Research Summary

Namrata Vaswani, Iowa State University

My research lies at the intersection of machine learning for high dimensional problems, signal and information processing and applications in video, big-data analytics and bio-imaging. In the last several years I have worked on designing and analyzing online algorithms for various high-dimensional structured data recovery problems and on demonstrating their usefulness in dynamic magnetic resonance imaging (MRI) and video analytics.

In the last two decades, the sparse recovery problem, or what is now more commonly referred to as compressive sensing (CS), has been extensively studied, see for example [1, 2, 3, 4, 5, 6] and later works. More recently various other structured data recovery problems, such as low-rank or low-rank plus sparse matrix recovery, have also been studied in detail. Sparse recovery or CS refers to the problem of recovering a sparse signal from a highly reduced set of its projected measurements. Low-rank matrix recovery (or matrix completion) refers to the problem of recovering a low-rank matrix from a subset of its entries. Low-rank plus sparse matrix recovery refers to recovering a sparse matrix and a low-rank matrix from their sum or from a subset of entries of their sum. Many medical imaging techniques image cross-sections of human organs non-invasively by acquiring their linear projections one at a time and then reconstructing the image from these projections. For example, in magnetic resonance imaging (MRI), one acquires Fourier projections one at a time. For all these applications, the ability to accurately reconstruct using fewer measurements directly translates into reduced scan times and hence sparse recovery methods have had a huge impact in these areas. A key motivating application for low-rank matrix recovery is the Netflix problem where the goal is to recover the movie ratings’ matrix when most users do not rate all movies. Low-rank plus sparse recovery occurs in video layering (separating video into foreground and background layers) which is a key first step to simplify various video analytics algorithms, as well as in outlier-robust recommendation system design, functional MRI based brain activity detection, or anomaly detection in dynamic networks.

In recent years, in my group, we have developed provably accurate online (recursive) solutions to two structured data recovery problems: (a) online (recursive) sparse matrix recovery, and (b) online (recursive) sparse plus low-rank matrix recovery. Problem (a) can also be understood as a problem of recursive recovery of sparse vector (signal) sequences. Problem (b) can be interpreted either as a problem of online robust principal components’ analysis (PCA) / online robust matrix completion / robust subspace tracking, or as a problem of online sparse recovery in large but structured (low-dimensional and dense) noise. We describe our recent work on the above problems in the next two sections.

1 Online Sparse Matrix Recovery (Recursive Recovery of Sparse Signal Sequences)

The goal of this work was to design and analyze recursive algorithms for causally reconstructing a time sequence of (approximately) sparse signals from highly undersampled measurements. The signals were assumed to be sparse in some transform domain referred to as the sparsity basis and their sparsity patterns (support set of the sparsity basis coefficients) could change with time. One key application where this problem occurs is dynamic MRI for real-time medical applications such as interventional radiology and MRI-guided surgery, or in functional MRI to track brain activation changes. MRI is a technique for cross-sectional imaging that sequentially acquires the 2D Fourier projections of the cross-section to be re-
constructed. Since MR data is acquired one Fourier projection at a time, the ability to accurately reconstruct using fewer measurements directly translates into reduced scan times. Shorter scan times along with online (causal) and fast (recursive) reconstruction can enable real-time imaging of fast changing physiological phenomena, thus making many interventional MRI applications practically feasible. Cross-sectional images of the brain, heart, larynx or other human organ images are piecewise smooth, and thus approximately sparse in the wavelet domain. In a time sequence, their sparsity pattern changes with time, but quite slowly [7, 8, 9]. This simple fact is the key reason why our proposed algorithms can achieve accurate reconstruction from much fewer measurements. In recent years, the static sparse recovery or compressive sensing (CS) problem has been thoroughly studied [1, 2, 3, 4, 5, 6]. But most existing algorithms for the dynamic problem, e.g. [10, 11], just use CS solutions to jointly reconstruct the entire time sequence in one go. This is a batch solution and as a result (a) it is very slow in recovering a long sequence and (b) its memory requirement increases linearly with the sequence length. The alternative - solving the CS problem at each time separately (simple-CS) - is online, fast and memory-efficient, but needs many more measurements (higher scan time).

To the best of our knowledge, our recent work [7, 8, 12, 13] proposed the first solutions for recursively reconstructing sparse signal sequences using much fewer measurements than those needed for accurate recovery using simple-CS methods. The computational and storage complexity of our proposed algorithms is only as much as that of simple-CS, but their reconstruction performance is much better. For example, in results reported in [13], for reconstructing simulated dynamic MRI sequences, our proposed algorithm (modified-CS) needed only about 16-19% measurements for accurate recovery while a simple-CS solution (basis pursuit solved using the L1-Magic software) needed about twice as many measurements for the same recovery error. Existing MR scanners use the inverse Fourier transform which needs 100% measurements.

Our work described below relies on the following easily verifiable assumptions.

1. The sparsity patterns (support set of the sparse basis vectors) of natural signal/image sequences usually change “slowly” over time [9, 13].

2. In most cases, the values of the nonzero coefficients also change gradually over time [14].

When using only assumption [1] above, the above problem can be reformulated as one of sparse reconstruction with partially “known” support. The support estimate from the previous time serves as the “known” part. We can further improve the proposed algorithm by also using assumption [2].

[KF-CS and LS-CS] The key idea of our first approach (LS-CS-residual or LS-CS) is to solve the \( \ell_1 \) minimization problem with the observation replaced by the least squares (LS) observation residual computed using the “known” part of the support [7, 12]. The LS residual measures a signal that has much fewer large components compared to the original signal (it is what can be called a “sparse-compressible” signal). As a result, when fewer measurements are available, the LS-CS reconstruction error is lower than that of simple-CS methods. By also using fact [2] we can replace the LS residual by the Kalman filtering residual (KF-CS) [7]. This improves the reconstruction particularly when the number of measurements is too few even for LS-CS.

[Modified-CS] Even though LS-CS and KF-CS improved reconstruction accuracy over simple-CS methods, they could not be used for “exact” reconstruction from fewer noise-free measurements. This led to our second and more powerful approach - modified-CS [13]. Denote the “known” part of the support by \( T \). Modified-CS tries to find the signal that is sparsest outside of \( T \) while satisfying the data constraint. If \( T \) has small error (few extras and misses), modified-CS can achieve provably exact reconstruction from very few measurements [13]. By also using slow signal value change, one can design regularized modified-CS which also constrains the change of the nonzero coefficient values along \( T \) [14, 15]. In numerical experiments as well as in experiments with simulated dynamic MR imaging, modified-CS significantly outperformed existing work at the time [13].
Under the practically valid assumption of slowly changing support, we have also been able to prove (a) exact reconstruction and (b) error stability over time, using fewer measurements than what simple-CS methods need. We explain these results next.

[Exact reconstruction] We have obtained exact reconstruction conditions for Modified-CS and have argued that it achieves exact reconstruction under much weaker sufficient conditions (using much fewer noise-free measurements) than those needed to provide the same guarantee for simple CS [13].

[Stability over time] For both LS-CS and for modified-CS for noisy measurements, we have shown the following. Under fairly mild assumptions (bounded noise, a large enough minimum coefficient magnitude increase rate, and weaker requirements on the number of measurements than what is needed for obtaining small bounds on simple-CS error, both the support recovery errors and the reconstruction errors, are “stable”, i.e. they remain bounded by time-invariant and small values at all times [12][16][17].

Stability is critical for any recursive algorithm since it ensures that the reconstruction error does not blow up over time. In practice, say in real-time MRI, provably stable and small error means that the reconstructed images seen by the interventional radiologist or surgeon are always a close approximation of reality.

2 Online Sparse + Low-Rank Matrix Recovery (Recursive or Online Robust PCA)

Principal Components Analysis (PCA) is a widely used tool for dimension reduction. Given a matrix of data $D$, PCA seeks to recover a small number of orthogonal directions that contain most of the variability of the data. This is typically accomplished by performing a singular value decomposition (SVD) of $D$ and retaining the left singular vectors corresponding to the largest singular values. A limitation of SVD is that it is highly sensitive to outliers. "Outlier" is a loosely defined term that refers to any corruption that is not small compared to the true data vector and that occurs occasionally. As suggested in [18], an outlier can be nicely modeled as a sparse vector whose nonzero entries can have any magnitude.

Robust PCA, which refers to the problem of PCA in the presence of outliers, has been a well-studied problem for a long time and many useful heuristics have been proposed for it, e.g., see [19] and references therein. However none of the practically useful algorithms from older literature come with performance guarantees. Moreover, they also fail to extract the outlier support when the outlier magnitude is not too large. In recent works [20, 21], Candés et. al. and Chandrasekaran et. al. posed robust PCA as a problem of separating a sparse matrix $S$ and a low-rank matrix $L$ from their sum, $Y := S + L$. They introduced the Principal Components Pursuit (PCP) program and obtained performance guarantees for it under mostly mild assumptions [20], [21]. Later work by Hsu et. al. [22] improved the result of [21]. Since then, there has been much later work on provably accurate robust PCA solutions but all of it has been for batch methods.

We consider an online or recursive version of the robust PCA problem where we seek to separate observed data vectors, $y_t$, into low dimensional components, $\ell_t$, and sparse components, $s_t$, as they arrive, using the previous estimates, rather than re-solving the entire problem at each time $t$. To start with, it is assumed that an accurate estimate of the subspace of $\ell_0$ is available. An application where this type of problem is useful is in video analysis [19]. Imagine a video sequence that has a distinct background and foreground. An example might be a surveillance camera where a person walks across the scene. If the background does not change very much, and the foreground is sparse (both practical assumptions), then separating the background and foreground can be viewed as a robust PCA problem. Various video analytics tasks such as video retrieval or activity recognition are significantly more reliable and quicker when applied to the extracted layer of interest rather than to the raw video. Other applications where the robust PCA problem occurs include detecting anomalous behavior in dynamic social networks; recommendation system design in the presence of outliers and missing data; detection of brain activation patterns from functional MRI (fMRI) sequences. In most of these applications, an online solution is desirable.

In recent work [23][24][25][26][27], we have introduced a novel online solution called ReProCS; obtained
a correctness result for it under mild assumptions; and shown that it significantly outperforms most existing batch as well as recursive methods for video layering of difficult videos. We explain this work below.

[ReProCS for online robust PCA] We introduce a novel online robust PCA solution called Recursive Projected CS (ReProCS) [23, 24, 26]. It uses two extra, but usually practically valid, assumptions, beyond what PCP [20, 21] and other batch methods need. First, it assumes that an accurate estimate of the subspace of $\ell_t$ is available at the initial time (easy to obtain using a short sequence of background-only video frames) and second, that its subspace is either fixed or changes slowly over time (valid assumption for slow changing backgrounds). Like PCP, it also needs a denseness (non-sparseness) assumption on the left singular vectors of $L$. Its key idea is as follows. At any time $t$, it first projects the measurement vector, $y_t$, perpendicular to the current estimate of the subspace of $\ell_t$. Because of the slow subspace change assumption, this approximately nullifies $\ell_t$. Recovering $s_t$ from these projected measurements then becomes a standard sparse recovery problem in small noise. The denseness assumption ensures that $s_t$ can be recovered accurately from the projected measurements by ell-1 minimization, or any other CS technique. The recovered $s_t$ is used to estimate $\ell_t = y_t - s_t$ by subtraction and this is then used to update the subspace estimate using a modification of standard PCA that we call projection-PCA. Because of how the sparse recovery is done, ReProCS requires a significantly weaker assumption on how frequently the support of the sparse vectors needs to change compared with PCP.

[Video Analytics Application] In foreground and background layering experiments with real video sequences containing large and slow moving or occasionally static objects (infrequent support change), as long as a short initial background-only sequence is available, ReProCS significantly outperforms batch robust PCA approaches, such as PCP and RSL [19], as well as online algorithms such as incremental RSL [28], and a recently introduced algorithm called GRASTA [29]. The experimental comparisons are available in [26, 30] and at [http://www.ece.iastate.edu/~hanguo/PracReProCS.html](http://www.ece.iastate.edu/~hanguo/PracReProCS.html).

[Correctness result for ReProCS] In recent work [27, 25], we have shown that, as long as the ReProCS algorithm parameters are set appropriately, a good-enough estimate of the initial subspace is available, slow subspace change (quantified in a certain way) holds, the basis vectors for the subspaces are dense enough, and there is a certain amount of support change at least every so often, then the support of $s_t$ can be exactly recovered with high probability. Moreover, the sparse and low-rank matrix columns can be recovered with bounded and small error; and the subspace recovery error decays to a small value within a short delay of a subspace change time. To the best of our knowledge, our result [27] is among the first correctness results for an online (recursive) robust PCA algorithm or equivalently for sparse plus low-rank matrix recovery.

Online algorithms are needed for real-time applications; and even for offline applications, they are faster and need less storage compared to batch techniques. Moreover, online approaches can provide a natural way to exploit temporal dependencies in the dataset. In our case, we show that ReProCS uses slow subspace change to allow for significantly more correlated support sets of the sparse vectors than do the various results for PCP [20, 21, 22]. Our result for ReProCS allows a constant fraction of nonzeros in any row of $S$, while the result for PCP from [21, 22] only allows this fraction to be proportional to $1/r$ where $r$ is the rank of $L$. The result for PCP from [20] uses an even stronger assumption - it requires uniformly randomly selected support sets. Of course this advantage for ReProCS comes at a cost; it needs a tighter bound on the rank-sparsity product compared to what is needed by the PCP result from [20].

For obtaining our result, we needed to develop new proof techniques and we expect that these can potentially be applied to various other problems involving a PCA step where the data and noise are correlated. New techniques were needed because all existing correctness results were only for batch robust PCA approaches. Also, our proof could not just be a combination of a sparse recovery result and a result for PCA, because in the PCA step for ReProCS, the error $e_t = \hat{\ell}_t - \ell_t$ is correlated with $\ell_t$. Almost all existing work on finite sample PCA assumes that the error between the measured and true data vectors is uncorrelated with the true data, see e.g. [31] and references therein.
References


