# Geometry in Signal Processing and Machine Learning

(Report on a Workshop Sponsored by the National Science Foundation)

Ali Pezeshki, Louis Scharf, and Douglas Cochran

12 September 2017

## Background

A three-day workshop to assess emerging roles for geometry in signal processing and machine learning was held in Estes Park, Colorado, between October 12th and 15th, 2016. The workshop was sponsored by the NSF Directorate for Computer and Information Science and Engineering (CISE) and was organized by the authors of this report.

The purpose of the Estes Park Workshop was to assemble mathematicians, statisticians, and engineers to explore problems in signal processing and machine learning, within the purview of CISE, that would be illuminated by a geometric view and be solved by methods of Riemannian, differential, symplectic, and algebraic geometry. The workshop was the latest in a sequence of workshops convened by Dr. John Cozzens of NSF:

1. **Keystone, CO, November 1991.** Engineers and applied scientists assembled to review future directions in wavelets, compression, and signal processing.

2. **Pacific Grove, CA, November 1995.** Mathematicians, statisticians, and engineers assembled after the Asilomar Conference on Signals and Systems to explore emerging ideas in inference in large data sets.

3. **Breckenridge, CO, March, 2007.** Engineers and mathematicians assembled to explore applications of differential geometry and algebraic topology to complex data analysis for emerging applications.

4. **Estes Park, CO, October 2016.** Mathematicians, physicists, and engineers assembled to explore the theory and application of geometry for signal processing and machine learning.

This report from the workshop clarifies how geometry enters naturally into signal processing and machine learning, as a way to code for constraints. Then a representative set of new problems is presented, as these serve as pulls on theory. A representative set of theoretical principles serves to push new ideas into emerging problems. From these pushes and pulls, several themes emerged from the workshop. These are reviewed, and followed by an account in case study format of where geometry has had significant impact on signal processing and machine learning. The closing section elaborates on the essential points of this document, and concludes with a recommendation to the National Science Foundation to establish an Initiative in Geometry for Signal Processing and Machine Learning. Information about the meeting structure and workshop participants is provided in the appendix.

# 1 Introduction

The mathematical foundations of modern signal processing and machine learning are harmonic analysis, linear algebra, multivariate statistics, optimization, and probability. The view is largely Euclidean, where measurements in a high-dimensional measurement space are mapped to features in a low-dimensional feature space, for subsequent decision, control, or visualization.

A common assumption is that measurements are cardinal and that they share a common time and/or space stamp for synchronization. The range of new problems presented by the internet, massive sensor nets, computational intelligence, and autonomous control impel the community to extend this corpus and to explore new mathematical frameworks for addressing these emerging problems.

# 2 How does Geometry Enter Signal Processing and Machine Learning?

Geometry enters signal processing, machine learning, and control whenever constraints are active, either as measurement constraints or as modeling constraints. In signal processing and control, modeling constraints are usually physically based, and these constraints are coded into an analysis-synthesis model. In machine learning, constraints are often used to contruct predictive models with a controllable number of parameters per measurement. The following examples are illustrative.

1. **Linear spaces** constrain mean values in first-order models, and covariances in second-order factor analysis models.
2. **Grassmannian manifolds** constrain models to sets of competing subspaces of a fixed dimension, as in subspace detection.
3. **Flag manifolds** constrain competing models, or models of increasing resolution, to telescoping sequences of subspaces of increasing dimensions, as in multiresolution analysis and hierarchical decision-making.
4. **Schubert varieties** impose intersection constraints on experimentally estimated subspaces, for regularization of subspace averages.
5. **Cones** constrain covariance matrices in statistical inference based on second-order methods.
6. **Convex sets** arising from equality and inequality constraints determine feasible solutions in model identification or predictive machine learning, as in structured matrix decompositions.
7. **Nonlinear transformations** map physical parameters into mathematical sets with known topological structures that may then be used for visualization and classification.
8. **Information geometry** speaks to the scheduling of sensors, signal processing, and machine learning under resource constraints to control mutual information, divergences, Fisher information, and related functions of probability distributions.

9. **Groups** are used to characterize the geometry of symmetries and invariances of physical and chemical systems, and to establish invariances in measurement and parameter spaces for statistical inference.

10. **Lie groups and homogeneous spaces** describe control action under constraints in robotics, and characterize admissible configurations in physical systems, chemical molecules, and biological structures.

11. **Invariant measures** are used to integrate on manifolds such as Stiefel and Grassmann, as in moment calculations and marginalizations with respect to Haar measure.

12. **Graphs and simplicial complexes** are used for representation of relational dependencies among algebraic or geometric objects, such as temporal and spatial relationships among data sets.

# 3 Problem Pulls and Theory Pushes

In a great number of fields of engineering and applied science the problem is to invert a signal, image, or data set, acquired from a sensor suite, measurement program, or survey. To invert is to identify the underlying field, object, or condition of state that produced the measurements. Certainly this describes the problem of identifying sources of information in signal intelligence, cyber attackers in data networks, radiating sources in radar, sonar, and remote sensing, objects of interest in CCTV records, and anomalies in medical images.

Much of the theory and many of the methods of signal processing and machine learning applied to these problems are based on the enforcement of subspace constraints in a Euclidean measurement space. Typically the methods are invariant to nonsingular linear transformations of measurements, and to scalings and rotations of signals within subspaces. But they do not serve when the problem is to detect and classify signals and images that have undergone more general impairments that may be given a group structure. They do not serve when measurements or parameters that constrain measurements are bound to a manifold.

New problems require new mathematical theory and methodology, and new mathematics suggests lines of attack on new problems. In the following subsections we outline pulls from problems and pushes from theory that suggest new methodologies for solving new problems.

## 3.1 Problem Pulls from Emerging Applications

1. **Attention and tracking** in images to classify objects (for example vehicles, people, faces) that change their location and articulation in a sequence of images. Invariant subspaces for manifolds of images subjected to transformation groups. Kernel articulation manifolds and optical flows for tracking in images.

2. **Feature extraction** as a problem in compression. Applies to classification, dictionary learning, and machine learning.

3. **Detection** in sensor networks for surveillance and control. Spacetime image processing for detection and tracking. Applies to anomaly detection for cyber security, resource identification in cognitive communication, and passive remote sensing.

4. **Invariant and equivariant image classification** under variable illumination and pose. Applies to face recognition and object classification.

5. **Manifold learning** for transforming high-dimensional data into low- dimensional sets, for inference and classification. Geodesic metrics on manifolds for constructing neighborhoods and decision boundaries.

6. **Coherence analysis** for measuring similarities in multi-channel and MIMO systems.

7. **Performance analysis** in parametric inference from compressed data, regularized inversions, and non-smooth optimizations.

8. **Uncertainty quantification** for inference on nonlinear dynamical systems.

9. **Generalized rate-distortion theory** for control of complexity and distortion under geometric constraints.

10. **Resource allocation** in sensing, processing, and control, using principles of partially observed Markov decision processes, and approximate dynamic programming. Geometry of submodular strings.

11. **Characterization of configuration space** in autonomous systems and in molecular chains.

12. **Characterization of large data sets** according to topological invariants, for visualization and inference.

13. **Computational linguistics** for study of language evolution, acquisition, and storage. Phylogenic classification of languages based on geometry of syntactic parameters.

14. **Analysis of ordinal data** for ranking and comparative judgments, based on partial information.

15. **Graphical signal processing** for computing global network states and dependencies from locally-computed statistics.

16. **Quantification of temporal evolution** in dynamical networks.

### 3.2 Theory Pushes from Geometry

1. **Computational differential geometry** for modeling large data sets. Manifold embeddings, such as Isomap and local linear embedding.

2. **Special groups** to characterize reachable reference frames.

3. **Riemannian, differential, symplectic, and algebraic geometries** for modeling shapes and drawing inference in constrained spaces. From extrinsic metrics to intrinsic (geodesic) metrics on manifolds.

4. **Optimization on manifolds** for detection, estimation, and model identification.

5. **Relationship between manifolds and statistical models** for inference in passive sensor networks and remote sensing.

6. **Use of Haar measure and geometric constraints** for Bayesian inference in multi-channel data sets, and integration on Stiefel and Grassmann manifolds.

7. **From subspaces to flags to Schubert varieties** for model identification at variable dimension.

8. **Replacement of Euclidean features** by geometrical objects such as orbits, subspaces, and flags of subspaces, when these arise from invariance or equivariance arguments.

9. **Persistent homology** for topological classification in large data sets of disparate types.

10. **Low-rank plus sparse decompositions** for robust dimension reduction.

11. **Comparison lemmas** for estimating and bounding performance of non-smooth optimizations.

12. **Information theoretic principles of source and channel coding** for analyzing the representation and compression of manifold-bound data and parameters.

13. **Information geometry** for resource allocation in sensing, processing, and control.

14. **Shortest path algorithms** as computational devices for approximating geodesics on a manifold.

15. **Continuum limits** for solving combinatorial optimization problems on graphs as geodesic problems on manifolds.

16. **Methods for non-metric multidimensional scaling (MDS)** and low-dimensional embeddings for completing ordinal matrices from partial information.

17. **Graph metrics** for structural analysis of dynamical graphs.

## 4 Emergent Themes

Several broad themes emerged from the Workshop: (1) a number of new and important problems in signal processing and machine learning do not fit traditional methodologies; (2) there exists a body of mathematical theory which, if widely disseminated and embraced, would equip engineers and applied scientists with new methods for addressing these problems; (3) these new problems will encourage the development of new theory and new computational methods in geometry; (4) a review of case studies where geometry has already had an impact argues in favor of continued successes for geometry. According to the dynamics of scientific development, the application of new methods will bring better solutions and new insights. This will lead to deeper and more general mathematical theory, and a broader range of applications for the theory.

1. **Statistical signal processing beyond subspaces.** Linear subspace models, and the corresponding linear subspace methods of signal processing and machine learning, have advanced the state-of-the-art in a vast number of applications where sensing and inference are required to invert measurements for information. But these models do not afford the tools required to tease out geometric structure in measurements that occupy more general spaces, such as smooth manifolds in high-dimensional ambient spaces.

2. **Statistical signal processing beyond subspaces, continued.** A theory of measurement and inference where prior physical models and posterior data models are fused into geometric constraints, more general than linear subspace constraints, will expand the reach of signal processing and machine learning. A resulting methodology comprised of geometrical principles and approaches, extending those of multivariate statistics, but not married to subspace models, would impact engineering, applied science, biology, and social science. Among the applications are tracking on manifolds such as SO(2) for phase tracking and on SO(3) for tracking of images.

3. **Geometrically-driven inference.** Signal processing and machine learning under manifold constraints on data or models raises the question of geometry for coding symmetries, invariances, and constraints in analysis and synthesis models, or in predictive models for which there are no such constructs.

4. **Geometrically-driven inference, continued.** For signal processing and machine learning, especially on large experimental data sets, multivariate statistical distributions and subspace constraints do not fully characterize data and model dependencies. To what extent can geometrical reasoning augment and inform probabilistic reasoning? To what extent do tractable manifold constraints model the physical or social mechanisms that generate measurements?

5. **Geometrically-driven inference, continued.** Attention to and tracking of objects in stationary backgrounds is a geometrical problem reminiscent of subspace tracking. But now the constraints are geometric, and new methods of analysis are required, based on algebraic methods of matrix decomposition and geometric principles on manifolds of images.

6. **Flag, Stiefel, and Grassmann manifolds; Schubert varieties.** Order determination speaks to model identification and dimension reduction. To what extent should the idea of model identification be replaced with the idea of model family identification, in a style reminiscent of Schubert varieties and multi-resolution analysis? The calculus of Schubert varieties provides a framework for solving rate-distortion problems, wherein the problem is to model data at variable dimension (rate) and variable distortion (some kind of metric). Fix the distortion and minimize dimension or fix the dimension and minimize distortion. Flags and varieties provide a framework for moving away from fixed order determination to modeling *at* dimension, so that models may be fit to data at various dimensions and processed at various dimensions for control of distortion.

7. **Flag, Stiefel, and Grassmann Manifolds; Schubert Varieties, continued.** In robotics and machine vision, low-dimensional subspace constraints are replaced by low-dimensional manifold constraints. When can these be known approximately from a physical model and when are they to be deduced from measurements? How are they to be deduced, and what is the impact of small data support on the quality of what can be learned and used?

8. **Invariant measures.** Invariance of probability models and related statistics under actions of groups plays a central role in deducing optimal hypothesis tests and estimators in statistical signal processing and in analyzing their performance. When the group involved possesses topological or geometric structure (e.g., a Lie group), the statistical problem inherits structure from its invariances. The essential characteristics of the data space, for example, are embodied in its quotient by the action of the invariance group. Recent work in multi-channel detection and estimation has demonstrated the value of this perspective in development and

optimization of generalized coherence tests. It offers potential for broad new geometric insight and methodology in statistical signal processing and machine learning.

9. **Geometry and Conditional Probability Measures; Symplectic Geometry.** Probability distributions conditioned on probability-zero events – e.g., random walks satisfying topological constraints like closure – are important in models throughout biology and physics, but sampling and analyzing them is a substantial challenge. Such distributions are implicitly geometric, since conditioning on a probability-zero event only makes sense in the context of something like Hausdorff measure on a Riemannian submanifold, but Riemannian techniques often are not useful when the ambient state space is itself non-Euclidean. In this setting, symplectic and algebraic geometry can provide both a theoretical framework and explicit coordinates for sampling and numerical integration when, for example, the conditional state space can be realized as a symplectic reduction or GIT quotient.

10. **Geometric signal processing.** The information content of nominally high-dimensional data often defines a curved, low-dimensional manifold. The geometry of this manifold and its intrinsic dimensionality and metric hold the key to extracting meaning from the data, even in the presence of substantial noise and, for dynamical systems, timing uncertainty. Mathematically rigorous dimensionality-reduction techniques now provide descriptions of data manifolds in terms of the eigenfunctions of well-known operators evaluated with respect to unknown, so-called induced metrics. Subject to general assumptions about the nature of the operations giving rise to the intrinsic manifold, however, it is often possible to identify the manifold metric. It is then possible to navigate purposefully on the intrinsic data manifold, i.e., to produce any desired state of the system from any starting point. In this approach, the intrinsic geometry of the data is used to extract information, even in the presence of overwhelming noise and timing uncertainty.

11. **Uncertainty quantification on manifolds.** In Euclidean space with multivariate normal measurements, quadratic measures of uncertainty lead to concentration ellipses as the geometric characterization of error concentration. It is a simple matter to compute volumes of these ellipses. Is there a natural analog of a concentration ellipse on a manifold and, if so, what is its shape and what is its volume? Generally, what do performance guarantees and confidence intervals look like on manifolds? How is a manifold to be quantized, and how are Voronoi cells to be computed?

12. **Low-dimensional geometrical models.** Despite the promise of low-dimensional geometrical models (such as manifolds) for image processing, computer vision, and machine learning tasks, their utility has been hamstrung in practice by two fundamental challenges. First, practical image manifolds are nonisometric to their underlying parameter space, while the state-of-the-art manifold modeling and learning frameworks assume isometry. Second, practical image manifolds are strongly perturbed by nuisance parameters such as illumination variations, occlusions, and clutter. Addressing these challenges requires the development of new metrics and representations that are more appropriate for images in the wild. To address the isometry challenge, for example, new distances, based on Optical Flow and the Earth Mover's Distance, are more natural metrics for inter-image distances than the standard Euclidean distance. These distances can be used to establish the isometry of manifolds generated by translations and rotations of a reference image. To address the nuisance transformation challenge, for example, an image representation based on local keypoint features can be used to define a keypoint articulation manifold, whose low-dimensional structure is robust to a

wide range of imaging transformations.

13. **Information Theory and Stochastic Modeling on Lie Groups.** Lie groups are a mathematical object that bridge geometry and algebra, and are relevant in describing both symmetries and configuration spaces of physical phenomena. Equipped with appropriate concepts of Gaussian distribution, convolution, and Fourier transform, Lie groups have all of the necessary components to extend information theory to physical domains. The interplay between statistical mechanics, stochastic processes on Lie groups, and generalizations of rate-distortion theory are applicable to physical information channels that arise when studying interacting robots and biomolecular docking and transcription problems.

14. **Geometric embeddings of combinatorial optimization.** One of the major roadblocks to making scientific progress in solving grand challenge problems is the curse of dimensionality, This problem is especially acute in combinatorial optimization where the behavior of the objective function under permutations and combinations has no obvious geometric structure. Remarkably, smooth geometric structure emerges as one allows the domain dimension to grow in many Euclidean combinatorial optimization problems including shortest path through a similarity graph and multiobjective pattern matching. This geometric embedding can lead to approximate solution of the combinatorial problem via solution of a simpler variational continuous optimization problem. Further progress in this field could lead to general combinatorial solvers that utilize the considerable machinery available in scientific computing, e.g., general ordinary differential equation (ode) and partial differential equation (pde) solvers. Grand challenge problems that could benefit from this research include: monitoring pandemics (path analysis on epidemic proximity graphs); energy and transportation (optimal routing); and adaptive drug design (computing Pareto frontiers); to name just a few.

15. **New measurement protocols and resulting data structures.** Inference from cardinal-valued data in a spacetime data structure now gives way to inference from ordinal-valued data on graphs. This is just one example of the data structures to be encountered in future applications of signal processing and machine learning. This raises the question of geometrical methods for organizing unstructured data and designing operators in a composite space of multi-type data sets.

16. **Topology for visualization and classification.** The ability to characterize and quantify recurrent structure in complex multi-scale time varying systems is fundamental in prediction and control. This is specially so when the collected data is multi-modal – e.g. numerical time series, video data, dynamic networks, etc – and there are no rigorous mathematical models that faithfully describe the process at hand. A promising avenue toward addressing this problem combines two approaches: (1) delay embeddings to reconstruct attractors in the state space underlying the observed dynamics, and (2) leveraging tools from computational topology – such as persistent homology – to quantify, determine and describe their topological structure. The resulting topological summaries provide features which can be used in machine learning frameworks for prediction, and the global description of attractors can be used as input for the design of data-driven models.

17. **Massive data, massive models, and machine learning.** With massive data comes the flexibility to identify massive parameters in massive predictive models. The extent to which deep learning constructs classifiers from complex training data is the extent to which massive parameters in massive models organize data into disjoint geometrical sets. The understanding of these sets and their construction is a problem in topology, geometry, and statistics.

18. **Massive data, massive models, and machine learning, continued.** Massive data sets serve as Monte Carlo simulations that can be used to solve combinatorial optimization problems. For example, a massive data set can serve as a Monte Carlo simulation of value-to-go in partially observed Markov decision processes that compute optimal policies for resource allocation and decision-making.

19. **Inference from sets of ordinal data.** The problem of finding Euclidean embeddings that respect ordinal rankings is a variation on multidimensional scaling. A complication is that matrices of ordinal data must be completed from partial orderings among subsets of the data, suggesting a connection with algebraic geometry.

20. **Graphs and simplicial complexes.** Graphs are the natural mathematical structures for modeling and analyzing networks. Simplicial complexes generalize graphs in ways that allow for modeling dependencies among sets of more than two nodes.

21. **Analysis of dynamical graphs.** Dynamic graphs provide abstract models to describe interaction processes between agents and time evolution of dynamical networks. For instance, such graphs can be used to analyze financial market networks, or quantify functional brain connectivity. The detection of such networks depends on the estimation of the coupling between the time-series that quantify the activity of the abstract agents (e.g., stocks, brain regions, etc.). The final stage of the analysis involves the inference of the graph that underpin the interaction processes. At this juncture, the practitioner (data analyst) is faced with the unpleasant surprise that existing tractable graph models (e.g., preferential attachment, exponential random graphs, stochastic block models, graphons, etc) – which lead to statistical inference – cannot replicate features of real-life networks. There is a pressing need to develop novel realistic graph models and the associated inferential methods. This raises the question of metrics for measuring the similarity of graphs in a sequence, and of designing algorithms for real-time computation.

22. **Computational geometry for engineers and applied scientists.** Calculus on manifolds is understood. But how well developed is the pedagogical framework for teaching modern geometry to engineers and applied scientists? How well-developed is the computational toolkit, especially in the context of integration, optimization, model identification, tracking, and Bayesian inference, under manifold constraints?

23. **Computational geometry for engineers and applied scientists, continued.** Modern graduate education in signal processing, machine learning, and control should include differential manifolds, Lie Groups and Riemannian Geometry. Our consensus is that a course should be developed jointly with mathematicians and engineers to ensure that essential abstract theory and computational methods make contact with physics and engineering.

## 5 Case Studies: Recent Advances and Open Problems in Geometry for Signal Processing and Machine Learning

There have been many triumphs for geometry in Signal Processing and Machine Learning. A number of these exploit the connection between graph theory and differential geometry. But it is fair to say that the connection between differential geometry and algorithm development remains in its infancy. As a consequence, geometry remains to be developed in the context of new and big and noisy data sets, where apparently high-dimensional measurements are actually constrained to low-dimensional manifolds. The problem is then to use experimental data to resolve the manifold into its tangent and normal bundles, compute its curvatures, approximate Laplace-Beltrami operators, use the identified manifold as a constraint set for inference, etc. Then algorithms become principled devices for adaptive detection, estimation, classification, and tracking under the constraints of the manifold.

In the paragraphs to follow, we outline a few of the outstanding successes to date of geometry for signal processing and machine learning. These portend a promising future for an initiative in Geometry in Signal Processing and Machine Learning that would build upon these sucesses and extend them to new domains of application. No comprehensive account of the literature is warranted in a document like this, so we constrain our references to a few seminal contributions, and a few recent results that appear to hold great promise.

## 5.1 From PCA and MDS to Nonlinear Dimensionality Reduction: ISOMAP, eigenmaps, LLE, and geometric diffusions for manifold learning.

- *Background and motivation.* The problem of manifold learning is to extract a low-dimensional set that accounts for most measurements in a high-dimensional ambient space, and to parameterize this set, itself a manifold, in sucn a way that measurements may be classified, or sequences of measurements may be formed into tracks. Principle components analysis (PCA), perhaps extended to kernelized functions of data, allows the data analyst to fit low-dimensional subspace models to data. Multi-dimensional scaling, first initiated in [33], [14], and subsequently extended in [24] and elsewhere, is a theory for extracting configurations from dissimilarity matrices which are *Euclidean.* These two methodolgies are the mainstays of analyses which aim to identify low-dimensional linear spaces and their configuration points within these linear spaces. ISOMAP, LLE, eigenmaps, and geometric diffusions [3], [4], [6], [12], [13], [39] extend these ideas to manifold learning, often by exploiting the connection between graphs and manifolds.

- *A geometric theory.* The essential problem is to identify a low-dimensional manifold that accounts for measurements in a high-dimensional ambient space. One of the first problems to be solved is the problem of determining intrinsic distance between distant points on a manifold, which appear to be neighboring points in the measurement space [39]. For distant points, geodesic distance can be approximated by adding up a sequence of short hops between neighboring points. These approximations are computed efficiently by finding shortest paths in a graph with edges connecting neighboring data points. Once geodesic distances are estimated, a distance matrix may be decomposed using MDS, which returns a basis for a low-dimensional manifold. The theory of ISOMAP is an imaginative admixture of local neighborhood analysis in a linear space to estimate a distance matrix of intrinsic distances on a manifold, followed by MDS.

  In contrast, local linear imbedding (LLE) [31] argues for the use of local measurements as a basis for the tangent space at another measurement, suggesting that the manifold is modeled as exactly flat, with its tangent space spanned by $K$ nearest neighbors of a given measurement. Then the measurement is approximated as a linear combination of neighbors. The weights should serve on the manifold, as well as they do for the putative tangent space. So local coordinates of the manifold are found locally by solvng for a set of $K$ vectors of dimension $d$ that would be well approximated by these same weights. These ideas have been extended, clarified, and connected to the spectral theory of graphs, by [3], [4], [6], [12], [13].

A key geometric idea in these references is that, "... under certain conditions, the graph Laplacian of a point cloud converges to the Laplace-Beltrami operator on the underlying manifold," [3], [4]. The work of [3], [4], [12], [13] establishes a fundamental link between the Graph Laplacian, the Laplace-Beltrami operator and diffusion maps. A close connection between the geometric theory of manifold learning and adaptive inference on learned manifolds remains to be developed. There remains the question of manifold learning from noisy data. Noise may perturb measurements in the manifold, or it may perturb measurements in the ambient space. Both types of perturbations raise the question of distributions in high-dimensional ambient spaces, or on special manifolds such as the Grassmann and Steifel manifolds.

- *Applications.* The theory and methodology of manifold learning has been applied widely to object recognition under conditions of coordinate transformation and variable illuminations, and to word clustering based on word counts in common articles. It remains a cornerstone of machine learning. When more closely coupled with inference, manifold learning becomes the adaptive step in adaptive manifold inference.

## 5.2 Darboux Frames and Various Curvatures from Second Order Covariances, Computed Locally

- *Background and motivation.* It is not too hyperbolic to say that a large body of theory and method in multivariate analysis and statistical signal processing consists in testing hypotheses about experimentally estimated covariance. In some cases this covariance is used to extract a first-order model. This basic idea is central to PCA, MDS, and null hypothesis testing, and it extends to factor analysis for detection and estimation. So the natural question is, "to what extent can second-order covariance analysis be extended to measurements which are known, or hypothesized, to be bound to a manifold?" This is a fundamental question in the application of geometry to signal processing and machine learning.

- *A geometric theory.* Let us say signal processing on manifolds consists of representing a manifold by its tangent and normal bundles. When the manifold is a linear subspace of $\mathcal{C}^n$, the bundles contain just one $r$-dimensional tangent space and one $(n-r)$-dimensional normal space. In this case these subspaces are identified with variations on principal component analysis, canonical correlation analysis, and factor analysis, based on a second-order sample covariance matrix. Perhaps second-order covariance can be computed locally on a Riemannian manifold, and used to extract local (Darboux) frames for the tangent and normal spaces, *and* estimate Gauss, Ricci, sectional, and scalar curvatures. The recent work of [42], extending and modifying the work of [8], and [35] gives formulas for decomposing covariance at a point $p$ on a manifold $\mathcal{M}_r(n)$ into local and global components. The local component codes for local tangent and normal spaces, and provides a Darboux frame for the tangent space. The global component codes for curvature. The key idea is to compute covariance over the intersection of a cylinder of radius $\epsilon$, aligned perpendicular to the tangent space $T_pM$. This covariance matrix is a function of even powers of $\epsilon$. Terms that scale with $\epsilon^2$ determine local subspaces and those that scale with $\epsilon^4$ determine curvature. The applications for these theoretical results are manifold: (1) at a resolution determined by $\epsilon$, a sampling theory based on curvature may be derived; (2) local tangent spaces determine local subspace detection and estimation; (3) local curvature determines geodesic moves for tracking subspaces; (4) local and global decompositions determine performance bounds for manifold-bound parameters [36], [37].

In summary, the insight here is that first-order modeling, based on second-order statistics, carries over to Riemannian manifolds. The key mathematical idea is that a flag of covariances, indexed by cylinder radius $\epsilon$, may be used to decompose covariance into local and global components, one of which codes for local subspaces, and one of which codes for curvature. Call this the *metrology of manifold learning.*

- *Applications.* This theory is so broad in its implications that it offers the basis for turning information geometry, as founded by Amari [2] on the original principles of Rao, into a statistical theory where experimental data is analyzed for its information geometry. In this application, the Fisher matrix is learned locally for non-Gaussian data, parameterized by manifold-bound parameters, even when the bounding manifold is unknown. Once a tangent bundle and normal bundle are determined, all the pieces are in place for adaptive signal processing on manifolds.

### 5.3 Optimization on Manifolds

- *Background and motivation.* Optimization in Euclidean space [9] is a well-developed, but still active field of applied mathematics, science, and engineering. The theory of convexity [30] and convex optimization [7] has produced new theory and standard methodology and algorithms for convex optimization. The work of [15], [27], [28] addresses the problem of optimization under manifold constraints. When the constraining manifold is a Lie Group, then optimization may be embedded in the corresponding Lie Algebra, as developed in [28].

- *A geometric theory.* Abstractly, optimization under orthogonality constraints [15] is a problem of designing gradient algorithms on locally updated tangent spaces, and in some cases this problem may be re-formulated as a problem of optimization on a Lie Algebra, such as the algebra of skew-symmetric matrices. But more generally, the problem of optimization on manifolds applies to the generalization of subspace tracking to manifold tracking. As in subspace tracking, where allowance is made for updating and downdating of model order, there is a need for a geometric theory of manifold tracking that allows for updating and downdating of manifold dimension. It is expected that the further development of the geometric theory of optimization on manifolds will produce algorithms and software that is matched to manifold optimization in the same way that numerical linear algebra and Euclidean optimization codes are matched to linear models.

- *Applications.* Applications for optimization on manifolds extend to basis and subspace identification in Grassmann and Stiefel manifolds, to manifold tracking, to tracking on Lie Groups. An example of the latter is estimation on $SO(n)$, as in tracking of frequency and wavenumber in communications, radar, and remote sensing. In [28] optimization on a Lie group has been applied with success to independent component analysis (ICA).

### 5.4 The geometry of imaging: 3-D images from 2-D projections

- *Background and motivation.* The work of [19] puts forward a general geometric theory for reconstructing 3-D scattering scenes from a series of their 2-D projections (snapshots), as the 3-D scene undergoes unknown rotational transformations. Each snapshot is represented by a vector of pixel intensities recorded by a 2-D detector. The collection of snapshots forms a point cloud that lies on a low-dimensional manifold embedded in a high-dimensional Euclidean space. In order to determine the rotation connecting any pair of snapshots, one needs to relate the metric of the data manifold produced by the collection of snapshots to the metric of the manifold of rotations.

  The work of [19] shows that the metric of the data manifold can be decomposed into two parts, an object-agnostic part with high symmetry, and an object-specific residual part with low symmetry. Using results from the throry of Lie groups, [19] shows that the eigenfunctions of the Laplacian associated with the high-symmetry part are directly related to those associated with the natural metric of the manifold of rotations, and the elements of the unknown rotation matrix can be expressed as linear combinations of These eigenfunction can be approximated directly from the data snapshots at low computational complexity,.

- *A geometric theory.* The basic principle of the approach of [19] is that the data acquisition process can be described as a manifold embedding $\Phi$ that maps the set of orientations of the object to the Hilbert space $\mathcal{H}$ of snapshots on the detector plane. As a result, the differential-geometric properties of the rotation group, $SO(3)$, formally carry over to the scattering dataset.

  The natural metric $h_0$ for $SO(3)$ is the metric tensor associated with the Killing form of the Lie algebra of $SO(3)$. This metric is bi-invariant under the so-called left and right multiplication maps that act on $SO(3)$, making $(\mathcal{H}, h_0)$ a Riemannian homogeneous space. The corresponding group of isometries has $SO(3) \times SO(3)$ structure. A key consequence of the symmetries of this metric is that the eigenfunctions of its Laplace-Beltrami operator (Laplacian) are the Wigner D-functions, which can be parameterized in terms of Euler angles. Nine of these eigenfunctions can be put into one-to-one correspondence with the elements of the rotation matrix that we wish to estimate. In particular, the elements of the rotation matrix can be expressed as a linear combination of these eigenfunctions. Thus, if eigenfunctions of the Laplace-Beltrami operator associated with $h_0$ could be computed directly from the observed snapshots on the data manifold, the representation for the rotation could be used to invert for the embedding map $\Phi$, by fitting the eigenfunctions computed from the data to the Wigner D-functions.

  Discrete approximations of Laplacian eigenfunctions, directly from the data, can be obtained using computationally efficient graph-theoretic algorithms involving diffusion maps [3],[4],[12],[13]. However, these algorithms do not provide the eigenfunctions of the Laplacian associated with the natural metric $h_0$. Rather they provide the eigenfunctions of the Laplacian for an induced metric $g$ (associated with the embedding $\Phi$), which in general is not invariant under arbitrary left or right multiplication maps.

  The breakthrough result of [19] is to show that the induced metric $g$ can be decomposed into a homogeneous part $h$ that has $SO(2)\times SO(3)$ symmetry and is independent of the object, plus a low-symmetry residual part $w$, which depends on the object. Moreover, the homogeneous part of $g$ corresponds to a well-known solution of general relativity (the so-called Taub solution), which has the important property that it preserves the Wigner D-functions as solutions of the Laplacian eigenvalue problem. This result argues for using the leading eigenfunctions of the Laplace-Beltrami operator associated with the induced metric $g$ (which can be obtained from the data snapshots using graph-theoretic algorithms of [12] and [13]) as approximations to the eigenfunctions of the Laplace-Beltrami operator associated with homogenous metric $h$, i.e., the Wigner D-functions. Fitting these approximations to the Wigner D-functions (which are parameterized by Euler angles) will identify the unknown rotation.

- *Applications.* The versatility of the geometric theory outlined above is demonstrated in [34], where the theory has been applied for reconstructing three-dimensional structures in X-ray diffraction and cryo-electron microscopy at low dosage and for recovering timing information in time-resolved experiments from random snapshots of dynamically evolving systems at low signal-to-noise ratios (SNRs). In particular, [34] demonstrates 3-D structure recovery from radiation-sensitive objects at doses at least an order of magnitude below previously established levels (SNR of $-16$ dB), and reconstruction of the temporal order of snapshots at per snapshot SNR values as low as $-21$ dB.

### 5.5 Detection, estimation, and tracking under manifold constraints, with application to pose-invariant tracking of objects in images and image sequences

- *Background and Motivation.* It is known from the work or [19], [34] that 2-D imaging of 3-D scenes may be given a group structure. The work of [17], [18] shows that fun-house mirror images may be represented by a finite-dimensional coordinate transformation, which determines an orbit or manifold of finite dimension. A basic scene, imaged under coordinate transformation, determines a finite-dimensional manifold unique to that scene. A sequence of functionals of the image produces a basis for a corresponding finite-dimensional subspace. Remarkably, the subspace is invariant to fun-house transformation, and the basis codes for the transformation. So detection of an image is, for example, invariant to translation, rotation, magnification, and so on. The basis of the subspace may be inverted for the image transformation. The fundamental idea is that a subspace, and not its basis, becomes the defining statistic for classification. This stands in contrast to standard methodolgies for multivariate analysis and subspace signal processing, where the basis is the statistic. As a consequence, all inference questions are posed on a Grassmannian manifold, where Veronoi regions on a manifold determine classification boundaries.

- *A geometric theory.* So the flow of ideas is this [17]: (1) an object is defined in an ambient space of large dimension; (2) this object and its infinity of versions under radiometric and geometric distortions determines an orbit or manifold of images under a transformation group; (3) a sequence of nonlinear maps returns a *feature vector*, which is in fact a *subspace* ; (4) the inference problem is then to detect and classify on a Grassmannian of subspaces of an ambient Euclidean space. The dimension of the ambient space is determined by dimensionality of the nonlinear maps. Each object type is represented by an invariant subspace, invariant to geometric and radiometric transformation. The intrinsic distance between a computed subspace and a candidate subspace on a Grassmanian serves as the classification statistic. The coordinate transformation from one basis of the subspace to another basis for the same subspace serves as the tracking statistic, as it codes for the parameters of the geometric deformation as the object works its way through a sequence of images.

  Globally, fun-house coordinate transformations lead to high-dimensional subspaces as models for a given object under transformation. Locally, the subspace model is low dimensional. So, a sampling theory is needed, wherein *way points* on a global orbit are used to approximate a global geometric deformation by a sequence of local lower-dimensional deformations. In the simplest case these may be affine. In other words, the sampling theory would address the question of replacing a global orbit, with its corresponding high-dimensional subspace, with a sequence of local orbits, each with a low-dimensional subspace for its corresponding affine approximation. The object representation at each way point is determined by averaging subspaces that correspond to observations local to a viewing perspective, to return a subspace of optimal dimension that best approximates the data. The question of subspace averaging has been addressed in [25] and extended in [32]. The extension has produced an automatic fitting rule for dimension of the average subspace. The theory of subspace averaging has recently been extended to constrained subspace averaging, wherein the average subspace matches a regularizing subspace in one or more dimensions. This is a geometric story in Schubert varieties, a story which generalizes the singular value decomposition (SVD) by constraining the returned subspace to match prescribed coordinates of a regularizing subspace.

Such a geometry for signal processing and machine learning would extend what is known about matched and adaptive subspace detectors to a theory of *matched manifold detectors*. This would require the design of embeddings that embed signals into manifolds, where objects like linear spaces become statistics for detecting and classifying on the basis of intrinsic distance between points on the manifold. Under a transformation group, the point remains an invariant representation of a signal.

*Applications.* Matched and adaptive subspace detectors have had an important impact on statistical signal processing. They are maximum likelihood, and uniformly most powerful invariant. Their virtue lies in their invariance to transformation groups that model scalings and subspace rotations. But their limitation is that they do not address more general impairments, modeled by transformation groups that code for geometric and radiometric transformations. The geometric theory outlined in the previous paragraph would address these limitations by extending matched subspace signal processing to *matched manifold signal processing*, where invariances to radiometric, geometric, and more general transformation groups are enforced. The applications are to dynamical scene analysis, and to anomaly classification in medical images.

### 5.6 Measure and integration on manifolds with application to detection and estimation

- *Background and motivation.* Problems of detection and estimation invariably confront the problem of nuisance parameters. There are at least three competing approaches: (1) the use of ordinary or generalized likelihood to replace nuisance parameters by their maximum likelihood estimates, (2) reduction by invariance to expose maximal invariants, followed by an application of Wijsman's Theorem [43], [44] to derive the likelihood ratio of these invariants, and (3) Bayesian marginalization with respect to non-informative priors that are uniform with respect to Haar measure. The first two alternatives are far better developed than the third, which requires integration on special manifolds, as in [20], [21].

- *A geometric theory.* Invariance of probability models and related statistics under actions of groups plays a central role in deducing hypothesis tests and estimators in statistical signal processing and in analyzing their performance. When the group involved possesses topological or geometric structure (e.g., a Lie group), the statistical problem inherits structure from its invariances. The essential characteristics of the data space, for example, are embodied in its quotient by the action of the invariance group. Bayesian marginalization with respect to nuisance parameters then reduces to a problem of integrating with respect to Haar measure on a manifold such as the Grassmanian (in the case of unknown subspaces) or a compact Stiefel manifold (in the case of unknown spacetime transmissions).

- *Applications.* So far these ideas have been applied to multi-channel coherence and to Bayesian detection for passive radar [20], [21]. But the use of non-informative priors on unknown parameters offers an alternative to generalized likelihood detection when there are unknown nuisance parameters to which one would like to be invariant. So the Bayesian theory for marginalizing the joint distribution of measurements and parameters, with respect to parameters whose prior distributions are uniform with respect to Haar measure, offers potential for new geometric insight and methodology in statistical signal processing and machine learning.

### 5.7 Second-order perturbations, modal analysis, and the geometry of the Grassmanian.

- *Background and motivation.* Modal analysis of spacetime fields aims to identify modes of oscillation and directions of arrival for propagating fields. Its domains of application are vibration analysis, radar, sonar, geophysics, remote sensing, and radio astronomy.

  The problem of modal analysis is distinguished from unconstrained subspace identification by the parameterization of the subspace by frequencies and wavenumbers, each bound to $(0, 2\pi]$. Least squares and maximum likelihood methods of identification lead to intractable, non-convex, optimization problems. So the standard methodology has been to replace a data matrix by its SVD (a linear algebraic idea) and then use variations on linear prediction (another linear algebraic idea) to extract frequencies and wavenumbers from a "noise-cleaned" data matrix.

- *A geometric theory.* The recent geometric analysis of Palka [26] and Vaccaro [41] of the Grassmannian structure of the modal analysis problem has determined that the SVD is optimum up to a second-order perturbation for approximating the underlying noise-free signal component of a noisy data matrix. But the real surprise is that a geometric analysis showed for the first time how the dominant and subdominant components of the SVD were to be used for subsequent identification of modes. Rather than discard the subdominant components of the SVD as if they carried no information about the frequency-wavenumber modes (the standard linear algebraic methodology), the geometry of the Grassmanian and its related second-order perturbation analysis give explicit formulas for retaining and weighting sub-dominant modes. The impact on detection and localization of radiating sources has been dramatic, extending performance thresholds, and resolving closely-spaced point sources not previously resolvable. The idea that the geometry of a manifold informs perturbation analysis has yet to be explored for other inverse problems.

  If $\mathbf{X} \in \mathcal{C}^{L\times M}$ is a measured data matrix with SVD $\mathbf{FKG}$, then the new result of [26], [41] is that the subspace of dimension $r$ nearest in intrinsic distance on the Grassmanian of dimension $r$ is spanned by the basis $\mathbf{F}_r + \mathbf{K}^{-1}\mathbf{Z}\mathbf{F}_{L-r}$. Geometric reasoning determines the computation of the matrix $\mathbf{Z}$. This subspace is rotationally invariant to first-order. No previous result has ever argued for the use of the sub-dominant subspace $\mathbf{F}_{L-r}$. This new subspace basis is now used for detection, tracking, beamforming, and so on.

- *Applications.* The recent results of [26], [41] will impact every domain of application where the problem is to identify oscillating modes of a time series or space series. The extension to space-time series remains to be developed. The obvious applications are to space-time imaging of radiating fields, as in radar, sonar, remote sensing, and radio astronomy. But whenever there is a constraining manifold for measurements or parameters, there is a problem of local modeling, where perturbation analysis applies. This perturbation analysis, informed by manifold geometry, determines a local approximation to the manifold.

## 5.8 Tracking on Lie Groups

- *Background and motivation.* Estimating the state of a dynamical system is a well-studied problem in connection with numerous applications. When the state space is $\mathcal{R}^n$ and the dynamical and measurement models are linear with gaussian noise, the Kalman/Kalman-Bucy filter provides an elegant solution that incorporates new measurements into the state estimate in a way that is computationally efficient and optimal is several estimation-theoretic senses. Due to both its mathematical elegance and ubiquity of applications, the Kalman filter is considered by many to be one of the most significant algorithmic advances of the last half of the twentieth century. Many variants have been introduced in seeking to carry forward its efficacy to situations characterized by weaker assumptions, but (although widely used in practice) these generally lack both the mathematical elegance and the optimality guarantees on the original algorithm.

  In many important dynamical estimation applications, including for example robotics and tracking of space debris, the natural state space is not $\mathcal{R}^n$, but is rather a nonlinear smooth manifold, often with algebraic structure (e.g., a Lie group or homogeneous space) in addition to its locally euclidean characteristics. Dynamical systems evolving on such spaces are intrinsically nonlinear.

- *A geometric theory.* Modeling stochastic dynamical systems on nonlinear manifolds entails development of suitable probability measures as well as dynamical equations. Although there may exist maximum entropy probablity measures on these spaces that play a role analogous to the gaussian distribution on $\mathcal{R}^n$, the maximum-entropy form is seldom preserved under propagation by the system dynamics; a state estimate posterior to a measurement can be taken as the maximum-entropy distribution on the state space whose parameters (e.g., moments, expected loss, etc.) match those of the actual posterior distribution, thus retaining state estimates that are always in the class of maximum-entropy distributions on the manifold. Determining representations of the probability laws (e.g., in some choice of coordinate atlas on the state space) in which the propagation of parameters under the dynamical model and re-mapping into the maximum-entropy class are parsimoniously expressed may be aided by the group structure of (or acting transitively on) the state space.

- *Applications.* Recent work [38] has developed this perspective for iterative estimation of the state of a dynamical system evolving on $SO(3)$ where the maximum-entropy distributions corresponding to a given "mean" matrix are von Mises-Fisher matrix distributions. It also points to obvious extensions to the euclidean group. The result of this work is a "tracker" for object orientation that appears to achieve an unprecedented combination of fidelity and computational efficiency. The primary application of interest in this work was navigation, though the compact $SO(3)$ and non-compact euclidean group state spaces are of broad interest in robotics and space systems. Possibilities of extending this geometric-statistical approach to other manifolds modeling other applications are open for research.

### 5.9 Multi-channel coherence for detection and source localization

- *Background and motivation.* There is now great interest in passive tracking from networks of accelerometers, acoustic sensors, radar antennas, and so on. The basic problem is to form an image of source locations that can account for multi-channel coherence between time series measured in a field of co-located or distributed sensors. Functions of multi-channel coherence serve as one-dimensional ambiguity surfaces in a volume indexed by time, frequency, and wavenumber. Virtually nothing is known about the geometry of this ambiguity surface for complex source geometries.

  Multi-channel coherence, as introduced in [11], is fundamentally a geometric principle for detecting linear dependence between measurements at a co-located or distributed array of sensors. The basic principle has been extended to multi-channel time series, and this extension has led to a principled definition of multi-channel spectral coherence [29]. The use of multi-channel coherence for acoustic source localization using a microphone array has been described in [5]. The invariances and null distribution of multi-channel coherence are well understood [11], but the geometry of its level curves has not been investigated.

- *A geometric theory.* The geometry of multi-channel coherence is understood, to the extent that its invariances are understood. But, when coupled with synchronization and complex modulation in frequency and wavenumber, multi-channel coherence is a statistic that maps a frequency-wavenumber-delay space to the real numbers. Its performance will not be fully understood until its level curves are better understood, and this is a question in geometry. This abstract geometry will, in turn, determine the array geometries that promote high resolution of radiating sources. This question of resolution bears on the question of related statistical measures, such as Kullback-Leibler Divergence, for resolving probabilistic measures of separation between hypotheses. How well does multi-channel coherence preserve KL Divergence? How does multi-channel coherence compare with ad-hoc methods that compute level curves for pairwise uses of sensors. These are geometric questions.

- *Applications.* Multi-channel coherence is the correct geometric principle for passive imaging in radar, sonar, geophysics, and remote sensing. But it remains to be coupled with delay and complex modulation to form a multi-channel ambiguity surface. It remains to be applied to the coherence analysis of internet traffic recorded at routers, where coherences are to be classified as benign or malignant.

## 6 Conclusions

Multivariate analysis, and the corresponding subspace methods of signal processing and machine learning, have advanced the state-of-the-art in a vast number of applications. This theory and its applications form the basis of principal component analysis, canonical correlation analysis, factor analysis, and subspace signal processing. These, in turn, are the foundations for model-based inference, where the problem is to distinguish information-bearing signals from interference and sensor noise. But subspace models do not afford the tools required to tease out geometric structure in measurements that occupy more general spaces, such as convex cones, Stiefel manifolds and Grassmanian manifolds.

Moreover, the aims of subspace methods are typically to distinguish signals from interference and sensor noise. Extant and emerging problems in signal processing, machine learning, and big data analytics present different challenges, because the major variabilities are no longer interference and sensor noise, but rather unknown variabilities of data collection and geometric transformations due to deformations of objects. In many cases these variabilities may be given a group structure, where a transformation group codes for variability. High-dimensional data and low-dimensional parameterizations are manifold-bound, and the structure of these manifolds is unknown. When a probabilistic parameterization is appropriate there remains the problem of identifying the probability law, estimating performance bounds, and deriving detection and estimation algorithms in support of classification, tracking, and predictive analytics.

**Recommended Initiative.** With the advent of the internet, social media, big-data, massive sensor nets, cybersecurity, computational intelligence, and autonomous control, the time is propitious to revisit the mathematical foundations of signal processing and machine learning. Based on the outcomes of this workshop, we conclude that geometry should be central to these foundations. An NSF-funded Center or Institute for the study and application of Geometry in Signal Processing and Machine Learning would advance our understanding of emerging problems of national significance.

**Intellectual merit.** The intellectual merit of the proposed initiative may be defined by its proto- and meta-objectives. The proto-objective is to identify and solve *geometric* problems of detection, estimation, tracking, clustering, and classification, as they arise in physics, chemistry, and biology; in radar, sonar, geophysics and remote sensing; in image processing; in big data and predictive analytics. These problems are geometric by virtue of data and parameter constraints and by virtue of symmetries and invariances that are imposed on admissable solutions. The meta-objective is to develop a geometric theory of signal processing and machine learning that is as broad in its scope and as precise in its methodologies as multivariate analysis and subspace signal processing . Such a theory will augment statistical reasoning with geometrical reasoning, and bring new mathematical methods into play. It will address in a rigorous way the processing of signals and data sets that are constrained to manifolds, or whose underlying parameters are so constrained. It is anticipated that the geometry of signal processing and machine learning will illuminate the vices and virtues of existing algorithms and lead to improved algorithms with definable attributes or optimalities. It will open up to analysis and design problems in signal, image, and data analysis previously inaccessible to subspace methods. It will encourage development of numerical methods that are matched to analysis on manifolds in the same way that numerical linear algebra is matched to subspace methods.

# Appendix 1: Contributed Paragraphs

**Prof. Richard Baraniuk: Exploiting signal geometry in the wild.** Despite the promise of low-dimensional geometrical models (such as manifolds) for image processing, computer vision, and machine learning tasks, their utility has been hamstrung in practice by two fundamental challenges. First, practical image manifolds are nonisometric to their underlying parameter space, while the state-of-the-art manifold modeling and learning frameworks assume isometry. Second, practical image manifolds are strongly perturbed by nuisance parameters such as illumination variations, occlusions, and clutter. Addressing these challenges requires the development of new metrics and representations that are more appropriate for images in the wild. To address the isometry challenge, for example, we have shown that new distances based on Optical Flow and the Earth Mover's Distance are more natural metrics for inter-image distances than the standard Euclidean distance. We have applied these distances to establish the isometry of manifolds generated by translations and rotations of a reference image. To address the nuisance transformation challenge, for example, we have developed an image representation based on local keypoint features and used it to define a new keypoint articulation manifold whose low-dimensional structure is robust to a wide range of imaging transformations.

**Prof. Greg Chirikjian: Information Theory and Stochastic Modeling on Lie Groups.** Lie groups are a mathematical object that bridge geometry and algebra, and are relevant in describing both symmetries and configuration spaces of physical phenomena. Equipped with appropriate concepts of Gaussian distribution, convolution, and Fourier transform, Lie groups have all of the necessary components to extend information theory to physical domains. The interplay between statistical mechanics, stochastic processes on Lie groups, and generalizations of rate-distortion theory are applicable to physical information channels that arise when studying interacting robots and biomolecular docking and transcription problems.

**Prof. Alfred Al Hero: Geometric embeddings for combinatorial optimization.** One of the major roadblocks to making scientific progress on grand challenge problems is the curse of dimensionality, This problem is especially acute in combinatorial optimization where the behavior of the objective function under permutations and combinations has no obvious geometric structure. Remarkably, smooth geometric structure emerges as one allows the domain dimension to grow in many Euclidean combinatorial optimization problems including shortest path through a similarity graph and multiobjective pattern matching. This geometric embedding can lead to approximate solution of the combinatorial problem via solution of a simpler variational continuous optimization problem. Further progress in this field could lead to general combinatorial solvers that utilize the considerable machinery available in scientific computing, e.g., general ordinary differential equation (ode) and partial differential equation (pde) solvers. Grand challenge problems that could benefit from this research include: monitoring pandemics (path analysis on epidemic proximity graphs); energy and transportation (optimal routing); and adaptive drug design (computing Pareto frontiers); to name just a few.

**Dr. Stephen Howard: Invariance of probability models and related statistics under actions of groups.** plays a central role in deducing optimal hypothesis tests and estimators in statistical signal processing and in analyzing their performance. When the group involved possesses topological or geometric structure (e.g., a Lie group), the statistical problem inherits structure from its invariances. The essential characteristics of the data space, for example, are embodied in its quotient by the action of the invariance group. Recent work in multi-channel detection and estimation has demonstrated the value of this perspective in development and optimization of generalized coherence tests. It offers potential for broad new geometric insight and methodology in statistical signal processing and machine learning.

**Prof. Hamid Krim: Bi-Sparsity Pursuit: A Paradigm for Robust Principal Subspace Reconstruction.** The success of sparse models in computer vision and machine learning is due to the distribution of high dimensional data as a union of low dimensional subspaces in many real-world applications. The underlying structure may, however, be adversely affected by sparse errors. Our novel approach of a bi-sparse model as a framework to analyze this problem yields a novel approach to recover the union of subspaces in the presence of sparse corruptions. The established convergence of this methodology promises its increasing importance in a number of applications using real-world vision data as has already been demonstrated.

**Prof. Francois Meyer: Statistical graph analysis.** Dynamic graphs provide abstract models to describe interaction processes between agents. For instance, such graphs can be used to analyze financial market networks, or quantify functional brain connectivity. The detection of such networks depends on the estimation of the coupling between the time-series that quantify the activity of the abstract agents (e.g., stocks, brain regions, etc.). The final stage of the analysis involves the inference of the graph that underpin the interaction processes.

At this juncture, the practitioner (data analyst) is faced with the unpleasant surprise that existing tractable graph models (e.g., preferential attachment, exponential random graphs, stochastic block models, graphons, etc) – which lead to statistical inference – cannot replicate features of real-life networks. There is a pressing need to develop novel realistic graph models and their associated inferential methods.

**Prof. Abbas Ourmazd: Geometric signal processing.** The information content of nominally high-dimensional data often defines a curved, low-dimensional manifold. The geometry of this manifold ? its intrinsic dimensionality and metric ? hold the key to extracting meaning from the data (1), even in the presence of substantial noise (2) and, for dynamical systems, timing uncertainty (3). Mathematically rigorous ?dimensionality-reduction? techniques now provide descriptions of data manifolds in terms of the eigenfunctions of well-known operators evaluated with respect to unknown, so-called induced metrics (4,5). Subject to general assumptions about the nature of the operations giving rise to the intrinsic manifold, however, it is often possible to identify the manifold metric (1). It is then possible to navigate purposefully on the intrinsic data manifold, i.e., to produce any desired state of the system from any starting point. In this approach, the intrinsic geometry of the data is used to extract information, even in the presence of overwhelming noise and timing uncertainty (3).

**Prof. Jose Perea: Sliding windows and persistence–some applications of topological methods to signal analysis.** The ability to characterize and quantify recurrent structure in complex multi-scale time varying systems is fundamental in prediction and control. This is specially so when the collected data is multi-modal – e.g. numerical time series, video data, dynamic networks, etc – and there are no rigorous mathematical models that faithfully describe the process at hand. A promising avenue toward addressing this problem combines two approaches: (1) delay embeddings to reconstruct attractors in the state space underlying the observed dynamics, and (2) leveraging tools from computational topology – such as persistent homology – to quantify, determine and describe their topological structure. The resulting topological summaries provide features which can be used in machine learning frameworks for prediction, and the global description of attractors can be used as input for the design of data-driven models.

**Prof. Clayton Shonkwiler: Geometry and conditional probability measures; symplectic geometry.** Probability distributions conditioned on probability-zero events – e.g., random walks satisfying topological constraints like closure – are important in models throughout biology and physics, but sampling and analyzing them is a substantial challenge. Such distributions are implicitly geometric, since conditioning on a probability-zero event only makes sense in the context of something like Hausdorff measure on a Riemannian submanifold, but Riemannian techniques often are not useful when the ambient state space is itself non-Euclidean. In this setting, symplectic and algebraic geometry can provide both a theoretical framework and explicit coordinates for sampling and numerical integration when, for example, the conditional state space can be realized as a symplectic reduction or GIT quotient.

# Appendix 2: Workshop Structure and Participants

### Logistics

The Workshop was held in the Stanley Hotel, Estes Park, Colorado starting on October twelfth and ending on October fifteenth, 2016. NSF sponsorship enabled the organizers to provide hotel accommodation and meals to participants for the workshop period; individual participants traveled to Colorado at their own expense.

### Participants

The organizers undertook an extensive literature search to identify researchers whose recent work falls within the scope of the Workshop. The organizers' networks of professional contacts was also instrumental in identifying prospective participants to invite.

The Workshop participants are listed in the following table. In this table, the participant marked with (*) attended virtually and those marked with (†) were graduate students who assisted with recording and note-taking functions.

| Name | Institution | Country |
|---|---|---|
| Richard Baraniuk | Rice University | USA |
| Gregory Chirikjian | Johns Hopkins University | USA |
| Edwin Chong | Colorado State University | USA |
| Doug Cochran | Arizona State University | USA |
| Mary Cook† | Arizona State University | USA |
| John Cozzens* | National Science Foundation | USA |
| Lauren Crider† | Arizona State University | USA |
| Philip du Toit | Numerica Corporation | USA |
| Babak Hassibi | California Institute of Technology | USA |
| Alfred Hero | University of Michigan | USA |
| Stephen Howard | DST Group | Australia |
| Michael Kirby | Colorado State University | USA |
| Brett Kotschwar | Arizona State University | USA |
| Gitta Kutyniok | Technische Universität Berlin | Germany |
| Hamid Krim | North Carolina State University | USA |
| Matilde Marcolli | California Institute of Technology | USA |
| Francois Meyer | University of Colorado | USA |
| Robert Nowak | University of Wisconsin | USA |
| Abbas Ourmazd | University of Wisconsin | USA |
| Randy Paffenroth | Rensselaer Polytechnic Institute | USA |
| José Perea | Michigan State University | USA |
| Chris Peterson | Colorado State University | USA |
| Ali Pezeshki | Colorado State University | USA |
| Louis Scharf | Colorado State University | USA |
| Clayton Shonkwiler | Colorado State University | USA |
| Michael Wakin | Colorado School of Mines | USA |

**Format**

The Workshop was organized into five half-day sessions comprised of presentations by selected participants interspersed with liberal discussion periods. It concluded on the third day with a comprehensive wrap-up session from which much of this report has been drawn. The presentations were as follows.

- Richard Baraniuk, *Image Articulation Manifolds*
- Hamid Krim, *Bi-Sparsity Pursuit: A Paradigm for Robust Subspace Recovery*
- Abbas Ourmazd, *Extracting and Understanding Signals by Geometry*
- Chris Peterson, *Grassmannians, Flags, and Data*
- Babak Hassibi, *Comparison Lemmas and Convex Optimization*
- Gregory Chirikjian, *Signal Detection and Sensor Calibration via Lie Groups*
- Alfred Hero, *Restoring Geometric Insights*
- Clayton Shonkwiler, *Applications of Geometry to Constrained Random Walks and Polymer Models*
- Matilde Marcolli, *Geometric Models for Linguistics*
- José Perea, *Sliding Windows and Persistence: Applications of topological methods to signal analysis*
- Ali Pezeshki, *Distribution of Fisher Information after Random Compression*
- Rob Nowak, *Theory and Applications of Ordinal Embedding*
- Stephen Howard, *The Geometry of Invariants for Generalized Coherence Tests*
- Franois Meyer, *A Graph Metric for the Structural Analysis of Dynamic Networks*

Two additional talks were scheduled, but had to be canceled due to the presenters' travel complications:

- João Xavier, *Some Applications of Riemannian Geometry in Signal Processing*
- Gunnar Carlsson, *The Shape of Complex Data*

## References


[1] P. -A. Absil, R. Mahony, and R. Sepulchre, "Riemannian geometry of Grassmann manifolds with a view on algorithmic computations," Acta Applicandae Mathematicae, vol 80, no 2, pp 199-220, Jan 2004.

[2] S. Amari, "Information Geometry and its Applications," Springer-Verlag, 2016.

[3] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," Neural Comput., vol. 13, pp 1373-1396, 2003.

[4] M. Belkin and P. Niyogi, "Towards a Theoretical Foundation for Laplacian-Based Manifold Methods," Neural Comput., vol. 13, pp 1373-1396, 2003.

[5] J. Benesty, J. Chen, and Y. Huang, Microphone Array Signal Processing, Springer-Verlag, 2008.

[6] M. Bernstein, V. de Silva, J. C. Langford, and J. B. Tenenbaum, "Graph approximations to geodesics on embedded manifolds," December 20, 2000

[7] S. Boyd and L. Vandenberghe, Convex Optimization, Cambridge University Press Cambridge, 2004.

[8] D. S. Broomhead, R. Indyk, A. C. Newell, and R. C. Rand, "Local adaptive Galerkin bases for large-dimensional dynamical systems," Nonlinearity, vol 4, no 2, p 159, 1991.

[9] E. K. P. Chong and S. H. Zak, An Introduction to Optimization, 4th Edition, Wiley, Hoboken, NJ, 2013.

[10] A. Clausen and D. Cochran, Non-parametric multiple channel detection in deep ocean noise, in Proc. Conf. Rec. of the 31st Asilomar Conference on Signal, Systems and Computers, vol. 1, Oct. 1997, pp. 850854.

[11] D. Cochran, H. Gish, and D. Sinno, A geometric approach to multiple- channel signal detection, IEEE Trans. Signal Process., vol. 43, no. 9, pp. 20492057, Sep. 1995.

[12] R. R. Coifman and S. Lafon, "Diffusion maps," Appl. Comp. Harmon. Anal., vol. 21, pp 5-30, 2006.

[13] R. R. Coifman, et al., "Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps, PNAS, vol. 102, no. 21, May 24, 2005.

[14] T. F. Cox and M. A. A. Cox, Multidimensional Scaling. London: Chapman and Hall, 1994.

[15] A. Edelman, T. A. Arias, S. T. Smith, The geometry of algorithms with orthogonality constraints," SIAM J. Matrix Anal. Appl. 20 (2) (1998) 303353.

[16] Fan R. K. Chung, Spectral Graph Theory, CBMS Conf on Recent Advances in Spectral Graph Theory, AMS, Providence, RI, 1994.

[17] R. R. Hagege and J. M. Francos, Parametric estimation of affine transformations: An exact linear solution, J. Math. Imag. Vis., vol. 37, no. 1, pp. 116, Jan. 2010.

[18] R. R. Hagege and J. M. Francos, "Universal manifold embedding for geometrically deformed functions," IEEE Trans. Inform. Th., vol. 62, no. 6, June 2016.

[19] D. Giannakis, P. Schwander, and A. Ourmazd, "The symmetries of image formation by scattering. I. Theoretical framework," Optics Express, vol. 20, no. 12, 4 June 2012.

[20] S. D. Howard and S. Sirianunpiboon, "Passive Radar Detection using multiple transmitters," Asiloar Conf on Signals, Systems, and Computers," Pacific Grove, CA, Nov 2013.

[21] S. D. Howard, S. Sirianunpiboon, and D. Cochran, "An exact Bayesian detector for multistatic passive radar," Asilomar Conf Signals, Systems, and Computeres, Nov. 2016.

[22] A. Leshem and A.-J. van der Veen, Multichannel detection of gaussian signals with uncalibrated receivers, IEEE Signal Process. Letters, vol. 8, pp. 120122, 2001.

[23] Y. Ma, A differential geometric approach to multiple view geometry in spaces of constant curvature, International Journal of Computer Vision 58 (1), pp 3753, 2004

[24] K. V. Mardia, J. T. Kent, and J. M. Bibby, Multivariate Analysis. New York: Academic, 1979.

[25] T. Marrinan, J. R. Beveridge, B. Draper, M. Kirby, C. Peterson, "Finding the subspace mean or median to fit your need," in *Proc. of Computer Vision and Pattern Recognition* (CVPR), 2014.

[26] T. A. Palka, "Bounds and algorithms for subspace estimation on Riemannian quotient submanifolds, Ph.D. Dissertation, University of Rhode Island, 2017.

[27] M. D. Plumbley, "Lie Group Methods for Optimization with Orthogonality Constraints."

[28] M. D. Plumbley, "Geometrical Methods for Non-negative ICA: Manifolds, Lie Groups and Toral Subalgebras," Department of Electronic Engineering, Queen Mary University of London, Mile End Road, London E1 4NS, UK.

[29] D. Ramirez, J. Via, I. Santamaria, and L. L. Scharf, Detection of spatially-correlated Gaussian time series, IEEE Trans Signal Processing, vol 58, no 10, pp 5006-5015, Oct 2010.

[30] R. T. Rockafellar, Convex Analysis, Princeton University Press, Princeton NJ, 1970.

[31] S. T. Roweis and L. K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science, vol. 290, no. 5500, pp. 23232326, Dec. 2000.

[32] I. Santamaria, L.Ł. Scharf, C. Peterson, M. Kirby, J. Francos,"An order fitting rule for optimal subspace averaging," 2016 IEEE Workshop on Statistical Signal Processing, Palma de Majorca, 26-29 June, 2016.

[33] I. J. Schoenberg, "Metric spaces and completely monotone functions," Annals of Mathematics, vol. 39, no. 4, Oct 1938

[34] P. Schwander, D. Giannakis, C.˝ Yoon, and A. Ourmazd, "The symmetries of image formation by scattering. II. Applications," Optics Express, vol. 20, no. 12, 4 June 2012.

[35] F. J. Solis, "Geometry of local adaptive Galerkin bases," Applied Mathematics and Optimization," vol 41, pp 331-342, 2000.

[36] S. T. Smith, L. L. Scharf, and L. T. McWhorter, "Intrinsic quadratic performance bounds on manifolds," 2006 IEEE Internl Conf on Acoust, Speech, and Signal Proc, Toulouse, FR.

[37] S. T. Smith, “Covariance, subspace, and intrinsic Cramer-Rao bounds,” IEEE Trans Signal Proc, vol 53, no 5, pp 1610-1630, 2005.

[38] S. Suvorova, S. Howard, and B. Moran, “Bayesian recursive estimation on the rotation group,” *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal processing*, pp. 6411–6415, May 2013.

[39] J. B. Tenenbaum, V. de Silva, and J. C. Langford, “A global geometric framework for nonlinear dimensionality reduction,” Science, vol. 290, no. 5500, pp. 23192323, Dec. 2000.

[40] P. Turaga, A. Veeraraghavan, A. Srivastava, and R. Chellapa,“ Statistical computations on Grassman and Steiefel manifolds for image and video-based recognition,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol 33, n0 11, pp 2273-2286, Nov 2011.

[41] R. J. Vaccaro, “Applications of Subspace Estimation in Array Signal Processing,” IEEE UASP Workshop in Uncerwater Acoustic Signal Processing,” West Greenwich, R. I. October 4-6, 2017.

[42] J. A. Vizoso, “Geometry of Manifolds from Covariance Analysis,” Ph.D. Preliminary Examination, CColorado State University, Fort Collins, CO, Sept 5, 2017.

[43] R. A. Wijsman, “Invariant Measures on Groups and their Use in Statistics,” Lecture Notes-Monograph Series, Inst Mathematical Statistics, Hayward, CA, 1990.

[44] R. A. Wijsman, “Proper action in steps, with application to density ratios of maximal invariants,” The Annals of Statistics, vol.13, no.1, pp.395402, 1985.