DEMIXING SPARSE SIGNALS FROM NONLINEAR OBSERVATIONS

Mohammadreza Soltani, Chinmay Hegde

ECpE Department, Iowa State University, Ames, IA, 50010

ABSTRACT

Signal demixing is of special importance in several applications ranging from astronomy to computer vision. The goal in demixing is to recover a set of signals from their linear superposition. In this paper, we study the more challenging scenario where only a limited number of nonlinear measurements of the signal superposition are available. Our contribution is a simple, fast algorithm that recovers the component signals from the nonlinear measurements. We support our algorithm with a rigorous theoretical analysis, and provide upper bounds on the estimation error as well as the sample complexity of demixing the components (up to a scalar ambiguity). We also provide a range of simulation results, and observe that the method outperforms a previous algorithm based on convex relaxation.

1. INTRODUCTION

1.1. Motivation

In several applications in signal and image processing, data analysis, and statistics, the problem of *demixing* assumes a special importance. In the simplest case, the goal is to recover a pair of signals from their linear superposition. Mathematically, consider the observation model:

$$x = \Phi w + \Psi z, \tag{1.1}$$

where $x \in \mathbb{R}^n$ represents the observations, $\Phi, \Psi \in \mathbb{R}^{n \times n}$ are orthonormal bases, and $w, z \in \mathbb{R}^n$ are the coefficients of the constituent signals. The demixing problem involves reliably recovering the constituents w and z from the observations x. This problem has been studied in several applications [1, 2]. In image processing and computer vision applications, demixing methods impact tasks such as background-foreground separation [3, 4], while in astronomical imaging, demixing methods can enable the separation of astronomical features (stars/galaxies) from background sky phenomena [5, 6].

In general, demixing is an ill-posed problem since the number of unknowns ($2n$) is greater than the number of equations (n). Reliable recovery of the constituent signals is very difficult in general, and one inevitably has to assume some sort of *incoherence* between the constituent signals (or more specifically, between the corresponding bases Φ and Ψ) However, even if we assume that the signals are incoherent enough, demixing poses additional challenges under other, more stringent observation models. Assume, for instance, a measurement model given by $y = A(\Phi w + \Psi z)$, where $A \in \mathbb{R}^{m \times n}$ denotes a linear measurement operator, and $w, z \in \mathbb{R}^n$. We focus on the highly under-determined case where $m \ll n$. The under-determined case has received considerable recent attention in signal processing and high-dimensional statistics [7, 8, 9].

In this case, it might seem impossible to recover the components x and z since A possesses a nontrivial null space. Once again, this problem is highly ill-posed, and additional information about the *structure* of the components is necessary. For example, in the application of separating foreground and background images, the foreground image can be modeled as sparse while the background image can be modeled as low-rank. Such modeling assumptions on the constituent signals have been shown to enable successful demixing [10, 11, 12, 13, 14].

In this paper, we focus on an *even* more challenging scenario, where the measurements y are *nonlinear* functions of the signal superposition. That is, $y = f(Ax)$ where f is a nonlinear function (sometimes called a *link* function), and x is the signal superposition. Nonlinear link functions have long been studied in the statistics literature, and have been the recent focus of attention in the context of signal acquisition and recovery [15, 16, 17, 18]. We are interested in the problem of signal demixing in such scenarios. Specifically, we consider the model of $y = f(Ax)$ where $A \in \mathbb{R}^{m \times n}$ is a random matrix with $m \ll n$ and $x = \Phi w + \Psi z$. As part of our structural assumptions, we suppose that the signal coefficients w and z are s-sparse in the bases Φ and Ψ, respectively (i.e., w and z contain no more than s nonzero entries). Furthermore, f may be non-smooth, non-invertible, or even unknown. Then, the goal is to recover (modulo a scale ambiguity[1]) an estimate of w and z, given measurements, $\{y_i\}_{i=1}^m$, and knowledge of the matrix A and the bases Φ and Ψ.

1.2. Our Contributions

In this paper, we provide a simple, fast algorithm (that we call ONESHOT) to demix the constituent signals w and z, given the nonlinear observations y, the measurement operator A, and the bases Φ and Ψ. Our algorithm is non-iterative, does not require explicit knowledge of the link function f, and works even in the case where $m \ll n$.

We support our algorithm with a rigorous theoretical analysis. Our analysis reveals upper bounds on the *sample complexity* of demixing with nonlinear observations. (Here, sample complexity denotes a sufficient number of observations for reliable recovery of w and z modulo a scaling factor). In particular, we prove that the sample complexity of ONESHOT to achieve a constant estimation error is given by $m = O(s \log \frac{n}{s})$. See Section 3 for details.

Moreover, we provide numerical evidence for the efficiency of our methods. In particular, we compare the performance of ONESHOT with a previous method proposed in [19] based on convex optimization. Simulation results show that ONESHOT outperforms this convex method significantly in both demixing efficiency as well as running time, and consequently makes it an attractive choice in large-scale problems.

Our analysis of ONESHOT is based on the pioneering approach of [19]. Our contribution is to extend this idea for the (more general) nonlinear demixing problem, and to characterize the role of incoherence and how it effects the recovery process. Our technique is based

[1] A scale ambiguity can be unavoidable for certain types of nonlinear observations; for example, quantization to single-bit measurements [15] removes all amplitude information of a given signal.

on a geometric argument, and leverages the *Gaussian mean width* for the set of sparse vectors, which is a statistical measure of complexity of a set of points in a given space. Due to page-limit constraints, we merely state our theoretical claims, and refer the reader to [20] for full proofs.

1.3. Prior Work

While signal demixing is a classical problem, it has been the focus of a lot of research in recent years. In a class of image processing techniques known as morphological component analysis (MCA), 2D images are treated as the superposition of structured components, and the goal is to recover these components using computational techniques [5, 6]. Similar demixing problems arise in applications such as restoration of clipped audio signals and identification of faulty computer memory [21].

Research in the linear demixing problem has also considered a variety of signal models. The *robust PCA* problem [10, 11, 12] involves the separation of low-rank and sparse matrices from their sum. This idea has been used in several applications ranging from video surveillance to sensor network monitoring. In machine learning applications, the separation of low-rank and sparse matrices has been used for latent variable model selection [22] as well as the robust alignment of multiple occluded images [23]. Another type of signal model is the *low-dimensional manifold* model. In [13, 14], the authors proposed a greedy iterative method for demixing signals, arising from a mixture of known manifolds by iterative projections onto the manifolds. We refer to [1] for a comprehensive discussion on linear demixing with various applications.

Linear demixing falls under the category of *linear inverse problems*, an instance of which is the focus of *compressive sensing* [7, 8, 9]. A generalization of this framework includes the recently proposed 1-*bit compressive sensing* problem [15]. Here, the linear measurements of a given signal are quantized in the extreme fashion such that the measurements are binary (± 1) and only comprise the sign of the signal coefficient; therefore, the amplitude of the signal is completely discarded by the quantization operator. Formally, the challenge is to recover the signal x, given observations $y = \text{sign}(Ax)$ where A is a measurement matrix. It turns out that if the signal x is sparse enough, then this recovery problem can be done efficiently using convex optimization [16, 17]. In a related (and somewhat more challenging) problem, the phase information of the signal coefficients may be lost, and we have only access to the amplitude information of the signal [18].

None of the above works have (explicitly) considered the problem of *demixing* signals from *nonlinear* observations. In this paper, we address this problem and provide a fast algorithm with provable guarantees.

2. PRELIMINARIES

In this section, we establish the formal mathematical model and introduce some definitions. Throughout, the symbol $\|\cdot\|$ refers to the ℓ_2-norm. Consider a signal $x \in \mathbb{R}^n$ that is the superposition of a pair of sparse vectors in different bases, i.e.,:

$$x = \Phi w + \Psi z\,, \tag{2.1}$$

where $\Phi, \Psi \in \mathbb{R}^{n\times n}$ are orthonormal bases, and $w, z \in \mathbb{R}^n$ such that $\|w\|_0 \le s_1$, and $\|z\|_0 \le s_2$. Consider the *Gaussian* observation model:

$$y = f(Ax)\,, \tag{2.2}$$

where $A \in \mathbb{R}^{m\times n}$ is a random matrix with i.i.d. standard normal entries, and f denotes a (possibly unknown) link function which is not necessarily smooth, invertible, or continuous. We define the following quantities:

$$\bar{x} = \frac{\Phi w + \Psi z}{\|\Phi w + \Psi z\|} = \alpha(\Phi w + \Psi z), \tag{2.3}$$

where $\alpha = \frac{1}{\|\Phi w+\Psi z\|}$. Also, we define the set of sparse vectors in the bases Φ and Ψ as follows:

$$K_1 = \{\Phi a \mid \|a\|_0 \le s_1\},$$
$$K_2 = \{\Psi a \mid \|a\|_0 \le s_2\},$$

and we define $K = \{a \mid \|a\|_0 \le s\}$.

A fundamental assumption made by several demixing algorithms is that the sparsifying bases are sufficiently *incoherent* with respect to each other. We quantify the incoherence assumption as follows:

Definition 2.1. *(ε-incoherence). The bases Φ and Ψ are said to be ε-incoherent if:*

$$\varepsilon = \sup_{\substack{\|u\|_0\le s,\ \|v\|_0\le s\\ \|u\|_2=1,\ \|v\|_2=1}} |\langle \Phi u, \Psi v\rangle|.$$

We note that the parameter ε is related to the more well-known *mutual coherence* of a matrix. Indeed, if we consider the matrix $\Gamma = [\Phi\ \Psi]$, then the mutual coherence of Γ is given by $\gamma = \max_{i\neq j} |(\Gamma^T\Gamma)_{ij}|$, and one can show that $\varepsilon \le s\gamma$ [9].

Finally, we define a statistical measure of complexity of a set of signals, following the approach of [19].

Definition 2.2. *(Local Gaussian mean width). For a given set $K \in \mathbb{R}^n$ and a scalar $t > 0$, the local Gaussian mean width (or simply, the local mean width) at level t is defined as:*

$$W_t(K) = \mathbb{E} \sup_{\substack{x,y\in K,\\ \|x-y\|\le t}} \langle g, x - y\rangle.$$

where $g \sim \mathcal{N}(0, I_{n\times n})$.

3. PROPOSED DEMIXING ALGORITHM

Having defined the above quantities, we now present our proposed demixing algorithm. Recall that we wish to recover components w and z (modulo a scale ambiguity), given the nonlinear measurements y and the matrix A. Our proposed algorithm, that we call ONESHOT, is described in pseudocode form below.

Algorithm 1 ONESHOT

Inputs: Basis matrices Φ and Ψ, measurement matrix A, measurements y, sparsity level s.
Outputs: Estimates $\widehat{x} = \Phi\widehat{w} + \Psi\widehat{z}$, $\widehat{w} \in K_1$, $\widehat{z} \in K_2$

$\widehat{x}_{\text{lin}} \leftarrow \frac{1}{m} A^T y$ {form linear estimator}
$b_1 \leftarrow \Phi^* \widehat{x}_{lin}$ {forming first proxy}
$\widehat{w} \leftarrow \mathcal{P}_s(b_1)$ {Projection on set K_1}
$b_2 \leftarrow \Psi^* \widehat{x}_{lin}$ {forming second proxy}
$\widehat{z} \leftarrow \mathcal{P}_s(b_2)$ {Projection on set K_2}
$\widehat{x} \leftarrow \Phi\widehat{w} + \Psi\widehat{z}$ {Estimating $\widehat{x}$}

Here and below, for simplicity we assume that the sparsity levels s_1 and s_2, specifying the sets K_1 and K_2, are equal, i.e., $s_1 = s_2 = s$. The algorithm (and analysis) transparently extends to the case of unequal sparsity levels. Also, we have used the following *projection* operators:

$$\widehat{w} = \mathcal{P}_s(\Phi^* \widehat{x}_{\text{lin}}), \quad \widehat{z} = \mathcal{P}_s(\Psi^* \widehat{x}_{\text{lin}}).$$

Here, $\mathcal{P}_s$ denotes the projection onto the set of (canonical) s-sparse signals K, and can be implemented by hard thresholding. Observe that ONESHOT is *not* an iterative algorithm; this fact enables us to achieve a fast running time.

The mechanism of ONESHOT is simple. At a high level, ONESHOT first constructs a *linear estimator* of the target superposition signal, denoting by $\widehat{x}_{\text{lin}} = \frac{1}{m} A^T y$, and then projects $\widehat{x}_{\text{lin}}$ onto the constraint sets K_1 and K_2. Finally, it combines these two projections to obtain the final estimate of the target superposition signal.

We now provide a performance analysis of ONESHOT. Our proofs follow the geometric argument provided in [19], specialized to the demixing problem. In particular, we derive an upper bound on the estimation error of the *component* signals w and z, modulo scaling factors. In our proofs, we use the following result from [19], restated here for completeness.

Lemma 3.1. *(Quality of linear estimator). Given the model in Equation 2.2, the linear estimator, $\widehat{x}_{lin}$, is an unbiased estimator of $\bar{x}$ up to constants. That is, $\mathbb{E}(\widehat{x}_{lin}) = \mu\bar{x}$ and: $\mathbb{E}\|\widehat{x}_{lin} - \mu\bar{x}\|_2^2 = \frac{1}{m}[\sigma^2 + \eta^2(n-1)]$, where $\mu = \mathbb{E}(y_1\langle a_1, \bar{x}\rangle)$, $\sigma^2 = Var(y_1\langle a_1, \bar{x}\rangle)$, $\eta^2 = \mathbb{E}(y_1^2)$.*

We now state our main result, with proof provided in [20].

Theorem 3.2. *(Main theorem). Let $A \in \mathbb{R}^{m\times n}$ be a random matrix with i.i.d. standard normal entries, and let y be given the nonlinear measurements. Let $\Phi, \Psi \in \mathbb{R}^{n\times n}$ be bases with incoherence parameter ε. Suppose we use* ONESHOT *to produce estimates $\widehat{w}$ and $\widehat{z}$. Then, the estimation error for the first component w for all $t > 0$ satisfies the upper bound:*

$$\mathbb{E}\|\widehat{w} - \mu\alpha w\|_2 \leq t + \frac{2}{\sqrt{m}}\left(4\sigma + \eta\frac{W_t(K)}{t}\right) + 8\mu\varepsilon \qquad (3.1)$$

The coefficients μ, σ, and η are given in Lemma 3.1. A similar bound satisfies for the second component z.

The authors of [16, 19] provide upper bounds on the local mean width $W_t(K)$ of the set of s-sparse vectors. In particular, for any $t > 0$ they show that $W_t(K) \leq Ct\sqrt{s\log(2n/s)}$ for some absolute constant C. By plugging in this bound and letting $t \to 0$, we can combine components $\widehat{w}$ and $\widehat{z}$ which gives the following:

Corollary 3.3. *With the same assumptions as Theorem 3.2, the error of nonlinear estimation incurred by the final output $\widehat{x}$ satisfies the upper bound:*

$$\mathbb{E}\|\widehat{x} - \mu\bar{x}\|_2 \leq \frac{4}{\sqrt{m}}\left(4\sigma + C\eta\sqrt{s\log(2n/s)}\right) + 16\mu\varepsilon. \quad (3.2)$$

The constants σ, η, μ depend on the nature of the nonlinear function f, and are often rather mild. For example, if $f(x) = \text{sign}(x)$, then explicit calculations reveal that:

$$\mu = \sqrt{\frac{2}{\pi}} \approx 0.8,\ \sigma^2 = 1 - \frac{2}{\pi} \approx 0.6,\ \eta^2 = 1.$$

Corollary 3.3 has some interesting implications. First, in contrast with demixing algorithms for traditional (linear) observation models, our estimated signal $\widehat{x}$ can differ from the true signal x by a scale factor. Next, suppose we fix $\delta > 0$ as a small constant, and suppose that the incoherence parameter $\varepsilon = c\delta$ for some constant c, and that the number of measurements scales as:

$$m = O\left(\frac{s}{\delta^2}\log\frac{n}{s}\right).$$

Then, the (expected) estimation error $\|\widehat{x} - \mu\bar{x}\| \leq O(\delta)$. In other words, the *sample complexity* of ONESHOT is given by $m = O(\frac{1}{\delta^2} s\log(n/s))$, which resembles results for the linear observation case [14, 19][2].

We also observe that the estimation error in (3.2) is upper-bounded by $O(\varepsilon)$. This is meaningful only when $\varepsilon \ll 1$, or when $s\gamma \ll 1$. Per the Welch Bound [9], the mutual coherence γ satisfies $\gamma \geq 1/\sqrt{n}$. Therefore, Corollary 3.3 provides non-trivial results only when $s = o(\sqrt{n})$. This is consistent with the *square-root bottleneck* in sparse approximation [24].

The main theorem obtains a bound on the *expected value* of the estimation error. We can derive a similar upper bound that holds with high probability, provided that the measurements are generated from a *sub-Gaussian* distribution. We defer these results to [20].

4. NUMERICAL RESULTS

In this section, we provide some representative numerical experiments for our proposed algorithm. We also compare its performance with a LASSO-type technique for demixing. This method, first proposed in [19], was not explicitly developed in the demixing context but is suitable for our problem. We call this method the *Nonlinear convex demixing with LASSO*, or the NLCDLASSO for short. Casting into our notation from Section 2 and 3, NLCDLASSO solves the following convex problem:

$$\begin{aligned} \min_{z,w} \quad & \left\|\widehat{x}_{\text{lin}} - [\Phi\ \Psi][w; z]\right\|_2 \\ \text{subject to} \quad & \|w\|_1 \leq \sqrt{s}, \quad \|z\|_1 \leq \sqrt{s}. \end{aligned} \qquad (4.1)$$

Here, $\widehat{x}_{\text{lin}}$ denotes the linear estimate of x, $[w; z]$ is a column vector with length $2n$ and s denotes the sparsity level of signals w and z in bases Φ and Ψ, respectively. The constraints in (4.1) are convex penalties reflecting the knowledge that w and z are s-sparse and have unit ℓ_2-norm (since the nonlinearity is unknown, we have a scale ambiguity, and therefore w.l.o.g. we can assume that the underlying signals lie in the unit ball). The outputs of this algorithm are the estimates $\widehat{w}$, $\widehat{x}$, and $\widehat{x} = \Phi\widehat{w} + \Psi\widehat{z}$.

To solve the optimization problem in (4.1), we have used the SPGL1 solver [25, 26]. This solver can handle large scale problems, which is the scenario that we have used in our experimental evaluations. We impose the joint constraint $\|t\|_1 \leq 2\sqrt{s}$ (where $t = [w; z]$ denoting the stacking vector) which is a slight relaxation of the constraints in problem 4.1. The upper-bound of $\sqrt{s}$ in the constraints is a worst-case criterion; for comparison, we also include simulation results with a constraint $\|t\|_1 \leq \varrho$, where ϱ has been tuned to the best of our ability.

Now, we precisely describe the setup of our simulations. First, we generate $w \in \mathbb{R}^n$ (likewise, z) with $n = 2^{20}$ by choosing a

[2]Here, we use the term "sample-complexity" as the number of measurements required by a given algorithm to achieve an estimation error δ. However, we must mention that algorithms for the linear observation model are able to achieve stronger sample complexity bounds that are independent of δ.

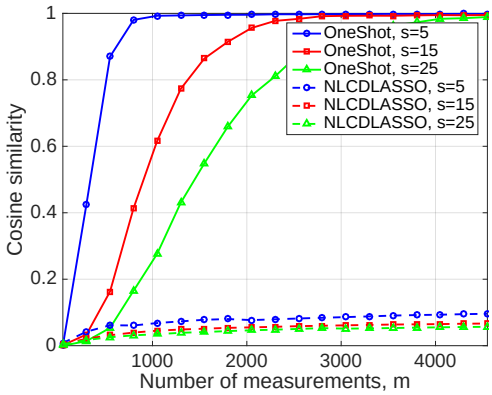

Fig. 1: *Performance of* ONESHOT *and* NLCDLASSO *with* $\|t\|_1 \leq 2\sqrt{s}$ *according to the* COSINE SIMILARITY *for different choices of sparsity level* s *for* $g(x) = sign(x)$.

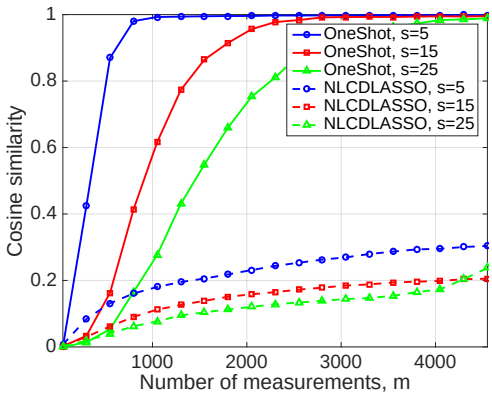

Fig. 2: *Performance of* ONESHOT *and* NLCDLASSO *with* $\|t\|_1 \leq \varrho$ *according to the* COSINE SIMILARITY *for different choices of sparsity level* s *for* $g(x) = sign(x)$.

random support with s nonzero elements and populating the nonzero entries with random ± 1 coefficients. As per the discussion in the Introduction, for successful recovery we require that the constituent signals are sufficiently incoherent. To achieve this, we consider the vector w to be s-sparse in the Haar wavelet basis, and the vector z to be s-sparse in the noiselet basis [27]. As the measurement operator A, we choose a partial DFT matrix. Such matrices are known to have similar recovery performance as random Gaussian matrices, but enable fast numerical operations [28].

As mentioned before, the scale (amplitude) of the underlying signal is irrevocably lost. To measure recovery performance in the absence of scale information, we use the *Cosine Similarity* criterion between x and $\widehat{x}$ to compare the performance of different methods. More precisely, suppose that ONESHOT (or NLCDLASSO) outputs $\widehat{w}$ and $\widehat{z}$, such that $\widehat{x} = \Phi\widehat{w} + \Psi\widehat{z}$. Then, the Cosine Similarity criterion between super position signals is defined as follows:

$$\cos(x, \widehat{x}) = \frac{x^T \widehat{x}}{\|x\|\|\widehat{x}\|}.$$

Figures 1 and 2 illustrate the performance of ONESHOT and NLCDLASSO according to the Cosine Similarity for different choices of sparsity level s when the nonlinear link functions are set to

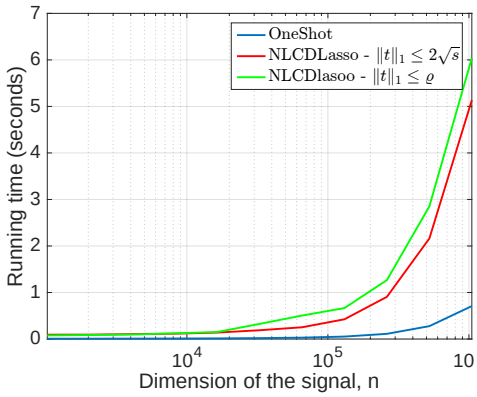

Fig. 3: *Comparison of running times of* ONESHOT *with* NLCDLASSO. *Our proposed algorithm is non-iterative, and in contrast with LASSO-based techniques, has a running time that is* nearly-linear *in the signal dimension* n *for certain structured sparsifying bases.*

$f(x) = \text{sign(x)}$ and we have used both $\|t\|_1 \leq 2\sqrt{s}$ and $\|t\|_1 \leq \varrho$ constraints. The horizontal axis denotes an increasing number of measurements. Each data point in the plot is obtained by conducting a Monte Carlo experiment over 20 trials and averaging the results; in each trial, we set the the parameters (m, n, s), generate a new random measurement matrix A, record the nonlinear observations, and reconstruct the underlying components using the differerent algorithms.

As we can see from the plots in Figs. 1 and 2, we obtain similar results for both choices of link functions. Notably, the performance of NLCDLASSO is worse than ONESHOT for any fixed choice of m and s. Even when the number of measurements increases (for example, at $m = 4550$), we see that ONESHOT outperforms NLCDLASSO by a significant degree. In this case, NLCDLASSO is at least 70% worse in terms of signal estimation quality, while ONESHOT recovers the (normalized) signal perfectly. This result indicates the inefficiency of NLCDLASSO in the context of nonlinear demixing.

Finally, we contrast the running time of both algorithms, illustrated in Figure 3. In this experiment, we measure the wall-clock running time of the two recovery algorithms (ONESHOT and NLCDLASSO), by varying signal size x from $n = 2^{10}$ to $n = 2^{20}$. Here, we set $m = 500$, $s = 5$, and the number of Monte Carlo trials to 20. Also, the nonlinear link function is considered as $g(x) = \text{sign}(x)$. As we can see from the plot, ONESHOT is at least 6 times faster than NLCDLASSO when the size of signal equals to 2^{20}. Overall, ONESHOT is efficient even for large-scale nonlinear demixing problems. We mention that in the above setup, the main computational costs incurred in ONESHOT involve a matrix-vector multiplication followed by a thresholding step, both of which can be performed in time that is *nearly-linear* in terms of the signal length n for certain choices of A, Φ, Ψ. In particular, we experimentally verified that varying the sparsity level does not have any effect in the running time.

5. ACKNOWLEDGEMENTS

This work was supported in part by the National Science Foundation under the grants CCF-1566281 and IIP-1632116.

6. REFERENCES

[1] M. McCoy and J. Tropp, "Sharp recovery bounds for convex demixing, with applications," *Foundations of Comp. Math.*, vol. 14, no. 3, pp. 503–567, 2014.

[2] M. McCoy, V. Cevher, Q. Dinh, A. Asaei, and L. Baldassarre, "Convexity in source separation: Models, geometry, and algorithms," *IEEE Sig. Proc. Mag.*, vol. 31, no. 3, pp. 87–95, 2014.

[3] H. Guo, C. Qiu, and N. Vaswani, "An online algorithm for separating sparse and low-dimensional signal sequences from their sum," *IEEE Trans. Sig. Proc.*, vol. 62, no. 16, pp. 4284–4297, 2014.

[4] C. Qiu, N. Vaswani, B. Lois, and L. Hogben, "Recursive robust PCA or recursive sparse recovery in large but structured noise," *IEEE Trans. Inform. Theory*, vol. 60, no. 8, pp. 5007–5039, 2014.

[5] M. Elad, J. Starck, P. Querre, and D. Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *Appl. Comput. Harmonic Analysis*, vol. 19, no. 3, pp. 340–358, 2005.

[6] J. Bobin, J. Starck, J. Fadili, Y. Moudden, and D. Donoho, "Morphological component analysis: An adaptive thresholding strategy," *IEEE Trans. Image Proc.*, vol. 16, no. 11, pp. 2675–2681, 2007.

[7] E. Candès, "Compressive sampling," in *Proc. Int. Congress of Math.*, Madrid, Spain, Aug. 2006.

[8] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[9] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*, vol. 1, Springer.

[10] E. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of the ACM*, vol. 58, no. 3, pp. 11, 2011.

[11] V. Chandrasekaran, S. Sanghavi, P. Parrilo, and A. S. Willsky, "Sparse and low-rank matrix decompositions," in *Proc. Allerton Conf. on Comm., Contr., and Comp.*, 2009, pp. 962–967.

[12] V. Chandrasekaran, S. Sanghavi, P. Parrilo, and A. S. Willsky, "Rank-sparsity incoherence for matrix decomposition," *SIAM J. Opt.*, vol. 21, no. 2, pp. 572–596, 2011.

[13] C. Hegde and R. Baraniuk, "SPIN : Iterative signal recovery on incoherent manifolds," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, July 2012.

[14] C. Hegde and R. Baraniuk, "Signal recovery on incoherent manifolds," *IEEE Trans. Inform. Theory*, vol. 58, no. 12, pp. 7204–7214, Dec. 2012.

[15] P. Boufounos and R. Baraniuk, "1-bit compressive sensing," in *Int. Conf. Info. Sciences and Systems (CISS)*. IEEE, 2008, pp. 16–21.

[16] Y. Plan and R. Vershynin, "Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach," *IEEE Trans. Inform. Theory*, vol. 59, no. 1, pp. 482–494, 2013.

[17] Y. Plan and R. Vershynin, "One-bit compressed sensing by linear programming," *Comm. Pure and Applied Math.*, vol. 66, no. 8, pp. 1275–1297, 2013.

[18] E. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval via wirtinger flow: Theory and algorithms," *IEEE Trans. Inform. Theory*, vol. 61, no. 4, pp. 1985–2007, 2015.

[19] Y. Plan, R. Vershynin, and E. Yudovina, "High-dimensional estimation with geometric constraints," *arXiv preprint arXiv:1404.3749*, 2014.

[20] M. Soltani and C. Hegde, "Fast algorithms for demixing sparse signals from nonlinear observations," *arXiv preprint arXiv:1608.01234*, 2016.

[21] C. Studer, P. Kuppinger, G. Pope, and H. Bölcskei, "Recovery of sparsely corrupted signals," *IEEE Trans. Inform. Theory*, vol. 58, no. 5, pp. 3115–3130, 2012.

[22] V. Chandrasekaran, P. Parrilo, and A. S. Willsky, "Latent variable graphical model selection via convex optimization," in *Proc. Allerton Conf. on Comm., Contr., and Comp.*, 2010, pp. 1610–1613.

[23] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, "Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 34, no. 11, pp. 2233–2246, 2012.

[24] J. Tropp, "On the conditioning of random subdictionaries," *Appl. Comput. Harmon. Anal.*, vol. 25, no. 1, pp. 1–24, 2008.

[25] E. van den Berg and M. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM J. Sci. Comp.*, vol. 31, no. 2, pp. 890–912, 2008.

[26] E. van den Berg and M. Friedlander, "SPGL1: A solver for large-scale sparse reconstruction," June 2007, http://www.cs.ubc.ca/labs/scl/spgl1.

[27] R. Coifman, F. Geshwind, and Y. Meyer, "Noiselets," *Appl. Comput. Harmonic Analysis*, vol. 10, no. 1, pp. 27–44, 2001.

[28] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, 2006.