PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 00, Number 0, Pages 000–000
S 0002-9939(XX)0000-0

THE MAX-LENGTH-VECTOR LINE OF BEST FIT TO A COLLECTION OF VECTOR SUBSPACES

DANIEL J. BATES, BRENT R. DAVIS, MICHAEL KIRBY, JUSTIN MARKS, AND CHRIS PETERSON

(Communicated by)

ABSTRACT. Let $\mathcal{C} = \{V_1, V_2, \ldots, V_k\}$ be a finite collection of nontrivial subspaces of a finite dimensional real vector space V. Let L denote a one dimensional subspace of V and let $\theta(L, V_i)$ denote the principal (or canonical) angle between L and V_i. We are interested in finding all lines that maximize the function $F(L) = \sum_{i=1}^{k} \cos\theta(L, V_i)$. Conceptually, this is the line through the origin that best represents $\mathcal{C}$ with respect to the criterion $F(L)$. A reformulation shows that L is spanned by a vector $v = \sum_{i=1}^{k} v_i$ which maximizes the function $G(v_1, \ldots, v_k) = ||\sum_{i=1}^{k} v_i||^2$ subject to the constraints $v_i \in V_i$ and $||v_i|| = 1$. Using Lagrange multipliers, the critical points of G are solutions of a polynomial system corresponding to a multivariate eigenvector problem. We use homotopy continuation and numerical algebraic geometry to solve the system and obtain the max-length-vector line(s) of best fit to $\mathcal{C}$.

1. INTRODUCTION AND MOTIVATION

Established geometric theory has inspired the development of algorithms for the purpose of understanding structure in large data sets [24, 8, 29, 7, 12, 15, 21, 28, 1]. Through methods such as the Singular Value Decomposition, one can model or capture features of a given data set with a subspace or with an ordered set of orthonormal vectors. Since Grassmann (resp. Stiefel) manifolds parametrize subspaces (resp. ordered sets of orthonormal vectors) of a given size, aspects of a data set can be captured with a single point on such a manifold. This has led to the consideration of algorithms on Grassmann and Stiefel manifolds as tools for the purposes of representation, classification, and comparison of data sets [17, 10, 25, 19, 11]. Collections of data sets lead naturally to collections of points. Given a collection/cluster of points on a manifold, procedures have been developed to determine a single point on the manifold which best represents the cluster with respect to various optimization criteria [6, 2, 20, 18, 27, 16]. Cluster representatives can then be used to reduce the cost of classification algorithms or to aid in hierarchical clustering tasks. In this context, we became interested in the following problem: *Given a finite collection, $\mathcal{C}$, of subspaces of a vector space V, find a line in V that best represents $\mathcal{C}$.*

Received by the editors October 10, 2013.
2010 *Mathematics Subject Classification.* 15A18, 65KXX, 65H20, 13PXX, 14QXX, 65FXX.
Key words and phrases. Numerical algebraic geometry, Homotopy continuation, Nonlinear optimization, Principal angles, Grassmannian.

Let $Gr(p,n)$ denote the Grassmann manifold whose points parametrize the set of all p-dimensional subspaces of $\mathbb{R}^n$. One can represent a given p-dimensional subspace of $\mathbb{R}^n$ as the column space of a full rank $n \times p$ matrix, M. Note that M and MA represent the same point on $Gr(p,n)$ for any $A \in GL(p)$. Letting $[M]$ denote the equivalence class of $n \times p$ matrices that have the same column space as M, points on $Gr(p,n)$ can be identified with equivalence classes of full rank $n \times p$ matrices. One could also represent a given p-dimensional subspace of $\mathbb{R}^n$ as the column space of an $n \times p$ orthonormal matrix, M, with the caveat that M and MA represent the same point on $Gr(p,n)$ for any $A \in O(p)$. For the purpose of computation, we typically use orthonormal matrix representatives for points on a Grassmannian. More precisely, an $n \times p$ orthonormal matrix Y will be used as a representative for the p-dimensional subspace $[Y]$ spanned by the columns of Y. An elementary computation shows that $Gr(p,n)$ has dimension $p(n-p)$ as a real manifold.

Given a finite cluster of points on $Gr(p,n)$, corresponding to a collection of p-dimensional subspaces of $\mathbb{R}^n$, a commonly used cluster representative is the Karcher mean, $[\mu_{KM}] \in Gr(p,n)$. Like all orthogonally invariant processes on a Grassmannian, the Karcher mean can be expressed in terms of *principal angles* between spaces. Given two subspaces $[X]$ and $[Y]$ of $\mathbb{R}^n$, of possibly different dimensions, there are $p = \min(\dim[X], \dim[Y])$ principal angles $0 \leq \theta_1(X,Y) \leq \theta_2(X,Y) \leq \ldots \leq \theta_p(X,Y) \leq \frac{\pi}{2}$ between $[X]$ and $[Y]$. If X and Y are orthonormal matrix representatives for $[X]$ and $[Y]$, then the principal angles between $[X]$ and $[Y]$ may be computed as the inverse cosine of the singular values of X^TY [9]. In the particular setting where $\dim([X_i]) = p$ for all i, $[\mu_{KM}]$ is identified as the p-dimensional subspace of $\mathbb{R}^n$ given by

$$[\mu_{KM}] = \operatorname*{argmin}_{[\mu]} \sum_{i=1}^{k} d([\mu],[X_i])^2,$$

where $d([X],[Y])^2 = \sum_{i=1}^{p} \theta_i(X,Y)^2$. The Karcher mean can be computed using an iterative procedure, involving local Exp and Log maps, which is guaranteed to converge for clusters of points located within a certain convergence radius [6].

In the following section, we describe a line representative for a collection of points lying on a disjoint union of Grassmannians. In other words, we describe this line representative for a collection of subspaces $\mathcal{C} = \{V_1, V_2, \ldots, V_k\}$ of a fixed vector space V with potentially differing dimensions. Due to a geometric interpretation of the construction, we call this representative the *max-length-vector line of best fit to* $\mathcal{C}$. While the Karcher mean is a solution to an optimization problem involving squares of principal angles, the max-length-vector line of best fit is based on an optimization problem that depends upon cosines of principal angles.

Section 3 describes the conversion of this optimization problem into a polynomial system, which can then be solved via numerical algebraic geometry (§4). In particular, we make use of a generalization of the parameter homotopy method of [13].

2. Formulations of the Optimization Problem

In this section, we provide several equivalent formulations of the optimization problem that leads to the max-length-vector line of best fit. The name used to describe this line is a consequence of a geometric interpretation given in §2.2.

2.1. **Optimization Problem Formulation.** Let $\mathcal{C} = \{V_1, V_2, \ldots, V_k\}$ be a finite collection of nontrivial subspaces of $\mathbb{R}^n$. Let d_i denote the dimension of V_i and let Y_i be an $n \times d_i$ orthonormal matrix whose column space is V_i (so we identify V_i with $[Y_i]$). Let L denote a one dimensional subspace of $\mathbb{R}^n$ and let $\theta(L, V_i)$ denote the principal angle between L and V_i. We wish to find a one dimensional subspace L_{MLV} via the following optimization problem

$$\begin{aligned} L_{MLV} = \underset{L}{\operatorname{argmax}} \quad & \sum_{i=1}^{k} \cos\theta(L, V_i) \\ \text{subject to} \quad & L \in \mathbb{R}^n \text{ is a one-dimensional vector space} \end{aligned} \tag{2.1}$$

Note that maximizing the function $F(L) = \sum_{i=1}^{k} \cos\theta(L, V_i)$ is equivalent to minimizing the function $G(L) = \sum_{i=1}^{k} \sin\theta(L, V_i)$ (see [14] for a striking use of a similar function in packing problems for linear subspaces).

If L is the span of a unit length vector ℓ, then $\cos\theta(L, V_i)$ is the singular value of $\ell^T Y_i$ [9]. Noting that the singular value of $\ell^T Y_i$ is simply the length of the projection of ℓ onto V_i, we see that finding a line L that maximizes the function $\sum_{i=1}^{k} \cos\theta(L, V_i)$ is equivalent to finding a unit length vector, ℓ, that maximizes the function $\sum_{i=1}^{k} ||\text{proj}_{V_i}\ell||$. Recall that $\text{proj}_{V_i}\ell$ is a vector which makes the smallest possible angle with ℓ subject to the constraint of lying in V_i. Since $||\text{proj}_{V_i}\ell|| = \cos\theta(L, V_i) = \ell^T v_i$ for some vector $v_i \in V_i$ of unit length, we conclude that (2.1) is equivalent to the optimization problem

$$\begin{aligned} \max_{\ell, v_i} \quad & \sum_{i=1}^{k} \ell^T v_i \\ \text{subject to} \quad & \ell^T \ell = 1, v_i \in V_i, \text{ and } v_i^T v_i = 1 \text{ for } 1 \le i \le k \end{aligned} \tag{2.2}$$

Let Y_i be an orthonormal matrix such that $[Y_i] = V_i$. Note that $v_i \in [Y_i]$ implies that $v_i = Y_i \alpha_i$ for some $\alpha_i \in \mathbb{R}^{d_i}$. If Y_i is an orthonormal matrix, then $1 = v_i^T v_i = \alpha_i^T Y_i^T Y_i \alpha_i = \alpha_i^T \alpha_i$, so we can reformulate (2.2) as

$$\begin{aligned} \max_{\ell, \alpha_i} \quad & \sum_{i=1}^{k} \ell^T Y_i \alpha_i \\ \text{subject to} \quad & \ell^T \ell = 1 \text{ and } \alpha_i^T \alpha_i = 1 \text{ for } 1 \le i \le k. \end{aligned}$$

Factoring out ℓ^T, we arrive at

$$\begin{aligned} \max_{\ell, \alpha_i} \quad & \ell^T \sum_{i=1}^{k} Y_i \alpha_i \\ \text{subject to} \quad & \ell^T \ell = 1 \text{ and } \alpha_i^T \alpha_i = 1 \text{ for } 1 \le i \le k. \end{aligned} \tag{2.3}$$

Note that ℓ and α_i can be computed independently. Let $v = \sum_{i=1}^{k} Y_i \alpha_i$. Since $\|\ell\| = 1$, and $\ell^T v = \|\ell\| \|v\| \cos\phi$ (where ϕ is the angle between ℓ and v), the optimal solution can be found by first determining the set of α_i that maximize the length of v and then choosing $\ell = v/\|v\|$ (i.e., choose ℓ to point in the direction of v). If we note that the problem of maximizing $\|v\|$ is the same as the problem of maximizing $\|v\|^2$, we can determine ℓ by finding the α_i that optimize

$$\begin{aligned} \max_{\alpha_i} \quad & \left\| \sum_{i=1}^{k} Y_i \alpha_i \right\|^2 \\ \text{subject to} \quad & \alpha_i^T \alpha_i = 1 \text{ for } 1 \leq i \leq k. \end{aligned} \tag{2.4}$$

then setting $v = \sum_{i=1}^{k} Y_i \alpha_i$ and setting $\ell = v/\|v\|$.

2.2. **Geometric interpretation of reformulation.** The vector $v = \sum_{i=1}^{k} Y_i \alpha_i$ described above has a nice geometric interpretation. The vector $Y_i \alpha_i$, in the linear combination, lies in the subspace $[Y_i]$. The constraint $\alpha_i^T \alpha_i = 1$ restricts the vector $Y_i \alpha_i$ to have length 1. Thus the vector v is the longest vector that can be obtained by adding together unit length vectors $v_1, v_2, \ldots, v_k$ with $v_i \in V_i$. Since L is the span of v, we use the name *max-length-vector line of best fit* for the one dimensional subspace that maximizes the function $F(L) = \sum_{i=1}^{k} \cos\theta(L, V_i)$.

3. Computing the max-length-vector line of best fit

In this section, we return to the optimization problem (2.4). We provide conditions for optimality and describe the solving method. In addition, a remark is made about a related optimization problem.

First notice that the feasible solutions to (2.4) form a compact set. The objective function is continuous, so the optimal solution to (2.4) will be achieved. The local (and global) optimal solutions to (2.4) can be found as solutions of the polynomial system resulting from the use of Lagrange multipliers. Let $\alpha_i = \alpha_{i,1}, \ldots, \alpha_{i,d_i}$ denote the variables in the α_i and let $\alpha^T = (\alpha_1^T, \ldots, \alpha_k^T)$ be a row vector. Employing

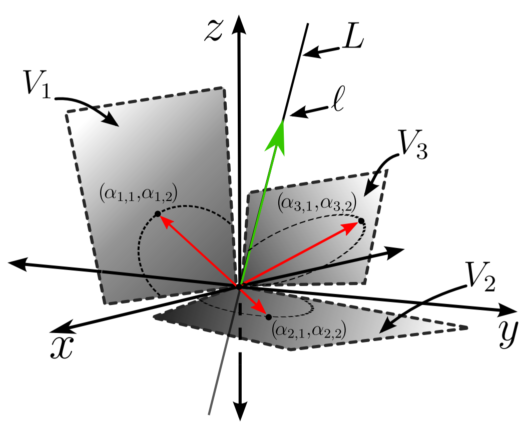

Figure 1. A geometric interpretation of the max-length-vector: Red vectors can rotate in their respective subspaces V_1, V_2, V_3. The length of the sum of these vectors is maximized.

the general formulation as in [3], solutions to (2.4) satisfy

$$\begin{aligned} \nabla_\alpha \left\| \sum_{i=1}^{k} Y_i \alpha_i \right\|^2 + \sum_{i=1}^{k} \lambda_i \nabla_\alpha (\alpha_i^T \alpha_i - 1) = 0 \\ \alpha_i^T \alpha_i - 1 = 0 \text{ for } 1 \leq i \leq k. \end{aligned} \tag{3.1}$$

Define the block matrix $\mathbf{Y} = (Y_1|Y_2|\cdots|Y_k)$. Since the Y_i are orthonormal,

$$\mathbf{Y}^T\mathbf{Y} = \begin{pmatrix} I_1 & Y_1^T Y_2 & \cdots & Y_1^T Y_k \\ Y_2^T Y_1 & I_2 & \cdots & Y_2^T Y_k \\ \vdots & \vdots & \ddots & \vdots \\ Y_k^T Y_1 & Y_k^T Y_2 & \dots & I_k \end{pmatrix}.$$

Using straightforward algebraic manipulation, the system of polynomial equations in (3.1) can be written compactly as

$$\begin{aligned} \left(\mathbf{Y}^T\mathbf{Y} + \text{diag}(\lambda_1^{d_1}, \lambda_2^{d_2}, \dots, \lambda_k^{d_k})\right) \cdot \alpha = 0 \\ \alpha_i^T \alpha_i = 1 \text{ for } 1 \leq i \leq k. \end{aligned} \tag{3.2}$$

The notation is as follows

- I_i is a $d_i \times d_i$ identity matrix where d_i is the dimension of V_i.
- $\text{diag}(\lambda_1^{d_1}, \lambda_2^{d_2} \dots, \lambda_k^{d_k})$ is a diagonal matrix that has variable λ_1 repeated d_1 times, λ_2 repeated d_2 times, etc.
- Y_i is a matrix with orthonormal columns such that $[Y_i] = V_i$.
- α is a $\left(\sum_{i=1}^k d_i\right) \times 1$ column vector where $\alpha^T = (\alpha_1^T, \dots, \alpha_k^T)$

Typically there is a unique max-length-vector line of best fit. However, there are several degenerate cases that lead to multiple or even infinitely many such lines. This can occur when too few conditions are imposed by the subspaces or when there is a lot of symmetry such as when the subspaces are mutually orthogonal. Three such cases are enumerated below:

i) If the V_i correspond to the coordinate axes in $\mathbb{R}^n$ then there are multiple (but finitely many) max-length-vector lines of best fit.
ii) If the V_i all intersect in a common subspace then any line in the common subspace is a max-length-vector line of best fit.
iii) The z-axis and the xy-plane in $\mathbb{R}^3$ determine infinitely many max-length-vector lines of best fit whose union determines a pair of cones meeting at a point.

All of these cases can be handled with the numerical irreducible decomposition, a standard tool of numerical algebraic geometry, but this is not the focus of this paper.

4. Numerical algebraic geometry

The local optimal solutions to (2.4) satisfy a straightforward system of polynomial equations. Numerical algebraic geometry, rooted in homotopy continuation methods for polynomial systems, can be used to find these local optima.

4.1. **Homotopy continuation and witness points.** Given a polynomial system $f(z)$, *homotopy continuation* provides a means to numerically approximate (to any level of accuracy) all isolated solutions of $f(z) = 0$. The process can be described in three steps:

(1) Construct an easily-solved polynomial system $g(z)$ based on the characteristics of $f(z)$, and solve $g(z) = 0$.
(2) Construct an appropriate homotopy function $H(z,t) = f(z)\cdot(1-t)+g(z)\cdot t$, so that $H(z,1) = g(z)$ (for which we know the solutions) and $H(z,0) = f(z)$ (for which we want the solutions).
(3) The solutions at $t = 1$ and $t = 0$ are connected by *homotopy paths* $z_i(t)$. Track each of these paths, starting from the solutions of $g(z) = 0$, using numerical predictor-corrector methods. The ends of those paths that do not diverge are the desired solutions of $f(z) = 0$.

Figure 2 provides a schematic depiction of this procedure, with paths beginning at the right and ending at the left. Through continuation, we can numerically solve polynomial systems of moderate size.

Homotopy continuation methods have been extended in many ways, one of which is a method for the computation of the positive-dimensional solution sets of a polynomial system. This computation uses hyperplane sections to reduce to the zero-dimensional case, leading to a cascade of homotopies whose output is the *numerical irreducible decomposition*. This decomposition includes *witness points*, i.e. numerical approximations to generic points, on each such irreducible component of the solution set of $g(z)$, along with the dimension and the degree of each component. Free, open-source software has been developed to reliably find all isolated solutions and compute the numerical irreducible decomposition of a polynomial system of equations [4, 30]. The general theory of numerically solving polynomial systems of equations using homotopy continuation is an important component of numerical algebraic geometry. Details for the methods used in this paper can be found in [26, 5].

4.2. **Parameter homotopy.** Equation 3.2 is a *multivariable eigenvalue problem*. An efficient method for solving the multivariable eigenvalue problem, based on

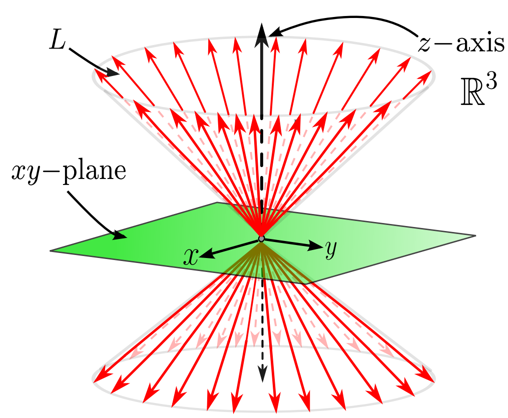

FIGURE 2. Infinitely many equivalent lines that form a pair of cones

homotopy continuation, is discussed in [13]. This is done by placing (3.2) in a parametrized system and deforming from a system with known solutions to the system given by (3.2). It is important that a general element in the system has the same behavior and the same number of solutions as the start system and that the start system is sufficiently general. An application of Bertini's theorem then applies to show that no singularities are encountered on the deformation path. In order to carry out this deformation, we use the software package *Bertini*. The output will be a numerical approximation of each solution to (3.2). Using the values of the α_i obtained from these solutions, we can build the vectors involved in (2.4) and can then determine which one attains the maximum length.

A start system, with easily determined solutions, that can be used in a parametrized system to solve the multivariable eigenvalue problem is

$$\begin{aligned} \left(\mathrm{diag}(z_1, z_2, \ldots, z_N) + \mathrm{diag}(\lambda_1^{d_1}, \lambda_2^{d_2}, \ldots, \lambda_k^{d_k})\right) \cdot \alpha &= 0 \\ \alpha_i^T \alpha_i = 1 \text{ for } 1 \le i \le k. \end{aligned} \tag{4.1}$$

where $N = \sum_{i=1}^{k} d_i$ and the z_i are random complex numbers. The system decomposes into blocks of equations of the form

$$\begin{aligned} \left(\mathrm{diag}(z_1, z_2, \ldots, z_{d_i}) + \mathrm{diag}(\lambda_i^{d_i})\right) \cdot \alpha_i &= 0 \\ \alpha_i^T \alpha_i &= 1 \end{aligned} \tag{4.2}$$

There are $2d_i$ solutions to (4.2). In the form $(\lambda_i, \alpha_{i,1}, \ldots, \alpha_{i,d_i})$, these solutions are $(-z_1, \pm 1, 0, \ldots, 0), (-z_2, 0, \pm 1, 0, \ldots, 0), \ldots, (-z_{d_i}, 0, \ldots, 0, \pm 1)$. There are $\prod_{i=1}^{k} 2d_i$ solutions to (4.1) that correspond to all of the different ways in which the solutions to the k different blocks of equations, of the form (4.2), can be combined. Let $G(z)$ correspond to the system given in (4.1) and let $F(z)$ correspond to the system given in (3.2). The parameter homotopy proceeds by constructing the homotopy function $H(z,t) = F(z) \cdot (1-t) + G(z) \cdot t$, so that $H(z,1) = G(z)$ (for which we know the

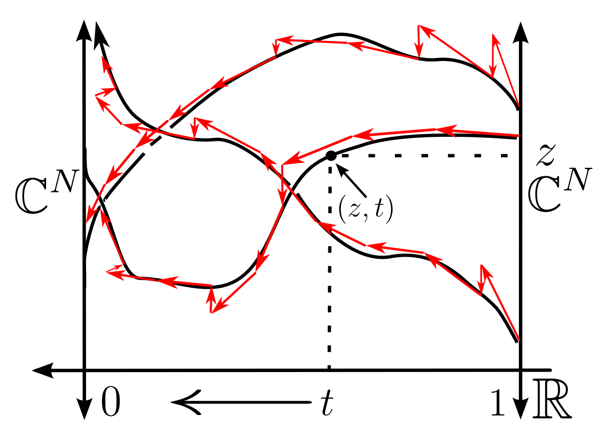

FIGURE 3. Schematic of homotopy paths with predictor-corrector steps. Please note that this is a vastly oversimplified depiction of the technique for illustrative purposes only, excluding essential details such as adaptive stepsize and precision, endgames, and more.

solutions) and $H(z, 0) = F(z)$ (for which we want the solutions). The solutions at $t = 1$ and $t = 0$ are connected by homotopy paths and we can track each of these paths, starting from the solutions of $G(z) = 0$, using numerical predictor-corrector methods. The ends of those paths that do not diverge are the desired solutions of $F(z) = 0$. Finally note that the cost function in Equation 2.4 has a natural $\pm$-symmetry. That is, $\|v\|^2 = \|-v\|^2$. This carries over along the homotopy, so we can recover all solutions by tracking only half of the paths.

5. Examples

For a first example, we considered five randomly generated subspaces $Y_1, \ldots, Y_5$ in $\mathbb{R}^{10}$ of dimensions $4, 3, 3, 2, 2$. The example was run using the parallel implementation of *Bertini* v1.3.1 using 18 (2.67 GHz Xeon-5650) compute nodes with the CentOS 6.4 operating system. Using the parameter homotopy routine, *Bertini* tracked 2304 paths in approximately 6 seconds. Among the 2304 paths, 1776 converged to finite isolated solutions, of which 86 were real. Post-processing of the data to compute v and ℓ was done in serial in negligible time.

Note that the complexity of path tracking does not depend on the ambient dimension n. With this in mind, we tried a second example consisting of nine randomly generated subspaces $Y_1, \ldots, Y_9$ in $\mathbb{R}^{100}$ of dimensions $4, 3, 3, 3, 3, 2, 2, 2, 2$. Using 272 compute notes, *Bertini* tracked $1, 327, 104$ paths in approximately 30 minutes. Among these 1.3 million paths, 2542 converged to finite real isolated values of the α_i. Again, computing v and ℓ can be done in negligible time. In this particular example, we found the max-length-vector to have length 4.27.

The length of the max-length-vector for a collection of k mutually orthogonal subspaces is $\sqrt{k}$ while the length of the max-length-vector for k subspaces which contain a common line is k. In general, the length of the max-length-vector for k subspaces is bounded between these two extremes. Thus, the length of the longest vector is a measure of the mutual orthogonality of the collection. Note that low dimension subspaces of a high dimensional space tend to be orthogonal.

6. Conclusions, Limitations, and Future Work

6.1. **Conclusions.** The preceding sections describe how to find a line that best represents a collection of subspaces with respect to a certain optimality criterion. A reformulation led to a characterization of the line as the span of the longest vector that can be obtained by summing a sequence of vectors chosen from a distinguished collection of hyperspheres. The longest vector was realized as a critical point of a length function. The critical points of the length function were shown to satisfy a multivariable eigenvector problem. The multivariable eigenvector problem could be solved efficiently by deforming a special polynomial system and tracking solutions of a homotopy function using numerical predictor-corrector methods. A data file and Bertini script implementing the procedure can be found at the site: `http://www.math.colostate.edu/~bates/preprints/MLV_computation_page.html`. The data file and script are in a form that can be modified to suit the interested reader's needs.

6.2. **Limitations.** Given a collection of subspaces $\mathcal{C} = \{V_1, V_2, \ldots, V_k\}$ of $\mathbb{R}^n$ with dimensions $d_1, d_2, \ldots, d_k$, we end up with a multivariate eigenvector problem that leads to a polynomial system involving $k + \sum_{i=1}^{k} d_i$ variables. When finding solutions to the system, the parameter homotopy described in Section 4 involves tracking $\prod_{i=1}^{k} 2d_i$ paths. In homotopy continuation, path tracking can be done in parallel in the sense that individual paths can be tracked on individual processors. With care, path tracking can be carried out on systems involving on the order of 1,000 variables. We note that systems involving many 1-dimensional subspaces become very easy since the orthogonality constraint becomes $x_{i1}^2 = 1$. Depending on the number of cores, speed, and memory available to the user and the size of the problem being considered, one can get a rough estimate of the feasibility of solving a given problem by estimating the time to track one path, multiplying by the number of paths, and dividing by the number of available cores. Using a parameter homotopy, restricting the start system to track a very small number of paths, and carrying out an explicit computation, one can obtain a reasonable estimate for the typical time to track a path.

6.3. **Future Work.** There are several different ways to extend the ideas in this paper to compute

- Weighted max-length-vector lines of best fit
- k-dimensional subspaces of best fit.
- Flags of best fit

To aid in the understanding of the structure in a collection, knowledge of the expected length of the max-length-vector for a random system with given parameters will be useful. The authors are working on the development of these topics, including algorithms and applications to problems in large scale data analysis. It will be useful to explore the iterative method of Liao and Zhang for computing the solutions to the multivariate eigenvector problem [23].

Acknowledgements. This research was partially supported by NSF awards DMS-1115668, DMS-1228308, DMS-1322508, by DARPA N66001-11-1-4184, and by the AFOSR. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF, DARPA, or AFOSR. The second author wishes to thank Zhaojun Bai for a very productive conversation.

References

[1] P. A. Absil, P. A. Mahony, and R. Sepulchre, *Riemannian geometry of Grassmann manifolds with a view on algorithmic computation*, Acta Appl. Math. **80** (2004), no. 2, 199–220.

[2] T. Arias, A. Edelman, and S. T. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl. **20** (1998), 303–353.

[3] P. Rostalski, I. A. Fotiou, D. J. Bates, G. A. Beccuti, and M. Morari, *Numerical algebraic geometry for optimal control applications*, SIAM J. Optim. **21** (2011), no. 2, 417–437.

[4] D. J. Bates, J. D. Hauenstein, A. J. Sommese, and C. W. Wampler, *Bertini: Software for numerical algebraic geometry*, 2006. Software available at `http://bertini.nd.edu`.

[5] D. J. Bates, J. D. Hauenstein, A. J. Sommese, and C. W. Wampler, *Numerically solving polynomial systems with Bertini*, SIAM, 2013.

[6] E. Begelfor and M. Werman, *Affine invariance revisited*, Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (Washington, DC, USA), CVPR '06, IEEE Computer Society, 2006, pp. 2087–2094.

[7] J. R. Beveridge, M. Kirby, and Y. M. Lui, *Action classification on product manifolds*, 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2010), 833–839.

[8] J.R. Beveridge, B. Draper, J.M. Chang, M. Kirby, H. Kley, and C. Peterson, *Principal angles separate subject illumination spaces*, IEEE Transactions on Pattern Analysis and Machine Intelligence **31** (2009), no. 2, 351–356.
[9] A. Bjorck and G. H. Golub, *Numerical methods for computing angles between linear subspaces*, Math. Comp. **27** (1973), 579–594.
[10] J.M. Chang, M. Kirby, H. Kley, C. Peterson, B. Draper, and J.R. Beveridge, *Recognition of digital images of the human face at ultra low resolution via illumination spaces*, Proceedings of the 8th Asian conference on Computer vision-Volume Part II, Springer-Verlag, 2007, pp. 733–743.
[11] R. Chellappa, A. Srivastava, P. Turaga, and A. Veeraraghavan, *Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition*, IEEE Trans. Pattern Anal. Mach. Intell. **33** (2011), no. 11, 2273–2286.
[12] Y. Chikuse, *Procrustes analysis on some special manifolds. statistical inference and data analysis (tokyo, 1997)*, Comm. Statist. Theory Methods **28** (1999), 885–903.
[13] M. T. Chu and J. L. Watterson, *On a multivariate eigenvalue problem: I. algebraic theory and power method*, SIAM J. Sci. Comput. **14** (1993), 1089–1106.
[14] J.H. Conway, R.H. Hardin, and Sloane N.J.A., *Packing lines, planes, etc.: Packings in Grassmannian space*, Experimental Mathematics **5** (1996), 139–159.
[15] D. Donoho, I. Drori, I. Rahman, P. Schroder, and V. Stodden, *Multiscale representations for manifold-valued data*, Multiscale Model. Simul. **4** (2005), no. 4, 1201–1232.
[16] B. Draper, M. Kirby, J. Marks, T. Marrinan, and C. Peterson, *A flexible flag representation for finite collections of subspaces of mixed dimensions*, 2013.
[17] S. Dutta, R. W. Heath, and B. Mondal, *Quantization on the Grassmann manifold*, IEEE Trans. Signal Process. **55** (2007), no. 8, 4208–4216.
[18] P. T. Fletcher, S. Venkatasubramanian, and S. Joshi, *The geometric median on Riemannian manifolds with application to robust atlas estimation*, NeuroImage **45** (2009), no. 1 Suppl, S143.
[19] J. Hamm, *Subspace-based learning with Grassmann kernels*, Ph.D. thesis, University of Pennsylvania, 2008.
[20] H. Karcher, *Riemannian center of mass and mollifier smoothing*, Communications on pure and applied mathematics **30** (1977), no. 5, 509–541.
[21] T.-K. Kim, J. Kittler, and R. Cipolla, *Learning discriminative canonical correlations for object recognition with image sets*, European Conference on Computer Vision (ECCV), 2006, pp. 251–262.
[22] H. W. Kuhn and A. W. Tucker, *Nonlinear programming*, Proceedings of the second Berkeley symposium on mathematical statistics and probability, vol. 5, 1951.
[23] L.-Z. Liao and L.-H. Zhang, *An alternating variable method for the maximal correlation problem*, J. Global Optim. **54** (2012), no. 1, 199–218.
[24] A. Shashua and L. Wolf, *Learning over sets using kernel principal angles*, J. Mach. Learn. Res. **4** (2003), 913–931.
[25] P. Shi and T. Wang, *Kernel Grassmannian distances and discriminant analysis for face recognition from image sets*, Pattern Recognition Letters **30** (2009), no. 13, 1161–1165.
[26] A. J. Sommese and C. W. Wampler, *The numerical solution of systems of polynomials: Arising in engineering and science*, World Scientific Publishing Company, 2005.
[27] A. Srivastava and E. Klassen, *Monte Carlo extrinsic estimators of manifold-valued parameters*, IEEE Transactions on Signal Processing **50** (2002), no. 2, 299–308.
[28] K. Turner, *Cone fields and topological sampling in manifolds with bounded curvature*, arXiv:1112.6160v1 (2012).
[29] K. Varshney and A. Willsky, *Linear dimensionality reduction for margin-based classification: high-dimensional data and sensor networks*, IEEE Trans. Signal Process. **59** (2011), 2496–2512.
[30] J. Verschelde, *Algorithm 795: PHCpack: A general-purpose solver for polynomial systems by homotopy continuation*, ACM Trans. Math. Softw. **25** (1999), no. 2, 251–276. Software available at `http://www.math.uic.edu/~jan`.

Dept of Mathematics, Colorado State University, Fort Collins, Colorado 80523
E-mail address: bates@math.colostate.edu

Dept of Mathematics, Colorado State University, Fort Collins, Colorado 80523
E-mail address: davisb@math.colostate.edu

Dept of Mathematics, Colorado State University, Fort Collins, Colorado 80523
E-mail address: kirby@math.colostate.edu

Dept of Mathematics, Colorado State University, Fort Collins, Colorado 80523
Current address: Dept of Mathematics, Bowdoin College, Brunswick, Maine, 04011
E-mail address: jmarks@bowdoin.edu

Dept of Mathematics, Colorado State University, Fort Collins, Colorado 80523
E-mail address: peterson@math.colostate.edu