DEMIXING SPARSE SIGNALS VIA CONVEX OPTIMIZATION

Yi Zhou⋆ *Yingbin Liang*⋆

⋆ Department of EECS, Syracuse University

ABSTRACT

We consider demixing a pair of sparse signals in orthonormal basis via convex optimization. Theoretically, we characterize the condition under which the solution of the convex optimization problem correctly demixes the true signal components. In specific, we introduce the local subspace coherence to characterize how a basis vector is coherent with a signal subspace, and show that the convex optimization approach succeeds if the subspaces of the true signal components avoid high local subspace coherence. Furthermore, we illustrate via examples that our condition for exact demixing is more fundamental than existing conditions. We then verify our theoretical finding through numerical experiments.

Index Terms— signal demixing, convex optimization, sparsity, local subspace coherence.

1. INTRODUCTION

Consider an observation model where we observe a signal $\mathbf{z}_0 \in \mathbb{R}^n$ that is a mixture of two indivisual components $\mathbf{x}_0, \mathbf{y}_0 \in \mathbb{R}^n$, i.e.,

$$\mathbf{z}_0 := \mathbf{x}_0 + \mathbf{y}_0. \tag{1}$$

Usually, the pair of components $(\mathbf{x}_0, \mathbf{y}_0)$ correspond to signals with different types of features, and an important problem is to demix the individual components $(\mathbf{x}_0, \mathbf{y}_0)$ from the observation $\mathbf{z}_0$. Many practical applications such as image feature decomposition [1, 2, 3], image denoising [4] and signal separation [5, 6, 7, 8] can be captured by this model. Of course, directly demixing the signal components based on eq. (1) is an ill-posed problem, since one can find infinitely many feasible decomposition pairs. Thus, we need to further exploit certain structure of the signal components to identify a unique and correct decomposition, and a useful one is the sparsity structure.

In this paper, we exploit the sparsity structure of the signal components in orthonormal basis, and demix them from the observation model in eq. (1) by solving a convex optimization problem. To be specific, consider a pair of orthonormal basis $\mathbf{\Psi}, \mathbf{\Phi} \in \mathbb{R}^{n\times n}$ with basis vectors denoted as $\mathbf{\Psi} := [\boldsymbol{\psi}_1^\top; ...; \boldsymbol{\psi}_n^\top]$ and $\mathbf{\Phi} := [\boldsymbol{\phi}_1^\top; ...; \boldsymbol{\phi}_n^\top]$, respectively. We assume that the signal components $\mathbf{x}_0, \mathbf{y}_0$ have sparse representations in basis $\mathbf{\Psi}$ and $\mathbf{\Phi}$, respectively, i.e., $\mathbf{\Psi}\mathbf{x}_0$ and $\mathbf{\Phi}\mathbf{y}_0$ are sparse. Then for some $\lambda > 0$, we aim to demix the signal components from eq. (1) by solving the following convex optimization problem

$$\min_{\mathbf{x},\mathbf{y}\in\mathbb{R}^n} \|\mathbf{\Psi}\mathbf{x}\|_1 + \lambda\|\mathbf{\Phi}\mathbf{y}\|_1, \ \text{s.t.} \ \mathbf{x}+\mathbf{y} = \mathbf{z}_0. \tag{P}$$

The form of the convex optimization problem in **(P)** exploits the fact that both signal components are sparse in their corresponding orthonormal basis. Thus, we seek for the feasible decomposition with minimum ℓ_1 norm, which is the tightest convex relaxation of sparsity – the non-convex ℓ_0 norm. Our goal in this paper is to characterize the *exactness condition* of (P), under which $(\mathbf{x}_0, \mathbf{y}_0)$ is the unique solution pair of **(P)**.

Demixing sparse signals via convex optimization has been well studied in the literature [9, 8, 7, 10, 11, 12, 13]. In specific, [8] shows that **(P)** is exact if the signal components are sparse enough, and gives an explicit characterization of the phase transition curve with respect to the sparsity of the signal components. [9] characterizes the exactness condition of **(P)** by introducing the mutual coherence between the pair of orthonormal basis, i.e.,

$$\text{(Mutual Coherence)}: \ \mu(\mathbf{\Psi}, \mathbf{\Phi}) := \max_{i,j} |\langle \boldsymbol{\psi}_i, \boldsymbol{\phi}_j\rangle|.$$

In a special case where $\mathbf{\Psi}$ is the discrete Fourier basis and $\mathbf{\Phi}$ is the canonical (discrete time) basis , they show that **(P)** is exact with high probability if the total sparsity[1] of the signal components are less than the order $\mathcal{O}(\frac{n}{\log n})$. Moreover, for general choices of orthonormal basis, the exactness condition of **(P)** requires the total sparsity be less than the order $\mathcal{O}(\frac{1}{\mu^2(\mathbf{\Psi},\mathbf{\Phi})\log^6 n})$. A drawback of this result is the mutual coherence barrier, i.e., it becomes trivial when $\mu(\mathbf{\Psi}, \mathbf{\Phi}) = 1$. In [7, 10], the exactness condition of **(P)** is characterized for $\mathbf{\Psi}, \mathbf{\Phi}$ being tight frames. Specially, they introduce the notion of cluster coherence, i.e.,

$$\text{(Cluster Coherence)}: \ \mu(\mathbf{\Psi}_\Omega, \mathbf{\Phi}) := \max_j \sum_{i\in\Omega} |\langle \boldsymbol{\psi}_i, \boldsymbol{\phi}_j\rangle|,$$

where $\Omega \in [n]$ is a subset of the index. It is shown that **(P)** can demix the signal components in the asymptotic regime (i.e.,

This work is supported in part by the grants AFOSR FA9550-16-1-0077 and NSF ECCS 16-09916.

[1] Number of non-zero coefficients of the orthonormal representation.

$j \to n$) near perfectly provided that the corresponding cluster coherence vanishes. [11, 12, 13] further consider demixing between a low rank matrix and a sparse corruption matrix. Specially, [13] characterizes the exactness condition of the robust PCA problem based on the leverage scores of the underlying low rank matrix. Other notions such as the level dependent coherence are proposed for compressive sensing problems [14].

Another closely related topic is finding sparse representations of signals [15, 16, 9, 17, 18, 19, 20, 21]. There, the focus is to characterize the conditions under which **(P)** returns the unique and sparsest decomposition of $\mathbf{z}_0$. The conditions are based on characterizing the so called uncertainty principle [22, 9, 16, 18] — a signal cannot be simultaneously sparse in two incoherent orthonormal basis.

Existing exactness condition of **(P)** either encounters the mutual coherence barrier [9], or depends on null space analysis but can only demix the signal in the asymptotic regime [7, 10]. Thus, the goal of this paper is to characterize a new exactness condition of **(P)**, and our contributions are of three fold: 1) we introduce the notion of local subspace coherence, which measures the coherence between a basis vector and the subspace spanned by the signal component; 2) we show that **(P)** is exact if the subspace spanned by the signal component distributes in a way that avoids high local subspace coherence; 3) By considering some examples, we show that our exactness condition for **(P)** recovers existing result and overcomes the mutual coherence barrier.

Throughout the paper, $\|\cdot\|_p$ denotes the Euclidean l_p norm and $\|\cdot\|$ denotes the spectral norm of an operator.

2. PRELIMINARY

Recall that $\mathbf{z}_0 = \mathbf{x}_0 + \mathbf{y}_0$, and $\mathbf{\Psi}, \mathbf{\Phi}$ are a pair of orthonormal basis. We denote $\boldsymbol{\theta}_x := \mathbf{\Psi}\mathbf{x}_0, \boldsymbol{\theta}_y := \mathbf{\Phi}\mathbf{y}_0$ as the representations of the signal components in the corresponding orthonormal basis, and also denote $\mathcal{S}_x, \mathcal{S}_y$ as the support sets of the representations, respectively. Then, we can identify the subspaces spanned by the signal components as

$$\mathcal{X} := \operatorname{span}\{\boldsymbol{\psi}_j, j \in \mathcal{S}_x\}, \;\; \mathcal{Y} := \operatorname{span}\{\boldsymbol{\phi}_j, j \in \mathcal{S}_y\}. \quad (2)$$

It follows that $\mathbf{x}_0 \in \mathcal{X}$, and $\mathbf{y}_0 \in \mathcal{Y}$. We also denote $\mathrm{P}_{\mathcal{X}}, \mathrm{P}_{\mathcal{Y}}$ as the projection operators onto the corresponding subspaces, respectively. Our exactness condition is characterized in a probabilistic way, and we adopt the following random model on the signal components.

Assumption 1. *The signal components $\mathbf{x}_0, \mathbf{y}_0$ satisfy:*

1. *The support set $\mathcal{S}_x$ is fixed. The support set $\mathcal{S}_y$ follows a Bernoulli model, i.e., for all $j \in [n], \mathbb{P}(j \in \mathcal{S}_y) \sim \operatorname{Bernoulli}(p_j)$.*
2. *The signs of the non-zeros of $\boldsymbol{\theta}_x, \boldsymbol{\theta}_y$ are distributed independently from their locations, and take values from $\{+1, -1\}$ with equal probability.*

The random support model in Assumption 1.1 has been adopted in [9] for developing a probabilistic argument. We note that [9] assumes that both of the support sets are uniformly picked at random, while we assume a "fixed + random Bernoulli" model. Hence, our goal is to characterize the exactness of **(P)** under the relationship between subspace $\mathcal{X}$ and the distribution $\{p_j\}_j$ of subspace $\mathcal{Y}$. Moreover, the randomness of the signs in Assumption 1.2 is introduced for convenience of the proof, and can be eliminated via a derandomization argument [12].

To characterize the relationship between subspace $\mathcal{X}$ and the distribution $\{p_j\}_j$ of subspace $\mathcal{Y}$, we next introduce the notion of local subspace coherence — a measure of coherence between $\mathcal{X}$ and the basis vectors $\{\boldsymbol{\phi}_j\}_j$.

Definition 1. *The local subspace coherence between the basis vectors $\{\boldsymbol{\phi}_j\}_j$ and the subspace $\mathcal{X}$ is defined as*

$$\mu(\mathcal{X}, \boldsymbol{\phi}_j) := \|\mathrm{P}_{\mathcal{X}}\boldsymbol{\phi}_j\|_2, \quad \forall j \in [n]. \quad (3)$$

The local subspace coherence has an intuitive explanation: it measures the length of a unit basis vector $\boldsymbol{\phi}_j$ onto the subspace $\mathcal{X}$. Thus, a higher local subspace coherence implies that the basis vector is more aligned with the subspace $\mathcal{X}$. Furthermore, this notion of coherence has a fully decomposable structure. To be specific, consider arbitrary orthogonal decomposition of the subspace $\mathcal{X} := \bigoplus_{l=1}^{p} \mathcal{X}_l$, then it follows by Definition 1 that

$$\mu^2(\mathcal{X}, \boldsymbol{\phi}_j) = \textstyle\sum_{l=1}^{p} \mu^2(\mathcal{X}_l, \boldsymbol{\phi}_j). \quad (4)$$

That is, the local subspace coherence is fully decomposable with respect to orthogonal decompositions under square addition. It collects the local subspace coherences with respect to the decomposed subspaces $\{\mathcal{X}_l\}_l$. Also, notice that $\dim(\mathcal{X}) = |\mathcal{S}_x|$ corresponds to the sparsity of $\mathbf{x}_0$. Hence, the local subspace coherence preserves the sparsity information of $\mathbf{x}_0$ via the following accumulation rule

$$\textstyle\sum_{j=1}^{n} \mu^2(\mathcal{X}, \boldsymbol{\phi}_j) = \dim(\mathcal{X}) = |\mathcal{S}_x|. \quad (5)$$

Our notion of local subspace coherence is a measure of coherence between a basis vector and a signal subspace. In comparison, the notion of mutual coherence [9] seeks for the maximal coherence among all possible pairs of basis vectors. Thus, the local subspace coherence can be low even a high mutual coherence occurs, and we illustrate this point in Example 2 later. On the other hand, The cluster coherence [7, 10] corresponds to the sum of coherences of a cluster of basis vectors, and only helps to characterize the exactness condition in the asymptotic regime. In summary, our local subspace coherence has finer decomposable structures as in eq. (4), and naturally exploits the sparsity information as in eq. (5). In the next section, we show the fundamental role that the local subspace coherence plays in characterizing the exactness condition of **(P)**.

3. MAIN RESULT AND DISCUSSION

With the notion of local subspace coherence, the exactness condition of **(P)** is characterized as follows. We outline the proof due to the page limitations.

Theorem 1. *Suppose Assumption 1 hold and set $\lambda = \frac{1}{\log n}$ in **(P)**. Then $(\mathbf{x}_0, \mathbf{y}_0)$ is the unique minimizer of **(P)** with probability at least $1 - n^{-\sqrt{C_0}}$, provided that for all $j \in [n]$*

$$1 - p_j \geq C_0 \mu(\mathcal{X}, \boldsymbol{\phi}_j) \log^2 n, \tag{6}$$

$$1 - p_j \geq \tfrac{1}{n^3}, \tag{7}$$

where C_0 is a universal positive constant.

Outline of the proof. The proof is to show exactness of **(P)** via a convex duality argument. Specifically, **(P)** is shown to be exact if there exists a dual vector $\boldsymbol{\nu} \in \mathcal{Y}^\perp$ that satisfies the following set of conditions

$$\|\mathrm{P}_{\mathcal{X}} - \mathrm{P}_{\mathcal{X}} \mathrm{R}_{\mathcal{Y}^\perp} \mathrm{P}_{\mathcal{X}}\| \leq \tfrac{1}{2} \tag{8}$$

$$\|\mathrm{P}_{\mathcal{X}}(\boldsymbol{\gamma}_x - \lambda \boldsymbol{\gamma}_y + \boldsymbol{\nu})\|_2 \leq \tfrac{\lambda}{n} \tag{9}$$

$$\|\boldsymbol{\Psi}_{\mathcal{S}_x^c}(\boldsymbol{\nu} - \lambda \boldsymbol{\gamma}_y)\|_\infty \leq \tfrac{1}{4} \tag{10}$$

$$\|\boldsymbol{\Phi}\boldsymbol{\nu}\|_\infty \leq \tfrac{\lambda}{4}, \tag{11}$$

where $\mathrm{R}_{\mathcal{Y}^\perp}$ is a weighted sampling operator satisfying $\mathbb{E}\mathrm{R}_{\mathcal{Y}^\perp} = \mathrm{P}_{\mathcal{Y}^\perp}$, $\boldsymbol{\gamma}_x = \boldsymbol{\Psi}^\top \mathrm{sgn}(\boldsymbol{\theta}_x)$, $\boldsymbol{\gamma}_y = \boldsymbol{\Phi}^\top \mathrm{sgn}(\boldsymbol{\theta}_y)$, and $\boldsymbol{\Psi}_{\mathcal{S}_x^c}$ is matrix $\boldsymbol{\Psi}$ restricted on the rows with the indexes in $\mathcal{S}_x^c$. The dual vector $\boldsymbol{\nu}$ is then constructed via the iterative "golfing scheme" in [23]. Lastly, by exploiting the conditions in eqs. (6) and (7) and introducing the weighted infinity norm in [24], $\boldsymbol{\nu}$ is shown to satisfy the above set of conditions with high probability. □

The condition in eq. (7) is just to avoid singularity in the proof. Hence, the exactness condition of **(P)** is mainly characterized by eq. (6), which essentially requires that

$$\mathbb{P}(\boldsymbol{\phi}_j \notin \mathcal{Y}) \propto \mu(\mathcal{X}, \boldsymbol{\phi}_j).$$

That is, if the basis vector $\boldsymbol{\phi}_j$ has a high local subspace coherence with subspace $\mathcal{X}$, then the subspace $\mathcal{Y}$ should better not be spanned by this basis vector. Intuitively, this condition avoids the signal subspaces $\mathcal{X}, \mathcal{Y}$ to be highly coherent with each other. Otherwise, it is difficult to distinguish between the signal subspaces, and hence fails to demix the signal components. In summary, eq. (6) characterizes the distribution of the subspace $\mathcal{Y}$ that avoids high coherence with subspace $\mathcal{X}$, and this is a general exactness condition that is applicable to all pairs of orthonormal basis. We note that the choice $\lambda = \frac{1}{\log n}$ is just to ease the proof details, and one can adopt other choices (such as $\lambda = 1$) in practical applications.

Next, we consider several specific examples to further understand the role that the local subspace coherence plays in the exactness condition. Also, we characterize the total sparsity of the signal components allowed under eq. (6), and show its advantage by comparing with other state-of-art results. Throughout, we consider the orthonormal basis: $\mathcal{F}, \mathcal{I}, \mathcal{H}$, which correspond to the discrete Fourier basis, canonical basis and bivariate Haar wavelet basis, respectively.

Example 1. $\boldsymbol{\Psi} = \mathcal{F}, \boldsymbol{\Phi} = \mathcal{I}$.

This example has been discussed in literature [9, 8, 16], where one aims at demixing sparse Fourier signal and sparse discrete-time signal from their mixture. A well known fact of the $\mathcal{F}$-$\mathcal{I}$ basis pair is the maximal incoherence property, i.e.,

$$\forall i, j \in [n], \;\; |\langle \boldsymbol{\psi}_i, \boldsymbol{\phi}_j \rangle| = \frac{1}{\sqrt{n}}. \tag{12}$$

Thus, by calculation we obtain that $\mu(\mathcal{X}, \boldsymbol{\phi}_j) \equiv \sqrt{\frac{|\mathcal{S}_x|}{n}}$. In this case, eq. (6) suggests that the distribution $\{p_j\}_j$ of $\mathcal{S}_y$ be uniform, coinciding with existing result [9]. Then, taking the sum of eq. (6) over all $j \in [n]$ and noticing that $\sum_j p_j = \mathbb{E}|\mathcal{S}_y|$, we further obtain that

$$\mathbb{E}|\mathcal{S}_y| + C_0 \sqrt{n|\mathcal{S}_x|} \log^2 n \leq n. \tag{13}$$

Hence, our exactness condition of **(P)** for Example 1 allows $\mathbb{E}|\mathcal{S}_y| = \mathcal{O}(n)$, and $|\mathcal{S}_x| = \mathcal{O}(\frac{n}{\log^4 n})$. The allowed total sparsity is comparable (up to certain logarithm factors) to the results in [9, Theorem 4.2, Theorem 5.1]. We note that $\mathbb{E}|\mathcal{S}_y|$ can be replaced by $|\mathcal{S}_y|$ by concentration of Bernoulli random variables.

Example 2. $\boldsymbol{\Psi} = \mathcal{H}, \boldsymbol{\Phi} = \mathcal{F}$.

Wavelets are suitable for representing singularities in the signal, and is a typical choice of basis to provide sparse representations of images. This pair of basis share a common basis vector $(\frac{1}{\sqrt{n}}, \cdots, \frac{1}{\sqrt{n}})$, and hence the mutual coherence is as high as $\mu(\mathcal{H}, \mathcal{F}) = 1$. Consequently, the exactness condition in [9, Theorem 5.1] allows a total sparsity of the trivial order $\mathcal{O}(\frac{1}{\log^6 n})$.

However, the coherence between the basis vectors vanishes geometrically fast in the asymptotic regime, and the mutual coherence does not exploit this asymptotic incoherence property. To characterize the coherence pattern specifically, we assume, without of loss of generality, that $n = 2^p$ for some positive integer p. Then, we partition the index set $[n]$ into $p+1$ dyadic levels, i.e., $L_0 = \{1\}, L_1 = \{2\}, L_2 = \{3, 4\}, \cdots, L_p = \{\frac{n}{2} + 1, \cdots, n\}$. For any $j \in [n]$ we also define $k(j)$ as the dyadic level that j belongs to, i.e., $j \in L_{k(j)}$. Then, the following pattern of coherence follows from [25, Lemma D.1]

$$\forall i, j \in [n], \;\; |\langle \boldsymbol{\psi}_i, \boldsymbol{\phi}_j \rangle| \leq 2^{-\frac{k(j) + |k(j) - k(i)|}{2}}. \tag{14}$$

The coherence pattern in eq. (14) implies that the Fourier basis vectors and the wavelet basis vectors are asymptotically incoherent (i.e., for large $k(i), k(j)$). Moreover, it helps to control the local subspace coherence by the sparsity of the signal component as follows.

Theorem 2. *Consider the coherence pattern in eq.* (14)*. Then for the subspace* $\mathcal{X}$ *with* $\dim(\mathcal{X}) = |\mathcal{S}_x|$*, we have*

$$\sum_{j=1}^{n} \mu(\mathcal{X}, \boldsymbol{\phi}_j) \leq \tfrac{2\sqrt{2}}{\sqrt{2}-1}|\mathcal{S}_x|. \tag{15}$$

Consequently, by summing over all $j \in [n]$*, the exactness condition in eq.* (6) *for Example 2 implies that*

$$\mathbb{E}|\mathcal{S}_y| + |\mathcal{S}_x| \leq \mathcal{O}(\tfrac{n}{\log^2 n}). \tag{16}$$

That is, the asymptotic incoherence pattern together with our exactness condition, allows a total sparsity on a near optimal order (up to certain logarithm factor) for Example 2. Hence, our exactness condition overcomes the mutual coherence barrier in [9, Theorem 5.1] via the local subspace coherence, which allows to exploit the pattern of coherence as in eq. (14). Thus, the local subspace coherence plays a more fundamental role in characterizing the exactness condition of **(P)** than the mutual coherence.

Example 3. $\boldsymbol{\Psi} = \boldsymbol{\Phi}$.

In this extreme example, the basis are exactly aligned, i.e.,

$$\forall i, j \in [n],\ |\langle \boldsymbol{\psi}_i, \boldsymbol{\phi}_j \rangle| = \mathbb{I}\{i = j\}, \tag{17}$$

where $\mathbb{I}\{\cdot\}$ is the binary indicator function. Consequently, for $C_0 \log^2 n \geq 1$, the exactness condition in eq. (6) requires that $\boldsymbol{\phi}_j \notin \mathcal{Y}$ whenever $\boldsymbol{\phi}_j \in \mathcal{X}$. This is to ensure that $\mathcal{X} \cap \mathcal{Y} = \emptyset$, since otherwise it is impossible to demix signals at a common coordinate in the same basis.

The three examples above imply that the local subspace coherence plays a more fundamental role in characterizing the exactness condition of **(P)**. We note that the discussion is not limited to the examples presented here. In general, one need to identify the coherence pattern of the pair of basis, and utilize that to further characterize the local subspace coherence.

4. EXPERIMENTS

In this section, we verify our theoretical finding via numerical experiment. Specifically, we randomly generate $\mathbf{x}_0, \mathbf{y}_0 \in \mathbb{R}^{512}$ with their sparsity $|\mathcal{S}_x|, |\mathcal{S}_y|$ ranging from $0\% - 60\%$ of the dimensionality. The support of $\boldsymbol{\theta}_x$ is generated uniformly at random, while the support of $\boldsymbol{\theta}_y$ is generated according to the exactness condition in eq. (6). We then solve **(P)** with $\lambda = 1$ for the solution $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ via the popular ADMM algorithm [26]. The details of the settings are summarized in Algorithm 1. All experiments are repeated 50 times, and we claim a success if the average of the relative error $\frac{\|\hat{\mathbf{x}}-\mathbf{x}_0\|_2+\|\hat{\mathbf{y}}-\mathbf{y}_0\|_2}{\|\mathbf{x}_0\|_2+\|\mathbf{y}_0\|_2}$ is less than 0.01.

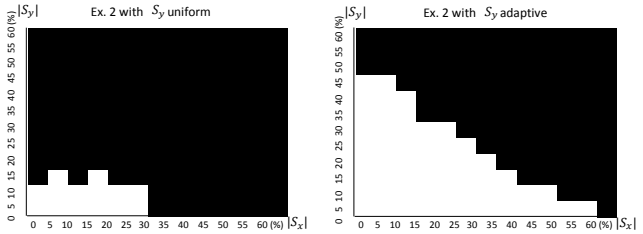
We consider Example 2 with $\boldsymbol{\Psi} = \mathcal{H}, \boldsymbol{\Phi} = \mathcal{F}$. For this example, we compare two distributions of $\mathcal{S}_y$: the uniform distribution required by the mutual coherence in [9] and the distribution adapted to the local subspace coherence in eq. (6). Figure 1 shows the success region with respect to the sparsity

Algorithm 1 Experiment setup

1). Specify a pair of orthonormal basis $\boldsymbol{\Psi}, \boldsymbol{\Phi}$.
2). Generate the support of $\boldsymbol{\theta}_x$ uniformly at random with cardinality $|\mathcal{S}_x|$; Evaluate the local subspace coherence $\{\mu(\mathcal{X}, \boldsymbol{\phi}_j)\}_j$, and normalize $\{1 - p_j\}_j$ in eq. (6) to a probability distribution; Generate the support of $\boldsymbol{\theta}_y$ according to

$$\mathbb{P}(j \notin \mathcal{S}_y) = 1 - p_j$$

with cardinality $|\mathcal{S}_y|$ (After each sample j we set $1 - p_j = 0$ and renormalize $\{1 - p_j\}_j$.).
3). Generate the magnitudes of $\boldsymbol{\theta}_x, \boldsymbol{\theta}_y$ from normal distribution, and set $\mathbf{x}_0 = \boldsymbol{\Psi}^\top \boldsymbol{\theta}_x, \mathbf{y}_0 = \boldsymbol{\Psi}^\top \boldsymbol{\theta}_y$.
4). Solve **(P)** with $\lambda = 1$ via ADMM for the solution $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$.
5). Claim a success if $\frac{\|\hat{\mathbf{x}}-\mathbf{x}_0\|_2+\|\hat{\mathbf{y}}-\mathbf{y}_0\|_2}{\|\mathbf{x}_0\|_2+\|\mathbf{y}_0\|_2} < 0.01$.



(a) (b)

Fig. 1. Comparison of success region between (a) $\mathcal{S}_y$ be uniformly at random and (b) $\mathcal{S}_y$ be adapted to local subspace coherence.

for different distributions of $\mathcal{S}_y$. Clearly from Figure 1.(a), **(P)** fails to demix the signal components for a major sparsity region when $\mathcal{S}_y$ is generated uniformly at random. This is because the highly coherent basis vectors in $\mathcal{H}$ and $\mathcal{F}$ likely span the signal subspaces under the uniform distribution, and lead to a high local subspace coherence. On the other hand, if $\mathcal{S}_y$ is generated according to our exactness condition, high local subspace coherence is avoided and hence **(P)** succeeds for a much wider sparsity region as shown in Figure 1.(b). Hence, the simulation confirmed that avoiding high local subspace coherence is the key factor that makes **(P)** exact.

5. CONCLUSION

In this paper, we consider the problem of demixing a pair of sparse signals via a convex optimization approach. We introduce the notion of local subspace coherence to characterize how a basis vector is coherent with a subspace. It turns out that convex demixing succeeds as long as the signal subspaces are distributed in a way that avoids high local subspace coherence. We hope that this exactness condition can provide a criterion for designing new models for demixing signals.

6. REFERENCES

[1] J. L. Starck, M. Elad, and D. L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1570–1582, 2005.

[2] A. Aldroubi, C. Chui, M. Elad, J.-L. Starck, P. Querre, and D.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.

[3] X. Huo, "Sparse image representation via combined transforms," *Ph.D. dissertation*, 1999.

[4] J. Wright and Y. Ma, "Dense error correction via ℓ_1 - minimization," *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3540–3560, 2010.

[5] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Processing*, vol. 81, no. 11, pp. 2353 – 2362, 2001.

[6] J. Bobin, J. L. Starck, J. Fadili, and Y. Moudden, "Sparsity and morphological diversity in blind source separation," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2662–2674, 2007.

[7] D. L. Donoho and G. Kutyniok, "Geometric separation using a wavelet-shearlet dictionary," in *SAMPTA'09*, May 2009, p. Special session on geometric multiscale analysis.

[8] M. McCoy, "A geometric analysis of convex demixing," *Ph.D. dissertation*, 2013.

[9] E. J. Candès and J. Romberg, "Quantitative robust uncertainty principles and optimally sparse decompositions," *Foundations of Computational Mathematics*, vol. 6, no. 2, pp. 227–254, 2006.

[10] D. L. Donoho and G. Kutyniok, "Microlocal analysis of the geometric separation problem," *Communications on Pure and Applied Mathematics*, vol. 66, no. 1, pp. 1–47, 2013.

[11] J. Wright, A. Ganesh, K. Min, and Y. Ma, "Compressive principal component pursuit," in *IEEE International Symposium on Information Theory*, July 2012, pp. 1276–1280.

[12] E. J. Candès., X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of the ACM*, vol. 58, no. 3, pp. 11:1–11:37, 2011.

[13] H. Zhang, Y. Zhou, and Y. Liang, "Analysis of robust pca via local incoherence," in *Advances in Neural Information Processing Systems*, pp. 1810–1818. 2015.

[14] B. Adcock, A.C. Hansen, C. Poon, and B. Roman, "Breaking the coherence barrier: asymptotic incoherence and asymptotic sparsity in compressed sensing," *arXiv*, 2013.

[15] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, pp. 797–829, 2004.

[16] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, 2006.

[17] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Transactions on Information Theory*, vol. 49, no. 12, pp. 3320–3325, 2003.

[18] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2558–2567, 2002.

[19] A. Feuer and A. Nemirovski, "On sparse representation in pairs of bases," *IEEE Transactions on Information Theory*, vol. 49, no. 6, pp. 1579–1581, 2003.

[20] J. J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 1341–1344, 2004.

[21] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[22] D. L. Donoho and P. B. Stark, "Uncertainty principles and signal recovery," *SIAM Journal on Applied Mathematics*, vol. 49, no. 3, pp. 906–931, 1989.

[23] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1548–1566, 2011.

[24] Y. Zhou, H. Zhang, and Y. Liang, "On compressive orthonormal sensing," *54th Annual Allerton Conference on Communication, Control, and Computing*, 2016.

[25] C. Boyer, J. Bigot, and P. Weiss, "Compressed sensing with structured sparsity and structured acquisition," *arxiv*, 2015.

[26] B. He and X. Yuan, "On the $\mathbf{O}(1/n)$ convergence rate of the douglasrachford alternating direction method," *SIAM Journal on Numerical Analysis*, vol. 50, no. 2, pp. 700–709, 2012.