Research Statement

Timothy Marrinan

Fields of Interest

My main field of interest is geometric data analysis, and in particular, low-dimensional approximation of high-dimensional data. This field connects to many areas including numerical differential geometry (optimization on smooth manifolds), nonnegative matrix factorization, and statistical signal processing; with applications to pattern recognition, data mining, computer vision, and image processing. These topics are unified by the idea that low-rank representations can identify and extract useful information about the relationships between data. Some low-rank models are inspired by domain-specific knowledge, so the rank of the solution is predetermined and enforced through constraints. In other situations the goal is to identify the unknown dimension of data within a noisy system, and thus a low-rank representation is uncovered as the result of optimization or analysis. I have had the privilege of working in many interesting fields during my PhD and postdoctoral positions. The breadth of my work has led to productive international collaborations with researchers in many disciplines. Each new topic has been accompanied by a transitional period of learning, but as a result, I excel in communicating precise mathematical concepts in approachable ways. My recent work can be split into three categories:

1. **Constrained optimization on smooth manifolds** - Data points are matrices (or equivalence classes) of a fixed rank. Optimization theory and algorithms must respect the manifold geometry while satisfying additional constraints. This was the main topic of my PhD research with Michael Kirby and Chris Peterson at Colorado State University.
2. **Nonnegative matrix factorization** - Nonnegative data matrices are approximated by the product of two low-rank nonnegative matrices. I develop identifiability theory for such factorizations and algorithms for computing them. This was the topic of my second postdoctoral position with Nicolas Gillis at Université de Mons.
3. **Rank-adaptive models** - Statistical or geometric criteria are used to estimate the *best* rank for an underlying data model in settings where the amount or type of samples make estimation difficult. This was a focus in my first postdoctoral position with Peter Schreier at Universität Paderborn, and applies to both previous topics as well.

Problem Descriptions and Recent Work

1. Constrained optimization on smooth manifolds: The focus of my PhD was optimization on Grassmann manifolds, or the set of k-dimensional subspaces of $\mathbb{R}^n$, denoted $\mathrm{Gr}(k, n)$. The Grassmann manifold has a Riemannian structure which makes it possible to compute gradients and Hessians, so optimization problems can be solved by adapting techniques created for vector spaces to this curved space[1]. The number of publications on Riemannian optimization has risen steadily over the past 15 years[2], in part because many applications can be represented on matrix manifolds like the Grassmann, the set of orthonormal k-frames in n dimensions, the set of fixed-rank positive semidefinite matrices, etc.

If $\mathcal{M}$ is a Riemannian manifold and $f : \mathcal{M} \to \mathbb{R}$ is a smooth function, we are interested in problems of the form: $\min_{\mathbf{x}\in\mathcal{M}} f(\mathbf{x})$. For example, given a set of points $\{\mathbf{x}_i\}_{i=1}^P \subset \mathrm{Gr}(k, n)$, let d be a metric and let f be the ℓ_2-norm of the vector of distances between a generic point $\mathbf{u} \in \mathrm{Gr}(k, n)$ and the points in this set, that is, $f(\mathbf{u}) := \sqrt{\sum_{i=1}^P d(\mathbf{u}, \mathbf{x}_i)^2}$. The point, $\mathbf{u}^*$, that minimizes f is the center of mass of $\{\mathbf{x}_i\}_{i=1}^P$ on $\mathrm{Gr}(k, n)$. Recently, we examine how modifications of the functions d and f in the above example lead to averages with different properties (see Publication 9 in the attached list).

One generalization of this type of problem is to consider subspaces, $\{\mathbf{x}_i\}_{i=1}^P$, that are of differing dimensions. Although these points do not naturally live on a single Grassmann manifold, we can map each of them to a related subset of points on $\mathrm{Gr}(k, n)$. Using a point-to-set similarity measure we recently demonstrated an efficient method for finding the center of the minimum enclosing ball of a collection of linear subspaces of different dimensions (see Publication 11).

Manifold optimization problems can also include constraints in addition to the ones that define the manifold, for example, $\min_{\mathbf{x}\in\mathcal{M}} f(\mathbf{x})$ such that $g(\mathbf{x}) \geq 0$ or $h(\mathbf{x}) = 0$. These are the types of constraints considered in the recent paper[3], where the authors formulate the augmented Lagrangian method and exact penalty method on Riemannian manifolds. In

[1] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

[2] Source: https://app.dimensions.ai Exported: September 09, 2020. Criteria: Text - 'riemannian' AND 'optimization' in full data.

[3] C. Liu and N. Boumal. "Simple algorithms for optimization on Riemannian manifolds with constraints". In: *Applied Math. & Opt.* (2019).

my dissertation I considered problems where the constraints took the form $g(\mathbf{x}) \doteq \dim(\mathbf{x} \cap \mathbf{s}) \geq \alpha$, for some subspace $\mathbf{s}$ and some $\alpha \leq \min\{\dim(\mathbf{x}), \dim(\mathbf{s})\}$. When there are multiple of these constraints and the subspaces $\{\mathbf{s}_i\}_{i=1}^M$ form a nested sequence, $\{\mathbf{0}\} \subseteq \mathbf{s}_1 \subseteq \cdots \subseteq \mathbf{s}_M \subseteq \mathbb{R}^n$, the set of points on $\mathrm{Gr}(k, n)$ satisfying the constraints is referred to as a Schubert variety. Schubert variety constraints are locally Lipschitz, but neither convex nor differentiable, making optimization difficult. My dissertation proposes an infeasible method to compute local solutions for this problem, and in Publication 5 we solved a convex relaxation that can be applied to direction-of-arrival estimation in beamforming.

2. Nonnegative matrix factorization: Given a nonnegative matrix, $X \in \mathbb{R}_+^{m\times n}$, and a rank, $r \leq \min\{m, n\}$, nonnegative matrix factorization (NMF) identifies $W \in \mathbb{R}_+^{m\times r}$ and $H \in \mathbb{R}_+^{r\times n}$ such that $X = WH$. Sufficient conditions exist under which this decomposition is identifiable, that is, unique up to permutation and scaling of the columns of W and the rows of H. Some of these conditions are easy to check, but rarely satisfied in practice. Other sufficient conditions are often satisfied in applications, but checking them is NP-hard[4]. One of my current research projects is to formulate sufficient conditions for identifiability that are commonly satisfied in practice and can be checked efficiently for large problems.

In many applications of NMF, the factorization rank, r, is small and assumed to be known. Thus we again have a low-rank model where the rank is known via domain-specific knowledge. NMF is posed as an nonconvex optimization problem in the nonnegative orthant of the form, $\min_{W,H} d(X, WH)$ such that $W, H \geq 0$, where d is a loss function like the squared Frobenius norm of the residual, $d(X, WH) = \|X - WH\|_F^2$. If the columns of W are considered to be the vertices of a polytope in $\mathbb{R}_+^n$ and the columns of H are normalized so that $H^T\mathbf{1}_r = \mathbf{1}_r$, then $d(X, WH) = 0$ implies that the samples (columns of X) are vectors within this polytope. However, *any* W that corresponds to a polytope containing the data will have a loss of zero, so one way to find an identifiable NMF is to penalize the volume of simplex defined by the columns of W and the origin. That is, we solve $\min_{W,H}\ d(X, WH) + \beta v(W)$ such that $W, H \geq 0$ and $H^T\mathbf{1}_r \leq \mathbf{1}_r$, where $v : \mathbb{R}^{m\times r} \to \mathbb{R}$ measures the volume of W and $\beta > 0$ is a penalty parameter. The constraint, $H^T\mathbf{1}_r \leq \mathbf{1}_r$, forces the estimates of X to be within the convex hull of the columns of W and the origin. In Publication 1 we considered a scenario where multiple nonnegative matrices, $\{X_i \in \mathbb{R}_+^{m_i\times n}\}_{i=1}^P$, are all approximated with a single endmember matrix, W, and we look to minimize the maximum error in these factorizations, $\min_{W,H_i} \max_{i=1,\ldots,n} \|X_i - WH_i\|_F^2 + \beta v(W)$ such that $W, H_i \geq 0$ and $H_i^T\mathbf{1}_r \leq \mathbf{1}_r$. For example, the X_i's could be a patches sampled from a larger hyperspectral image, in which case the proposed modification does a good job of identifying materials that are only present in a small number of pixels.

3. Rank-adaptive models: Unlike the first two problems, signal processing problems often have an unknown underlying model that we would like to detect from data. In signal processing, model order is the term used for the number of parameters of the data model, and model-order selection is the task of detecting the model order from samples. For example, in canonical correlation analysis (CCA) data is assumed to follow a linear mixing model, $x = A_x s_x + \nu_x$, and $y = A_y s_y + \nu_y$, where the signals, $\mathbf{s}_x \in \mathbb{R}^{m_x}$ and $\mathbf{s}_y \in \mathbb{R}^{m_y}$, are jointly Gaussian random vectors. The model-order selection problem for CCA is to estimate the number of nonzero correlation coefficients. In Publication 3 we proposed a method based on random projections for model-order selection when the number of samples is small. In Publication 2 and Publication 4 we considered model-order selection for multiset canonical correlation analysis (MCCA), where the correlation structures are more complicated. Performing model-order selection as a pre-processing step leads to improved performance in tasks where the rank of the model is an input parameter, such as the ones mentioned in the first two topics or in joint blind source separation which attempts to estimate the actual signals $\mathbf{s}_x$ and $\mathbf{s}_y$ in the CCA problem.

A related problem arises in manifold optimization. When subspaces are of different dimensions, like in the example described above, it is not clear which Grassmann manifold should be used for analysis. In Publication 11 we proposed a geometric criteria to identify the value of k for which subspace averaging on $\mathrm{Gr}(k, n)$ best approximated the data without over-fitting. In Publication 10 we took a different tactic and represented a collection of subspaces by an average flag, or nested sequence of subspaces, such that each element of the flag was the best approximation of the data for the given dimension. We go on to show that this flag-based average can be used effectively in a facial recognition pipeline.

Research Project Proposal:
Low-dimensional models for pattern recognition and signal processing

With respect to the preceding topics, I propose a short and medium-term research plan with the following main objectives: (1) Study theoretical issues, (2) Develop tests, tools, and algorithms, and (3) Connect the theory to practical applications. Below are some examples of how these objectives relate to each topic - along with an estimated timescale.

---

[4] K. Huang, N.D. Sidiropoulos, and A. Swami. "Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric decomposition". In: *IEEE Trans. Signal Processing* 62.1 (2013), pp. 211–224.

Theory *(Short-term)*:

1. The optimization theory for Riemannian manifolds is growing, but relatively few of the related papers address optimization with additional constraints. Additionally, it is not always straightforward to generalize the optimality conditions or convergence guarantees to these structured spaces, which creates an opportunity for further research. Examples include: (i) Provide globally optimal solutions to the Schubert-variety constrained problems discussed above, and (ii) Prove convergence of first-order optimization methods on the flag manifold with appropriate metrics.
2. I am particularly interested in the aspects of NMF theory that relate to the identifiability of a factorization. Since it is known that the most general version of the problem is not identifiable, we need to propose conditions under which the solution is unique that are often satisfied, easy to check, or both. Examples include: (i) Identify weaker sufficient conditions for checking whether a given solution satisfies the so-called *sufficiently scattered conditions* for identifiability in polynomial time, and (ii) Identify conditions under which online updates of NMF will converge to the same solution as the factorization computed with simultaneous access to all samples.
3. In model-order selection we again need identifiability conditions. Examples include: (i) Formulate and solve model-order selection problems for NMF. Current methods rely on NMF models that may have limited applicability, and (ii) Perform model-order selection for CCA with provable bias correction for dependent data.

Tests, Tools, and Algorithms *(Short-term)*:

1. Euclidean techniques can sometimes be adapted to non-flat spaces but require more structure and subsequently more computation, so this work centers around creating tools that are practical and efficient. Examples include: (i) Developing a projected gradient method for the Schubert variety constrained problem on the Grassmann manifold, and (ii) Creating discriminative metrics for the flag manifold and first-order optimization methods that employ them.
2. State-of-the-art methods for NMF can be efficient, but still struggle with very large sample sizes or streaming data. Examples include: (i) Solving large-scale NMF problems using randomization and batch-processing, and (ii) Proposing an online update technique with error bounds that are proportional to the size of the updates.
3. Techniques for rank-adaptive models are often tailored to a particular task, so they go hand-in-hand with the practical applications. Examples include: (i) Designing a method for joint dimensionality reduction and model-order selection for MCCA, and (ii) Formulating a generalized likelihood ratio test for independent vector analysis/MCCA.

Applications *(Medium-term)*:

1. One hurdle in practical applications of manifold optimization is in identifying domains whose invariant structures match the geometry of a known manifold. Once the invariants have been identified that connect data with a particular manifold, implementation is often straightforward. Examples include: (i) Improving pivot creation/selection for clustering of action videos using Schubert variety constrained optimization, and (ii) Designing an efficient sparse subspace clustering method by minimax optimization on the Grassmann manifold.
2. NMF problems naturally provide sparse and interpretable solutions so the applications are far-reaching. Here I propose hyperspectral image applications, but the connections to image processing, climate modeling, topic modeling, and others are straightforward. Examples include: (i) Detecting small targets in hyperspectral images with minimax NMF, and (ii) Identifying optically clear chemicals and submerged targets in hyperspectral videos via nonlinear NMF.
3. Rank-adaptive models can be used in exploratory data analysis where "ground truth" often comes from domain experts. Examples include: (i) Performing large scale MCCA with joint dimensionality reduction for biomedical data, eg., breast cancer tumor cells, EEG data from schizophrenia patients, and physiological data, and (ii) Improving techniques for face and action recognition in noisy video data by machine learning on flag manifolds.

Summary:

I have proposed a research plan to investigate **low-dimensional models for pattern recognition and signal processing** that focuses on three topics: (1) Constrained optimization on smooth manifolds, (2) Nonnegative matrix factorization, and (3) Rank-adaptive models. These topics have broad applicability and there is growing interest in the underlying theory. My international collaborations have led to a strong publication record[5] and prepared me well for independent research.

[5] Source: `https://scholar.google.com/citations?hl=en&user=98T4wnUAAAAJ`. Exported: September 23, 2020.