Overview:

How can video clips be analyzed to describe organic human actions automatically? How can the long-wave infrared spectrum of an underwater object be identified when the water is distorting the signature of the target? If you had the reflection of an image from a fun-house mirror, could you recognize the original image? My research focuses on geometric analysis of high dimensional data, and includes forays into optimization, computational mathematics, and pattern recognition. A common thread between my projects lies in exploiting linear algebra and differential geometry to create tools for averaging and optimization on Riemannian manifolds. In particular, the Grassmann manifold, or the set of all $k$-dimensional subspaces of an $n$-dimensional vector space, has grown in popularity as a setting for data analysis in the past decade. Belhumeur and Kriegman showed that the set of images of a convex object under different illumination conditions can be well approximated by a low-dimensional linear subspace [4]. Additive combinations of images of the object with different lighting conditions will create new examples of the object with possibly unseen illuminations. Using a subspace that spans variation as input in a recognition model helps avoid over-fitting to a training set. Representations of videos and images as points on Grassmann manifolds are used for activity modeling and recognition [26], shape analysis [22], appearance recognition [21], action classification [17], face recognition [15], noisy image classification [19], chemical detection in hyperspectral images [20], and general manifold clustering [13, 3].

Figure 1: Comparison between the flag mean and an eigenvector decomposition of a pool of 20 images of 5 people. The eigenvectors on the left show an average face effect. The flag vectors on the right were computed by creating four subspaces, one for each male subject, and then adding one picture of the female subject to each group. Since flag vectors model what is common among groups, they model the appearance of the woman and discount the men.

All of these applications rely on mathematical methods for averaging, clustering, and optimization. My contributions thus far have been to develop better and more versatile tools for achieving these same goals, and in finding maps between data in domain applications and points on these structured manifolds. One example of such a tool is the flag mean. The flag mean defines a nested sequence of vector spaces that best represents the collection based on a natural optimization criterion [8]. It is an example of an extrinsic mean for a collection of points on a union of Grassmann manifolds. It distinguishes itself from other techniques like principal component analysis (PCA) or eigenfaces, in that the flag mean behaves like a median [12, 15]. An example of the difference between a flag mean and an eigenface representation of the
same collection of images can be seen in Figure 1. The figure shows how common features shared between the subspaces, represented by the four black ovals on the right, are preserved in the leading elements of the flag. On the left, the eigenvector decomposition of the complete data set identifies the mean features of all the faces in the collection.

**Background:**

This section describes the background mathematics and defines the notation used in my current research. If you are already familiar with Grassmann manifolds and their metrics, feel free to skip to the section on my novel contributions.

Many signal processing and computer vision systems represent data as a set of linear subspaces of a fixed dimension [25, 26, 10, 3]. This structure allows the data to be treated as a collection of points on a single Grassmann manifold. The Grassmann manifold $\text{Gr}(k, \mathbb{V})$ is a manifold whose points parametrize the subspaces of dimension $k$ inside the vector space $\mathbb{V}$. In this document, we will assume that $\mathbb{V}$ is a finite dimensional real vector space and thus we can identify $\mathbb{V}$ through its dimension, $n$. For the remainder of this paper, we denote by $\text{Gr}(k, n)$ the Grassmann manifold of $k$ dimensional subspaces of $\mathbb{R}^n$, $\text{GL}(k)$ denotes the general linear group of invertible $k \times k$ matrices and $O(k)$ denotes the orthogonal group of $k \times k$ orthogonal matrices.

Let $\mathbb{R}^{n \times k}$ denote the vector space of $n \times k$ matrices with real entries and let $(\mathbb{R}^{n \times k})^o$ denote the open submanifold of full rank $n \times k$ matrices. For each $Y \in (\mathbb{R}^{n \times k})^o$, let $[Y]$ denote the column space of $Y$. There is a natural surjective map $\phi: (\mathbb{R}^{n \times k})^o \to \text{Gr}(k, n)$ given by $\phi(Y) = [Y]$ (with $[Y]$ identified with its corresponding point on $\text{Gr}(k, n)$). It is clear that $\phi(X) = \phi(Y)$ if and only if there exists an $A \in \text{GL}(k)$ such that $XA = Y$. Thus a point $q$ on $\text{Gr}(k, n)$ corresponds to a $k$-dimensional subspace $V_q$ of $\mathbb{R}^n$ and can be represented by any element of a $\text{GL}(k)$ orbit of a full rank $n \times k$ matrix, $Y$, whose column space $[Y]$ is equal to $V_q$. For the purposes of computation, we utilize a representative with orthonormal columns (note that if $Y$ is a representative with orthonormal columns and if $B \in O(k)$ then $BY$ will be another representative with orthonormal columns). Since a matrix representative, with orthonormal columns, for a point on a Grassmann manifold is only unique up to right multiplication by an orthogonal matrix, it is important that the output of any algorithm is invariant to such a multiplication.

Let $d: \text{Gr}(k, n) \times \text{Gr}(k, n) \to \mathbb{R}$ be a metric. The metric, $d$, is said to be orthogonally invariant if for every $[X], [Y] \in \text{Gr}(k, n)$ and every $A \in O(n)$, $d([AX], [AY]) = d([AX], [AY])$. One commonly used distance measure on Grassmannians is the metric induced from the projection Frobenius norm (denoted by $d_{p,F}$). It is an elementary exercise to show that $d_{p,F}$ is an orthogonally invariant metric on $\text{Gr}(k, n)$. The projection Frobenius norm arises from the identification of points in $\text{Gr}(k, n)$ with $n \times n$ projection matrices of rank $k$. If $X, Y$ are full rank $n \times k$ matrices with orthonormal columns, then the distance between $[X], [Y] \in \text{Gr}(k, n)$ is computed as a constant times the Frobenius norm of the difference between the projection matrix representations of the points: $d_{p,F}([X], [Y]) = 2^{-k} ||X^TX - YY^T||_F$. As shown by Edelman et al., this distance can also be computed as the $\ell_2$-norm of the vector of the sines of the principal angles between $[X]$ and $[Y]$ [9]. If $X$ (resp. $Y$) are orthonormal matrix representatives for $[X]$ (resp. $[Y]$) then the cosines of the principal angles between $[X]$ and $[Y]$ are the singular values of $X^TY$ [5]. Note that if $A, B \in O(k)$, then the singular values of $X^TY$ are the same as the singular values of $(XA)^T(YB)$.

In many applications, it can be natural and advantageous to represent aspects of data through subspaces lying in a fixed ambient space that are of differing dimensions. In such applications, a set of subspaces live naturally on a collection of Grassmann manifolds rather than on a single Grassmann manifold. Suppose that $[X] \in \text{Gr}(k_1, n)$ and $[Y] \in \text{Gr}(k_2, n)$ for $k_1 < k_2$. As illustrated in Bjork and Golub’s foundational paper, there will be $k_1$ principal angles between $[X]$ and $[Y]$ [5] and we define $d_{p,F}([X], [Y])$ as the $\ell_2$-norm of the vector of the sines of the $k_1$ principal angles between $[X]$ and $[Y]$. Note that $d_{p,F}$ is no longer a metric due to the possibility of $d_{p,F}([X], [Y]) = 0$ while $[X] \neq [Y]$ (for instance, if $[X]$ is a proper subspace of $[Y]$).

**Current Research:**

Leveraging these ideas my collaborators and I recently characterized a method for computing a flag of best fit for a point cloud living on a disjoint union of Grassmann manifolds [8]. This work was funded in part by grants from DARPA, http://www.math.colostate.edu/~marrinan
Figure 2: Cartoon depicting \( u_1 \), the first element of the flag mean of the subspaces \([X_1], [X_2], [X_3]\) represented by the three cones. The vectors \( k_1, k_2, k_3 \) represent the elements from within each subspace that are closest to \( u_1 \).

the NSF and the U.S. Air Force, and was published in *Linear Algebra and Its Applications* \(^1\). Let \( \tilde{Q} = \{q_1, q_2, \ldots, q_P\} \) be an ordered set of integers such that \( q_1 < q_2 < \ldots < q_P \). A flag in \( \mathbb{R}^n \) of type \( \tilde{Q} \) is a nested sequence of subspaces \( S_1 \subset S_2 \subset \cdots \subset S_P \) where \( \dim(S_i) = q_i \). It is important to note that if \( \tilde{Q} = \{q_1\} \) is a single integer, then the flag manifold of type \( \tilde{Q} \) is just the Grassmann manifold, \( \text{Gr}(q_1, n) \). The method that we described, referred to as the flag mean, then computes the sequential 1-dimensional optimizers, \([u^{(j)}]\), of

\[
\begin{align*}
\arg \min_{[u^{(j)}] \in \text{Gr}(1,n)} & \sum_{i=1}^P d_{PF}([u^{(j)}], [X_i])^2 \\
\text{subject to} & \quad u^{(j)}^T u^{(j)} = 1 \\
& \quad u^{(j)}^T u^{(k)} = 0 \quad \text{for } k < j,
\end{align*}
\]

where \( [X_i] \in \text{Gr}(m_i, n) \) for \( m_i \in \mathbb{N} \) and \( d_{PF}([X_i], [Y]) \) is the projection Frobenius norm between two linear subspaces. Figure 2 contains a cartoon that shows a candidate for the 1-dimensional solution to the flag mean optimization problem. While the vector \( u_1 \) in Figure 2 lies in the span of the subspaces \([X_1], [X_2], [X_3]\), it is not necessarily an element of any of them. The result of this optimization, the flag \( \mu_{PF} \), is a nested sequence of linear subspaces that is built by spanning the 1-dimensional optimizers, \([u^{(j)}]\), to create

\[
\mu_{PF} = \text{span}\{u^{(1)}\} \subset \text{span}\{u^{(1)}, u^{(2)}\} \subset \cdots \subset \text{span}\{u^{(1)}, \ldots, u^{(r)}\} \quad (2)
\]

where \( r \) is the dimension of \( \text{span}\{[X_1], [X_2], \ldots, [X_P]\} \).

When the data to be averaged is produced by a mixture of processes or contains outliers, the flag mean has been shown to model the dominate process and thus approximate the geometric median for linear subspaces \([18]\). The median-like property of this average has thus been exploited in multiple applications. A more thorough discussion of the properties of the flag mean, and a comparison to other Fréchet means for points on Grassmann manifolds can be found in our publication in the proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition \([18]\).

One of our applications of the flag mean is an algorithm for detecting chemical signatures in long-wave infrared hyperspectral images. The research for this algorithm was funded by a grant from the National Science Foundation under the Algorithms for Threat Detection program\(^2\). In a typical digital photograph, the image is represented by a 3-way array where each sheet corresponds to the response of the scene to red, green, or blue visible light. Similarly, a hyperspectral image consists of the response of a scene to numerous bands of infrared light for wavelengths in the 8-11 micron range. Figure 3a depicts the identification of ‘pixels’ in the hyperspectral image and vectors in \( \mathbb{R}^n \).

A common model for the observed spectrum of an optically clear chemical in a hyperspectral image is the sum of the spectrum of chemical, the spectrum of the background, and Gaussian noise. Most techniques for detecting chemicals rely

\(^1\)The relevant grant identifiers are DARPA: N66001-11-1-4184, NSF: CDS&E-MSS-1228308, and DOD-USAF: FA9550-12-1-0408 P00001.

\(^2\)The grant identifier is NSF: DMS-1322508.
(a) One pixel of an hyperspectral image corresponds to a high-dimensional vector. The value of a single element in the vector is the absorption of that location in the scene to a small range wavelengths of long-wave infrared light.

(b) Triethelyne phosphate detection map using the flag mean based algorithm. Warmer colors indicate a higher probability of gas presence.

Figure 3: A flag mean based detection algorithm for chemical signatures in long-wave infrared hyperspectral images.

on such a model and attempt to isolate the spectrum of interest by whitening the data with respect to the estimated covariance of the background. One well known algorithm that follows this path the Adaptive Cosine or Coherence Estimator (ACE) [16]. It is a common baseline algorithm because its results are well understood. ACE measures the cosine of the angle between a library radiance spectrum for a gas of interest and the whitened HSI pixel to detect signatures. One of the biggest bottlenecks in this procedure is obtaining an accurate estimate of the background covariance, and then inverting that rank-deficient matrix.

Our method, which has been accepted for publication in SPIE Defense, Security, and Sensing, avoids estimating the background covariance by cuing the detector with an image of the same scene from a time known to be free of gas. A subspace is formed from the span of each pixel in the frame of interest and the corresponding pixel in the gas-free frame. Spatially adjacent pixels are then averaged using the flag mean, and the angle is measured between the early elements in the resulting flag and the laboratory spectrum for the gas. The flag mean pushes shared information towards the initial elements of the flag, boosting the signature of the gas if it is present in adjacent pixels. An example of such a detection map can be seen in Figure 3b. The relative difference in the detection statistics for the gas plume and the background in the figure make it easy to visually distinguish the gas on the horizon. These qualitative results were verified on synthetic data where we could compute ROC curves. Our algorithm significantly outperforms the ACE algorithm, particularly in cases where the magnitude of the gas plume spectrum is weak relative to that of the background. In comparisons the ACE algorithm was given the true sample covariance of the background to ensure that the comparison was as equitable as possible. Our flag-based gas detection algorithm includes a ‘snapshot method’ for scenarios when a gas-free image of the same scene is not available, and it also shows impressive results. Hyperspectral videos are becoming more common in applications like monitoring factory emissions and some military uses, however, so plume-free scenes should be relatively common.

The most recent direction for my research is to solve optimization problems for a class of functions on the Grassmann manifold when the solution is restricted to live on a Schubert variety. This is motivated in part by an application from my collaborators in the electrical engineering department, involving estimation of geometric transformations of images. Briefly, given an n-dimensional vector space $\mathcal{V}$, and a flag $F = \{0\} \subset [U_1] \subset \cdots \subset [U_P] \subset \mathcal{V}$, and a vector of integers $a = [a_1, \ldots, a_P]$ with $0 < a_1 < \cdots < a_P \leq k$, the associated Schubert variety on $Gr(k, \mathcal{V})$ is defined as

$$\Omega_{F,a} = \{[Q] \mid [Q] \in Gr(k, \mathcal{V}) \text{ and } \text{dim}([Q] \cap [U_i]) \geq a_i, \forall i\}.$$ (3)

Thus the Schubert variety $\Omega_{F,a}$ contains all points on the Grassmann manifold that intersect the $i$th subspace in the flag $F$ in at least $a_i$ dimensions. In the context of our application, the flag $F$ is a theoretical or ‘oracle’ solution to a geometric transformation of an image. Given a set of observed sample subspaces, $[Q_i] \in Gr(k, \mathcal{V})$ for $i \in \mathbb{N}$ that are actual images
taken after corresponding geometric transformations of the objects being photographed, our intention is to average the
set $\{[Q_i]\}$ with the added constraint that the solution be restricted to the Schubert variety $\Omega_{F,a}$. The solution will then
contain information about the actual observed samples without disregarding information about the theoretical solution.

I have worked on many more projects during my time at Colorado State University than I have space to discuss here.
I also spent the past summer working as a PhD intern for a group of mathematicians and computer scientists at the
Pacific Northwest National Laboratory in the National Security Directorate. There I was able to broaden my research
into applications in cyber security from a graph theory perspective and visual analytics which led to a paper that will
appear in the proceedings for the IEEE Symposium on Technologies for Homeland Security [13]. If you would like further
information about my current research or past projects, I would be happy to provide more details.

New Directions:

I have numerous short-term research goals related to the mathematics described in this document. First, one of the main
benefits of the flag mean over other averages for points on Grassmann manifolds is that the resulting flag is structured so
that the more important dimensions are pushed to the front. Much research is already done on discovering low-dimensional
structure in high-dimensional data [6, 7, 2]. In that vein, it is fairly common in applications for points sampled from a
particular Grassmannian to have their mutual information contained within a lower dimensional subspace. When this
is true, we would like an average of those subspaces to fit the dimension of the relevant information and discard the
dimensions which contain noise. For the flag mean this is already a feature, but for the well known Karcher mean or the
intrinsic center of mass on a Grassmann manifold, it is not. I would like to use the tools that I have developed to create
the flag mean and to solve the Schubert variety constrained optimization to build a new flag that solves a relaxation of
the Karcher mean optimization problem. This would give scientists a way to use the optimal subspace from within the
Karcher mean as a representative for the mutual information in their data.

Another short term goal is to broaden the scope of my Schubert variety optimization. We are close to solving the
problem algebraically for one particular cost function, and to generalizing it to a larger class of functions if we solve it
via conjugate gradient. I would like to analyze the convergence in the conjugate gradient case, and complete a further
relaxation that would turn it into a convex problem that can be solved by existing semidefinite programming methods.

Over a longer time frame, I would like to integrate my work on Schubert variety constrained optimization into methods
like Allard et al’s geometric multi-resolution analysis [2], and to applications in unsupervised attribute discovery for
computer vision tasks [24, 23, 14]. Allard and his collaborators fit subspaces of appropriate dimension to subsets of
data from a high-dimensional cloud to approximate the underlying structure of the data. I think that such work has
a natural overlap with the applications of Schubert variety constrained optimization that my current collaborators and
I are tackling. On the other hand, attribute discovery is currently approached most often from a machine learning
perspective. I think there is an opening for geometric techniques to identify attributes defined by overlap with linear
subspaces spanning appropriate features. The recent success of deep convolutional neural networks like the one designed
by Jia et al. suggests to me that there may be identifiable attribute subspaces within the high-dimensional feature space
computed by their algorithm [11]. The ease of use of their algorithm and availability of data make explorations with Jia’s
architecture very simple.

I would also like to expand my research to include the young but growing area of ‘algebraic vision’, in which mathe-
maticians are trying to solve age-old computer vision problems related to 3D scene reconstruction with projective geometry
and linear algebra. Computer scientists have created ever-improving engineering solutions to scene reconstruction, but it
is possible that we will be able to solve some of these problems algebraically in the near future. A nice paper describing
the translation of the problem into the language of computational algebraic geometry comes from Aholt et al. [1].

References


http://www.math.colostate.edu/~marrinan


http://www.math.colostate.edu/~marrinan