My research focuses on the interaction among the fields of signal processing, data analysis and machine learning, using tools from approximation theory, harmonic analysis and optimization. Signal processing and data analysis have always relied on utilizing prior knowledge of the signals (or data) being processed, from Shannon-Nyquist sampling of bandlimited signals, to compressive sensing of sparse signals, to recovering low-rank matrices from incomplete measurements. On the other hand, the recent advances in signal processing and machine learning also heavily depend on the development of optimization techniques and the availability of massive amounts of computational resources. As the dimensionality increases, the curse of dimensionality brings new challenges both statistically and computationally. To meet the challenges of big data, we rely on exploiting inherent structures, developing scalable computational tools as well as optimization-oriented problem analysis (such as the optimization geometry for nonconvex problems).

Completed Research

Nonconvex Optimization

Unlike the objective functions of convex optimizations that have simple landscapes, the objective functions of general nonconvex programs have much more complicated landscapes which prevents a number of efficient iterative algorithms from converging to a global minimum. Recent advances in optimization imply a number of iterative optimization algorithms (such as alternating minimization, gradient descent, and the the trust region method) can either converge to a global minimum with a good initialization or a local minimum from a random (or an arbitrary) initialization given that the objective function satisfies the so-called strict saddle property [1–3].

We have considered a set of low-rank matrix optimizations where we factorize the unknown matrix \( X \) into \( UV^T \), and optimize over the \( n \times r \) and \( m \times r \) matrices \( U \) and \( V \) rather than the \( n \times m \) matrix \( X \). In particular, for a general and yet well-conditioned objective function whose restricted strong convexity and smoothness constants are comparable, we show that the factored objective function has no spurious local minima and obeys the strict saddle property if the original problem admits a low-rank solution [4]. General nuclear-norm regularized optimization [5] has a similar geometric property which ensures global convergence of a number of iterative optimization algorithms. If we apply these results to the matrix sensing problem, the number of measurements needed matches the information-theoretically optimal sampling complexity. We also have shown [6] that the matrix factorization and sensing problems further satisfy the so-called robust strict saddle property, ensuring global convergence of many local search algorithms in polynomial time.

Subspace Modeling on the Continuum

In problems such as time of arrival estimation in matched filtering, radar signal processing and feature detection, the signals obey a parameterized subspace model in which the signals of interest are inherently low-dimensional and live in a union of subspaces, but the choice of subspace is controlled by a small number of continuous-valued parameters.

The fast Slepian transform. A fundamental and representative element in signal processing is the sampled sinusoid

\[ e_f := [e^{j2\pi f_0} e^{j2\pi f_1} \cdots e^{j2\pi f(n-1)}]^T \in \mathbb{C}^n \]

which is a length-\( n \) vector parameterized by the digital frequency \( f \in [-\frac{1}{2}, \frac{1}{2}] \). The optimal subspace for representing a set of sampled sinusoids \( \{e_{f,W} : -W \leq f \leq W\} \) (which appears naturally in the discrete vector one obtains when collecting a finite set of uniform samples from a bandlimited analog signal) in the least-squares sense is spanned by the time-limited version of the discrete prolate spheroidal sequences (DPSS’s) [7].

However, due to the high computational complexity of projecting onto the DPSS basis (also known as the Slepian basis) by naive matrix-vector multiplication, this representation is often overlooked in favor of the fast Fourier transform (FFT). We [8, 9] have shown that there exist fast constructions for computing approximate

---

1By the factorization approach, the number of degrees of freedom in \( U \) and \( V \) almost match the number of degrees of freedom for \( n \times m \) rank-\( r \) matrices.
projections onto the leading DPSS basis elements. The complexity of the resulting algorithms is comparable to the FFT.

**ROAST: Rapid orthogonal approximate Slepian transform.** Due to the fact that windowing in the time domain will spread out the spectrum in the frequency domain, the discrete Fourier transform (DFT) suffers from “frequency leakage” when used to represent a finite-length vector arising from a bandlimited signal with narrowband spectrum, or even a pure sinusoid. We \[10, 11\] have shown that this “frequency leakage” can be easily overcome by constructing an orthobasis of few elements which together with the partial DFT matrix compactly captures most of the energy in oversampled bandlimited signals. We also give non-asymptotic results to guarantee that the proposed basis not only provides a very high degree of approximation accuracy in an MSE sense for bandlimited sample vectors, but also that it can provide high-quality approximations of all sampled sinusoids within the band of interest.

**Approximating sampled multiband signals.** It is natural to use a dictionary formed by concatenating a collection of modulated DPSS vectors for representing the discrete vector one obtains when collecting a finite set of uniform samples from a multiband analog signal. But is this dictionary the optimal one? We \[12\] studied the angle between the subspaces spanned by this dictionary and an optimal dictionary, and we concluded that the multiband modulated DPSS dictionary—which is simple to construct and more flexible than the optimal dictionary in practical applications—is nearly optimal for representing multiband sample vectors.

As a stylized application in through-the-wall radar imaging, we \[13–15\] have shown that the multiband modulated DPSS dictionary can not only efficiently mitigate the wall clutter with the compressive measurements, but also help detect the extend target positions.

**Super-resolution in SAR imaging: Analysis with the atomic norm.** Motivated by through-the-wall radar imaging, we \[16\] have investigated the estimation of point target positions via atomic norm minimization, which provides higher resolution than traditional back-projection imaging and sparse reconstruction techniques based on discretized dictionaries.

**The eigenvalue distribution of discrete time-frequency limiting operators.** A slightly different alternative to the DPSS vectors are the periodic discrete prolate spheroidal sequences (PDPSSs), which form an optimal orthobasis for the subspace parameterized by the frequency in the DFT domain. We \[17\] established new nonasymptotic results on the eigenvalue distribution of the corresponding discrete time-frequency localization operators. We also characterized the spectrum of submatrices of the DFT matrix, which is of independent interest in signal processing.

**Asymptotic equivalence of circulant and Toeplitz matrices.** A basic question involved in using local subspace approximations is what is the appropriate dimension of the subspace fit. Via the Karhunen-Loève (KL) transform, we know that the effective subspace dimension is determined by the effective rank of a covariance matrix corresponding to a randomly drawn signal over the parameters in the range of interest. In some cases, the covariance matrix turns out to be a Toeplitz matrix. For this case, we \[18\] provided practical and efficiently computable estimates of individual eigenvalues of Toeplitz matrices by showing the individual asymptotic convergence of the eigenvalues of Toeplitz and circulant matrices. Our results suggest that instead of directly computing the eigenvalues of a Toeplitz matrix, one can compute a fast spectrum approximation using the FFT.

**Dictionary Learning and Sensing Matrix Optimization**

The availability of rich, large-scale datasets has made it possible to learn a dictionary or transform from the representative data (aka training dataset) when the explicit information about the signal to be proceeded is not known a priori. In dictionary learning (which belongs to unsupervised feature learning in machine learning), a widely used prior for the data is the notion of “sparsity”, which has different implications for different models including the synthesis model, analysis model and transform model. In the synthesis model, a signal is said to be sparse under a dictionary $\Psi$ if the signal can be expressed as a linear combination of few atoms from the dictionary, while the other two models focus more on the transform coefficients (i.e., $\Psi^H x$).

Motivated by the importance of mutual coherence in sparse signal recovery and compressive sensing, we \[19, 20\] considered learning a dictionary which is incoherent and simultaneously provides a sparse rep-
representation for the training data. In [21], we demonstrated the underlying equivalence between analysis dictionary learning and transform learning.

A random sensing matrix is widely utilized in compressive sensing and dimensionality reduction mostly because it admits strong theoretical guarantees and is also easy to generate. For a given dictionary, we demonstrated it is also possible and beneficial to design a sensing matrix adaptive to the dictionary based on minimizing the mutual coherence.

Future Work

Nonconvex optimization. Tremendous progress has recently been made for nonconvex optimization in both computational tools and theoretical guarantees for many optimization algorithms. But there are still many challenging open problems in the area of nonconvex optimization. For example, most of the recent theoretical guarantees are established for gradient-descent based algorithms (like [1]); what guarantees can we establish for other type of algorithms like coordinate descent (including alternating minimization which is widely utilized by practitioners)? How to efficiently solve constrained optimization problems with similar theoretical guarantees as for the unconstrained ones? Also, numerical results show simple gradient descent can not only escape the saddle points but rather escape the “stuck region” near saddle points and converge to a local minimum efficiently for strict saddle functions. Based on the current result in [1], can we obtain further convergence guarantees for gradient descent by characterizing the neighborhoods of saddle points? It is also observed that a number of gradient descent based algorithms (like accelerated gradient descent) can converge to a local minimum with random initializations.

On the other hand, in some cases, the nonconvex approach could even yield better sampling complexity than that needed via convex approach. For example, in recovering a length-$n$ vector with $s$ non-zero elements from phaseless measurements, it is proved that $O(\frac{s^2}{\log n})$ measurements are needed via convex relaxation schemes [23]. The fact that the sampling complexity does not match the information theoretical lower bound (which is $O(s)$) is mainly because using convex relaxation can do no better for simultaneously structured models (like sparse and low-rank matrices) [24]. For the sparse phase retrieval problem, it is proved [25] that the nonconvex approach can efficiently recover the signal from a number of measurements on the order of the number of degrees of freedom. From this point of view, applying a nonconvex approach for many other structured models has beneficial both in computational complexity and information-theoretically optimal sampling complexity.

Unified framework for subspace modelling on the continuum. Rather than attempt to discretize the parameters and use existing signal processing techniques (a program potentially fraught with difficulty, as the sparsity generally does not translate directly into the discrete domain), our essential research goal is to develop a general framework for natural parameterized subspace models. In particular, we aim to 1) construct a subspace for approximating (almost) all the signals controlled by a small number of continuous-valued parameters ranging within some certain interval; 2) develop rigorous, theoretically-backed techniques for computing projections onto and orthogonal to these subspaces. By developing an appropriate basis to economically represent the signals of interest, one can apply effective tools developed for subspace models and the sparse recovery framework for signal processing. In the processing of building local subspace fits, we will also provide an answer to the effective dimensionality of such signals.

Learning compressive sensing systems. As an adaptive dictionary gives better sparse representation for a set of signals or data, it is expected that learning a sensing matrix from training data would also give better performance for dimensionality reduction and compressive sensing. An interesting potential approach is to utilize the deep learning framework for designing compressive sensing systems, where the dictionary can be either explicit or hidden within the sparse recovery procedure. Aside from the mean squared error, many other measures such as mutual incoherence can be incorporated into the deep learning framework to avoid some bad local solutions and speed up the training stage. This reminds us that it would be beneficial to incorporate the notion of models and many other structures into the deep learning framework to get better results, especially when the available resources (such as training data and computational power) are limited.
References


