR 1342832 (96f:65013)

[8] Peter Petersen, *Riemannian geometry*, Graduate Texts in Mathematics, p. 242, Springer, 2010.

[9] Stefan Sommer, François Lauze, Søren Hauberg, and Mads Nielsen, *Manifold valued statistics, exact principal, geodesic analysis and the effect of linear, approximations*, Proceedings of the 11th European Conference on Computer Vision: Part VI (Berlin, Heidelberg), ECCV'10, Springer-Verlag, 2010, pp. 43–56.

[10] ———, *Optimization over geodesics for exact principal geodesic analysis*, Adv. Comput. Math. **40** (2014), no. 2, 283–313. MR 3194707