A Fast Algorithm for Demixing Signals with Structured Sparsity

Chinmay Hegde
Electrical and Computer Engineering
Iowa State University
Ames, IA, USA 50010
Email: chinmay@iastate.edu

***Abstract*—Demixing, or source separation, involves disentangling a complicated signal into simpler, more informative components. Algorithms for signal demixing impact several applications, ranging from interference cancellation in communication signals, to foreground-background separation in images and video, and outlier suppression in machine learning. Central to several modern demixing algorithms is an assumption of low-dimensional structure being present in the signal components. However, the majority of these modern algorithms are based on convex optimization, and their computational complexity can be high (polynomial) in terms of the signal size.**

In this paper, we propose a new algorithmic approach for signal demixing based on structured sparsity models. Our approach leverages recent advances in constructing sparsity models via appropriately chosen graphs; this graphical approach can be shown to model a diverse variety of low-dimensional structures in signals. Despite being highly nonconvex, our algorithm exhibits a nearly-linear running time, and therefore is scalable to very high-dimensional signals. We supplement our proposed algorithm with a theoretical analysis, providing sufficient conditions for provable reconstruction of the underlying components. Finally, we demonstrate the validity of the method via numerical experiments on real 2D image data.

I. Introduction

A. Motivation

Demixing, or source separation, refers to the process of separating out a pair of signals from a (possibly noisy) observation of their superposition. Demixing methods are of special importance in diverse applications spanning audio signal analysis [1], interference cancellation in medical imaging [2], image processing in astronomy [3], [4], surveillance video analysis and compression [5], and machine learning [6].

As succinctly described in the recent survey article [7], modern approaches for demixing rely on two main assumptions: (i) that the component signals have a simple, or *low-dimensional*, representation relative to the ambient signal dimension, and (ii) that the components are sufficiently different-looking, or *incoherent*, with respect to each other. A flexible (and convenient) mathematical model for signal simplicity presupposes that each component signal is *sparse* with respect to a given basis or dictionary.

Sparsity models form the foundation of numerous advances in signal compression [8], compressive sensing [9], [10], and machine learning [11]. However, recovering sparse signals from noisy measurements has long been acknowledged as being a very challenging combinatorial optimization problem even for very simple cases [12], and the typical solution is to relax the problem into a convex optimization formulation [13]. Such a relaxation is advantageous in many respects. First, one can immediately use algorithms for convex optimization in a black-box fashion to demixing problems. Second, precise and rigorous theory that can characterize the performance of these algorithms is also available [14], [15].

An added advantage is that by appropriately modifying the convex penalties in the optimization problem, one can model much more refined structures *beyond sparsity* in the components. Such a "structured sparsity" assumption is applicable in several settings. For example, the (overlapping) group-sparsity penalty is used in cases where the basis coefficients of the components are additionally expected to be clustered. A general framework for constructing penalties tailored to various structures can be developed via the notion of *atomic norms* [15]–[17].

B. Our contributions

In this paper, we depart from the above trends, and propose an alternative approach for demixing signals with structured sparsity. Our signal demixing approach relies on the *graph-structured sparsity* framework proposed in [18]. The graph-based approach is widely applicable, and can model diverse structures in high-dimensional data including group structures, hierarchical structures, and spatio-temporal correlations [19]. We describe this model in greater detail in Section II below.

The core of our approach is a novel, *nonconvex* algorithm for demixing the source components. Our algorithm can be interpreted as a variant of [20], specialized to demixing. A key feature of our algorithm is that despite its non-convex nature, the algorithm enjoys a *nearly-linear* running time for a large class of demixing problems. To the best of our knowledge, this is the first nearly-linear time algorithm for demixing with structured sparsity. This is particularly important for massive data applications where convex algorithms that exhibit even a (low-degree) polynomial running time can become intractable.

We supplement our proposed algorithms with a rigorous theoretical analysis, providing sufficient conditions under which the methods provably reconstruct the underlying components. Our analysis reveals that the algorithm is stable to noise, and additionally exhibits a *linear* rate of convergence for a range of signal sparsity parameters. We also provide numerical experiments on 2D image data, and obtain promising results.

C. Techniques

At a high level, our demixing approach is an application of the graph-structured sparsity framework of [18]. There, the authors proposed a framework for solving general linear inverse problems where the support of the unknown coefficient vector can be modeled as a union of small connected components over a pre-defined graph. However, the main focus of that work is on recovering structured signals from under-sampled linear measurements, and the algorithm proposed in that paper is somewhat different. In contrast, our proposed algorithm (that we call STRUCTDEMIX) is more closely related to the approximate model-iterative thresholding (AM-IHT) algorithm developed in [20], [21]. Our theoretical analysis of STRUCTDEMIX is an immediate consequence of the derivation provided in [20].

Due to the nature of the demixing problem, the implementation of the algorithm appears somewhat different and can be achieved via a series of alternating projections. (As opposed to traditional alternating projection methods, our method relies on projections onto *nonconvex* constraint sets.) The key algorithmic challenge here is to design the update rules within each iteration such that linear convergence to the desired solution occurs even for this challenging case. We achieve this via a sequence of *approximate projections* onto the feasible signal sets, coupled with suitable gradient steps. Section II details precisely how these approximate projections are defined. Executing a specific sequence of projection steps seem to be crucial for proving algorithm convergence.

We instantiate our approach in the context of demixing 2D images. We consider as input a sparse image with spatially clustered nonzeros that is corrupted by sinusoidal noise. We model the image domain as a 2D *lattice* graph, and consider the image as belonging to a graph-sparsity model. The combinatorial optimization approach of [18] readily gives us the approximate projection steps required to establish convergence. Moreover, each of the approximate projections provably have a nearly-linear running time; therefore, iterating a logarithmic number of times gives us the desired running time of our overall algorithm. Section IV contains the details.

II. SETUP

A. Notation

A vector $x \in \mathbb{R}^n$ is said to be s-sparse if it contains no more than s nonzero coefficients. The support of a vector is the set of indices corresponding to its nonzero entries, i.e., $\mathrm{supp}(x) = \{i : |x_i| > 0\}$. The symbol $\|\cdot\|$ denotes the ℓ_2-norm unless explicitly specified.

In order to express structures in signals beyond sparsity, we will use the *model-based* approach of [22]. Specifically, let $\mathbb{M} = \{S_1, S_2, \ldots, S_q\}$ denote a family of supports such that $S_i \subseteq [n]$. Then, the set of all vectors supported on any of the S_i is called the *sparsity model* induced by $\mathbb{M}$:

$$\mathcal{M} = \{x \in \mathbb{R}^n : \mathrm{supp}(x) \subseteq S_i \text{ for some } S_i \in \mathbb{M}\}.$$

B. Modeling assumptions

For the demixing problem, we use the *MCA signal model* of [3]. We express the observed signal $y \in \mathbb{R}^n$ as a (possibly noisy) superposition of two structured signals:

$$y = Ax + Bz + e,$$

where $A, B \in \mathbb{R}^{n \times n}$ are (known) orthonormal bases, x and z represent the unknown coefficient vectors, and e denotes the noise. For convenience, let us suppose that x corresponds to the "signal" of interest, and z corresponds to the "interference". By an appropriate change of coordinates, we can assume without loss of generality that A is the canonical (identity) basis. Hence, we rewrite the observation model as:

$$y = x + Fz + e, \tag{1}$$

for some known orthonormal basis F. We will assume that F is ε-*incoherent* with the canonical basis, i.e., $|F_{ij}| \leq \varepsilon$ for $1 \leq i, j \leq n$. The goal is to investigate reliable and efficient demixing methods that recover the signal representations x and z from the observations y.

Equation (1) provides an *under-determined* system of linear equations with n observations and $2n$ unknowns, and as such cannot be solved without additional information. Therefore, we will make two extra modeling assumptions. First, we will assume that the coefficient vector z is at most k-sparse for some parameter k, i.e., $z \in \Sigma_k$, where $\Sigma_k = \{x \in \mathbb{R}^n : |\mathrm{supp}(x)| \leq k\}$.

Further, we will assume that x belongs to a *graph-sparsity model* [18], formally defined as follows. Consider a graph $G = (V, E)$ with n nodes and m edges, where the nodes correspond to coordinates of vectors in $\mathbb{R}^n$. Then, we can identify supports $S \subseteq [n]$ with subgraphs of G. Let $\mathrm{cc}(S)$ denote the number of *connected components* in the subgraph of G corresponding to the support S. Consider the set of supports $\mathbb{M}$, defined as:

$$\mathbb{M} = \{S \subseteq [n] : |S| \leq s,\ \mathrm{cc}(S) \leq g\},$$

Then, the corresponding model induced by $\mathbb{M}$ is called the graph-sparsity model $\mathcal{M}(G, s, g)$. We will assume that $x \in \mathcal{M}(G, s, g)$.

The choice of the graph G depends on the particular application, and reflects the structural interactions between signal coefficients. Different choices of G yield different sparsity models. For example, a natural choice of G for modeling 1D time series is the *line graph*, and the corresponding sparsity model can be used to express sparse signals whose coefficients are clustered as contiguous groups [23]. Another common choice of G for modeling 2D images is the *2D lattice graph*, which can be used to model both "thin" as well as "thick" clusters of pixel intensity values over a two-dimensional spatial domain [24], [25].

C. Comparison with prior work

Before we present our approach, let us acknowledge that sparsity-constrained linear inverse problems of the form (1) have a very long history, dating back to [12] and earlier. The work of Elad et al. [3] and Bobin et al. [4] posed the demixing problem as an instance of *morphological components analysis* (MCA), and formalized the observation model (1). These authors posed the recovery problem in terms of a convex optimization procedure, such as the LASSO [11]:

$$(\widehat{x}, \widehat{z}) = \min_{x,z} \ \|y - x - Fz\|_2^2 \qquad (2)$$
$$\text{s.t. } \|x\|_1 \leq \tau_x, \ \|z\|_1 \leq \tau_z.$$

where τ_x and τ_z are user-specified parameters. These works provided upper bounds on successful recovery as a function of the problem dimensions n, s, k. The work of Pope et al. [26] analyzed somewhat more general conditions under which stable demixing could be achieved.

More recently, the work of McCoy et al. [14], [15] showed a curious phase transition behavior in the performance of the convex optimization methods, under a random model on the interference basis F. Specifically, they demonstrated a sharp statistical characterization of the achievable and non-achievable parameters for which successful demixing of the signal components can be achieved. Moreover, they extended the demixing problem to a large variety of signal structures beyond sparsity via the use of general *atomic norms* in place of the ℓ_1-norm in the above optimization [16]. See [7], [15] for an in-depth discussion of atomic norms, their statistical and geometric properties, and their applications to demixing.

While a very good statistical understanding of demixing algorithms is now available, the computational implications of these algorithms are somewhat less clear. One could apply standard black-box convex optimization solvers (such as interior-point methods) to solve (2), but these are typically too slow for large problem sizes. The survey article [7] advocates the use of the alternating direction method of multipliers (ADMM) in conjunction with appropriate proximal operators for the corresponding atomic norms. However, the rate of convergence for this algorithm can be *sublinear* in general, and no better bounds seem to be available for our problem.

A related paper by Rao et al. [17] also consider extensions of (2) to more general atomic norms, and develop a variant of the Frank-Wolfe algorithm to efficiently solve this problem. However, this algorithm also seems to exhibit a sublinear rate of convergence in the worst case. Moreover, while atomic-norm methods can (in principle) be used in the context of the graph-structured sparsity model that we have defined above, the proximal operators can incur a high-degree polynomial running time; for a detailed discussion of this matter, see the appendix of [18].

The majority of modern approaches for demixing rely on convex relaxation procedures. In contrast, the work of [27] proposes an alternative, nonconvex algorithm called SPIN. In that work, the signal and interference vectors are modeled as belonging to a pair of incoherent submanifolds of $\mathbb{R}^n$, and exact *projection* operators onto these submanifolds are assumed to be available. Under some relatively mild conditions, a simple iterative projection algorithm is shown to converge *linearly* to a solution that is close to the best possible. However, this method heavily relies on efficient, exact projection operators. In the graph-sparsity case, such an exact projection is known to be NP-hard due to a reduction from the Steiner Tree problem [28], and projection operators with even a polynomial running time are unlikely to exist.

Algorithm 1 STRUCTDEMIX

Input: $y \in \mathbb{R}^n$
Outputs: $\widehat{x}, \widehat{z} \in \mathbb{R}^n$
Parameters: s, g, G, k, number of iterations t.
$\widehat{x}_0 \leftarrow 0$, $\widehat{z}_0 \leftarrow 0$
for $i \leftarrow 0, \ldots, t-1$ **do**
  $b \leftarrow \widehat{x}_i + H(y - \widehat{x}_i - F\widehat{z}_i)$
  $\widehat{x}_{i+1} \leftarrow T(b)$
  $b' \leftarrow \widehat{z}_i + P(F^* y - F^* \widehat{x}_i - \widehat{z}_i)$
  $\widehat{z}_{i+1} \leftarrow P(b')$
return $\widehat{x} \leftarrow \widehat{x}_t, \widehat{z} \leftarrow \widehat{z}_t$

III. A Nonconvex Demixing Algorithm

We now describe our proposed demixing algorithm. Suppose we are given an input signal y obeying (1) where $z \in \Sigma_k$ and $x \in \mathcal{M}(G, s, g)$ where k, s, g, G are known parameters. Our goal is to produce estimates of coefficient vectors $\widehat{x}, \widehat{z}$ such that $\|x - \widehat{x}\|$ and $\|z - \widehat{z}\|$ are comparable to the noise level, i.e., $\max(\|x - \widehat{x}\|, \|z - \widehat{z}\|) \leq C\|e\|$, for some constant C. In order to achieve this, we assume availability of the following (nonconvex) projection oracles:

1) *Exact projection onto* Σ_k: There exists an oracle $P(\cdot)$ such that for any arbitrary $w \in \mathbb{R}^n$, the oracle returns a vector in Σ_k such that:
$$\|w - P(w)\| \leq \min_{w' \in \Sigma_k} \|w - w'\|.$$
2) *Approximate tail projection onto* $\mathcal{M}(G, s, g)$: There exists an oracle $T(\cdot)$ such that for any arbitrary $w \in \mathbb{R}^n$, the oracle returns a vector in $\mathcal{M}(O(s), g, G)$:
$$\|w - T(w)\| \leq c_T \min_{w' \in \mathcal{M}(s,g,G)} \|w - w'\|,$$
where $c_T > 1$ is a constant.
3) *Approximate head projection onto* $\mathcal{M}(G, s, g)$: There exists an oracle $H(\cdot)$ such that for any arbitrary $w \in \mathbb{R}^n$, the oracle returns a vector in $\mathcal{M}(O(s), O(g), G)$:
$$\|H(w)\| \geq c_H \max_{w' \in \mathcal{M}(s,g,G)} \|w - w'\|,$$
where $c_H < 1$ is a constant.

The exact projection oracle $P(\cdot)$ is equivalent to a *hard-thresholding* operator. The approximate tail and head projection oracles, $T(\cdot)$ and $H(\cdot)$, can be implemented using the *prize-collecting Steiner Forest* (PCSF) framework of [18].

Given these projection oracles, we propose an alternating projection algorithm to produce the estimates $\widehat{x}$ and $\widehat{z}$. The algorithm, that we call StructDemix, is described in pseudocode form as Algorithm 1. Intuitively, the algorithm iteratively builds up an approximation of $\widehat{x}$ and $\widehat{z}$ by considering the *residual* at the current iteration, $r = y - \widehat{x}_i - F\widehat{z}_i$, and performing an appropriate sequence of projections. Operating upon the residual r is equivalent to processing the *gradient* of the squared-loss function $f(x,z) = \|y - x - Fz\|_2^2$.

Therefore, in this sense, this method can be viewed as a form of block co-ordinate descent (BCD), coupled with projections onto the feasible sets. A rather similar BCD-like method was also used in [27]. However, the key differences are that we perform *two* projection operations in each iteration (instead of one) for updating the estimate of $\widehat{x}$ as well as $\widehat{z}$. Moreover, we merely use approximate projection oracles (instead of exact projection oracles). These refinements appear to be crucial in order to achieve provable convergence for demixing using the graph-sparsity model.

We can bound the running time of the overall algorithm by multiplying the number of iterations t by the time taken per iteration. The main computational challenges in each iteration involve implementing the three projection oracles, plus the time taken for matrix-vector multiplication with the orthonormal basis F. For a large class of problems, each of the three projection oracles listed above can be implemented (for arbitrary $w \in \mathbb{R}^n$) in nearly-linear time. The hard thresholding operator $P(\cdot)$ can be implemented using a (generalized) median finding routine in $O(n)$ time. For graph-sparsity models $\mathcal{M}(s,g,G)$, the approximate tail and head projection oracles can be implemented in $O(n\log^3 n)$ time in the case of bounded-degree graphs G, as detailed in [18].

Multiplying with the orthonormal basis matrix F can take $O(n^\omega)$ time for general F, where ω is the matrix multiplication constant. However, several structured matrices, such as the discrete cosine transform (DCT) matrix, or the Walsh-Hadamard Transform (WHT) matrix, satisfy the incoherence condition specified above, support fast matrix-vector multiplications in $O(n\log n)$-time, and also can model interference signals of various kinds. For these specific cases, the per-iteration cost of StructDemix is nearly-linear in the input size. We still need to bound the number of iterations of the overall algorithm, which we show for some special cases below.

IV. Case Study: Demixing for 2D Images

For illustration purposes, we consider a concrete setting where our signal is a 2D foreground image $x \in \mathbb{R}^n$ obeying a graph-sparsity structure. This signal is superimposed with an interference image Fz where $F \in \mathbb{R}^{n\times n}$ is the DCT basis and $z \in \mathbb{R}^n$ is an arbitrary k-sparse vector. Finally, the composite signal is corrupted with an additive noise vector $e \in \mathbb{R}^n$.

We prove the following theoretical result:

Theorem 1. *There exist constants $c, C > 0$ for which the following statement is true: if $k + s < c\sqrt{n}$, then Algorithm 1 produces a signal coefficient vector $\widehat{x} \in \mathcal{M}(5s,g,G)$ and an interference coefficient vector $\widehat{z}$ such that:*

$$\max(\|x-\widehat{x}\|, \|z-\widehat{z}\|) \leq C\|e\|.$$

Moreover, the running time of the algorithm is $O\left(n\log^3 n \log\frac{\|y\|}{\|e\|}\right)$.

Proof sketch. The proof is a straightforward concatenation of existing results. First, we observe that the incoherence parameter ε of the DCT basis F equals $\sqrt{2/n}$ [29]. This is nothing but the *mutual coherence* of the composite matrix $Q = [I \;\, F]$. For a given sparsity parameter p, the *restricted isometry constant* of Q is defined as the smallest nonnegative number δ such that:

$$(1-\delta)\|u\|^2 \leq \|Qu\|^2 \leq (1+\delta)\|u\|^2 ,$$

for any p-sparse vector $u \in \mathbb{R}^{2n}$. By an application of Gershgorin's disc theorem, we can show that $\delta < p\epsilon = p\sqrt{2/n}$. Assuming that $p < c\sqrt{n}$ for sufficiently small c, we can upper-bound the restricted isometry constant of Q.

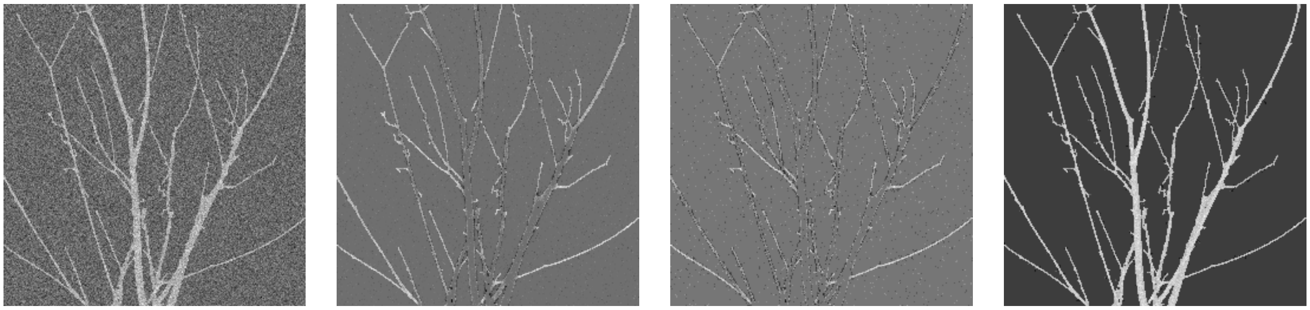
Next, we can adapt Theorem 11 of [20], which details the convergence of an algorithm for recovering signals belonging to arbitrary sparsity models from linear observations. Suppose we are given a matrix Q with restricted isometry constant δ for sparsity p, and linear observations $y = Qu + e$. From [18], we also possess approximate projection oracles onto the signal model $\mathcal{M}(s,g,G)$. Then, the AM-IHT algorithm returns a sequence of estimates $\{u_i\}_{i=1}^t$ such that:

$$\|u - u_i\| \leq \alpha\|u - u_{i-1}\| + \beta\|e\| ,$$

for fixed α and β that depend on the approximation constants c_T and c_H. Iterating $t = O(\log\frac{\|y\|}{\|e\|})$ times, we can reduce the error $\|u - u_t\|$ down to $C\|e\|$ for some constant C. Invoking this fact for $Q = [I \;\, F]$ and $u = [x^* \;\, z^*]^*$, we obtain the desired convergence result of Alg. 1. $\square$

We have implemented and tested our algorithm on realistic 2D image data. Due to space constraints, we only show a single representative result, and will report more comprehensive experiments in a future manuscript. We consider a sparse grayscale image (x) of the silhouette of a tree with size $n = 256 \times 256$, sparsity level $s = 6500$, and number of connected components $g = 2$ with respect to the 2D lattice graph G. This image is corrupted by an interference image (Fz) that is sparse in the DCT basis F with sparsity parameter $k = 7800$, as well as a small amount of additive white Gaussian noise (e), to yield the input image (y). This image is shown as Fig. 1(a).

We applied three demixing algorithms on this input image: basis pursuit denoising (BPDN) [7], [13], a version of SPIN that only enforces sparsity constraints in both x and z [20], and Alg. 1 (StructDemix) with the approximate tail and head projection oracles of [18]. All algorithms were assumed to possess oracle knowledge of the "right" parameter settings for



(a) Input image | (b) BPDN [7], [13] (RNSR = 2.11dB) | (c) SPIN [27] (RSNR = 0.87dB) | (d) Alg. 1 of this paper (RSNR = 19.66dB)

Fig. 1. Comparison of StructDemix (Alg. 1 of this paper) with previous approaches. (a) Corrupted input image with parameters $n = 256 \times 256$, $s = 6500$, $g = 2$, $k = 7800$. (b,c,d) Recovered images using various algorithms. Our proposed algorithm leverages the implicit clustering of the nonzero coefficients of the true image in the 2D plane, and is able to achieve superior results compared to standard sparsity-based techniques that ignore this additional structure.

respectively achieving best performance. A rigorous method for automatically choosing the best possible algorithm parameters is currently unavailable for our approach, and is an important point of consideration for future work.

We measure algorithm performance using the recovery SNR (or RSNR) metric, calculated as $\|x - \widehat{x}\|/\|x\|$, and expressed in decibels. Figures 1(b), (c), and (d) display the recovered images $\widehat{x}$ produced by the three different demixing algorithms. We observe that enforcing the graph-sparsity structure using our method results in dramatic benefits, resulting in over 17dB improvement over the previous two methods. Moreover, despite the nonconvex nature of our method, we have observed very rapid convergence to the final solution. For this experiment, the signal sparsity s is much larger than $\sqrt{n}$, indicating that our above theoretical results are somewhat pessimistic, and that there is considerable scope for improvement.

Acknowledgements

This work was supported in part by the National Science Foundation under the grant CCF-1566281.

References

[1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*, vol. 46, 2004.
[2] M. Lustig, D. Donoho, J. Santos, and J. Pauly, "Compressed sensing MRI," *IEEE Sig. Proc. Mag.*, vol. 25, no. 2, pp. 72–82, 2008.
[3] M. Elad, J-L. 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.
[4] J. Bobin, J.-L. Starck, J. Fadili, Y. Moudden, and D. Donoho, "Iological component analysis: An adaptive thresholding strategy," *IEEE Trans. Image Proc.*, vol. 16, no. 11, pp. 2675–2681, 2007.
[5] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Adv. Neural Inf. Proc. Sys. (NIPS)*, 2009, pp. 2080–2088.
[6] H. Xu, C. Caramanis, and S. Sanghavi, "Robust pca via outlier pursuit," in *Adv. Neural Inf. Proc. Sys. (NIPS)*, 2010, pp. 2496–2504.
[7] M. McCoy, V. Cevher, Q.T. 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.
[8] S. Mallat, *A wavelet tour of signal processing: the sparse way*, Academic press, 2008.
[9] E. Candès, "Compressive sampling," in *Proc. Int. Cong. Math.*, Madrid, Spain, 2006, vol. 3, pp. 1433–1452.
[10] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, 2006.
[11] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal Statist. Soc B*, vol. 58, no. 1, pp. 267–288, 1996.
[12] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inform. Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.
[13] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1998.
[14] M. McCoy, *A geometric analysis of convex demixing*, Ph.D. thesis, California Institute of Technology, 2013.
[15] M. McCoy and J. Tropp, "Sharp recovery bounds for convex demixing, with applications," *Foundations of Comput. Math.*, vol. 14, no. 3, pp. 503–567, 2014.
[16] V. Chandrasekaran, B. Recht, P. Parrilo, and A. Willsky, "The convex geometry of linear inverse problems," *Foundations of Comput. Math.*, vol. 12, no. 6, pp. 805–849, 2012.
[17] N. Rao, P. Shah, and S. Wright, "Forward-backward greedy algorithms for signal demixing," in *Proc. Asilomar Conf. Sig. Sys. Comput.*, 2014, pp. 437–441.
[18] C. Hegde, P. Indyk, and L. Schmidt, "A nearly linear-time framework for graph-structured sparsity," in *Proc. Int. Conf. Machine Learning (ICML)*, July 2015.
[19] C. Hegde, P. Indyk, and L. Schmidt, "Fast algorithms for structured sparsity," *Bulletin of the EATCS*, , no. 117, pp. 197–228, Oct. 2015.
[20] C. Hegde, P. Indyk, and L. Schmidt, "Approximation algorithms for model-based compressive sensing," *IEEE Trans. Inform. Theory*, vol. 61, no. 9, pp. 5129–5147, Sept. 2015.
[21] C. Hegde, P. Indyk, and L. Schmidt, "Approximation-tolerant model-based compressive sensing," in *Proc. ACM Symp. Discrete Alg. (SODA)*, Jan. 2014.
[22] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inform. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.
[23] V. Cevher, P. Indyk, C. Hegde, and R. Baraniuk, "Recovery of clustered sparse signals from compressive measurements," in *Proc. Sampling Theory and Appl. (SampTA)*, May 2009.
[24] S. Vicente, V. Kolmogorov, and C. Rother, "Graph cut based image segmentation with connectivity priors," in *IEEE Intl. Conf. Comp. Vision and Pattern Recog. (CVPR)*, 2008, pp. 1–8.
[25] V. Cevher, M. Duarte, C. Hegde, and R. Baraniuk, "Sparse signal recovery using Markov Random Fields," in *Adv. Neural Inf. Proc. Sys. (NIPS)*, Dec. 2008.
[26] 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.
[27] C. Hegde and R. Baraniuk, "Signal recovery on incoherent manifolds," *IEEE Trans. Inform. Theory*, vol. 58, no. 12, pp. 7204–7214, Dec. 2012.
[28] R. Karp, *Reducibility among combinatorial problems*, Springer, 1972.
[29] P. Dragotti and Y. Lu, "On sparse representation in Fourier and local bases," *IEEE Trans. Inform. Theory*, vol. 60, no. 12, pp. 7888–7899, 2014.