Research Statement

Praneeth Narayanamurthy

1 Research Overview

There has been an explosion in the amount of data being generated across numerous domains such as healthcare, bioinformatics, commerce etc. This explosion can be attributed to the accelerated development of efficient acquisition, transmission, storage, and computational mechanisms. Eliciting *useful* information from this "data-deluge" has been the quest of high-dimensional statistical signal processing, and machine learning (ML). A common observation in this quest is the fact that although the observed data is high-dimensional, typically, most real world data (approximately) lies in a significantly lower-dimensional ambient space. Motivated by this observation, the ML community has actively developed provable, and efficient (in terms of sample-complexity, robustness, and computational-complexity) algorithms to learn this underlying latent space. However, there are two key challenges that need to be addressed in the data-processing pipeline (i) *data corruption:* either in terms of missing data, or in terms of outliers that seep in; and (ii) *cost of data acquisition.* The overarching goal of my Ph.D. research was to address the first limitation: I developed provable, robust, and efficient algorithms for low-dimensional structural recovery problems. I exploited the temporal nature of data (typically observed in domains such as healthcare, and commerce) and thus utilized the rich-literature on time-series analysis. In my current postdoctoral work, I am working to address the second challenge: to *leverage* this inherent low dimensionality of data to design efficient sampling schemes for problems such as source-localization and derivative-free function optimization.

2 Subspace Tracking

The problem of *estimating and tracking* a low-dimensional linear subspace from time-series data has garnered significant interest in the signal processing and automatic control communities in the past three decades [1–3]. However, to the best of our knowledge, all convergence results for this problem were either asymptotic, assumed a *single* underlying subspace, or only provided partial guarantees. My graduate research provided key advances towards resolving this long standing problem. Furthermore, I designed and analyzed provable, and non-asymptotic algorithms that can deal with multiple subspaces, and are also robust.

Model-Based Robust Subspace Tracking. I first considered a linear superposition of a low-rank (r-dimensional subspace in n dimensions) and sparse structure for spatiotemporal data [4, 5]. In the offline setting, this is commonly referred to as Robust Principal Component Analysis [6] (RPCA) and has received significant attention in the literature. The dynamic version of this problem is referred to as the *Robust Subspace Tracking* (RST) problem [7], but there were no complete, provable, guarantees for the problem . This model is applicable for problems such as Video Layering (separating a video into foreground and background layers), social-network structure identification, and recommendation system design to name a few. Here, I assumed that the underlying subspaces can change every so often (in a piecewise constant fashion), but imposed a constraint on *how* the changes occur . Formally, I assume that whenever the subspace changes, only 1 out of the r directions changes (see Fig. 1 for a simple schematic). I developed an algorithm dubbed simple-Recursive Projected Compressive Sensing (s-ReProCS) based on the ReProCS framework [8] to track the (a) sparse outliers, (b) the *true* low-dimensional data, and (c) the underlying subspace. I showed that using a "good enough" initialization, and under standard RPCA/RST assumptions: incoherence of the subspaces, a lower-bound on most outlier magnitudes, mild statistical assumptions on the subspace coefficients, and the subspace change model mentioned above, s-ReProCS is able obtain ε-accurate estimates[1] (of the low-dimensional and sparse vectors, and the underlying subspaces) using just $\mathcal{O}(r \log n \log(1/\varepsilon))$ samples. Additionally, I showed that by exploiting the statistical assumptions, s-ReProCS can tolerate a larger fraction of outliers per-row (increase from $\mathcal{O}(1/r)$ to $\mathcal{O}(1)$) in the sparse matrix. Finally, the running time of our algorithm is equal to (upto constant factors) running a rank r-vanilla SVD on the data matrix.

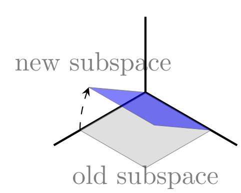

Figure 1: Subspace change example in 3D with $r = 2$.

[1] in this document, $\varepsilon \in (0, 1)$ is the final desired accuracy

Nearly Optimal Robust Subspace Tracking. Next, I relaxed the restrictive subspace change assumption by proposing a modified algorithm [9–11] referred to as Nearly Optimal RST (NORST). The key insight that allows one to relax this assumption is that "re-estimating" an r-dimensional subspace is just as good as "adding" one changed direction to a (well estimated) r-dimensional subspace. Here, I assumed that the subspace either (a) follows a piecewise constant model but when the subspace does change, it can do so arbitrarily or (b) the subspace is allowed to change at each time, but only by a *little* and have abrupt changes at certain times. I showed that under standard RST assumptions, NORST is able to obtain ε-accurate estimates of the underlying subspaces using just $\mathcal{O}(r \log n \log(1/\varepsilon))$ samples. Even with perfect data, estimating a r dimensional subspace in n dimensions requires r samples, and thus our upper bound is only logarithmic factors away from the lower bound. Akin to the previous results, our proposed method has an improved outlier tolerance, and the running time (upto constant factors) is the same as running a vanilla SVD on the data matrix.

A critical component of the proof involved developing finite sample guarantees for Principal Components Analysis (PCA) in data-dependent noise. Although PCA has been exhaustively studied in the last several decades, most results assume that the noise is uncorrelated (if not independent) with the true data. I considered the setting where the noise can be correlated with the data and provided finite sample guarantees for the SVD solution. In particular, I assumed that the noise depends linearly on the data. A key application of the PCA in (sparse) data-dependent noise is in ReProCS based RST. I built upon [12] and provided improved sample-complexity analysis [13, 14].

Robust Subspace Tracking with Missing Entries. In the next generalization, I studied the above problem in a setting with missing data [15–17] that can occur due to a plethora of reasons such as transmission and communication failures. The *static* version of this problem is commonly referred to as *Matrix Completion* (*Robust Matrix Completion* in the presence of outliers). While (Robust) Matrix Completion has been extensively studied in the literature, to the best of our knowledge, there were no finite-sample, *complete* guarantees for the Subspace Tracking with missing entries (STMiss) problem. I showed that through a simple modification of my approach for solving RST, the proposed method can also deal with missing data. In particular, I showed that under mild and easily interpretable assumptions, the proposed method is fast, sample efficient, and provably correct. Furthermore, while most Matrix Completion methods require that the set of observed entries follow uniform random sampling scheme (i.e., each entry is observed independently of all others with a fixed probability), my algorithm can tolerate deterministic patterns. The tradeoff is that my method requires a larger number of observed entries[2]. The critical component in the proof was a careful application of the previously developed results and a result for sparse recovery with partial support knowledge.

Federated Over-Air Subspace Learning. Implicitly, in all previous work, I assumed that all the data is available at a single computing center. However, in several practical settings, it is more natural to consider a decentralized setting such that the data is collected in a distributed fashion. Owing to the sheer magnitude, sharing the raw data to a central server is communication-inefficient, and also raises privacy concerns. To alleviate this, I analyzed the previously discussed RST-miss problem, but in a federated, over-air setting [18]. Federated Learning [19] refers to a paradigm wherein the data is distributed across K peer nodes and the nodes can only share *summary statistics* of their raw data with the central server. For the communication protocol, I considered the newly developed wireless over-air transmission modality that allows for synchronous transmission by the peer nodes as it is K times time- and bandwidth- efficient. However, the central server only receives a sum (superposition) of the individual transmissions and the received sum is corrupted by additive channel noise. I develop an algorithm dubbed Federated Over-Air Subspace Tracking with Missing data (Fed-OA-STMiss) to solve RST while obeying the constraints of federated, over-air communication. In particular, I showed that under standard RST assumptions and i.i.d. Gaussian *iteration noise*, with high probability, Fed-OA-STMiss computes an ε-accurate subspace estimate (an r dimensional subspace in n dimensions) using just $\mathcal{O}(r \log n \log(1/\varepsilon))$ samples. I also showed that the running time is equal to (upto constant factors) that of performing a rank-r vanilla SVD on the data matrix. To the best of our knowledge, this is the first provable, and robust, unsupervised learning algorithm in a federated setting. In an ongoing research project, I am endeavoring to derive a guarantee for differentially private RST in a distributed setting.

Low-Rank Phase Retrieval. While the bulk of my research has focused on *linear inverse* problems, I also have examined the Low-Rank Phase Retrieval (LRPR) problem. Phase Retrieval [20] refers to the problem of recovering an unknown n-dimensional signal from *magnitudes* of its linear measurements. It is an important problem encountered in several applications such as X-ray crystallography, spectroscopy, and Fourier ptychography. The primary approach to developing sub-linear (in terms of time- and sample-complexity) is to impose a structure on the unknown signal. The most commonly used structures are sparsity and low-rankness. The applicability of

[2] An equivalent tradeoff can also observed for RPCA wherein, if the support of the sparse matrix is chosen in a probabilistic manner, the tolerable fraction of outliers is larger.

both these structures are justified in problems such as Ptychography of live-biological specimens from microscopic systems. In our work [21, 22], we considered the latter. Low Rank Phase Retrieval is defined as follows: recover an $n \times q$ matrix of rank r from a different and independent set of m phaseless (magnitude-only) random linear projections of each of its columns. Our proposed solution, Alternating Minimization for Low-Rank Phase Retrieval (AltMinLowRaP), is provably fast, and sample efficient. Our guarantee shows that AltMinLowRaP solves LRPR to ε-accuracy, with high probability, as long as $mq = \mathcal{O}(nr^4 \log(1/\varepsilon))$, the measurement matrices are i.i.d. standard Gaussian, and the ground truth matrix is right-incoherent. To the best of our knowledge, this is the *first* provable algorithm for LRPR.

3 Ongoing Work and Future Research Agenda

Ongoing Work. The problem of *designing active sampling strategies* has been a burgeoning research area in the statistics and Machine Learning community [23, 24]. While there is a vast literature in this area, there are several key domains where this strategy has yet been unexplored, and that is the focus of my current postdoctoral work. Specifically, the problem of localizing a source from very few energy measurements [25] has been *reduced* to a matrix completion problem with unimodality constraints. I am exploring how to provably localize the source with the fewest possible *intelligently sampled* energy measurements.

I argue that an understanding of the *active source localization* also provides a way of performing derivative-free optimization [26] by querying just a few function values. Identifying critical points of (highly non-convex) functions is a fundamental problem in Machine Learning optimization that could potentially impact future directions of research.

Future Work. My research experience places me in a unique situation to tackle numerous unexplored research problems. A few key directions are the following.

1. **Extending the Notion of Robustness:** A promising future direction is to extend the notion of robustness that goes beyond the well analyzed model of *gross outliers* – adversarial robustness [27, 28], and connections to more insightful generalization bounds [29]. Extending our work thus far to handle these types of *noise* in an unexplored territory that is challenging and also has utility in understanding the fundamental hardness of low-dimensional structure recovery problems.

2. **Exploit Modern Low-Dimensional Structures:** Another direction that is closely related to my research is to consider more general, and possibly non-linear underlying structures – a natural assumption in the last few years has been to assume that data (approximately) lies in a nonlinear, low-dimensional space modeled as the *range of Generative Networks* [30]. This non-linearity adds a challenging layer of complexity to the problems previously considered.

3. **Deep Neural Network Training:** Finally, I am also interested in understanding the theoretical properties of training deep neural networks. Although there has been a deluge of papers in this area [31,32], the direction that I personally find interesting is based on recent work [33] that connects proof ideas from the low-rank matrix recovery to provably training shallow networks.

Lab and Potential Funding Sources. I intend to apply for grants from the National Science Foundation and in particular the Communications and Information Foundations track. In my research career, I have assisted Prof. Namrata Vaswani in the writing of successful NSF proposals. I intend to apply for the NSF CRII and CAREER grants based on the directions previously outlined and I believe that my background and technical expertise puts me in an excellent position to successfully secure such opportunities. I intend to pursue funding opportunities from defense sources such as AFOSR, ARO, and ONR. Finally, I also intend to develop industrial collaborations and endeavor to pursue early-career faculty grants from companies such as Google, Facebook, Amazon, Nvidia, and others.

References

[1] P. Comon and G. H. Golub, "Tracking a few extreme singular values and vectors in signal processing," *Proceedings of the IEEE*, vol. 78, no. 8, pp. 1327–1343, 1990.

[2] B. Yang, "Asymptotic convergence analysis of the projection approximation subspace tracking algorithms," *Signal Processing*, vol. 50, pp. 123–136, 1996.

[3] N. Vaswani, T. Bouwmans, S. Javed, and **P. Narayanamurthy**, "Robust subspace learning: Robust pca, robust subspace tracking, and robust subspace recovery," *IEEE Signal Processing Magazine*, vol. 35, no. 4, pp. 32–55, 2018.

[4] **P. Narayanamurthy** and N. Vaswani, "Provable dynamic robust pca or robust subspace tracking," in *IEEE International Symposium on Information Theory*, 2018, pp. 376–380.

[5] **P. Narayanamurthy** and N. Vaswani, "Provable dynamic robust pca or robust subspace tracking," *IEEE Transactions on Information Theory*, vol. 65, no. 3, pp. 1547–1577, 2019.

[6] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. ACM*, vol. 58, no. 3, 2011.

[7] N. Vaswani and **P. Narayanamurthy**, "Static and dynamic robust pca and matrix completion: A review," *Proceedings of the IEEE*, vol. 106, no. 8, pp. 1359–1379, 2018.

[8] C. Qiu, N. Vaswani, B. Lois, and L. Hogben, "Recursive robust pca or recursive sparse recovery in large but structured noise," *IEEE Transactions on Information Theory*, pp. 5007–5039, August 2014.

[9] **P. Narayanamurthy** and N. Vaswani, "A fast and memory-efficient algorithm for robust pca (merop)," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2018, pp. 4684–4688.

[10] **P. Narayanamurthy** and N. Vaswani, "Nearly optimal robust subspace tracking," in *International Conference on Machine Learning*, 2018, pp. 3701–3709.

[11] **P. Narayanamurthy** and N. Vaswani, "Fast robust subspace tracking via pca in sparse data-dependent noise," *to appear, IEEE Journal of Special Areas in Information Theory*, 2020.

[12] N. Vaswani and H. Guo, "Correlated-pca: Principal components' analysis when data and noise are correlated," in *NIPS*, 2016.

[13] N. Vaswani and **P. Narayanamurthy**, "Finite sample guarantees for pca in non-isotropic and data-dependent noise," in *Allerton Conf. on Commun., Control, and Comput.*, 2017.

[14] N. Vaswani and **P. Narayanamurthy**, "Pca in sparse data-dependent noise," in *IEEE International Symposium on Information Theory*, 2018, pp. 641–645.

[15] **P. Narayanamurthy**, V. Daneshpajooh, and N. Vaswani, "Provable subspace tracking with missing entries," in *International Symposium on Information Theory*, 2019, pp. 1867–1871.

[16] **P. Narayanamurthy**, V. Daneshpajooh, and N. Vaswani, "Provable memory-efficient online robust matrix completion," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2019, pp. 7918–7922.

[17] **P. Narayanamurthy**, V. Daneshpajooh, and N. Vaswani, "Provable subspace tracking from missing data and matrix completion," *IEEE Transactions on Signal Processing*, pp. 4245–4260, 2019.

[18] **P. Narayanamurthy**, N. Vaswani, and A. Ramamoorthy, "Federated over-the-air subspace learning from incomplete data," *(Submitted to IEEE Transactions on Signal Processing (arXiv preprint arXiv:2002.12873)*, 2020.

[19] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong, "Federated machine learning: Concept and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 2, pp. 1–19, 2019.

[20] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," in *NIPS*, 2013, pp. 2796–2804.

[21] S. Nayer, **P. Narayanamurthy**, and N. Vaswani, "Phaseless pca: Low-rank matrix recovery from column-wise phaseless measurements," in *International Conference on Machine Learning*, 2019, pp. 4762–4770.

[22] S. Nayer, **P. Narayanamurthy**, and N. Vaswani, "Provable low rank phase retrieval," *IEEE Transactions on Information Theory*, 2020.

[23] B. Settles, "Active learning literature survey," *Tech. Report – University of Wisconsin-Madison Department of Computer Sciences*, 2009.

[24] D. Cohn, L. Atlas, and R. Ladner, "Improving generalization with active learning," *Machine learning*, vol. 15, no. 2, pp. 201–221, 1994.

[25] J. Chen and U. Mitra, "Unimodality-constrained matrix factorization for non-parametric source localization," *IEEE Transactions on Signal Processing*, vol. 67, no. 9, pp. 2371–2386, 2019.

[26] A. R. Conn, K. Scheinberg, and L. N. Vicente, *Introduction to derivative-free optimization*, SIAM, 2009.

[27] N. Carlini, A. Athalye, N. Papernot, W. Brendel, J. Rauber, D. Tsipras, I. Goodfellow, A. Madry, and A. Kurakin, "On evaluating adversarial robustness," *arXiv preprint arXiv:1902.06705*, 2019.

[28] J. Cohen, E. Rosenfeld, and Z. Kolter, "Certified adversarial robustness via randomized smoothing," in *International Conference on Machine Learning*. PMLR, 2019, pp. 1310–1320.

[29] D. Stutz, M. Hein, and B. Schiele, "Disentangling adversarial robustness and generalization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6976–6987.

[30] A. Bora, A. Jalal, E. Price, and A. G. Dimakis, "Compressed sensing using generative models," in *International Conference on Machine Learning*. PMLR, 2017, pp. 537–546.

[31] H. Sedghi and A. Anandkumar, "Provable methods for training neural networks with sparse connectivity," *arXiv preprint arXiv:1412.2693*, 2014.

[32] Kenji Kawaguchi, "Deep learning without poor local minima," *arXiv preprint arXiv:1605.07110*, 2016.

[33] B. Neyshabur, Z. Li, S. Bhojanapalli, Y. LeCun, and N. Srebro, "Towards understanding the role of over-parametrization in generalization of neural networks," *arXiv preprint arXiv:1805.12076*, 2018.