Visualizing Data Sets on the Grassmannian Using Self-Organizing Mappings

Michael Kirby
Department of Mathematics
Colorado State University
Fort Collins, Colorado 80523-1874
Email: Michael.Kirby@Colostate.edu

Chris Peterson
Department of Mathematics
Colorado State University
Fort Collins, Colorado 80523-1874
Email: Chris.Peterson@Colostate.edu

***Abstract*—We extend the self-organizing mapping algorithm to the problem of visualizing data on Grassmann manifolds. In this setting, a collection of k points in n-dimensions is represented by a k-dimensional subspace, e.g., via the singular value or QR-decompositions. Data assembled in this way is challenging to visualize given abstract points on the Grassmannian do not reside in Euclidean space. The extension of the SOM algorithm to this geometric setting only requires that distances between two points can be measured and that any given point can be moved towards a presented pattern. The similarity between two points on the Grassmannian is measured in terms of the principal angles between subspaces, e.g., the chordal distance. Further, we employ a formula for moving one subspace towards another along the shortest path, i.e., the geodesic between two points on the Grassmannian. This enables a faithful implementation of the SOM approach for visualizing data consisting of k-dimensional subspaces of n-dimensional Euclidean space. We illustrate the resulting algorithm on a hyperspectral imaging application.**

I. Introduction

In the visualization of data we often resort to the computation of a mean, or in high dimensions, the centroid, of a collection of data. It is then natural to visualize these centroids, and the associated neighborhood data, using dimensionality reduction techniques such as Self-Organizing Mappings (SOMs) [15], [16], [17], [18]. This approach has proven to be a valuable tool for the low-dimensional visualization of data, see, e.g., [19], [24] in addition to an extended bibliography indicating the widespread applications of this methodology [14]. The key ingredient of this idea is that points that are neighbors in high-dimensional space are also represented as neighbors in the low-dimensional index space, a feature that arises through the self-organizing properties of the algorithm.

There is now considerable evidence that subspace approaches for data analysis are extremely effective at capturing the variability in data that often confounds pattern recognition, or classification systems, see, e.g., [22], [12], [23], [21]. The basic idea behind the approach is to compare an unlabeled observation to sets of data, each consisting of several patterns of a given class. Using a set of labeled data captures the variability of the data set over variations such as, e.g., illumination [4], [7], [2], [5]. It has been observed that the pattern set framework can greatly enhance the robustness of pattern recognition algorithms. For example, the resolution of images can be greatly reduced without sacrificing classification accuracy [6].

One attractive feature of set-to-set pattern recognition is the existence of the geometric mathematical framework known as the Grassmannian, or Grassmann manifold. The Grassmannian is a manifold whose points parameterize the k-dimensional subspaces of a fixed n dimensional vector space. As such, it provides a setting for comparing distances between subspaces using a variety of metrics. Due to their invariance under the natural action of the orthogonal group, the most widely used class of metrics are functions of the principal angles between the subspaces. Further, it is possible to transition from one subspace to another along a shortest path, or geodesic of k-dimensional subspaces. These ingredients make it possible to convert the SOM algorithm on vector spaces to an analogous algorithm on Grassmannians.

In this paper we develop the Grassmannian-SOM algorithm, i.e., we modify the standard SOM algorithm to operate in the setting of the Grassmann framework. We use the resulting approach to visualize subspaces of data in high dimensional spaces on low-dimensional index sets.

II. The Mathematics of the Grassmannian

The real Grassmann manifold $Gr(k, n)$ is a parameterization of all k-dimensional linear subspaces of real Euclidean n-dimensional space $\mathbb{R}^n, 0 < k \leq n$. We use the geometric structure of the Grassmann manifold to represent sets of points as subspaces. The framework allows us to compute distances between subspaces and to generate a curve of points (orthonormal matrices) between any two given points, i.e., subspaces. A point on $Gr(k, n)$, i.e., a k-dimensional subspace, can be non-uniquely represented by a basis, for instance by an $n \times k$ matrix $\mathbf{U}$ with orthonormal columns ($\mathbf{U}^T\mathbf{U} = \mathbf{I}_k$, the identity matrix) [13]. Two $n \times k$ full rank matrices correspond to the same point on $Gr(k, n)$ if they span the same subspace. The distances between points on the Grassmannian are measured in terms of the principle angles between the subspaces.

A. Overview of Angles

We provide a short summary of the computation of angles between subspaces initially described in [3]. Let X and Y be two vector subspaces of $\mathbb{R}^n$ such that

$$p = \dim(X) \geq \dim(Y) = q \geq 1,$$

where $\dim$ denotes the dimension of the subspace. Then the *principal angles* $\theta_k \in [0, \frac{\pi}{2}], 1 \leq k \leq q$ between X and Y are

978-1-5090-6638-4/17/$31.00

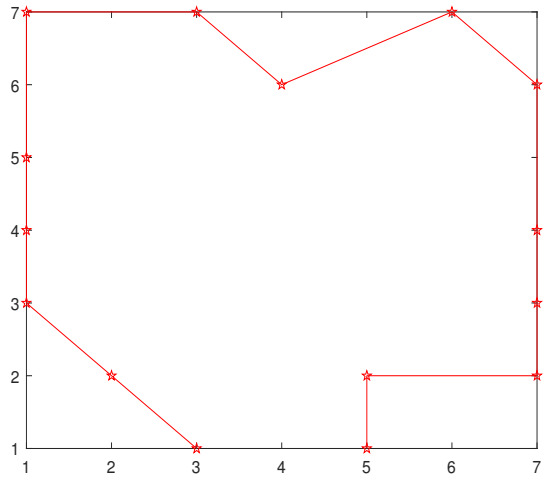

Fig. 1: This figure shows the approximate geodesic path of a sequence of 15 two-dimensional subspaces of 20-dimensional space.

Alfalfa	46
Corn-notill	1428
Corn-mintill	830
Corn	237
Grass-pasture	483
Grass-trees	730
Grass-pasture-mowed	28
Hay-windrowed	478
Oats	20
Soybean-notill	972
Soybean-mintill	2455
Soybean-clean	593
Wheat	205
Woods	1265
Buildings-Grass-Trees-Drives	386
Stone-Steel-Towers	93

TABLE I: The classes of the Indian Pines data. The bold denotes the data that was used in our experiment.

defined recursively by first defining θ_1 by

$$\cos(\theta_1) = \max_{\substack{u\in X \\ ||u||_2=1}} \max_{\substack{v\in Y \\ ||v||_2=1}} u^T v \tag{1}$$

Thus θ_1 is the smallest angle achieved between vectors lying in X and vectors lying in Y. If θ_1 is the angle between $u_1 \in X$ and $v_1 \in Y$, then θ_2 is defined as the smallest angle between vectors lying in X which are orthogonal to u_1 and vectors lying in Y which are orthogonal to v_1. Continuing in this manner, additional principal angles can be found by requiring additional orthogonality constraints. This leads to q principal angles which satisfy $0 \leq \theta_1 \leq \theta_2 \leq \ldots \theta_q \leq \frac{\pi}{2}$. Henceforth, $\theta = (\theta_1, \ldots, \theta_q)$ will denote the principal angle vector. Note that we have abused notation somewhat in using X to represent both a subspace and an orthonormal matrix which spans this space. For additional details related to algorithms for the computation of principal angles please see [3].

B. Metrics on the Grassmannian

Let X, Y be two points on the Grassmannian $Gr(k,n)$. Again, we are thinking of these points equivalently as subspaces, or orthonormal matrices that span the subspaces. The geodesic distance between these two points is given by

$$d_g(X,Y) = \|(\theta_1, \ldots, \theta_k)\|_2 \tag{2}$$

Other metrics are possible, e.g., the chordal distance

$$d_g(X,Y) = \|(\sin(\theta_1), \ldots, \sin(\theta_k))\|_2 \tag{3}$$

We note that it is possible to show that the Grassmannian may be isometrically embedding in Euclidean space when the chordal metric is employed and this is not the case for the geodesic metric [11]; see also [9].

Principal angles between subspaces are defined regardless of the dimensions of the subspaces. Thus, inspired by the Riemannian geometry of the Grassmannian, we may define, for any vector subspaces A, B of $\mathbb{R}^n$ the *geodesic* distance

$$d_\ell(A,B) = \|(\theta_1, \ldots, \theta_\ell)\|_2,$$

for any $\ell \leq \min\{\dim X, \dim Y\}$. While d_ℓ is not, strictly speaking, a metric (for example, if $\dim A \cap B \geq \ell$, then $d_\ell(A,B) = 0$), it nevertheless provides an efficient and useful tool for analyzing configurations in $\cup_{k\geq\ell} Gr(k,n)$. The geometry driving these distance measures is captured by the notion of a special kind of Schubert variety $\bar{\Omega}_\ell(W) \subseteq Gr(k,n)$. Let W be a subspace of $\mathbb{R}^n$, then we define

$$\bar{\Omega}_\ell(W) = \{E \in Gr(k,n) \,|\, \dim(E \cap W) \geq \ell\}.$$

With this notation, $d_\ell(A,B)$ simply measures the distance between A and $\bar{\Omega}_\ell(B)$, i.e. $d(A, \bar{\Omega}_\ell(B)) = min\{d_k(A,C) \,|\, C \in \bar{\Omega}_\ell(B)\}$ (it is worth noting that under this interpretation, $d_\ell(A,B) = d_\ell(B,A)$).

For $l = 1$ we employ only the smallest angle to determine the distance between two subspaces. This pseudo-metric was used to generate the Grassmannian SOM shown in Figures 2 and 3.

C. Geodesics

In this section we follow the results obtained in [1] for computing the geodesic path between two points $X, Y \in Gr(k,n)$ given by

$$G(t) = XV\cos\Theta t + U\sin\Theta t \tag{4}$$

We observe that

$$X = G(0)$$

and

$$Y = G(1)$$

and the trajectory $G(t)$ traces out the path of shortest distance on $Gr(k,n)$ in terms of the geodesic metric given by Equation (2). The quantities U, Σ and V are found by computing the singular value decomposition of the projection of

$$M = Y(X^TY)^{-1} \tag{5}$$

onto the orthogonal complement of X, i.e.,

$$U\Sigma V^T = (I - XX^T)Y(X^TY)^{-1} \tag{6}$$

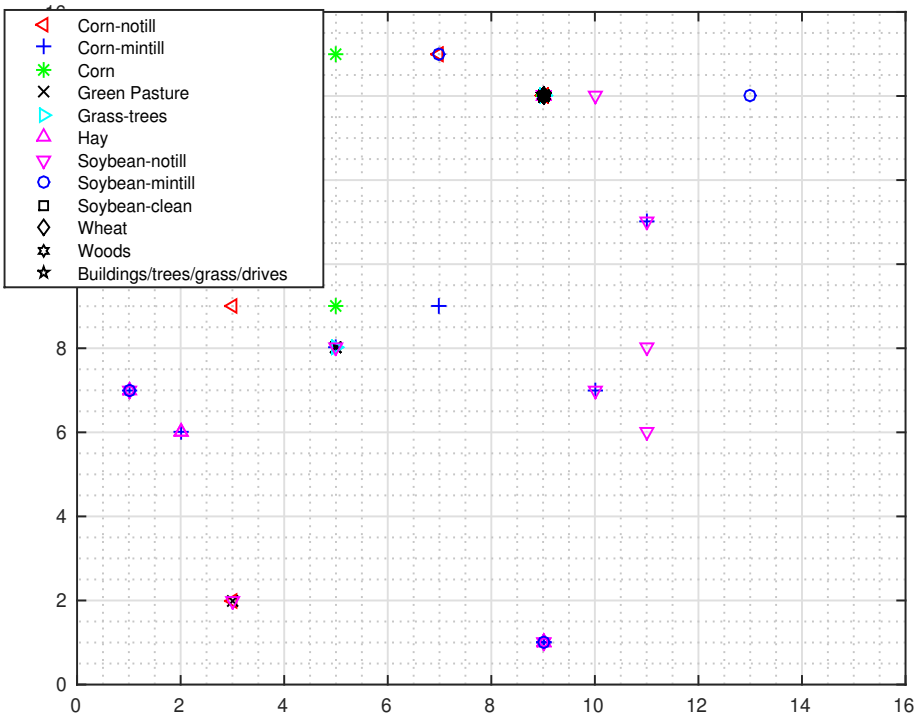

Fig. 2: Initial mapping of the classes to a 15 by 15 index set. We see that the data is not organized and that different classes are mapped to the same index.

where X and Y are given and the inverse of X^TY exists. Further, it can be shown that

$$\Theta = \text{atan}(\Sigma) \tag{7}$$

to complete the requirements of computing the geodesic between two subspaces X and Y as prescribed in Equation (4). This formula is a key ingredient for extending the self-organizing mapping algorithm on vector spaces to Grassmannians. See [1] for additional details.

III. Self-Organizing Mappings on $Gr(k, n)$

Here we extend the SOM algorithm on vector spaces to the setting of the Grassmannian, a parameterization of all k-dimensional subspaces of a fixed n-dimensional space. The general setting is that there is a collection of training patterns $x^{(\mu)}, \mu = 1, \ldots, P$ and an initial set of centers $\{c_i\}$ where the subscript i is the label of the spatial index I_i. Since the algorithm iteratively updates these initial centers we add a superscript m to identify the value of c_i at the mth iteration. The standard SOM update equation is given by

$$c_i^{m+1} = c_i^m + \epsilon_n h(d(i, i^*))(x - c_i^m) \tag{8}$$

where i^* is the winning center associated to pattern x, i.e.,

$$i^* = \arg\min_i \|x - c_i\|_2$$

where the distance between the point x and the center c_i is given by the standard Euclidean norm; see [15], [16], [17], [18] for additional details. We also take the localization function as the standard

$$h(s) = \exp(-s^2/\sigma^2)$$

and d is a metric that induces the topology on the index set. For simplicity, in this paper we will restrict our attention to the case

$$d(i, j) = \|\mathcal{I}_i - \mathcal{I}_j\|_2$$

where the indices are enumerated by the subscripts, i.e., the index set consists of $I_1, I_2, \ldots, I_N$. In this paper we use $I_1 = (1, 1), I_2 = (1, 2), \ldots, I_{225} = (15, 15)$.

On the Grassmannian the points are no longer elements of n-dimensional Euclidean space, but points $X, Y \in Gr(k, n)$, i.e., k-dimensional subspaces of $\mathbb{R}^n$. For a given subspace X we identify the center, i.e., from the set of subspaces that represent centers $\{C_i\}$, that is closest via

$$i^* = \arg\min_i d_g(X, C_i)$$

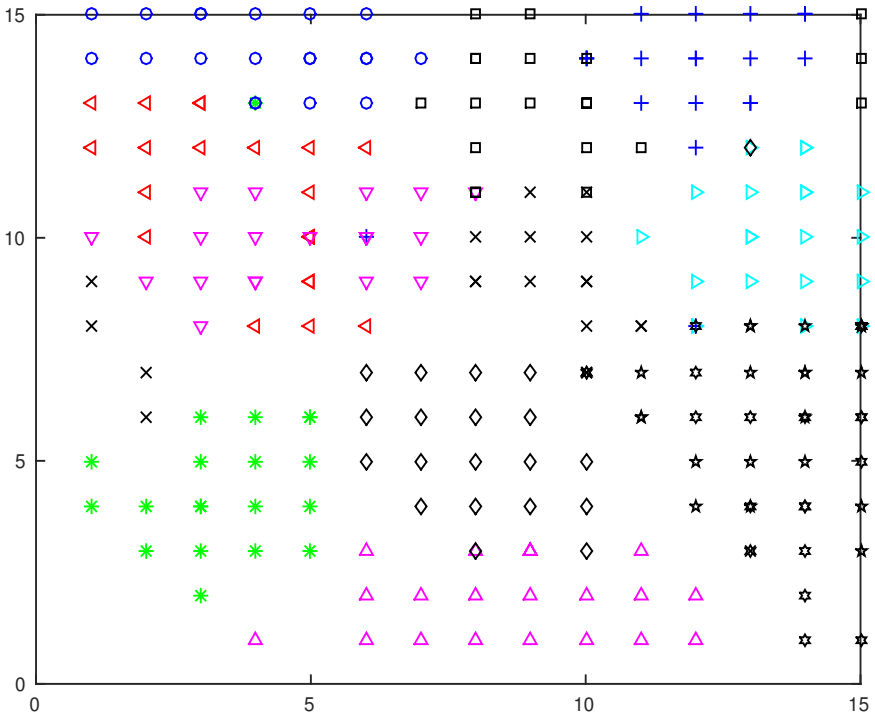

Fig. 3: This figure shows the final configuration of the points as mapped to the 2D index set. In contrast to Figure 2 we see clear organization of the class labels.

where the metric d_g is given by Equation (2). To move the centers towards the pattern subspace X according to the SOM update we compute the geodesic between each subspace center C_i and subspace pattern X

$$U_i \Sigma V_i^T = (I - C_i C_i^T) X (C_i^T X)^{-1} \tag{9}$$

Our localization term now becomes

$$t = t_0 h(d(i, i^*)) \tag{10}$$

where h and d are the same as for standard SOM but now $0 \leq t \leq 1$ is the amount we move each center C_i along the geodesic between it and the subspace X. We take $t_0 = 1/10$ as an initial distance along the geodesic that is annealed over the course of the iterations along with the width parameter σ. The updated centers are found using

$$C_i^{m+1} = C_i^{m+1} V_i \cos \Theta_i t + U_i \sin \Theta_i t \tag{11}$$

which moves each center towards the presented pattern X, i.e., point on the Grassmannian.

This pseudo-code for this process is summarized in Algorithm 1.

Algorithm 1: Grassmannian Self-Organizing Mapping

Input Data: Load class labeled data matrices $\{X_i^j\} \in \mathbb{R}^{n \times d}$ where k is the number of samples in each subspace and n is the dimension of the data, j is the class index, i is the matrix index.

Output Data: Final centers and indices of each data subspace.

Result: Representation of points on $Gr(k, n)$ as indices of SOM centers.

Initialization: Set the number of samples per subspace k, the number of centers N, initialize centers C_i as random d-dimensional subspaces and select the index set.

Define: Select the (pseudo)-metric on Grassmannian and compute the distance matrix between all pairs of subspaces.

Step 1: Present a random subspace to the network.

Step 2: Move all the centers C_i proportionally towards the presented subspace along the appropriate geodesic.

Step 3: Repeat until convergence.

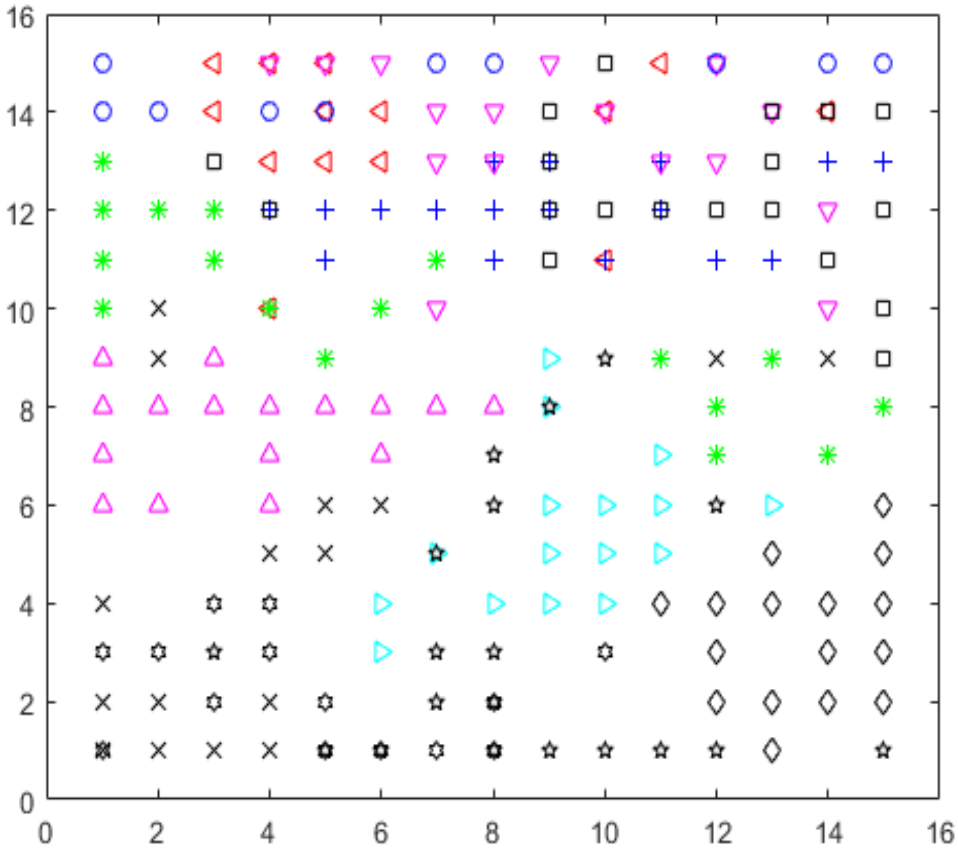

Fig. 4: The self-organizing mapping of the Indian Pines data set using Standard SOM.

IV. Numerical Results

A. Grassmannian Geodesic

In our first experiment we compute a geodesic path of two-dimensional subspaces of twenty dimensions using random matrices X, Y, Z. We form a geodesic along the path $X \to Y \to Z \to X$. Each segment has five sampled subspaces. See Figure 1 for a 2 dimensional visualization of this geodesic path that is traversing 20 dimensions. The fact that the line is not self-intersecting indicates that the nearest neighbor property on the path is preserved in two dimensions.

B. Indian Pines

To illustrate the utility of the proposed method for visualizing data on Grassmannians, we apply it to the well-known Indian Pines hyperspectral image [20]. We have considered this data set before in the context of the band selection problem [8] and the persistent homology for signal detection on Grassmannians [10]. A related visualization application invokes the technique of multi-dimensional scaling and sparse support vector machines [9].

In this application we selected the 12 classes that were large enough to give 20 subspaces of dimension ten. Since this application is merely intended to illustrate the model we made no attempt to optimize our parameters. However, our previous work suggests these dimensions are reasonable [9]. Thus we are visualizing 240 labeled points in 220 dimensions by first constructing sets of 10-dimensional subspaces in 220-dimensions using the SVD.

We initialized the centers for Grassmannian-SOM by selecting 225 ten dimensional subspaces at random. This was done by computing the singular value decomposition of matrices of size 220 by 10 from the uniform distribution. Associated with these centers is a 15 by 15 grid of integers that serve as the index set for the SOM. The initialization of the method is shown in Figure 2. We see the points are not organized and different classes are mapping to the same indices. In contrast, in Figure 3, we see the results of the Grassmannian-SOM algorithm, i.e., iterating Equation (11), where points in the same class have been organized to have similarly valued indices.

This data set is well-known as a challenging classification problem. For example, there are classes which are inherently very similar such as corn, corn-notill and corn-mintill. We see that these three classes are all located in a neighborhood of the index set with some overlap. Similarly, the three soybean

classes all clump together in a region reflecting the fact that this data has strong similarities. Classes such as green pasture, grass and trees and wheat are clean clusters.

It is interesting to compare this approach to the standard self-organizing mapping given by Equation (8). We see less coherent organization in the clustering in the standard SOM approach.

V. Conclusion

We have presented an extension of the self-organizing mapping algorithm to the setting of the Grassmann manifold. In the same spirit as the original SOM proposed in [15], the Grassmannian SOM moves centers towards patterns presented to the network moving proportionally along the geodesic, or shortest path between two elements of $Gr(k, n)$. We illustrate the method by showing that the algorithm organizes the data in the index space and separates ten dimensional subspaces of 220 dimensional space. This approach exploits the ability of subspaces to capture the variability of a family of patterns and is seen to produce more coherent organized structure than standard SOM on the hyperspectral example shown in the paper. Additional data sets need to be examined to determine if this behavior is exhibited more generally.

Although the algorithm was applied to subspaces of equal dimension, i.e., the mathematical setting of the Grassmannian, it is a straight-forward procedure to visualize subspaces of differing and possibly high dimensions.

Acknowledgment

This paper is based on research partially supported by the National Science Foundation under Grants Nos. DMS-1322508 and IIS-1633830. 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 National Science Foundation.

References

[1] P-A Absil, Robert Mahony, and Rodolphe Sepulchre. Riemannian geometry of grassmann manifolds with a view on algorithmic computation. *Acta Applicandae Mathematica*, 80(2):199–220, 2004.

[2] J.R. Beveridge, Bruce Draper, Jen-Mei Chang, Michael Kirby, Holger Kley, and Chris Peterson. Principal angles separate subject illumination spaces in ydb and cmu-pie. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 29, 2008.

[3] Ake Björck and Gene H Golub. Numerical methods for computing angles between linear subspaces. *Mathematics of computation*, 27(123):579–594, 1973.

[4] Jen-Mei Chang, J.R. Beveridge, Bruce Draper, Michael Kirby, Holger Kley, and Chris Peterson. Illumination face spaces are idiosyncratic. In *IPCV'06*, volume 2, pages 390–396. CSREA Press, June 2006.

[5] Jen-Mei Chang, Michael Kirby, Holger Kley, Chris Peterson, J.R. Beveridge, and Bruce Draper. Examples of set-to-set pattern classification. In *Mathematics in Signal Processing Conference Digest*, pages 102–105, Royal Agricultural College, Cirencester, U.K., December 2006. The Insititute for Mathematics and its Applications.

[6] Jen-Mei Chang, Michael Kirby, Holger Kley, Chris Peterson, Bruce Draper, and J Ross Beveridge. Recognition of digital images of the human face at ultra low resolution via illumination spaces. In *Computer Vision–ACCV 2007*, pages 733–743. Springer, 2007.

[7] Jen-Mei Chang, Michael Kirby, and Chris Peterson. Set-to-set face recognition under variations in pose and illumination. In *2007 Biometrics Symposium*, Baltimore, MD, September 2007.

[8] Sofya Chepushtanova, Christopher Gittins, and Michael Kirby. Band selection in hyperspectral imagery using sparse support vector machines. In Miguel Velez-Reyes and Fred A. Kruse, editors, *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XX*, volume 9088 of *Proc. of SPIE*, 2014.

[9] Sofya Chepushtanova and Michael Kirby. Sparse grassmannian embeddings for hyperspectral data representation and classification. *IEEE Geoscience and Remote Sensing Letters*, 14(3):434–438, 2017.

[10] Sofya Chepushtanova, Michael Kirby, Chris Peterson, and Lori Ziegelmeier. An application of persistent homology on grassmann manifolds for the detection of signals in hyperspectral imagery. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 2015*, 2015.

[11] John H Conway, Ronald H Hardin, and Neil JA Sloane. Packing lines, planes, etc.: Packings in grassmannian spaces. *Experimental mathematics*, 5(2):139–159, 1996.

[12] Bruce Draper, Michael Kirby, Justin Marks, Tim Marrinan, and Chris Peterson. A flag representation for finite collections of subspaces of mixed dimensions. *Linear Algebra and its Applications*, 451:15–32, 2014.

[13] A. Edelman, T.A. Arias, and S.T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 20:303–353, 1998.

[14] Samuel Kaski, Jari Kangas, and Teuvo Kohonen. Bibliography of self-organizing map (som) papers: 1981–1997. *Neural computing surveys*, 1(3&4):1–176, 1998.

[15] T. Kohonen. Self-organized formation of topologically correct feature maps. *Biol. Cybern.*, 43:59, 1982.

[16] Teuvo Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, 1990.

[17] Teuvo Kohonen. The self-organizing map. *Neurocomputing*, 21(1):1–6, 1998.

[18] Teuvo Kohonen. Essentials of the self-organizing map. *Neural Networks*, 37:52–65, 2013.

[19] Teuvo Kohonen, Erkki Oja, Olli Simula, Ari Visa, and Jari Kangas. Engineering applications of the self-organizing map. *Proceedings of the IEEE*, 84(10):1358–1384, 1996.

[20] D Landgrebe. Aviris nw indianas indian pines 1992 data set, 1992.

[21] Tim Marrinan, J Ross Beveridge, Bruce Draper, Michael Kirby, and Chris Peterson. Flag manifolds for the characterization of geometric structure in large data sets. In *Numerical Mathematics and Advanced Applications-ENUMATH 2013*, pages 457–465. Springer, 2015.

[22] Tim Marrinan, Bruce Draper, J Ross Beveridge, Michael Kirby, and Chris Peterson. Finding the subspace mean or median to fit your need. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 1082–1089. IEEE, 2014.

[23] Timothy Marrinan, J. Ross Beveridge, Bruce Draper, Michael Kirby, and Chris Peterson. Flag-based detection of weak gas signatures in long-wave infrared hyperspectral image sequences. In *Proceedings SPIE Defense + Security*, pages 98401N–98401N. International Society for Optics and Photonics, 2016.

[24] Helge Ritter and Teuvo Kohonen. Self-organizing semantic maps. *Biological cybernetics*, 61(4):241–254, 1989.