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

Manuscript Number: JMVA-14-443R1

Title: Scale and Curvature Effects in Principal Geodesic Analysis

Article Type: Regular Article

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

Corresponding Author: Mr. Drew Lazar,

Corresponding Author's Institution: Central Michigan University

First Author: Drew Lazar

Order of Authors: Drew Lazar

Abstract: There is growing interest in using the close connection between differential geometry and statistics to model smooth, manifold valued data. In particular, much work has been done recently to generalize principal component analysis (PCA), the method of dimension reduction in linear spaces, to Riemannian manifolds. One such generalization is known as principal geodesic analysis (PGA). This paper, in a novel fashion, obtains Taylor expansions in scaling parameters introduced in the domain of objective functions in PGA. It is shown this technique not only leads to better closed-form approximations of PGA but also reveals the effects that scale and curvature have on solutions to PGA and on their differences to first-order tangent space approximations. This approach should be able to be applied not only to PGA but also to other generalizations of PCA and more generally to other intrinsic statistics on Riemannian manifolds.

RESPONSE TO REVIEWS
SCALE AND CURVATURE EFFECTS IN PRINCIPAL GEODESIC ANALYSIS

DREW M. LAZAR

1. General Responses to both reviews

I rewrote the manuscript, expanding the part of the initial manuscript that both reviewers thought was interesting and novel, that is, "the analysis of curvature effects with increasing ϵ radius of the data" to the whole paper. I couldn't fully address or resolve the reviewers concerns about the RPGA construction so that was essentially left out of the rewrite.

As a bit of background, I became interested in this subject after a professor gave a talk at my university which involved applications of manifold valued statistics to biostatistics. I read the paper [2] and thought there was no accounting for the displacement of the mean after applying left multiplication. I also thought there might be a better choice of isometry for variance removal after I read [3] to learn how to do computations in $SO(3)$. That is how I came to the formulation of RPGA.

Introducing the scaling parameter and obtaining Taylor expansions of solutions in this parameter was a way I tried to justify and understand RPGA and PGA in general. As reflected in the reviewer's comments I think this idea is very useful and I believe I demonstrated that quite well in my revised manuscript. I expanded the result in section 7.3, page 15 of my initial manuscript to a proposition that can be used to obtain the expansion of all PGA directions in symmetric spaces immersed in $\mathbb{R}^n$ (section 2.1, pages 4-6). I obtained such expansions in three important symmetric spaces and tested and demonstrated their uses in each space.

I also used the expansion argument to take a look at the formulation of PGA in [2] and to examine and improve on the *linear difference indicators* of [4]. I briefly addressed the possibility of recentering data (line 7 from the bottom, page 20) in my revision as in RPGA.

I believe that researchers using manifold valued statistics will be convinced that this is a important way to justify and understand their methods.

Note: I don't address all of the reviewers points as my revision is much different (and improved, I believe) from the reviewed manuscript. RPGA is not included, so I obviously don't address all the points that concern it.

Also, as of August 14, 2015 I haven't added all the code I intend to at: https://github.com/DMLazar/PGAScale. I need to neaten the code up and add comments, etc. I will be adding more code shortly but the wrist rotation data and the simulated data from the Von Mises-Fisher distribution is already there.

2. Responses to review #1

(1) I didn't include RPGA in my revised manuscript. In my revision when I introduced each symmetric space, however, I mention how data analysis is useful in the space in question (for example, the first paragraph on page 7).
(2) I consider three important symmetric spaces here (I didn't just limit it to one) and the results in my revision can be applied to other symmetric spaces in the same manner.
(3) That is a disadvantage and I couldn't find a data set or justification to overcome it. I think the effect on the intrinsic mean when encoding data in this fashion is important to understand and quantify, however. I address this with expansions ((44) and (45), page 20)

and suggest a different isometry (in only $SO(n)$ but a similar analysis can be applied to $P(n)$) to minimize that effect. I also quantify the rotational effect of removing variability (the equation 11 lines from the bottom, page 19).

(4) N/A (as there is no RPGA this time)

(5) I used simulated data for testing but I also included a real rotational data set and simulated data from the Von Mises-Fisher Distribution (section 3.1, pages 9-11) which is a very important distribution in directional statistics and has paramaters that relate directly to my results. I also used higher dimensional unit spheres S^{10} (second to last paragraph, page 9) and S^{15} (section 3.1, pages 9-11) in testing and simulation. I used $P(3)$ because that is useful in applications.

(6) Both of the papers you mention, [3] and particulary [5] were very useful to me in writing code for computations and I cited them both. I didn't include details of how I did computations this time as it is somewhat redundant (it is covered in [5]) and the revised manuscript is already 29 pages. Also, in my conclusions on (the first paragraph in my conclusions, page 25), I included a reference and brief discussion of [6] as you suggested and related it to the work in my revised manuscript.

3. Responses to review #2

(1) I didn't include RPGA this time. However, perhaps the motivation for RPGA is clearer this time as given in section 5.3, page 19. I believe the issues addressed in that section are important.

(2) I included three types of data this time and used simulated and real data.

(3) "The exploration of the link between curvature and linearized/non-linearized operations" makes up the whole of my revision.

(4) I didn't use the phrase "non-linear constraints" this time but instead used "smooth manifold valued data". I think it will be clear to a reader what I am referring to.

(5) I didn't compare methods of dimension reduction in my revision to each other or LPGA so this isn't an issue. I did use the normal distribution in the tangent space for testing, however, as I've seen it used for this purpose (for example in [1]) and it makes it simple to clearly separate directions of variability in the tangent space. In the case when I used simulated data distributed this way for comparison (along with a real data set) I used your suggestion of taking angles with the eigenvectors (Section 5.3, table 2, page 21).

Thank you reviewers for your consideration, careful comments and suggestions.

Drew Lazar, PhD.

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] 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.

[3] Salem Said, Nicolas Courty, Nicolas Le Bihan, and Stephen J. Sangwine, *Sangwine s.: Exact principal geodesic analysis for data on so(3*, InProc.ofthe 15th European Signal Processing Conference (EUSIPCO 2007, 2007.

[4] 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.

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

[6] Miaomiao Zhang and P. Thomas Fletcher, *Probabilistic principal geodesic analysis.*, NIPS (Christopher J. C. Burges, Lon Bottou, Zoubin Ghahramani, and Kilian Q. Weinberger, eds.), 2013, pp. 1178–1186.

Highlights (for review)

- Expansions of PGA directions in scale in symmetric spaces immersed in R^n .
- Expansions reveal effects of curvature on first-order approximations to PGA.
- Applied to important symmetric spaces such as the n-spheres and rotation group.
- Improved initial formulation of PGA in Lie groups in SO(n) using expansions.
- Improved measures of accuracy of first-order approximations before computation.

Scale and Curvature Effects in Principal Geodesic Analysis

Drew M. Lazar

Abstract

There is growing interest in using the close connection between differential geometry and statistics to model smooth, manifold valued data. In particular, much work has been done recently to generalize principal component analysis (PCA), the method of dimension reduction in linear spaces, to Riemannian manifolds. One such generalization is known as principal geodesic analysis (PGA). This paper, in a novel fashion, obtains Taylor expansions in scaling parameters introduced in the domain of objective functions in PGA. It is shown this technique not only leads to better closed-form approximations of PGA but also reveals the effects that scale and curvature have on solutions to PGA and on their differences to first-order tangent space approximations. This approach should be able to be applied not only to PGA but also to other generalizations of PCA and more generally to other intrinsic statistics on Riemannian manifolds.

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

1. Introduction

Principal component analysis (PCA) is an important statistical method for dimension reduction and exploration of the variance structure of data in a linear space. PCA has been generalized to data in smooth manifolds in various principal geodesic procedures in which projection is done to geodesic subspaces which are contained entirely within the manifolds and which serve as non-linear analogues of the linear subspaces of PCA.

Principal geodesic analysis (PGA), as introduced in [10], successively identifies orthogonal explanatory directions in the tangent space at the intrinsic mean of data and then exponentiates the span of the results to form explanatory submanifolds. In [10] first-order tangent space approximations of PGA were formulated. Subsequently methods for exact computation of PGA in specific manifolds were offered as in [26] and [17]. Then in [28], using the derivative of the exponential map and ODEs if necessary in gradient descent algorithms, procedures to find exact solutions in a general class of manifolds were outlined.

As pointed out in [27], however, exact computation of PGA can be computationally complex and time-intensive and thus there is interest in determining the accuracy and effectiveness of first-order approximations to PGA. This will depend on the distribution of data and its dispersion from the tangent space, the curvature and shape of the manifold in question and the interaction of these factors.

In this paper we quantify these effects by introducing scaling parameters on projections of data to the tangent space and by obtaining Taylor expansions of solutions to PGA procedures in these parameters. Leading terms will be corresponding solutions in the tangent space and next-order terms will demonstrate how local curvature and scale interact to contribute to differences between first-order approximations and exact solutions. This will not only allow for more accurate closed-form approximations to PGA but also contribute to a better understanding of the parts of PGA and corresponding statistics. In this paper data in three types of symmetric spaces which have regular application are considered. Also using [26], [17] and [28] we can compute exact solutions in these spaces which allows for comparison and testing.

1.1. Outline

Section 2 includes notations and definitions. In section 2.1 a proposition which will allow the expansion of PGA directions in this paper is stated and proved. In section 3 we review the geometry of the unit n-spheres and obtain and test expansions using our proposition. We also carry out experiments on data sampled from the Von Mises-Fisher distribution to show improved approximations. In section 3 we review the geometry of the space of positive definite matrices and obtain expansions using our proposition and computer algebra. In section 5 we review the geometry of the special orthogonal group and obtain expansions of PGA in this space. Also, in section 5.3 we take a closer look at PGA in Lie groups in [11] to show how expansions can give insight into the formulation of such intrinsic manifold statistics. In section 6, using expansions, we obtain improvements of the *linear difference indicators* introduced in [27]. In section 7 we discuss the results and consider their applications in similar contexts.

2. Notations and Definitions

Let M be a Riemannian manifold with metric tensor $p \to \langle \quad , \quad \rangle_p$ for $p \in M$. Given $p \in M, \mathrm{T}_pM$ is the *tangent space at p*. The unit sphere in T_pM is then $\mathrm{S}_pM = \{X \in \mathrm{T}_pM; \langle X, X\rangle_p = 1\}$. The *Riemannian exponential* and *Riemannian log maps*, are denoted by $\mathrm{Exp}_p : \mathrm{T}_pM \to M$ and $\mathrm{Log}_p : M \to \mathrm{T}_pM$, respectively. Given smooth manifolds M_1 and M_2, $p \in M_1$ and smooth mapping $\lambda : M_1 \to M_2$ we denote the *differential of* λ *at* p by $d_p\lambda$. Then given smooth function $f : M \to \mathbb{R}$ and $p \in M$ *the gradient of* f *at* p is denoted $\nabla_p f$ so that $\langle \nabla_p f, X\rangle_p = d_pf(X)\ \forall X \in \mathrm{T}_pM$. Differential geometry texts [6] and [24], provide a background for and definitions of these concepts.

All the manifolds we will deal with in the paper will be of the class defined below.

Definition 1 (Symmetric Space)**.** *Let M be a connected Riemannian manifold. M is a symmetric space if and only if for every $p \in M$ there exists an isometry $\phi_p : M \circlearrowleft$, such that*

$$\phi_p(p) = p \textit{ and } d_p\phi_p(X) = -X \quad \textit{for all } X \in \mathrm{T}_pM.$$

The following definition generalizes orthogonal projection in an inner product space to projection to submanifolds.

Definition 2 (Projection Operator)**.** *Let $p, \mu \in M$, $V_k = \{v_1, \ldots, v_k\} \subset \mathrm{T}_\mu M$ and $H(V_k) = \mathrm{Exp}_\mu(\mathrm{span}(V_k))$.*

$$\pi_{H(V_k)}(p) = \underset{x \in H(V_k)}{\mathit{argmin}}\, \mathrm{d}(x, p) \tag{1}$$

is the projection of p to $H(V_k)$.

We will assume throughout that for each V_k there is existence and uniqueness of projection which as in [16] will occur almost everywhere.

The first intrinsic statistic we formulate is a generalization of the arithmetic mean in an inner product space.

Definition 3 (Intrinsic Mean)**.** *Let $D = \{p_1, \ldots, p_N\} \subset M$.*

$$\mu(D) = \underset{x \in M}{\mathit{argmin}}\, \frac{1}{N} \sum_{i=1}^{N} \mathrm{d}(x, p_i)^2 \tag{2}$$

is the intrinsic mean of D.

The intrinsic mean is used as an offset in the following definition.

Definition 4 (Intrinsic Variance)**.** *Given data $D = \{p_1, \ldots, p_N\} \subset M$ with intrinsic mean $\mu(D)$*

$$\sigma^2(D) = \frac{1}{N} \sum_{i=1}^{N} \mathrm{d}(\mu(D), p_i)^2 \tag{3}$$

is the intrinsic variance of D.

Principal geodesic analysis (PGA), as introduced in [10], is generalization of principal component analysis (PCA) and is defined below.

Definition 5 (PGA directions)**.** *Let $D = \{p_1, \ldots, p_N\} \subset M$ and $K = \mathit{dim}(\mathrm{T}_{\mu(D)}M)$. PGA locates $\{v_1, \ldots, v_K\} \subset \mathrm{T}_{\mu(D)}M$ such that*

$$\begin{aligned}
v_1 &= \underset{\|v\|=1}{\mathit{argmin}}\ \frac{1}{N} \sum_{i=1}^{N} \mathrm{d}(p_i, \pi_{H_1(v)}(p_i))^2 \ \textit{ with } H_1(v) = \mathrm{Exp}_{\mu(D)}(\mathrm{span}(v)) \\
v_2 &= \underset{\|v\|=1,\, v \in C_1}{\mathit{argmin}} \frac{1}{N} \sum_{i=1}^{N} \mathrm{d}(p_i, \pi_{H_2(v)}(p_i))^2 \\
&\ \ \vdots \qquad\qquad \vdots \\
v_K &= \underset{\|v\|=1,\, v \in C_{K-1}}{\mathit{argmin}} \frac{1}{N} \sum_{i=1}^{N} \mathrm{d}(p_i, \pi_{H_K(v)}(p_i))^2
\end{aligned}$$

where $V_j = \{v_1, \ldots, v_j\}, C_j = \operatorname{span}(V_j)^\perp$ *and* $H_j(v) = \operatorname{Exp}_{\mu(D)}(\operatorname{span}(\{V_{j-1} \cup v\}))$ *for* $j = 2, \ldots, K$.

Objective functions in PCA are sum of squared distances of data to their orthogonal projections to linear subspaces. The symmetric, linear operator defined below is the gradient of the objective function in PCA.

Definition 6 (Covariance Operator)**.** *Given* $\{q_1, \ldots, q_N\} \in \mathrm{T}_\mu M$ *define*

$$\mathrm{L} : \mathrm{T}_u M \to \mathrm{T}_u M, \ \mathrm{L}(v) = \frac{1}{N} \sum_{i=1}^{N} \langle q_i, v \rangle_\mu q_i.$$

Throughout we assume $\{q_1, \ldots, q_N\} \in \mathrm{T}_\mu M$ is distributed so that L is diagonalizable. Let $u_1, \ldots, u_K$ be the eigenvectors of L with corresponding eigenvalues $\beta_1, \ldots, \beta_K$ given in descending order by magnitude.

2.1. Expansion of PGA directions

Let S^n, $P(n)$ and $SO(n)$ denote the unit n-sphere, the space of positive definite matrices and the special orthogonal group formulated as Riemannian manifolds as in sections 3.1, 4.1, and 5.1.

Assume M in the proposition in this section is one of these spaces. We will apply the proposition to these spaces in this paper.

Proposition (Expansion of PGA directions)**.** *Let* $\mu \in M$*,* $q_i \in \mathrm{T}_\mu M$ *and* $p_{i,\epsilon} = \operatorname{Exp}_\mu(\epsilon q_i)$ *for* $i = 1, \ldots, N$*. Also, letting* $D_\epsilon = \{p_{i,\epsilon}\}_i$ *assume* $\mu(D_\epsilon) = \mu$ *and that* $V_{K,\epsilon} = \{v_1(\epsilon), \ldots, v_K(\epsilon)\}$ *are the PGA directions of* D_ϵ $\forall \epsilon \neq 0$*.*

Further, let $f_j(v, \epsilon)$ *be the objective function for* $v_j(\epsilon)$ *in definition 5 for* $j = 1, \ldots, K$*. That is,*

$$f_j(v, \epsilon) = \frac{1}{N} \sum_{i=1}^{N} \mathrm{d}(p_{i,\epsilon}, \pi_{H_j(v)}(p_{i,\epsilon}))^2.$$

Also let $g_j(v, \epsilon)$ *be as* $f_j(v, \epsilon)$ *with* $u_1, \ldots, u_{j-1}$ *in place of the previous PGA directions and let*

$$\alpha_j = (1/2)\nabla_{u_j} g_{j,4}(v) \tag{4}$$

for $j = 1, \ldots, K$*. Then*

$$v_1(\epsilon) = v_{1,0} + v_{1,2}\epsilon^2 + O(\epsilon^4)$$

where $v_1(\epsilon) = u_1$ *and* $v_{1,2} = \sum_{j>1}^{K} c_j u_j$ *with*

$$c_j = \langle \alpha_1, u_j \rangle / (\beta_j - \beta_1) \text{ for } j > 1.$$

For $k > 1$

$$v_k(\epsilon) = v_{k,0} + v_{k,2}\epsilon^2 + O(\epsilon^4)$$

where $v_{k,0} = u_k$ *and* $v_{k,2} = \sum_{j=1, j \neq k}^{K} c_j u_j$ *with*

$$c_j = \begin{cases} \langle \alpha_j, u_k \rangle / (\beta_j - \beta_k) & \text{if } 1 \leq j < k \\ \langle \alpha_k, u_j \rangle / (\beta_j - \beta_k) & \text{if } k < j \leq K. \end{cases}$$

Proof. The proof is by induction. The base case can be shown in a similar manner as the induction step. Thus we let $k > 1$, assume the proposition holds for the first $k-1$ PGA directions $V_{k-1,\epsilon} = \{v_1(\epsilon), \ldots, v_{k-1}(\epsilon)\}$ and then show it holds for $v_k(\epsilon)$.

As given in sections 3.1, 4.1, and 5.1, M is a symmetric space and thus the mapping

$$\iota : M \circlearrowleft, \; \iota(p) = \mathrm{Exp}_\mu\big(-\mathrm{Log}_\mu(p)\big) \quad \text{for } p \in M \tag{5}$$

is an isometry. Thus, provided projection, as in (2), is unique, $f_j(v, \epsilon)$ is even in ϵ and we expand

$$f_j(v, \epsilon) = f_{j,2}(v)\epsilon^2 + f_{j,4}(v)\epsilon^4 + O(\epsilon^6). \tag{6}$$

for $j = 1, \ldots, K$.

As given in (19), (28), and (41) in M the Riemannian exponential map converges to the Euclidean exponential map (and more generally, as in [22], for any immersed submanifold in Euclidean space) and thus the geodesic distances converge to Euclidean distances in the tangent space. Thus we have

$$f_{k,2}(v) = \frac{1}{N}\sum_{i=1}^{N}\left(\langle q_i, q_i\rangle - \sum_{j=1}^{k-1}\langle q_i, v_j\rangle^2 - \langle q_i, v\rangle^2\right). \tag{7}$$

as the leading term of $f_k(v, \epsilon)$. Taking a derivative

$$\nabla_v f_{k,2}(v) = -\frac{2}{N}\sum_{i=1}^{N}\langle q_i, v\rangle\, q_i = -2\mathrm{L}(v)$$

so that

$$\nabla_v f_k(v, \epsilon) = -2\mathrm{L}(v)\epsilon^2 + \nabla_v f_{k,4}(v)\epsilon^4 + O(\epsilon^6) \tag{8}$$

Using Lagrange multipliers $(-2\lambda_1, \ldots, -2\lambda_{k-1}, -\lambda_k)$ we have

$$\nabla_{v_k(\epsilon)} f_k(v, \epsilon) = -2\lambda_1 v_1(\epsilon) - \cdots - 2\lambda_k v_k(\epsilon) \tag{9}$$

with constraints $\langle v_1(\epsilon), v_k(\epsilon)\rangle = \cdots = \langle v_{k-1}(\epsilon), v_k(\epsilon)\rangle = 0, \langle v_k(\epsilon), v_k(\epsilon)\rangle = 1$. Expand $v_k(\epsilon)$ in ϵ

$$v_k(\epsilon) = v_{k,0} + v_{k,2}\epsilon^2 + O(\epsilon^4). \tag{10}$$

Substituting this expansion and the expansions of $\{v_1(\epsilon), \ldots, v_{k-1}(\epsilon)\}$ in the constraints and equating coefficients in orders of ϵ gives

$$\begin{aligned} &\langle v_{k,0}, v_{j,0}\rangle = 0, \langle v_{k,0}, v_{j,2}\rangle = -\langle v_{k,2}, v_{j,0}\rangle \text{ for } j = 1, \ldots, k-1, \\ &\langle v_{k,2}, v_{k,0}\rangle = 0 \text{ and } \langle v_{k,0}, v_{k,0}\rangle = 1. \end{aligned} \tag{11}$$

Expand the Lagrange multipliers

$$\lambda_j = \lambda_{j,0} + \lambda_{j,2}\epsilon^2 + \lambda_{j,4}\epsilon^4 + O(\epsilon^6) \text{ for } j = 1, \ldots, k. \tag{12}$$

In computations in sections 3.2, 4.3, 5.2 we have $\nabla_{v_k(\epsilon)} f_{k,4} = \nabla_{u_k} g_{k,4} + O(\epsilon^6)$. Using this and substituting expansions (10), (12) and of $\{v_1(\epsilon), \ldots, v_{k-1}(\epsilon)\}$, in (9), using (8) and equating coefficients in orders of ϵ gives

$$\lambda_{1,0} v_{1,0} + \cdots + \lambda_{k,0} v_{k,0} = 0 \ (\implies \lambda_{1,0} = \cdots = \lambda_{k,0} = 0) \tag{13}$$

$$\lambda_{1,2} v_{1,0} + \cdots + \lambda_{k,2} v_{k,0} = \mathrm{L}(v_{k,0}) \tag{14}$$

$$\sum\nolimits_{j=1}^{k} \lambda_{j,2} v_{j,2} + \lambda_{j,4} v_{j,0} = -(1/2) \nabla_{v_{k,0}} g_{k,4}(v) + \mathrm{L}(v_{k,2}). \tag{15}$$

As L is a symmetric linear operator, (14) and (11) give

$$\lambda_{j,2} = \langle \mathrm{L}(v_{k,0}), v_{j,0} \rangle = \langle v_{k,0}, \mathrm{L}(v_{j,0}) \rangle = \beta_j \langle v_{k,0}, v_{j,0} \rangle = 0 \tag{16}$$

for $j = 1, \ldots, k-1$.

Thus by (14) $\mathrm{L}(v_{k,0}) = \lambda_{k,0} v_{k,0}$ and $v_{k,0}$ is a normalized eigenvector of L with eigenvalue $\frac{1}{N} \sum_{i=1}^{N} \langle v_{k,0}, q_i \rangle^2$. Further, as $f(v_{k,\epsilon}, \epsilon) \leq f(w, \epsilon)$ $\forall \epsilon \neq 0$ and $\forall w \in SV_{k-1}^{\perp}$ using (7)

$$\sum_{i=1}^{N} \langle q_i, v_{k,0} \rangle^2 \geq \sum_{i=1}^{N} \langle q_i, w \rangle^2 \ \ \forall w \in SV_{k-1}^{\perp}.$$

Thus $v_{k,0}$ is the dominant normalized eigenvector of L in $SV_{k-1}^{\perp}$ so that $v_{k,0} = u_k$ and $\beta_k = \lambda_{k,2}$ (as they must be as u_k is the solution in the tangent space).

By above, (15), (16) and the definition of α_k in (4) we have

$$(\mathrm{L} - \beta_k) v_{k,2} = \sum_{j=1}^{k} \lambda_{j,4} u_j + \alpha_k. \tag{17}$$

Consider the orthonormal expansion of $v_{k,2}$,

$$v_{k,2} = \sum_{\substack{j=1 \\ j \neq k}}^{K} c_j u_j \tag{18}$$

For $j < k$, using (11) and the form of the expansions of $\{v_1(\epsilon), \ldots, v_{k-1}(\epsilon)\}$

$$c_j = \langle v_{k,2}, u_j \rangle = -\langle u_k, v_{j,2} \rangle = \langle \alpha_j, u_k \rangle / (\beta_j - \beta_k).$$

Also, substituting (18) into (17) gives

$$\sum_{\substack{j=1 \\ j \neq k}}^{K} c_j (\beta_j - \beta_k) u_j = \sum_{j=1}^{k} \lambda_{j,4} u_j + \alpha_k$$

so that for $j > k$

$$c_j = \langle \alpha_k, u_j \rangle / (\beta_j - \beta_k).$$

Thus, $v_k(\epsilon)$ has the form given in the proposition and the proposition holds.

3. PGA in S^n

We first examine the role of scale in PGA in the unit n-spheres which we denote by S^n. Spherical data occurs in directional statistics and in preshapes in shape analysis as in [21] and [17]. We obtain the expansions of PGA directions in S^n according to the proposition in section 2.1, test these expansions and apply them to simulated data from the Von Mises-Fisher spherical distribution.

3.1. S^n as a Riemannian manifold

We denote the unit sphere in $\mathbb{R}^{n+1}$ by S^n. We have the identification

$$\mathrm{T}_p S^n \equiv \{v \in \mathbb{R}^{n+1}; \langle v, p\rangle = 0\}.$$

On S^n we use the Riemannian metric induced by the embedding $S(n) \hookrightarrow \mathbb{R}^{n+1}$, i.e., for any $p \in S(n)$ and $X, Y \in \mathrm{T}_p S(n)$, $\langle X, Y\rangle$ is the dot product. The geodesics in $S(n)$ are great circles and for $p \in S^n$ the Riemannian exponential map is directly computed as below.

Riemannian exponential map on $S(n)$.

$$\mathrm{Exp}_p(X) = \cos(\|X\|)p + \sin(\|X\|)\frac{X}{\|X\|} \tag{19}$$

for $X \in \mathrm{T}_p S(n); \|X\| < \pi\}$.

Then

$$\mathrm{Log}_p(q) = \arccos\left(\langle p, q\rangle\right) \frac{q - \langle p, q\rangle\, p}{\|q - \langle p, q\rangle\, p\|} \quad \text{for } q \in S^n, |\langle q, p\rangle| < 1 \tag{20}$$

Given $p, q \in S(n)$

$$\mathrm{d}(p, q) = \left\|\mathrm{Log}_p(q)\right\| = \arccos\left(\langle p, q\rangle\right) \tag{21}$$

that is, the distance between p and q is the ordinary spherical distance.

As in [7], S^n is a symmetric space with the symmetry at any point $p \in S^n$ provided by reflection over the line containing p in $\mathbb{R}^{n+1}$.

3.2. Expansion of PGA directions in S^n

As in [17] the projection operator in S^n has closed form. Given $p, \mu \in S^n$ and orthonormal set $V_k = \{v_1, \ldots, v_k\} \subset \mathrm{T}_\mu P(n)$, set $v_0 = \mu$. With $p \notin \mathrm{span}(\mu \cup V_k)^\perp$ so that projection is unique then

$$\pi_{H(V_k)}(p) = w / \|w\| \tag{22}$$

where $w = \sum_{j=0}^{k} \langle v_j, p\rangle\, v_j$. That is, projection of p is first done to $\mathrm{span}(\mu \cup V_k)$ in $\mathbb{R}^{n+1}$ and the result normalized to obtain projection to the hypersphere $H(V_k)$.

Let $q \in \mathrm{T}_\mu S^n$, $\epsilon \neq 0$, $p_\epsilon = \mathrm{Exp}_\mu(\epsilon q) = \cos(\|q\|\,\epsilon)\mu + \sin(\epsilon\,\|q\|)q/\,\|q\|$ and

$$t_j(\epsilon) = \left\langle \mathrm{Log}_\mu(\pi_{H(V_k)}(p_\epsilon)), v_j \right\rangle \text{ for } j = 1, \ldots, k$$

so that

$$\pi_{H(V_k)}(p_\epsilon) = \mathrm{Exp}_\mu\left(\sum_{j=1}^{k} t_j(\epsilon) v_j\right).$$

By using (20) and (22), taking an inner product and computing Taylor expansions we obtain

Expansion of projection coefficients in S^n.

$$\begin{aligned} t_a(\epsilon) &= \mathrm{acos}\left(\frac{\cos(\epsilon\,\|q\|)}{\sqrt{\cos^2(\epsilon\,\|q\|) + \sin^2(\epsilon\,\|q\|)\left(\sum_{j=1}^{k}\langle q, v_j\rangle^2\right)}}\right) \cdot \\ &\qquad \frac{\langle q, v_a\rangle}{\sqrt{\sum_{j=1}^{k}\langle q, v_j\rangle^2}} \\ &= \langle q, v_a\rangle + (1/3)\langle q, v_a\rangle\left(\langle q, q\rangle^2 - \sum_{j=1}^{k}\langle q, v_j\rangle^2\right)\epsilon^3 + O(\epsilon^5) \\ &= \cos\theta_a\,\|q\|\,\epsilon + (1/3)\cos\theta_a\left(1 - \sum_{j=1}^{k}\cos^2\theta_j\right)\|q\|^3\,\epsilon^3 + O(\epsilon^5) \end{aligned} \tag{23}$$

for $a = 1, \ldots, k$ where θ_a the angle formed by q and v_a for $a = 1, \ldots, k$.

Now as in the proposition in section 2.1 assume $\mu \in S^n$, $q_i \in \mathrm{T}_\mu M$ and $p_{i,\epsilon} = \mathrm{Exp}_\mu(\epsilon q_i)$ for $i = 1, \ldots, N$. Let $V_{k-1,\epsilon} = \{v_1(\epsilon), \ldots, v_{k-1}(\epsilon)\}$ be the first $k-1$ PGA directions. Using the proposition we find $v_k(\epsilon)$. Let $v \in \mathrm{S}_\mu S^n$ and

$$\begin{aligned} t_{i,j} &= \left\langle \mathrm{Log}_\mu(\pi_{H(V_\epsilon)}(p_{i,\epsilon})), v_j \right\rangle \text{ and} \\ t_{i,k} &= \left\langle \mathrm{Log}_\mu(\pi_{H(V_\epsilon)}(p_{i,\epsilon})), v \right\rangle \end{aligned}$$

for $i = 1, \ldots, N$ and $j = 1, \ldots, k-1$.

Our objective function is $f_k(v,\epsilon) =$

$$= \frac{1}{N}\sum_{i=1}^{N} \operatorname{acos}\left(\left\langle \operatorname{Exp}_\mu(t_{i,1}v_1 + \ldots + t_{i,k}v), \cos(\epsilon\,\|q_i\|)\mu + \sin(\epsilon\,\|q_i\|)\frac{q_i}{\|q_i\|}\right\rangle\right)^2$$

$$= \frac{1}{N}\sum_{i=1}^{N} \operatorname{acos}\Bigg(\cos\left(\sqrt{\sum\nolimits_{j=1}^{k} t_{i,j}^2}\right)\cos(\epsilon\,\|q\|) +$$

$$\sin(\epsilon\,\|q\|)\sin\left(\sqrt{\sum\nolimits_{j=1}^{k} t_{i,j}^2}\right)\frac{\sum_{j=1}^{k}\langle v_{i,j}, q_i\rangle}{\sqrt{\sum_{j=1}^{k} t_{i,j}^2}}\Bigg)^2$$

From which by taking Taylor expansions in ϵ we obtain

$$f_{k,4}(v) = \frac{1}{3N}\sum_{i=1}^{N}\Bigg(\left(\sum\nolimits_{j=1}^{k-1}\langle q_i, v_j\rangle^2 + \langle q_i, v\rangle^2\right)\cdot \left(\sum\nolimits_{j=1}^{k-1}\langle q_i, v_j\rangle^2 + \langle q_i, v\rangle^2 - \langle q_i, q_i\rangle\right)\Bigg) \tag{24}$$

Taking a derivative and evaluating gives

Expansion of $v_k's$ in S^n.

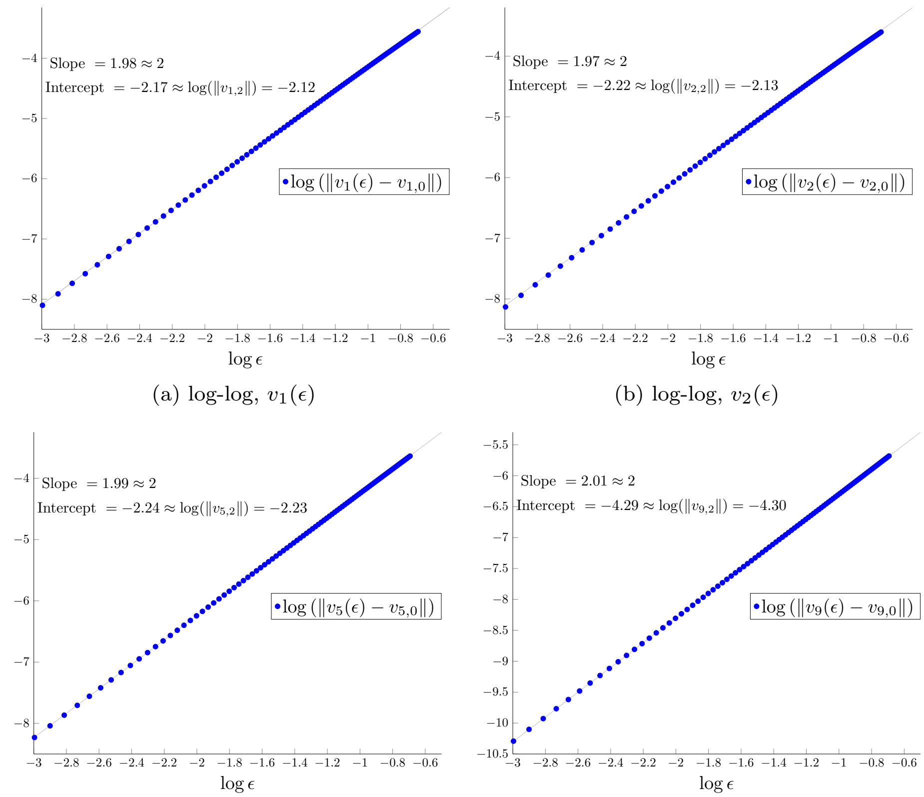
$$\begin{aligned}\alpha_k &= (1/2)\nabla_{u_k} g_{k,4}(v)\\ &= \frac{1}{3N}\sum_{i=1}^{N}\left(2\sum\nolimits_{j=1}^{k}\langle q_i, u_j\rangle^2 - \langle q_i, q_i\rangle\right)\langle q_i, u_k\rangle\, q_i\end{aligned}$$

which is used in the proposition to obtain expansion $v_k(\epsilon) = v_{k,0} + v_{k,2}\epsilon^2 + O(\epsilon^4)$.

In figure 1, the expansions of PGA directions obtained above are tested in S^{10}. With $\mu \in S^{10}$, 50 tangent vectors are sampled from $\mathrm{T}_\mu S^{10}$ with the entries of $\{q_i\}$ having normal distributions and variances varying by entry so that principal geodesic directions can be identified. We then take $D_\epsilon = \{p_{i,\epsilon}\}_i = \{\operatorname{Exp}_\mu(\epsilon q_i)\}_i$. PGA directions are located in MATLAB by using fixed-point algorithms such as those in [17] used to compute *principal component geodesics*. Log-log plots are shown for the $v_1(\epsilon), v_2(\epsilon), v_4(\epsilon)$ and $v_9(\epsilon)$. Similar plots were obtained for the other PGA directions.

3.3. Simulated Data from the Von-Mises Fisher Distribution

The Von-Mises Fisher distribution on S^n is important in directional statistics and applications exist in geology, text mining, cluster analysis and others



(a) log-log, $v_1(\epsilon)$

(b) log-log, $v_2(\epsilon)$

(c) log-log, $v_5(\epsilon)$

(d) log-log, $v_9(\epsilon)$

Figure 1: Tests of PGA expansions in S^n

(see [21], [2],and [14] for examples and definitions). The Von-Mises Fisher distribution has two parameters, μ and κ which are the expected mean and concentration parameters, respectively. A higher value of κ gives greater expected concentration about μ with $\kappa = 0$ giving an expected uniform distribution about S^n. Thus, κ should be closely and negatively correlated with the dispersion of the data from μ.

Below, R package movMF in [14] is used to simulate data in S^{15}. We fix μ and vary κ to measure the effect that scale plays on the accuracy of the leading- and next-order approximations of PGA.

In table 1 for each value of κ 100 data points are sampled in 5 runs. For each run all 14 PGA directions are computed. Also, their leading- and next-order approximations in the scale of the data are located, that is, $v_{k,0}$ (using PCA in the tangent space) and $v_{k,0} + v_{k,2}$, respectively, with $\epsilon = 1$ as the norms of the logs of the data scale the approximations. Across all runs and for each value of κ the means of the angles (in radians) the approximations make (m.est.θ_0 and m.est.θ_2) with the exact PGA directions and the mean of the norms of the logs of the data (m.scale) are computed.

Also, as proposed in [28], an iterative algorithm to locate PGA direction v_k can be initialized by doing PCA with the projections of the logs of the data into $\text{span}(V_{k-1}^{\perp})$. I compare this method with initialization by projection of the next-order approximations into $\text{span}(V_{k-1}^{\perp})$ by computing the means of the angles these initializations make (m.init.θ_0 and m.init.θ_2, respectively) with the $v_j's$ across the 5 runs for each value of κ.

κ	.5	2.5	5	10	20	30	40	50
m.scale	1.4695	1.3822	1.2273	1.0118	0.7772	0.6483	0.5631	0.5098
m.est.θ_0	0.7018	0.5957	0.5428	0.4156	0.2502	0.1381	0.1085	0.1025
m.est.θ_2	0.6527	0.5273	0.3997	0.2657	0.1434	0.0521	0.0297	0.0352
m.init.θ_0	0.4938	0.4751	0.4151	0.2838	0.1665	0.0911	0.0714	0.0669
m.init.θ_2	0.5058	0.3723	0.3715	0.1868	0.0935	0.0337	0.0190	0.0255

m.scale = mean of norms of logs
m.est.θ_0, est.θ_2 = mean angles of leading and next order estimates w/$v_j's$
m.init.θ_2, init.θ_3 = mean angles of leading and next order initializations w/$v_j's$

Table 1: Estimates of PGA in S^{15} with Von Mises-Fisher samples

At $\kappa = .5$ the data is nearly uniformly distributed which is reflected in m.scale = 1.4695 nearly $\pi/2$. Even at $\kappa = .5$ the next-order estimates are an improvement over the leading-order estimates for this distribution. Note the next-order estimates holding for data dispersed significantly from the the mean is also reflected in the plots with nearly correct intercepts at $log(\epsilon) = 0 \implies \epsilon = 1$ in figure 2. As κ increases and the samples draw in towards their computed means both estimates improve with the next-order estimates improving more sharply as they should.

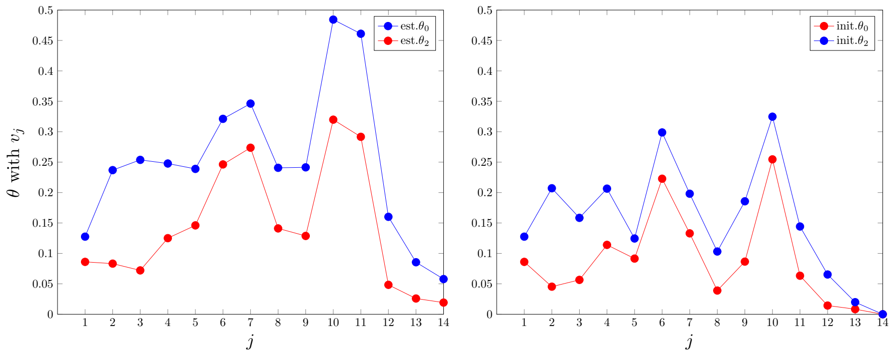
Also, projecting the next-order estimates into $\text{span}(V_{k-1}^{\perp})$ provides an improvement here over doing PCA in $\text{span}(V_{k-1}^{\perp})$ with m.init.$\theta_2 <$ m.init.θ_0 for all values of κ except the first where m.init.θ_2 is only slightly greater.

In figure 2, $\kappa = 20$ and for each PGA direction we plot the values of est.θ_0, est.θ_2 and init.θ_0, init.θ_2 which are means across 5 runs. Similar plots were obtained for other values of κ.

4. PGA in $P(n)$

We denote the space of positive definite matrices by $P(n)$. The imaging technology diffusion tensor MRI as in [3] produces data known as diffusion tensors in $P(3)$. In [10] PGA in definition 5 was proposed to allow the proper analysis of statistical variability of diffusion tensor data by formulating $P(3)$ as a manifold and generalizing PCA.

In this section using the proposition in section (2.1) and computer algebra we obtain expansions of the first two PGA directions in $P(n)$. We test the



(a) Estimates of $v_j's$ (b) Initializations of $v_j's$

Figure 2: Estimates of PGA in S^{15} with $\kappa = 20$

expansions and we take a closer look at the geometry of the first projection coefficient.

4.1. $P(n)$ as a Riemannian manifold

As in [20], $P(n)$ is an open set in the vector space of $n \times n$ symmetric matrices. Thus, we have the identification

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

Consider the following

Action on $P(n), \varphi$.

$$\begin{aligned} &\varphi : GL(n) \times P(n) \to P(n), \\ &\varphi(g, p) = \varphi_g(p) = gpg^\top \end{aligned} \tag{25}$$

where $GL(n)$ is the general linear group.

As in [4] this action is transitive and for $g \in GL(n)$ we have the following Riemannian metric up to a scalar multiple for which φ_g is an isometry.

Riemannian Metric on $P(n)$.

$$\langle X, Y \rangle_p = (1/2) tr \left(p^{-1} X p^{-1} Y \right) \tag{26}$$

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

As in [4] setting $M = P(n)$ and $\phi_p(q) = pq^{-1}p$ for $q \in P(n)$ in definition 1 then makes $P(n)$ a symmetric space.

For $p \in P(n)$ and $x, y \in \mathrm{T}_p P(n), [x, y] = xy - yx$ is the *commutator* of x, y. As $P(n)$ is a symmetric space, by [24], for $p \in P(n)$ and $x, y, z \in \mathrm{T}_p P(n)$ we have

$$\mathrm{R}_{x,y}\, z = [z, [x, y]] \tag{27}$$

where R is the *Riemannian curvature tensor*.

As in [4], at $\mathrm{I} \in P(n)$, $\mathrm{Exp}_{\mathrm{I}}$ is the matrix exponential function.

Riemannian Exponential Map on $P(n)$.

$$\mathrm{Exp}_{\mathrm{I}}(X) = \sum_{i=0}^{\infty} X^i/i! \quad \text{for } X \in \mathrm{T}_{\mathrm{I}} P(n). \tag{28}$$

Then for any $p = gg^\top \in P(n)$ where $g \in GL(n)$ we have

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

for $Y \in \mathrm{T}_p P(n)$.

Also in [4], for all $p \in P(n)$, Exp_p is bijective and thus Log_p is well-defined.

4.2. Computer Algebra

Expansions in this section and in sections 5.3 and 6 were obtained with the help of the Maxima computer algebra system. A matrix Taylor function and a function with the distributive and cyclic properties of the trace were coded to compute expansions involving $\mathrm{Log}_{\mathrm{I}}$ and $\mathrm{Exp}_{\mathrm{I}}$ in ϵ. This code along with the data sets used in this paper are available at https://github.com/DMLazar/PGAScale.

4.3. Expansion of PGA directions in $P(n)$

By employing the transitive action by isometries in (25) PGA in $P(n)$ can be carried out in the tangent space at the identity. Let $v \in \mathrm{S}_{\mathrm{I}} P(n), q \in \mathrm{T}_{\mathrm{I}} P(n)$ and $p_\epsilon = \mathrm{Exp}_{\mathrm{I}}(\epsilon q)$ for $\epsilon \neq 0$. To project p_ϵ to $H(v) = \mathrm{Exp}_{\mathrm{I}}(\mathrm{span}(v))$ find

$$t(\epsilon) = \underset{s \in \mathbb{R}}{\operatorname{argmin}}\ \mathrm{d}(\mathrm{Exp}_{\mathrm{I}}(sv), \mathrm{Exp}_{\mathrm{I}}(\epsilon q))^2 \tag{29}$$

Define $O_{\epsilon,s}(l)$ as

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

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 $h(s,\epsilon)$ as the cost function in (29) and expanding

$$\begin{aligned} h(s,\epsilon) &= \mathrm{d}(\mathrm{Exp}_{\mathrm{I}}(sv), \mathrm{Exp}_{\mathrm{I}}(\epsilon q))^2 \\ &= (1/2)tr\left(\mathrm{Log}_{\mathrm{I}}(g(s,\epsilon))\mathrm{Log}_{\mathrm{I}}(g(s,\epsilon))\right) \\ &= \frac{\epsilon^2 tr\,(qq)}{2} + s^2 - \epsilon s\, tr\,(qv) + \frac{tr\left(q^2v^2 - (qv)^2\right)\epsilon^2 s^2}{12} + O_{\epsilon,s}(6). \end{aligned} \tag{30}$$

Using (5) $t(\epsilon)$ is odd in ϵ and we expand

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

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

Solving for t_1 and t_3 in $\frac{\partial h_s(t(\epsilon),\epsilon)}{\partial s} = 0$ gives

$$t(\epsilon) = \frac{1}{2}tr\,(qv)\,\epsilon + \frac{1}{24}tr\,(qv)\left(tr\left((qv)^2 - q^2v^2\right)\right)\epsilon^3 + O(\epsilon^5). \tag{31}$$

Further, letting $K(q,v)$ be the sectional curvature of span$(\{q,v\})$, using (27) we have

$$\begin{aligned} K(q,v) &= \frac{\langle \mathrm{R}_{q,v}\, v, q\rangle_{\mathrm{I}}}{\langle q,q\rangle_{\mathrm{I}}\langle v,v\rangle_{\mathrm{I}} - \langle q,v\rangle_{\mathrm{I}}^2} \\ &= \frac{tr\left(vqvq - v^2q^2\right)}{(1/2)tr\,(qq) - ((1/2)tr\,(qv))^2}. \end{aligned} \tag{32}$$

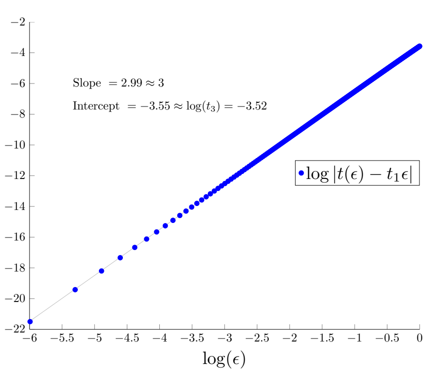
Using (32) in (31) we have

Expansion of projection (to a geodesic) coefficient in $P(n)$.

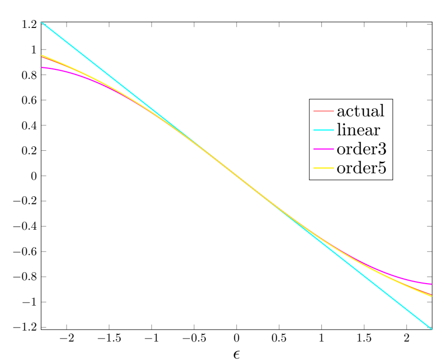
$$\begin{aligned} t(\epsilon) &= \frac{1}{2}tr\,(qv)\,\epsilon + \frac{1}{24}tr\,(qv)\left(tr\left((qv)^2 - q^2v^2\right)\right)\epsilon^3 + O(\epsilon^5) \\ &= \cos\theta\,\|q\|\,\epsilon + \frac{1}{12}\cos\theta\sin^2\theta K(v,q)\,\|q\|^3\,\epsilon^3 + O(\epsilon^5) \end{aligned} \tag{33}$$

where θ is the angle formed by q and v.

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



(a) Log-log plot



(b) Approximations

Figure 3: Tests of expansion of $t(\epsilon)$

As in figure 4 below, letting $\|q\| = 1$ 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 approximate Euclidean triangle in the tangent space and the geodesic triangle in the manifold. Further, as in in [4] $P(n)$ is of non-positive sectional curvature and thus $\frac{1}{12}\cos\theta\sin^2\theta K(v,q)$ will be non-positive for acute angle θ.

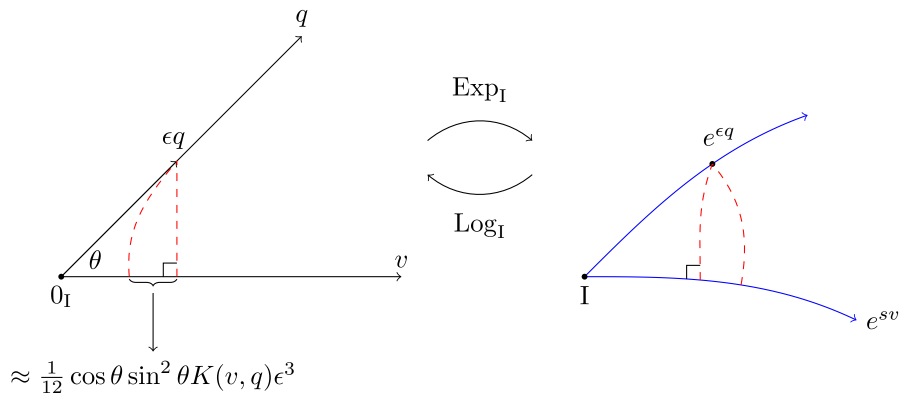


Figure 4: Approximation of projection coefficient

We apply the proposition to compute the expansions of the first two PGA directions in $P(n)$. First we have

$$f_1(v,\epsilon) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{d}(\mathrm{Exp}_{\mathrm{I}}(t_i(\epsilon,v)v), \mathrm{Exp}_{\mathrm{I}}(\epsilon q_i))^2 \tag{34}$$

Using (31) in (34) and expanding in ϵ gives

$$f_{1,4}(v) = \frac{1}{48N}\sum_{i=1}^{N} tr\left(q_i v\right)^2 \left(tr\left(q_i^2 v^2\right) - tr\left(q_i v q_i v\right)\right). \tag{35}$$

Taking a derivative and evaluating gives

Expansion of $v_1(\epsilon)$ in $P(n)$.

$$\begin{aligned}\alpha_1 = \frac{1}{96N}\sum_{i=1}^{N} & 2tr\left(q_i u_1\right) tr\left(q_i^2 u_1^2 - (q_i u_1)^2\right) q_i \\ & + tr\left(q_i u_1\right)^2 \left(q_i^2 u_1 + u_1 q_i^2 - 2 q_i u_1 q_i\right)\end{aligned} \tag{36}$$

which can be used in the proposition to obtain the expansion of $v_1(\epsilon)$.

Let orthonormal set $V_2 = \{v_1, v_2\} \subset \mathrm{S_I}P(n)$, $q \in \mathrm{T_I}P(n)$ and $p_\epsilon = \mathrm{Exp}_{\mathrm{I}}(\epsilon q)$ for $\epsilon \neq 0$. To project p_ϵ to $H(V_2) = \mathrm{Exp}_{\mathrm{I}}(\mathrm{span}(V_2))$ find

$$(t_1(\epsilon), t_2(\epsilon)) = \underset{s,r\in\mathbb{R}^2}{\mathrm{argmin}}\ \mathrm{d}(\mathrm{Exp}_{\mathrm{I}}(sv_1 + rv_2), \mathrm{Exp}_{\mathrm{I}}(\epsilon q))^2 \tag{37}$$

Taking partial derivatives of the objective function in (37), setting them to zero and solving the resulting system of equations gives

Expansion of $t_1(\epsilon), t_2(\epsilon)$ in $P(n)$.

$$t_1(\epsilon) = \frac{1}{2} tr\,(qv_1)\,\epsilon + \frac{1}{48}\Big(2tr\,(qv_1)\,tr\,\big((qv_1)^2 - q^2v_1^2\big) + tr\,(qv_2)\,tr\,(2qv_1qv_2 - qqv_2v_1 - qqv_1v_2)\Big)\epsilon^3 + O(\epsilon^5)$$

which gives $t_2(\epsilon)$ by symmetry.

Fixing v_1 and substituting these expansions into the cost function for $v_2(\epsilon)$ gives

$$f_2(v,\epsilon) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{d}(\mathrm{Exp_I}(t_{1,i}(\epsilon,v)v_1 + t_{2,i}(\epsilon,v)v), \mathrm{Exp_I}(\epsilon q_i))^2. \tag{38}$$

Expanding in ϵ, taking a derivative and evaluating gives

Expansion of v_2 in $P(n)$.

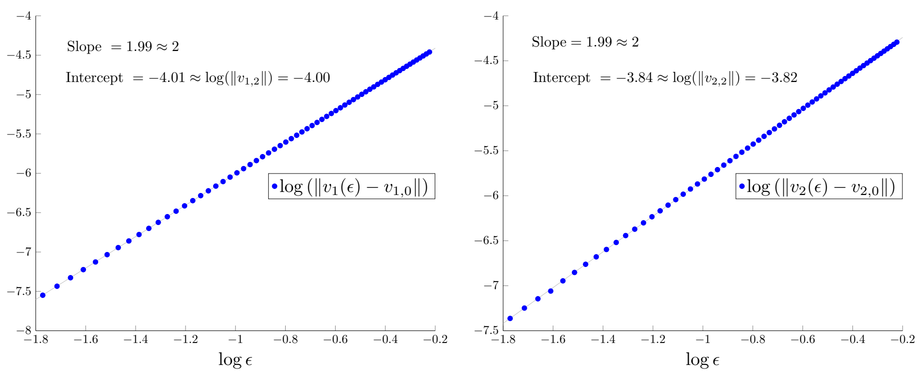
$$\begin{aligned}\alpha_2 = \frac{1}{48}\sum_{i=1}^{N} & tr\,(q_iu_2)^2\,(q_iq_iu_2 + u_2q_iq_i - 2q_iu_2q_i) \\ & + tr\,(q_iu_1)\,tr\,(q_iu_2)\,(q_iq_iu_1 + u_1q_iq_i - 2q_iu_1q_i) \\ & + 2tr\,(q_iu_2)\,tr\,(q_iq_iu_2u_2 - q_iu_2q_iu_2)\,q_i + \\ & \quad tr\,(q_iu_1)\,tr\,(q_iq_iu_2u_1 + q_iq_iu_1u_2 - 2q_iu_1q_iu_2)\,q_i\end{aligned}$$

which can be used with α_1 in (36) and the proposition to obtain the expansion of $v_2(\epsilon)$.

In figure 5, the expansions of $v_1(\epsilon)$ and $v_2(\epsilon)$ in $P(n)$ are tested. 75 matrices $\{q_i\}_i$ are sampled from $\mathrm{T_I}P(3)$ with entries distributed as in the test in figure 1 and the data is set as $D_\epsilon = \{p_{i,\epsilon}\}_i = \{\mathrm{Exp_I}(\epsilon q_i)\}_i$. The computations of the projection operator and principal geodesic directions are done using MATLAB minimization routines and user-supplied gradients as formulated in [28] with the derivative of the matrix exponential map provided by [23, Theorem 4.5].

5. PGA in $SO(n)$

We denote the *special orthogonal group* by $SO(n)$. $SO(n)$ is the group of rigid rotations of $\mathbb{R}^n$. In particular it includes *the rotation group* $SO(3)$ in which data



(a) log-log, $v_1(\epsilon)$

(b) log-log, $v_2(\epsilon)$

Figure 5: Tests of PGA expansions in $P(n)$

naturally arises in robotics, computer vision and others (see [13], [1] and [29]). We obtain and test expansions in $SO(n)$ and apply them to the formulation of PGA in [11] for Lie groups.

5.1. $SO(n)$ *as a Riemannian manifold*

As in [5] $SO(n)$ is a closed and bounded subset of $\mathbb{R}^{n^2}$ and thus a compact set. Also, we have the identification

$$\mathrm{T_I}SO(n) \equiv \mathfrak{so}(\mathfrak{n}) = \{n \times n \text{ skew-symmetric matrices}\}.$$

$SO(n)$ is a matrix Lie group and we have the following transitive action by left multiplication

Action on $SO(n), \varphi$.

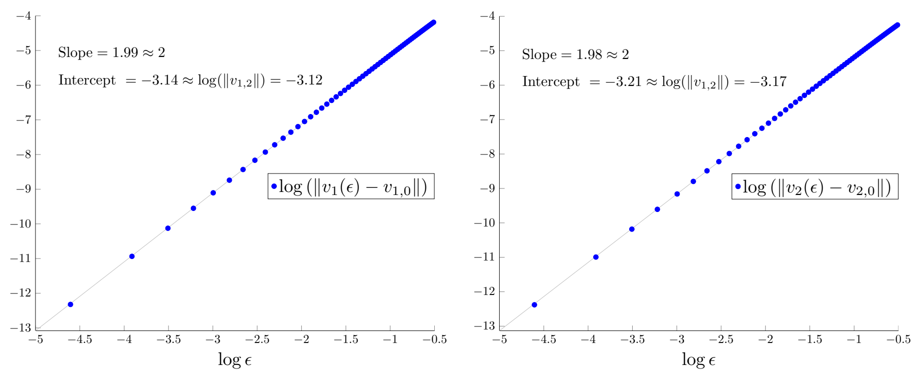
$$\begin{aligned} &\varphi : SO(n) \times SO(n) \to SO(n), \\ &\varphi(g,p) = \varphi_g(p) = gp. \end{aligned} \tag{39}$$

We then have the Riemannian metric, up to a positive scalar multiple, for which φ_g is an isometry

$$\langle X, Y \rangle_p = -(1/2)tr\left(p^{-1}Xp^{-1}Y\right) \tag{40}$$

As in $P(n)$ setting $M = SO(n)$ and $\phi_p(q) = pq^{-1}p$ for $q \in SO(n)$ in definition (1) makes $SO(n)$ a symmetric space.

Further, as $SO(n)$ is a symmetric space we have (27) as in $P(n)$.



(a) log-log, $v_1(\epsilon)$

(b) log-log, $v_2(\epsilon)$

Figure 6: Tests of PGA expansions in $SO(n)$

As in [8], at $\mathrm{I} \in SO(n)$, $\mathrm{Exp}_{\mathrm{I}}$ is the matrix exponential function.

Riemannian Exponential Map on $SO(n)$.

$$\mathrm{Exp}_{\mathrm{I}}(X) = \sum_{i=0}^{\infty} X^i/i! \quad \text{for } X \in \mathrm{T}_{\mathrm{I}}SO(n) \text{ with } \|X\| < \pi. \tag{41}$$

Also, for any $p \in SO(n)$, $Y \in \mathrm{T}_p SO(n)$ with $\|Y\| < \pi$

$$\mathrm{Exp}_p(Y) = \phi_p\left(\mathrm{Exp}_{\mathrm{I}}\big(d_p\phi_{p^{-1}}(Y)\big)\right) = p\mathrm{Exp}_{\mathrm{I}}\big(p^{-1}Y\big) \text{ with}$$

Note that using the results in [12], for any p, Exp_p diffeomorphism on $\{X \in \mathrm{T}_p SO(n); \|X\| < \pi\}$.

5.2. Expansion in $SO(n)$

Both $P(n)$ and $SO(n)$ are symmetric spaces that have the matrix exponential as the Riemannian exponential map at I. At I, the inner products of $P(n)$ and $SO(n)$ are scalar multiples of the trace of matrix products of tangent vectors. Also, the action in (25), with $SO(n)$ replacing $GL(n)$ and $P(n)$, is an action by isometries which is transitive as any $p \in SO(n)$ can be decomposed as $p = gg$ where $g = \mathrm{Exp}_{\mathrm{I}}((1/2)\mathrm{Log}_{\mathrm{I}}(p))$. Thus, the expansions in ϵ in $SO(n)$ are the expansions in section 4.3 for $P(n)$ with the inner product of $SO(n)$ replacing the inner product of $P(n)$.

Below, in the manner of figure 5 we test the expansions of the first and second PGA directions in $SO(3)$. To carry out PGA we use the identification of $SO(3)$ with $SU(2)$ as in [26] and fixed-point algorithms as in [17].

5.3. Lie group PGA

Using expansions in ϵ, in the case of $SO(n)$, we take a look at the first iteration of PGA in [11]. In a Lie group such as $SO(n)$, PGA was formulated in [11] as recursive variance removal through left multiplication as below.

Definition 7 (PGA alternative)**.** *Given Lie group M and $D = \{p_1, \ldots, p_N\} \subset M$ with $\mu(D) = \mu$, PGA directions $\{v_1, \ldots, v_m\} \subset \mathrm{S}_{\mu(D)}M$ are given*

Set $k = 1$.

I. *Find v_k s.t. $v_k = \underset{\|v\|=1}{argmin} \sum \mathrm{d}(\pi_{H(v)}(p_i), p_i)^2$ with $H(v) = \mathrm{Exp}_\mu(\mathrm{span}(v))$*

II. *Set $D' = \{g_1^{-1}p_1, \ldots, g_N^{-1}p_N\}$ where $g_i = \pi_{H(v)}(p_i)\ \forall k$*

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

In $SO(n)$ we can decompose any projection $g_i = \mathrm{Exp}_\mu(t_i v_k)$ above as

$$g_i = \mathrm{Exp}_\mu(at_i v_k)\mathrm{Exp}_\mu(bt_i v_k) \text{ where } a + b = 1.$$

Then $\gamma_{a,b} : p \rightarrow \mathrm{Exp}_\mu(-a(t_i v_k))\, p\, \mathrm{Exp}_\mu(-b(t_i v_k))$ is an isometry as

$$d_\mu\gamma(X) = \mathrm{Exp}_\mu(-a(t_i v_k))\, X\, \mathrm{Exp}_\mu(-b(t_i v_k)) \text{ for } X \in \mathrm{T}_\mu SO(n)$$

and through direct computation and using (40) we have

$$\langle d_\mu\gamma(X), d_\mu\gamma(Y)\rangle_\mu = \langle X, Y\rangle_\mu \text{ for } X, Y \in \mathrm{T}_\mu SO(n).$$

Of these isometries we should choose the one that "minimizes energy" in moving parts of data along a geodesic from g_i to μ. Assuming D has been demeaned so that $\mu = \mathrm{I}$, letting $p_i = \mathrm{Exp}_\mathrm{I}(\epsilon q_i)$ and expanding $\mathrm{Log}_\mathrm{I}(\gamma_{a,b}(\mathrm{Exp}_\mathrm{I}(\epsilon q_i))) =$

$$= (q_i - \frac{1}{2}tr\,(q_i v_k)\, v_k)\epsilon + \frac{1}{4}tr\,(q_i v_k)\,(b-a)(v_k q_i - q_i v_k)\epsilon^2 + O(\epsilon^3).$$

Thus for $a \neq b$ we have additional movement in the orthogonal complement of explanatory direction v_k. Note that accordingly, making the identification of $SO(3)$ with $SU(2)$, it is straightforward to show that with $a = b$, $\gamma_{a,b}$ corresponds to a simple plane rotation.

Also to consider in this method of PGA is the displacement of the mean as a result of curvature after removing variability in an explanatory direction. We quantify this effect in $SO(n)$ by letting $D_\epsilon = \{p_{1,\epsilon}, \ldots, p_{N,\epsilon}\} \subset SO(n)$ be such that $\mu(D_\epsilon) = \mathrm{I}\ \forall \epsilon, v_k \in \mathrm{S_I}SO(n)$ and

Case 1. $D'_\epsilon = \{g_1^{-1/2}p_{1,\epsilon}g_1^{-1/2}, \ldots, g_N^{-1/2}p_{N,\epsilon}g_N^{-1/2}\}$ using a=b=(1/2) and

Case 2. $D'_\epsilon = \{g_1^{-1}p_{1,\epsilon}, \ldots, g_N^{-1}p_{N,\epsilon}\}$ using a=1, b=0 as in the definition and by obtaining the expansion of $x(\epsilon) = \|\mathrm{Log}_{\mathrm{I}}(D'_\epsilon)\|$ in both cases.

Note that by [19] for function $f(y) = \mathrm{d}(y,p)^2 = \left\|\mathrm{Log}_y(p)\right\|^2$ we have

$$\nabla_y f = -2\mathrm{Log}_y(p_i). \tag{42}$$

Using this gradient in definition (3) $x(\epsilon)$ is s.t.

$$\sum_{i=1}^{N} \mathrm{Log}_{\mathrm{I}}(\mathrm{Exp}_{\mathrm{I}}(-x(\epsilon)/2)\mathrm{Exp}_{\mathrm{I}}(\epsilon r_i)\mathrm{Exp}_{\mathrm{I}}(-x(\epsilon)/2)) = 0 \tag{43}$$

with

1. $\mathrm{Exp}_{\mathrm{I}}(\epsilon r_i) = \mathrm{Exp}_{\mathrm{I}}(-t_i(\epsilon, v_k)v_k/2)\mathrm{Exp}_{\mathrm{I}}(\epsilon q_i)\mathrm{Exp}_{\mathrm{I}}(-t_i(\epsilon, v_k)v_k/2)$ and
2. $\mathrm{Exp}_{\mathrm{I}}(\epsilon r_i) = \mathrm{Exp}_{\mathrm{I}}(-t_i(\epsilon, v_k)v_k)\mathrm{Exp}_{\mathrm{I}}(\epsilon q_i)$

for $i = 1, \ldots, N$ in case 1 and 2, respectively.

Letting $x(\epsilon) = x_1\epsilon + x_2\epsilon^2 + x_3\epsilon^3 + x_4\epsilon^4 + O(\epsilon^5)$, substituting into (43) and solving for x_1, x_2, x_3 and x_4 gives the expansion

1. $x(\epsilon) = x_3\epsilon^3 + O(\epsilon^5)$ where
$$\begin{aligned} x_3 = \sum_{i=1}^{N} &(tr\,(q_i v_k)^2/96)(2v_k q_i v_k - q_i v_k v_k - v_k v_k q_i) - \\ &\quad (tr\,(q_i v_k)/24)(2q_i v_k q_i - q_i q_i v_k - v_k q_i q_i) \\ &\qquad + (tr\,(q_i v_k)/24)\big(tr\,(q_i q_i v_k v_k) - tr\,(q_i v_k q_i v_k)\big)v_k \end{aligned} \tag{44}$$

2. $x(\epsilon) = x_2\epsilon^3 + O(\epsilon^3)$ where
$$x_2 = \sum_{i=1}^{N} (1/4)tr\,(q_i v_k)\,(v_k q_i - q_i v_k) \tag{45}$$

in case 1 and case 2, respectively. Note that in general, for any a, b s.t. $a + b = 1$ we will have $x(\epsilon) = x_2\epsilon^2 + O(\epsilon^3)$ where

$$x_2 = \sum_{i=1}^{N} (1/4)tr\,(q_i v_k)\,((1-2a)q_i v_k + (1-2b)v_k q_i).$$

In case 1 we have $\|x(\epsilon)\| = O(\epsilon^3)$ and in case 2 we have $\|x(\epsilon)\| = O(\epsilon^2)$.

Considering these expansions, $a = b = 1/2$ gives less displacement of the mean and $a = b = 1/2$ is preferable. Still, the use of a centering or normalization step in this form of PGA might be considered, particulary when there is still significant variability in the data and if some degree degeneracy in explanatory directions is observed.

In table 2 in 500 runs I generated 50 vectors in $\mathrm{T}_{\mathrm{I}}SO(3)$ with entries having a standard Gaussian distribution and differing variances. I compare the mean angles over the 500 runs that located explanatory directions, both for $a = 1, b = 0$ (as in the definition 7) and $a = b = 1/2$, make with the eigenvectors of covariance operator L (θ w/ e.v.'s) and with PGA directions in definition 5 (θ w/ PGA). I also measure the displacement of the intrinsic mean after removing an explanatory direction (μ disp.) and the *average reconstruction error*, i.e., the intrinsic variability remaining in the data after removing explanatory directions (R.E.) under $a = b = 1/2$ and $a = 0, b = 1$.

I repeat the experiment for data given in [25] which was collected to investigate the variability in six subject's movements while completing a drilling task. The data I use is motion capture data of the rotation of the first subject's wrist. This data has small variability which is claimed in [25] is common in human kenetics studies and accordingly in [25] tangent space methods are used for analysis.

		θ w/ e.v.'s	θ w/ PGA	μ disp.	R.E.
$a = b = \frac{1}{2}$	first dir.	0.0844	0.00	0.0109	0.4640
	second dir.	0.1180	0.0021	0.0057	0.2530
$a = 1, b = 0$	first dir.	0.0844	0.00	0.0196	0.4640
	second dir.	0.6322	0.5880	0.0112	0.2610

Simulated data, 500 runs, mean intrinsic var. = 1.4878

		θ w/ e.v.'s	θ w/ PGA	μ disp.	R.E.
$a = b = \frac{1}{2}$	first dir.	1.308e-3	0.00	110.418e-6	43.853e-3
	second dir.	2.01e-3	27.997e-6	42.035e-6	19.926e-3
$a = 1, b = 0$	first dir.	1.308e-3	0.00	437.222e-6	43.853e-3
	second dir.	63.526e-3	61.940e-3	189.840e-6	20.247e-3

Wrist rotation data, intrinsic var. = 0.2912

Table 2: Comparisons of alt-PGA

The tables show better agreement with the eigenvectors of L and variability as identified by PGA for $a = b = 1/2$. Also, there is less displacement of the mean for $a = b = 1/2$ in agreement with the expansions above. There is also slightly less reconstruction error for $a = b = 1/2$ in these experiments. These effects are greater in magnitude for the simulated data which has greater tangent space variability.

6. Linear Difference Indicators

In [27] the difference between exact solutions to PGA and tangent space approximations was explored. To this end the authors of [27] introduced measures of the accuracy of approximations of the projection operator and of approximations of explanatory directions obtained by orthogonal projection and PCA in the tangent space, respectively. Given $D = \{p_1, \ldots, p_N\} \subset M$ with $\mu(D) = \mu$, PGA directions $V_{k-1} = \{v_1, \ldots, v_{k-1}\}$ and $v \in SV_{k-1}^{\perp}$ the *average projection* *difference* is formulated as

$$\tau_H = \frac{1}{N}\sum_{i=1}^{N} \mathrm{d}(p_i, \hat{\pi}_{H_v}(p_i))^2 - \mathrm{d}(p_i, \pi_{S_v}(p_i))^2 \tag{46}$$

where $H(v) = \mathrm{Exp}_\mu(\mathrm{span}(V_{k-1} \cup v))$ and $\hat{\pi}_{H_v}(p_i)$ is the exponentiation of the orthogonal projection of $q_i = \mathrm{Log}_\mu(p_i)$ in $\mathrm{T}_\mu M$ to $\mathrm{span}(V_{k-1} \cup v)$.

Then setting $v = v_k$ where v_k is the k-th PGA direction and letting $\hat{v}$ be its approximation obtained in $\mathrm{T}_\mu M$ the *average residual difference* is formulated as

$$\rho = \frac{1}{N}\sum_{i=1}^{N} \mathrm{d}(p_i, \pi_{H_{\hat{v}}}(p_i))^2 - \mathrm{d}(p_i, \pi_{H_v}(p_i))^2 \tag{47}$$

In [27] *difference indicators* which are shown to be correlated to these statistics and which can be computed before carrying out exact PGA are proposed. The indicator for the average projection difference is given as

$$\tau_H \approx \tilde{\tau}_{H(v)} = \frac{2}{N}\sum_{i=1}^{N} \left\| \nabla_{\hat{\pi}_{H(v)}(p_i)} f \right\| \tag{48}$$

where $f(y) = \mathrm{d}(p_i, y)^2$. Using (42), $\left\| \nabla_{\hat{\pi}_{H(v)}(p_i)} f \right\|$ can be computed as the magnitude of the component of $-2\mathrm{Log}_{\hat{\pi}_{H_v}(p_i)}(p_i)$ in $\mathrm{T}_{\hat{\pi}_{H(v)}(p_j)} M$.

The indicator for the average residual difference is given as the standard deviation of the difference between tangent space distances to $\hat{v}$ and manifold distances to $H(\hat{v})$. That is,

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \left(\left\| q_i - \mathrm{Log}_\mu(\hat{\pi}_{H(\hat{v})}(p_i)) \right\| - \mathrm{d}(p_i, \hat{\pi}_{H(\hat{v})}(p_i)) - m \right)^2}$$

where m is the mean of the differences between the distances.

6.1. Expansions of τ_H and ρ

We can obtain expansions in scaling parameter ϵ of τ_H and ρ. These will provide good approximations of these two statistics. Note that although $\tilde{\tau}_{H(v)}$ and σ show correlation only $\tilde{\tau}_{H(v)}$ is an approximation. These expansions will also give a better understanding of how scale and curvature terms affect τ_H and ρ which then can be used to make type of decisions about the use of linear approximations demonstrated in [27]. We will obtain these expansions for $k = 1$ which will allow us to carry out the type of experiments done in [27]. Expansions for $k > 1$ can be obtained in a similar manner.

Let $p_{i,\epsilon} = \mathrm{Exp}(\epsilon q_i)$ and $t_i(\epsilon) = t_{i,1}\epsilon + t_{i,3}\epsilon^3 + O(\epsilon^5)$ be the expansion of the projection coefficient to a geodesic as section 4.3. Using $t_{i,1} = \langle q_i, v\rangle$ and (21) in (46), in S^n we have

$$\begin{aligned}\tau_H &= \frac{1}{N}\sum_{i=1}^{N} \arccos(\cos(t_{1,i}v)\cos(\epsilon\,\|q_i\|) + \sin(t_1 v)\sin(\epsilon\,\|q\|)t_{i,1})^2 \\ &\qquad - \arccos(\cos(t_i(\epsilon))\cos(\epsilon\,\|q_i\|) + \sin(t_i(\epsilon)v)\sin(\epsilon\,\|q\|)t_{i,1})^2 \\ &= \frac{1}{N}\sum_{i=1}^{N}(t_{i,3})^2\epsilon^6 + O(\epsilon^8)\end{aligned}$$

In $P(n)$ or $SO(n)$ we use the expansion of the objective function in (30) and of $t_i(\epsilon)$ given in (31) to also obtain

$$\begin{aligned}\tau_H &= \frac{1}{N}\sum_{i=1}^{N} h_i(t_{i,1},\epsilon) - h_i(t_i(\epsilon),\epsilon) \\ &= \frac{1}{N}\sum_{i=1}^{N}(t_{i,3})^2\epsilon^6 + O(\epsilon^8)\end{aligned}$$

Consider the cost function $f_1(v,\epsilon)$ in (6) and the expansion $v_1(\epsilon) = v_{1,0} + v_{1,2}\epsilon^2 + v_{1,4}\epsilon^4 + O(\epsilon^6)$. We have

$$\rho = f_1(v_{1,0},\epsilon) - f_1(v_1(\epsilon),\epsilon) \tag{49}$$

Given the constraint $\|v_1(\epsilon)\| = 1$ and that for $v_1^a(\epsilon) = \sum_{j=0}^{a} v_{2j}$, $\|v_1^a(\epsilon)\|$ is not necessarily 1 for any a, consider the expansion

$$\frac{1}{\|v_1^a\epsilon\|} = \frac{1}{\sqrt{1+\langle v_{1,2},v_{1,2}\rangle\,\epsilon^4 + O(\epsilon^8)}} = 1 - \frac{\langle v_{1,2},v_{1,2}\rangle}{2}\epsilon^4 + O(\epsilon^8)$$

and set

$$\begin{aligned}\tilde{v}_1^a(\epsilon) &= v_1^a(\epsilon)/\,\|v_a\epsilon\| \\ &= v_{1,0} + v_{1,2}\epsilon^2 + \left(v_{1,4} - v_{1,0}\frac{\langle v_{1,2},v_{1,2}\rangle}{2}\right)\epsilon^4 + O(\epsilon^6)\end{aligned}$$

Then we have $v_1(\epsilon) = \lim_{a\to\infty}\tilde{v}_1^a(\epsilon)$ and as in (7) we have $(f_{1,2}(v_0) - f_{1,2}(v_1(\epsilon)))\epsilon^2 =$

$$\begin{aligned}&= \left(\frac{1}{N}\sum_{i=1}^{n}\langle q_i, v_1(\epsilon)\rangle^2 - \langle q_i, v_{1,0}\rangle^2\right)\epsilon^2 \\ &= \frac{1}{N}\sum_{i=1}^{N}\Big(\langle q_i, v_{1,2}\rangle^2 - \langle q_i, v_{1,0}\rangle^2\langle v_{1,2},v_{1,2}\rangle\Big)\epsilon^6 + O(\epsilon^8)\end{aligned} \tag{50}$$

with the second equality using

$$\frac{1}{N}\sum_{i=1}^{N}\langle q_i, v_{1,0}\rangle\langle q_i, v_{1,k}\rangle = \langle \mathrm{L}(v_{1,0}), v_{1,k}\rangle = \beta_1\langle u_1, v_{1,k}\rangle = 0$$

for $k = 2, 4$.

Using (24), in S^n we have $(f_{1,4}(v_{1,0}) - f_{1,4}(v_1(\epsilon)))\epsilon^4 =$

$$\begin{aligned}
&= \frac{1}{3N}\sum_{i=1}^{N}\left(\langle q_i, v_{1,0}\rangle^2\right)\left(\langle q_i, v_{1,0}\rangle^2 - \langle q_i, q_i\rangle\right) \\
&\qquad - \left(\langle q_i, v_1(\epsilon)\rangle^2\right)\left(\langle q_i, v_1(\epsilon)\rangle^2 - \langle q_i, q_i\rangle\right)\epsilon^4 \\
&= \frac{1}{3N}\sum_{j=1}^{N} 2\langle q_i, v_{1,0}\rangle\langle q_i, v_{1,2}\rangle\langle q_i, q_i\rangle - 4\langle q_i, v_{1,0}\rangle^3\langle q_i, v_{1,2}\rangle\epsilon^6 + O(\epsilon^8)
\end{aligned} \tag{51}$$

Using (35), in $P(n)$ we have $(f_{1,4}(v_{1,0}) - f_{1,4}(v_1(\epsilon)))\epsilon^4 =$

$$\begin{aligned}
&= \frac{1}{48N}\sum_{i=1}^{N}\Big(tr\,(q_i v_{1,0})^2\left(tr\left(q_i^2 v_{1,0}^2\right) - tr\,(q_i v_{1,0} q_i v_{1,0})\right) \\
&\qquad - tr\,(q_i v_1(\epsilon))^2\left(tr\left(q_i^2 v_1(\epsilon)^2\right) - tr\,(q_i v_1(\epsilon) q_i v_1(\epsilon))\right)\Big)\epsilon^4 \\
&= \frac{1}{48N}\sum_{i=1}^{N}\Big(tr\,(q_i v_{1,0})^2\, tr\,(2 q_i v_{1,0} q_i v_{1,2} - q_i q_i (v_{1,0} v_{1,2} + v_{1,2} v_{1,0})) \\
&+ 2tr\,(q_i v_{1,0})\, tr\,(q_i v_{1,2})\, tr\,(q_i v_{1,0} q_i v_{1,0} - q_i q_i v_{1,0} v_{1,0})\Big)\epsilon^6 + O(\epsilon^8)
\end{aligned} \tag{52}$$

In $SO(n)$ we just take the negative of this expression.

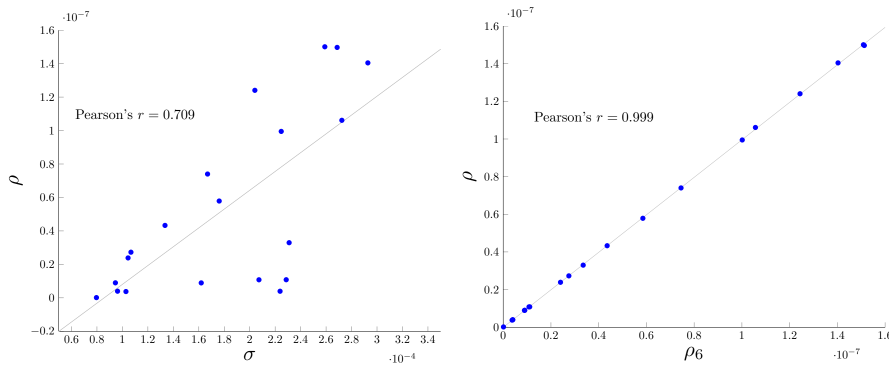
Using (49) the expansion of ρ in S^n is the sum of (50) and (51) and the expansion of ρ in $P(n)$ or $SO(n)$ is the sum of (50) and (52).

6.2. Experiments comparing difference indicators and expansions

We apply the difference indicators and the expansions to two data sets in experiments similar to the ones in [27]. The expansion of ρ is given as $\rho = \rho_6\epsilon^6 + O(\epsilon^8)$ and the expansion of τ_H as $\tau_H = \tau_{H,6}\epsilon^6 + O(\epsilon^8)$. The first data set is the wrist rotation data set in $SO(3)$ from section 5.3. The second is a synthetic data set in $P(n)$ with a sample of 36 from a distribution as in the test in figure 5. As in [27] for each experiment we draw a random sample of 8 from the data set 20 times and compute the relevant statistics each time for comparison. For the wrist rotation data in $SO(3)$ τ_H is not computed as projection has closed form in this case.

Experiment 1. Wrist Rotation Data

As indicated in section 5.3 this data set has little variability. Accordingly, ρ_6 in figure 7 is a very close estimate of ρ and provides a nearly perfect picture of the penalty (very small in this case) of using the first-order approximation of PGA. At the same time σ has a lower correlation with ρ and is of a much different scale and is thus of less value in assessing the use of a first-order approximation of PGA.



(a) Difference Indicator, σ

(b) Expansion of ρ

Figure 7: ρ for wrist rotation

Experiment 2. Synthetic Data

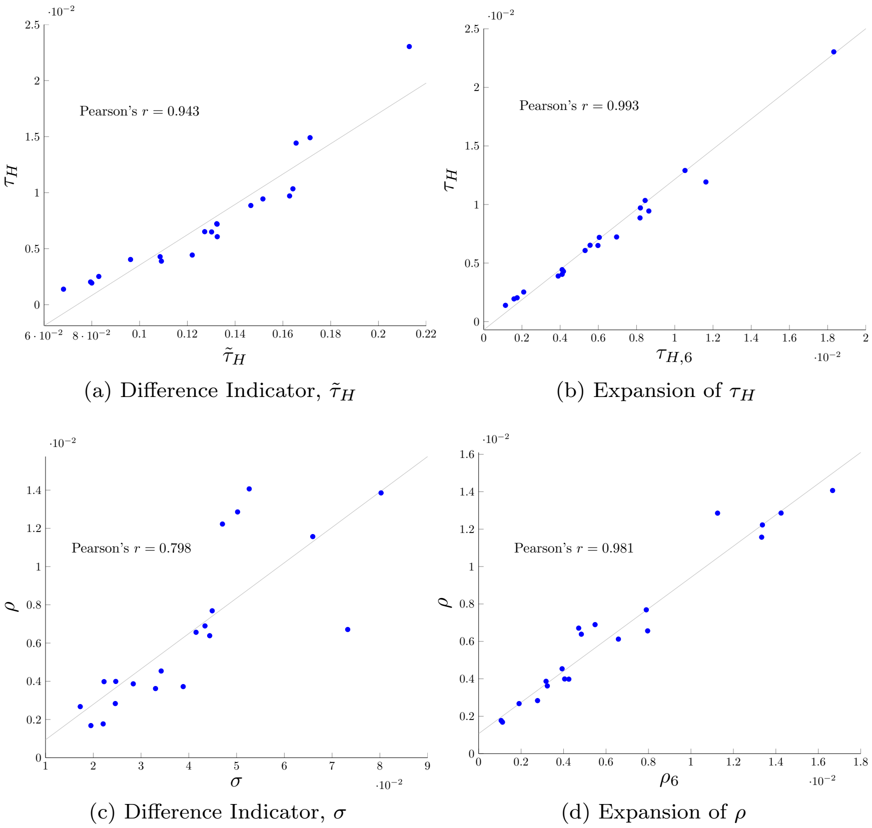
As in [27] the first PGA direction v_1 is set to v in (46) and used to compute τ_H. In figure 8 both $\tau_{H,6}$ and ρ_6 provide good estimates of τ_H and ρ with correlation coefficents of .993 and .981, respectively. Also to note, one might *square* the norm of the gradient in (48) to use the Pythagorean theorem in each $\mathrm{T}_{\hat{\pi}_{H(v)}(p_i)}M$ to get another improved estimate of τ_H over $\tilde{\tau}_H$.

7. Discussion and Conclusions

In [30] PGA is formulated as a probability model (PPGA) in which data is distributed according to a manifold generalization of the normal distribution. Explanatory directions are included as parameters to be estimated by maximum likelihood. Also, a location parameter and scaling parameters for the explanatory directions and for the variability or dispersion of the data are fit. Thus, as an advantage, not only are the the mean and explanatory directions jointly estimated but the dispersion of the data is also taken into account.

In the descriptive setting, in this paper, consideration of a dispersion or scaling factor was shown to be an essential element in revealing the underlying structure of solutions to PGA. In the proposition in section 2.1, for example, we see how the share of variability in the tangent space, accounted for by eigenvector u_k and measured by eigenvalue β_k, weights the curvature terms in $v_{k,2}$ to determine the difference, at least locally, between the first-order approximation and the exact solution. Note that as in (33) in section 4.3 all expansions in this paper can be expressed and interpreted strictly in geometric terms of angles, sectional curvatures and Riemannian curvature tensors.

Also, we see in this paper, at least experimentally, that the approximations obtained by expansion hold for data significantly dispersed from the tangent space. In figure 3, for example, with q, v uniformly distributed in $\mathrm{S_I}P(n)$ the



(a) Difference Indicator, $\tilde{\tau}_H$

(b) Expansion of τ_H

(c) Difference Indicator, σ

(d) Expansion of ρ

Figure 8: τ_H, ρ for data in $P(3)$

third- and fifth-order approximations in ϵ hold for $\epsilon > 2$. In this case, one can readily obtain bounds on the ratios of the coefficients in the expansion and derive expected values of those ratios to explain this plot. Such an approach should be able to be taken with other expansions and distributions of data as well. Also, in a bounded manifold the variability of data and thus ϵ is restricted. In S^n or $SO(n)$, for example, we take $\epsilon < \pi / \|q_i\|$ for data $p_{i,\epsilon} = \mathrm{Exp}_\mu(\epsilon q_i)$ so that the expansions only need to hold for these values of ϵ.

Since PGA was introduced in [11] and then [10] a number of other methods to analyze the variability of manifold valued data have been proposed. For example, [18] accounts for non-geodesic variability in spheres and [17] projects to geodesics that intersect orthogonally at a mean on the first geodesic that best fits the data. One could use the approach of this paper in these contexts, for example, an expansion of the difference between the mean located in [17] and the intrinsic mean might be obtained.

In addition, this paper and others use a definition of PGA that minimizes residual error while PGA was defined in [11] as maximizing projected variability. Expansions of solutions of PGA using the latter definition might be obtained and the higher-order terms compared to determine what accounts for the differences and how these differences might be taken into account in deciding which definition to employ. Also, the approach of this paper might be used to quantify differences between other generalizations of linear statistics such as *intrinsic MANOVA* in [15] and *geodesic regression* in [9] and their local, linear approximations.

[1] *Effective Sampling and Distance Metrics for 3D Rigid Body Path Planning*, IEEE, May 2004.

[2] Arindam Banerjee, Inderjit S. Dhillon, Joydeep Ghosh, and Suvrit Sra, *Clustering on the unit hypersphere using von Mises-Fisher distributions*, J. Mach. Learn. Res. **6** (2005), 1345–1382. MR 2249858

[3] P.J. Basser, J. Mattiello, and D. LeBihan, *{MR} diffusion tensor spectroscopy and imaging*, Biophysical Journal **66** (1994), no. 1, 259 – 267.

[4] Martin Bridson and André Haefliger, *Metric spaces of non-positive curvature*, Grundlehren der mathematischen Wissenschaften, 319, pp. 315–316, Springer, 1999.

[5] R. W. R. Darling, *Differential forms and connections*, Cambridge University Press, Cambridge, 1994. MR 1312606 (95j:53038)

[6] Manfredo P. do Carmo, *Differential geometry of curves and surfaces*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1976, Translated from the Portuguese. MR 0394451 (52 #15253)

[7] P. Jost-Hinrich Eschenburg, *Lecture notes on symmetric spaces*, January 1997, http://myweb.rz.uni-augsburg.de/ eschenbu/symspace.pdf.

[8] P. Thomas Fletcher, *Lecture notes on symmetric spaces*, January 2009, http://www.sci.utah.edu/fletcher/RiemannianGeometryNotes.pdf.

[9] ______, *Geodesic regression and the theory of least squares on Riemannian manifolds*, Int. J. Comput. Vis. **105** (2013), no. 2, 171–185. MR 3104017

[10] 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).

[11] 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.

[12] Jean Gallier and Dianna Xu, *Computing exponentials of skew symmetric matrices and logarithms of orthogonal matrices*, International Journal of Robotics and Automation **18** (2000), 10–20.

[13] Berthold K. P. Horn, *Relative orientation*, International Journal of Computer Vision (1990), 59–78.

[14] Kurt Hornik and Bettina Grn, *movmf: An r package for fitting mixtures of von mises-fisher distributions*, Journal of Statistical Software **58** (2014), no. 10, 1–26.

[15] Stephan Huckemann, Thomas Hotz, and Axel Munk, *Intrinsic manova for riemannian manifolds with an application to kendall's space of planar shapes.*, IEEE Trans. Pattern Anal. Mach. Intell. **32** (2010), no. 4, 593–603.

[16] ———, *Intrinsic shape analysis: geodesic PCA for Riemannian manifolds modulo isometric Lie group actions*, Statist. Sinica **20** (2010), no. 1, 1–58. MR 2640651 (2011f:62064)

[17] 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)

[18] Sungkyu Jung, Ian L. Dryden, and J. S. Marron, *Analysis of principal nested spheres*, Biometrika **99** (2012), no. 3, 551–568. MR 2966769

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

[20] Kenneth Lange, *Optimization*, Springer Texts in Statistics, 95, p. 36, Springer, 2013.

[21] Kanti V. Mardia and Peter E. Jupp, *Directional statistics*, Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd., Chichester, 2000, Revised reprint of ıt Statistics of directional data by Mardia [MR0336854 (49 #1627)]. MR 1828667 (2003b:62004)

[22] M. G. Monera, A. Montesinos-Amilibia, and E. Sanabria-Codesal, *The Taylor expansion of the exponential map and geometric applications*, Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Math. RACSAM **108** (2014), no. 2, 881–906. MR 3249982

[23] 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)

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

[25] Denis Rancourt, Louis-Paul Rivest, and Jérôme Asselin, *Using orientation statistics to investigate variations in human kinematics*, J. Roy. Statist. Soc. Ser. C **49** (2000), no. 1, 81–94. MR 1817876

[26] Salem Said, Nicolas Courty, Nicolas Le Bihan, and Stephen J. Sangwine, *Sangwine s.: Exact principal geodesic analysis for data on so(3*, InProc.ofthe 15th European Signal Processing Conference (EUSIPCO 2007, 2007.

[27] 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.

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

[29] Bryan Stanfill, Heike Hofmann, and Ulrike Genschel, *rotations: An r package for so(3) data*, The R Journal **6** (2014), 68–78.

[30] Miaomiao Zhang and P. Thomas Fletcher, *Probabilistic principal geodesic analysis.*, NIPS (Christopher J. C. Burges, Lon Bottou, Zoubin Ghahramani, and Kilian Q. Weinberger, eds.), 2013, pp. 1178–1186.