A GREEDY PURSUIT ALGORITHM FOR SEPARATING SIGNALS FROM NONLINEAR COMPRESSIVE OBSERVATIONS

Dung Tran, Akshay Rangamani, Trac D. Tran *

Johns Hopkins University Electrical and Computer Engineering Department Baltimore, MD, USA

ABSTRACT

In this paper we study the unmixing problem which aims to separate a set of structured signals from their superposition. In this paper, we consider the scenario in which the mixture is observed via nonlinear compressive measurements. We present a fast, robust, greedy algorithm called *Unmixing Matching Pursuit* (UnmixMP) to solve this problem. We prove rigorously that the algorithm can recover the constituents from their noisy nonlinear compressive measurements with arbitrarily small error. We compare our algorithm to the *Demixing with Hard Thresholding* (DHT) algorithm [1], in a number of experiments on synthetic and real data.

Index Terms— Unmixing, sparse recovery, compressed sensing, nonlinear measurements

1. INTRODUCTION

The problem of *unmixing* two signals from their superposition involves finding the unknown constituents through observations of the mixed signal. In general, this is a challenging problem. Without making further assumptions, this problem is ill posed and we cannot hope to reliably separate the unknown component signals from their superposition. Instead, if the constituent signals do not look similar to each other we can imagine being able to separate them. To formalize this assumption, we assume that each component signal can be linearly expressed by a dictionary of atoms. The dictionary atoms share some common structures that appear in the corresponding component signal, but do not appear in the other constituents of the mixture. In this scenario, we say that the unknown mixture components can be represented by *incoherent* dictionaries [2], [3], [4], [5]

We consider an observation model where the mixture signal is observed via linear compressive samples which are passed through a smooth, monotonic, nonlinear operator $h : \mathbb{R} \to \mathbb{R}$. Furthermore, the observations can be corrupted by dense additive noise. In particular, we wish to recover

Sang (Peter) Chin[†]

Boston University Department of Computer Science Boston, MA, USA

constituent signals from a limited number of *noisy*, *nonlinear*, *compressive* measurements of their superposition:

$$\boldsymbol{y} = h(\boldsymbol{A}(\boldsymbol{u} + \boldsymbol{v})) + \boldsymbol{e}, \tag{1}$$

where $A \in \mathbb{R}^{m \times N}$ is a sensing matrix. The number of observations m is typical far less than the ambient dimension N. $y \in \mathbb{R}^m$ is the observation vector, and $e \in \mathbb{R}^m$ is a zeromean noise vector. We assume that each coordinate of the noise vector is a Gaussian random variable with variance σ^2 , and the coordinates are independent of each other. Since this problem is ill posed in general, we assume that the constituent signals u, v have sparse representations in their dictionaries. We assume that the sensing matrix A, the nonlinear operator h, and the incoherent dictionaries sparsely representing the components are known.

Our contributions: In this paper, we propose a fast and robust iterative algorithm called UnmixMP to unmix component signals under the observation model (1). At a high level, each iteration of the algorithm consists of two main steps. First, it aims to identify a true dictionary atom for each component signal. Second, we refine our estimate of each constituent signal based on those chosen atoms and all corresponding dictionary atoms previously selected. Our algorithm is in the class of greedy pursuit algorithm which have received lots of attention in sparse recovery literature [2], [6], [7], [8], [9]. We prove convergence for UnmixMP in the noisy and noiseless (e = 0) cases. We also prove that the sample complexity to achieve this linear convergence rate is upper bounded by $\mathcal{O}\left(r\log\frac{N}{r}\right)$, where r is the total sparsity level of the component signals. In addition, we support our theoretical analysis by experiments on both synthetic and real image data. We demonstrate that our algorithm is significantly more robust than state-of-the-art unmixing algorithms in this nonlinear setting.

Related Work: The unmixing problem has been studied extensively in signal processing and statistics literature. Examples include morphological component analysis (MCA) in image processing, source separation in audio signals [5], and sparse noise correction in Robust PCA [10]. These classical problems assume a linear observation model in which the

^{*}This work is partially supported by NSF (CCF-1422995).

[†]This work is partially supported by AFOSR (FA9550-12-1-0136).

constituent objects are assumed to be sparse in some weakly correlated dictionaries. In fact when the observation model is linear, McCoy et. al. ([11], [12]) show that it is possible to reliably recover constituent signals from their compressive measurements. Furthermore, a majority of them formulate unmixing as a convex optimization problem, which are typically sensitive to parameter tuning and slower than greedy pursuit methods. Blumensath and Davies propose a gradient pursuit algorithm that is similar to ours in [13], but their algorithm is tailored to minimize the squared loss between linear measurements and the estimated signal, whereas our loss function is different (as explained in Section 2). They do not explore its application to recovery from nonlinear compressive measurements.

The work that is most closely related to ours is that of Soltani et. al. [1]. In their work, the author proposed a variant of the popular iterative hard thresholding (IHT) method to demix component signals in the nonlinear model, and achieve state-of-the-art performance. However, their algorithm requires both the knowledge of sparsity level of component signals and step-size parameter. Furthermore, as shown in Section 4, our algorithm is significantly more robust than theirs in unmixing incoherent signals from Sigmoid $\left(h(x) = \frac{1}{1+e^{-x}}\right)$ and ReLU $(h(x) = \max(0, x))$ compressive observations. While both UnmixMP and the method from [1] are only guaranteed to converge when the nonlinear function is smooth, we show in experiments that UnmixMP is able to recover components from even ReLU measurements. **Notation:** For the sensing matrix A, we let a_j^T denote its j^{th} row. For a dictionary **D**, we refer to it as both the set of dictionary atoms and a matrix whose columns corresponding to the atoms. We denote D_{Ω} as the matrix whose columns are those from D, indexed by Ω . For a set Ω , the number of elements is given by $|\Omega|$. Finally, the projection of a vector z on the subspace spanned by the columns of a matrix Dindexed by Ω is denoted by $\mathcal{P}_{D_{\Omega}} z$.

2. THE UNMIXMP ALGORITHM

In this section, we describe our proposed algorithm, called Unmixing Matching Pursuit (UnmixMP). We first briefly introduce some quantities associated with sparse representation of signals. The sparsity of a signal z is given by $||z||_{0,D}$, which is the smallest number of columns in D that can exactly represent z by a linear combination. We let $\operatorname{supp}_D(z)$ denote the index set of the atoms in D constituting z.

Definition 1. The mutual coherence of Φ , Ψ is given by

$$\mu(\mathbf{\Phi}, \mathbf{\Psi}) = \sup_{||\boldsymbol{x}||_2 = 1, ||\boldsymbol{y}||_2 = 1} |\langle \boldsymbol{\Phi} \boldsymbol{x}, \boldsymbol{\Psi} \boldsymbol{y} \rangle|.$$
(2)

We assume that the constituent signals u and v in the observation model (1) are k and s sparse w.r.t. some dictionaries Φ and Ψ , respectively. We also assume that the two dictionaries Φ , Ψ , are *mutually incoherent* (which means that the mutual coherence is small). In our setup, we assume that the dictionaries are known. One approach to solve the unmixing problem is thus to identify the dictionary atoms constituting the component signals. That can be done by solving the following optimization problem with sparsity constraints [1]:

$$\min_{\boldsymbol{u},\boldsymbol{v}} \quad f(\boldsymbol{u},\boldsymbol{v}) = \frac{1}{m} \sum_{j=1}^{m} \Gamma(\boldsymbol{a}_{j}^{T}(\boldsymbol{u}+\boldsymbol{v})) - y_{j} \boldsymbol{a}_{j}^{T}(\boldsymbol{u}+\boldsymbol{v})$$

s.t. $\|\boldsymbol{u}\|_{0,\boldsymbol{\Phi}} \leq k, \quad , \|\boldsymbol{v}\|_{0,\boldsymbol{\Psi}} \leq s.$ (3)

Here, the real-value function $\Gamma(\cdot)$ is the integral of h, i.e., $\Gamma(t) = \int_{-\infty}^{t} h(z) dz$.

We note that this is not the regular squared loss function that is typically considered. The problem in 3 can be considered as an empirical version of the following problem:

$$\min_{\boldsymbol{u},\boldsymbol{v}} \mathbb{E}\left[\Gamma(\boldsymbol{a}^T(\boldsymbol{u}+\boldsymbol{v})) - y\boldsymbol{a}^T(\boldsymbol{u}+\boldsymbol{v})\right]$$
(4)

which matches the Gaussian noise assumption, as pointed out in [1]. The partial gradients of the loss function in 3 have a nice, easy to compute, closed form:

$$\nabla_{\boldsymbol{u}} f(\boldsymbol{u}, \boldsymbol{v}) = \nabla_{\boldsymbol{v}} f(\boldsymbol{u}, \boldsymbol{v}) = \frac{1}{m} \boldsymbol{A}^T \left(h(\boldsymbol{A}\boldsymbol{u} + \boldsymbol{A}\boldsymbol{v}) - \boldsymbol{y} \right)$$
(5)

To solve the optimization problem (3), we propose a fast and robust greedy pursuit algorithm. Each iteration of the algorithm involves first computing a proxy g which encodes useful information from previous iterations. This proxy is chosen to be the partial gradient of the objective function f(u, v) evaluated at the estimated solution from the previous iteration. As shown in Section 3, the proxy g is most aligned with one of the atoms in each dictionary. We thus project g onto the dictionaries, and extract an atom from each one that is most correlated to it. Finally, we estimate the demixed components by minimizing the loss function, restricting the search space to the dictionary atoms extracted so far. This procedure is detailed in Algorithm 1. We analyze its performance guarantees in the next section.

Remark. Step 2 of the algorithm can be interpreted as a selection step. It is akin to the selection step shared by a majority greedy pursuit algorithms in sparse recovery literature. It can be inferred from Lemma 2 in the next section that each dictionary atom extracted at this step significantly reflects the structures of the corresponding constituent.

Step 4 in the algorithm is an Update step. Lemma 1 in Section 3 implies that the estimated components at this step look more similar to the correct constituents than the previous estimates. Intuitively, this is due to the fact that the selection step reveals more structures in the component signals. This step can be solved with any projected gradient descent type procedure.

Algorithm 1 Unmixing Matching Pursuit (UnmixMP)

Input: Mixture \boldsymbol{y} , sensing matrix \boldsymbol{A} , dictionaries $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, nonlinear operator h, sparsity (k, s) or stopping criterion TOL

Initialization: t = 0, $\Omega_{u}^{0} = \emptyset$, $\Omega_{v}^{0} = \emptyset$ while not converged do 1. $g = \frac{1}{m} A^{T} (h(Au^{t} + Av^{t}) - y)$ 2. $i_{u} = \operatorname{argmin}_{l} ||\mathcal{P}_{\psi_{l}}g||_{2}$ $i_{v} = \operatorname{argmin}_{l} ||\mathcal{P}_{\psi_{l}}g||_{2}$ 3. $\Omega_{v}^{t+1} = \Omega_{u}^{t} \cup \{i_{u}\}$ $\Omega_{v}^{t+1} = \Omega_{v}^{t} \cup \{i_{v}\}$ 4. $(u^{t+1}, v^{t+1}) = \operatorname{argmin}_{u,v} f(u, v)$ s.t. $u \in \operatorname{span}(\Phi_{\Omega_{u}^{t+1}}),$ $v \in \operatorname{span}(\Psi_{\Omega_{v}^{t+1}})$ 5. t = t + 1end while

3. THEORETICAL ANALYSIS OF UNMIXMP

This section rigorously analyzes the performance guarantee of the UnmixMP algorithm. In particular, we first show that when the loss function f(u, v) satisfies certain restricted strong convexity (RSC) and restricted strongly smoothness (RSS) properties, the demixed estimates converge linearly to the optimal solution of (3). We first introduce the following definition of RSC and RSS in the context of unmixing.

Definition 2 ((k, s)-**RSC/RSS**). Let S_k^u and S_s^v be the union of all subspaces spanned by all subsets of k columns of Φ and s columns of Ψ , respectively. A function f satisfies (k, s)-*RSC/RSS* with parameters $m_{k,s}$, $M_{k,s}$ if:

$$m_{k,s} \left(\|\boldsymbol{u}' - \boldsymbol{u}\|_{2}^{2} + \|\boldsymbol{v}' - \boldsymbol{v}\|_{2}^{2} \right) \leq f(\boldsymbol{u}', \boldsymbol{v}') - f(\boldsymbol{u}, \boldsymbol{v}) - \langle \nabla_{\boldsymbol{u}} f(\boldsymbol{u}, \boldsymbol{v}), \boldsymbol{u}' - \boldsymbol{u} \rangle - \langle \nabla_{\boldsymbol{v}} f(\boldsymbol{u}, \boldsymbol{v}), \boldsymbol{v}' - \boldsymbol{v} \rangle$$
(6)
$$\leq M_{k,s} \left(\|\boldsymbol{u}' - \boldsymbol{u}\|_{2}^{2} + \|\boldsymbol{v}' - \boldsymbol{v}\|_{2}^{2} \right)$$

for all $u', u \in \mathcal{S}_k^u$, and $v', v \in \mathcal{S}_s^v$.

This property plays a key role in our analysis. We are now ready to state our first result.

Theorem 1 (Convergence of Algorithm 1). Suppose the loss function f(u, v) satisfies the (2k, 2s)-RSC/RSS property with parameter $m_{2k,2s}$ and $M_{2k,2s}$, respectively. Let (u^*, v^*) be an optimal solution of (3). If $1 < \frac{M_{2k,2s}}{m_{2k,2s}} < \frac{1+\sqrt{5}}{2}$, the unmixing error satisfies

$$\|\boldsymbol{u}^{t+1} - \boldsymbol{u}^*\|_2 + \|\boldsymbol{v}^{t+1} - \boldsymbol{v}^*\|_2$$

$$\leq \eta^t (\|\boldsymbol{u}^0 - \boldsymbol{u}^*\|_2 + \|\boldsymbol{v}^0 - \boldsymbol{v}^*\|_2) + C\sigma \left(\sqrt{\frac{k}{m}} + \sqrt{\frac{s}{m}}\right),$$
(7)
with a convergence rate $\eta = \frac{\sqrt{M_{2k,2s}(M_{2k,2s} - m_{2k,2s})}}{s} < 1.$

with a convergence rate $\eta = \frac{m_{2k,2s}}{m_{2k,2s}} < 1$, and C is a small, positive constant depending on $m_{2k,2s}$ and $M_{2k,2s}$. We recall that σ^2 is the variance of the noise e. Theorem 1 implies that when the RSS and RSC constants of the objective function f satisfies the condition stated in the theorem, the unmixing error decays geometrically at each iteration. Furthermore, in the noiseless case, this implies that the unmixed estimates converge linearly to the optimal solution of (3). To achieve ϵ -close solutions for the unmixed components, we need $\mathcal{O}\left(\log \frac{\|u^0 - u^*\|_2 + \|v^0 - v^*\|_2}{\epsilon}\right)$ iterations.

Theorem 1 relies on certain convexity and smoothness properties of the loss function. When the derivative of the nonlinear operator h is bounded, and the dictionaries are sufficiently incoherent, these properties of the loss function with a relatively low sample complexity.

Theorem 2 (Sample complexity). Suppose that the rows of the sensing matrix **A** are zero mean Gaussian vectors, the absolute value of the derivative of h is bounded within a positive interval, and $\mu(\Phi, \Psi)$ is sufficiently small. If $m = \mathcal{O}\left((s+k)\log\frac{N}{s+k}\right)$, with high probability, the loss function f(u, v) satisfies the (k, s)-RSC/RSS properties with parameters $m_{2k,2s}$ and $M_{2k,2s}$.

Proof Sketch for Theorem 2: Let $\nabla_{2k+2s}^2 f(u, v)$ be any $(2k+2s) \times (2k+2s)$ submatrix of the Hessian matrix of the loss function. As the derivative of h is bounded away from zero, it can be seen that in order for f achieve the desired RSC/RSS properties, it suffices to show the minimum and maximum eigenvalues of any $\nabla_{2k+2s}^2 f(u, v)$ are bounded within a positive interval. Similar to the proof in Soltani et. al. [1], if $m = \mathcal{O}\left((k+s)\log\frac{N}{k+s}\right)$, for some $0 < \alpha < 1$, this holds with high probability which yields the desired result.

The proof for Theorem 1 consists of two main steps, which guarantee that the update and the selection steps yield good constituent estimates and dictionary atoms, respectively. These two insights are summarized in Lemma 1 and Lemma 2 respectively. We state these lemmas without proof due to space constraints.

Lemma 1 (Update step). Let

$$\begin{split} \kappa_{2k}^{\boldsymbol{u}} &= \max_{|\mathcal{S}| \leq 2k} \|\mathcal{P}_{\boldsymbol{\Phi}_{\mathcal{S}}} \nabla_{\boldsymbol{u}} f(\boldsymbol{u}^*, \boldsymbol{v}^*)\|_2 \\ \kappa_{2s}^{\boldsymbol{v}} &= \max_{|\mathcal{S}| < 2s} \|\mathcal{P}_{\boldsymbol{\Psi}_{\mathcal{S}}} \nabla_{\boldsymbol{v}} f(\boldsymbol{u}^*, \boldsymbol{v}^*)\|_2 \end{split}$$

then,

$$\begin{aligned} \|\boldsymbol{u}^{t+1} - \boldsymbol{u}^*\|_2 + \|\boldsymbol{v}^{t+1} - \boldsymbol{v}^*\|_2 \\ &\leq \sqrt{\frac{m_{2k,2s}}{M_{2k,2s}}} \left(\|\mathcal{P}_{\boldsymbol{\Phi}_{\Omega_{\boldsymbol{u}}^{t+1}}} \boldsymbol{u}^* - \boldsymbol{u}^*\|_2 + \|\mathcal{P}_{\boldsymbol{\Psi}_{\Omega_{\boldsymbol{v}}^{t+1}}} \boldsymbol{v}^* - \boldsymbol{v}^*\|_2 \right) \\ &+ C_1(\kappa_{2k}^{\boldsymbol{u}} + \kappa_{2s}^{\boldsymbol{v}}) \end{aligned}$$

$$(8)$$

where C_1 is a small constant.

Lemma 2 (Selection step). Let us define $\Delta_u = u^t - u^*$ and $\Delta_v = v^t - v^*$. Then

$$\begin{aligned} \|\mathcal{P}_{\boldsymbol{\Phi}_{i_{\boldsymbol{u}}}}\boldsymbol{\Delta}_{\boldsymbol{u}} - \boldsymbol{\Delta}_{\boldsymbol{u}}\|_{2} + \|\mathcal{P}_{\boldsymbol{\Psi}_{i_{\boldsymbol{v}}}}\boldsymbol{\Delta}_{\boldsymbol{v}} - \boldsymbol{\Delta}_{\boldsymbol{v}}\|_{2} \\ &\leq \sqrt{\frac{M_{2k,2s} - m_{2k,2s}}{m_{2k,2s}}} \left(\|\boldsymbol{\Delta}_{\boldsymbol{u}}\|_{2} + \|\boldsymbol{\Delta}_{\boldsymbol{v}}\|_{2}\right) + C_{2}(\kappa_{2k}^{\boldsymbol{u}} + \kappa_{2s}^{\boldsymbol{v}}) \end{aligned}$$

$$\tag{9}$$

where C_2 is a small constant.

Proof Sketch for Theorem 1. Applying Lemma 1 and 2 leads to

$$\begin{aligned} \| \boldsymbol{u}^{t+1} - \boldsymbol{u}^* \|_2 + \| \boldsymbol{v}^{t+1} - \boldsymbol{v}^* \|_2 \\ &\leq \eta \left(\| \boldsymbol{\Delta}_u \|_2 + \| \boldsymbol{\Delta}_v \|_2 \right) + C' \left(\kappa_{2k}^{\boldsymbol{u}} + \kappa_{2s}^{\boldsymbol{v}} \right), \end{aligned}$$
(10)

for some small constant C'. Applying this unmixing error bound iteratively and applying Khintchine inequality to bound κ_{2k}^{u} and κ_{2s}^{v} yields (7).

4. EXPERIMENTAL RESULTS

We perform some numerical experiments to demonstrate the effectiveness of demixMP. We compare our algorithms to the Demixing with Hard Thresholding (DHT) algorithm presented in Soltani et. al [1]. We tested our algorithms on both synthetic data and real images.

First we show results from some synthetic experiments. We generate the constituent signals u v of length $N = 2^{10}$ using the Identity and Fourier bases (Φ , Ψ). The measurement matrix A was chosen to be a random Gaussian matrix with normalized rows. These linear measurements were fed into a nonlinear function to generate the final measurements y. We tested our algorithm with both the Sigmoid $\left(h(x) = \frac{1}{1+e^{-x}}\right)$ and ReLU $(h(x) = \max(0, x))$ nonlinearities.

The sparsity of the signals was varied from s = 5 to s = 300, and the number of measurements was varied from m = 50 to m = 200. We measured the *Cosine Similarity* between the recovered signal and original signal. We ran 10 different iterations of the experiment for each setting of m and s, and counted the number of successful recoveries of x. A successful recovery was declared if the Cosine Similarity was > 0.95. The phase transition curves for Sigmoid and ReLU are shown in the figures 1 and 2. We observe that UnmixMP outperforms DHT in terms of better recovery at higher sparsity levels. We also see that DHT performs much worse on ReLU measurements as compared to Sigmoid measurements. Even though both DHT and UnmixMP are only guaranteed to work for smooth nonlinearities (like sigmoids), we note that UnmixMP seems to be able to handle non-differentiable functions like ReLU.

For our experiments on real images we corrupted some common test images (Boats, Barbara) [14] which were 64×64 in size, by adding a sparse (40 non-zero entries) matrix of 1s with randomly chosen support. We then tried to separate



Fig. 1. Phase transition diagrams for ReLU measurements



Fig. 2. Phase transition diagrams for Sigmoid measurements

the image from the sparse noise from m = 2000 compressed Sigmoid measurements. We use a discrete cosine transform (DCT) matrix and an identity matrix as dictionaries for the image and noise, respectively. We compared UnmixMP to DHT in terms of PSNR of the recovered image. In both cases we were able to perform better than DHT. These results are reported in Table 4.

Image	DHT	UnmixMP
Boats	13.8 dB	15.1 dB
Barbara	14.1 dB	14.9 dB

 Table 1. PSNR of image recovered from Sigmoid compressive measurements

5. CONCLUSION

We present a greedy pursuit algorithm UnmixMP, for unmixing the components of a signal from nonlinear compressive measurements. We also prove its convergence, and give bounds on its sample complexity. We also present experiments that show the superiority of UnmixMP to other recent methods [1], especially with popular nonlinearities like Sigmoid and ReLU. We would like to explore algorithms to learn the incoherent dictionaries as well as the sparse components. We would also like to extend our theoretical results to be able to prove convergence in the case of measurements made using non-smooth functions like ReLU.

6. REFERENCES

- Mohammadreza Soltani and Chinmay Hegde, "Fast algorithms for demixing sparse signals from nonlinear observations," *arXiv preprint arXiv:1608.01234*, 2016.
- [2] Stéphane G Mallat and Zhifeng Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on signal processing*, vol. 41, no. 12, pp. 3397– 3415, 1993.
- [3] Emmanuel Candes and Justin Romberg, "Sparsity and incoherence in compressive sampling," *Inverse problems*, vol. 23, no. 3, pp. 969, 2007.
- [4] David L Donoho, Michael Elad, and Vladimir N Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Transactions on information theory*, vol. 52, no. 1, pp. 6–18, 2006.
- [5] Michael Elad, J-L Starck, Philippe Querre, and David L Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis (mca)," *Applied and Computational Harmonic Analysis*, vol. 19, no. 3, pp. 340–358, 2005.
- [6] Joel A Tropp and Anna C Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on information theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [7] Wei Dai and Olgica Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.
- [8] Deanna Needell and Joel A Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.
- [9] Dung Tran, Tao Xiong, Sang Chin, and Trac Tran, "Nonnegative matrix factorization with gradient vertex pursuit," in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015.
- [10] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright, "Robust principal component analysis?," *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 11, 2011.
- [11] Michael B McCoy and Joel A Tropp, "The achievable performance of convex demixing," *arXiv preprint arXiv:1309.7478*, 2013.

- [12] Michael B McCoy and Joel A Tropp, "Sharp recovery bounds for convex demixing, with applications," *Foundations of Computational Mathematics*, vol. 14, no. 3, pp. 503–567, 2014.
- [13] Thomas Blumensath and Mike E Davies, "Gradient pursuits," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2370–2382, 2008.
- [14] W Dong, G Shi, X Li, Y Ma, and F Huang, "Compressive sensing via nonlocal low-rank regularization," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3618–3632, 2014.