SEPARATION OF A SUBSPACE-SPARSE SIGNAL: ALGORITHMS AND CONDITIONS

Arvind Ganesh *Zihan Zhou* *Yi Ma*

Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign

ABSTRACT

In this paper, we show how two classical sparse recovery algorithms, Orthogonal Matching Pursuit and Basis Pursuit, can be naturally extended to recover block-sparse solutions for subspace-sparse signals. A subspace-sparse signal is sparse with respect to a set of subspaces, instead of atoms. By generalizing the notion of *mutual incoherence* to the set of subspaces, we show that all classical sufficient conditions remain exactly the same for these algorithms to work for subspace-sparse signals, in both noiseless and noisy cases. The sufficient conditions provided are easy to verify for large systems. We conduct simulations to compare the performance of the proposed algorithms.

Index Terms— subspace sparse, subspace incoherence, subspace matching pursuit, subspace base pursuit.

1. INTRODUCTION

Sparse signal recovery has received a lot of attention recently, following the pioneering work of Donoho [1] and Candes [2]. The basic assumption is that the observed signal is a sparse linear combination of vectors (also called atoms) in a given dictionary. More generally, however, we may assume that the observed signal is in the direct sum of a small number of subspaces among a large set of possible subspaces. For example, an audio signal could be mixed from multiple sources (say speakers), where each source can be modeled by a linear subspace. We call such a signal *subspace-sparse*. More precisely, the observation vector $\boldsymbol{y}$ is expressed as:

$$\boldsymbol{y} = \sum_{i=1}^{m} A_i \, \boldsymbol{x}_i, \tag{1}$$

where A_i's are bases for the m subspaces in consideration and $\boldsymbol{x}_i$'s are the corresponding coefficient vectors. For a subspace-sparse $\boldsymbol{y}$, at most k of the m $\boldsymbol{x}_i$'s are non-zero. We are interested in the problem of separating these sparse components $\boldsymbol{y}_i = A_i\boldsymbol{x}_i$, or equivalently $\boldsymbol{x}_i$. One may treat this as a sparse signal representation problem: recovering $\boldsymbol{x}$ from $\boldsymbol{y} = A\boldsymbol{x}$ with $A \doteq [\,A_1 \;\, A_2 \ldots A_m\,]$ and $\boldsymbol{x} \doteq [\,\boldsymbol{x}_1^T \;\, \boldsymbol{x}_2^T \ldots \boldsymbol{x}_m^T\,]^T$. However, the desired solution $\boldsymbol{x}$ now has special structure, as the nonzero coefficients appear in blocks and sparsity within each block (subspace) is no longer important. This property has also been termed *block sparsity* in the literature [3].

The separation of a subspace-sparse signal, particularly by convex optimization methods, has been investigated in [3] and [4]. They have provided certain sufficient conditions under which the algorithms are guaranteed to find the correct block-sparse solution. However, the sufficient conditions given are computationally intractable to verify for large linear systems.

To alleviate some of that difficulty, in this paper, we generalize the notion of *mutual incoherence* to a set of subspaces, which is easy to compute, and show how the two classical sparse recovery algorithms, namely Orthogonal Matching Pursuit [5, 6] and Basis Pursuit [7], can be naturally extended to recover block-sparse solutions for subspace-sparse signals. In terms of the new subspace incoherence, we show that all classical sufficient conditions for these algorithms to work [8] remain exactly the same for the subspace-sparse problem, in both noiseless and noisy cases.

2. SUBSPACE INCOHERENCE

Given a set of subspaces $S_1, S_2, \ldots, S_m$ that do not intersect with each other, we define *mutual subspace incoherence* μ as a measure of the smallest angle between any two subspaces from the given set.

$$\mu \doteq \max_{\substack{i,j\in[m] \\ i\neq j}} \left\{ \max_{\substack{\boldsymbol{x}\in S_i \\ \boldsymbol{y}\in S_j}} \frac{|\langle \boldsymbol{x}, \boldsymbol{y}\rangle|}{\|\boldsymbol{x}\|_2 \; \|\boldsymbol{y}\|_2} \right\}. \tag{2}$$

We define the *spark* of the set of subspaces $S_1, \ldots, S_m \subseteq \mathbb{R}^d$, denoted spark($S_1, \ldots, S_m$), as the smallest integer r such that

$$\phi_{\lambda_1} + \phi_{\lambda_2} + \cdots + \phi_{\lambda_r} = 0,$$

where $\phi_i \neq 0, \phi_i \in S_i$ and the λ_i's are distinct indices from $[m] \doteq \{1, \ldots, m\}$. This is a natural generalization to the spark of a matrix, the smallest number of columns that are linearly dependent [8].

Computing the spark of a set of subspaces is combinatorial, while the incoherence can be computed relatively easily. Similar to Lemma 1 in [8], we can bound the spark by incoherence as follows:

Lemma 1. For any collection of subspaces $S_1, S_2, \ldots, S_m$, we have

$$\mathrm{spark}(S_1, \ldots, S_m) \geq 1 + \frac{1}{\mu(S_1, \ldots, S_m)}.$$

Proof. The proof is similar to the proof of Lemma 1 in [8], except for the definition of the matrix A. In our case, the matrix $A = [\phi_1 \; \phi_2 \ldots \phi_m]$, where ϕ_i is any vector from S_i. Thus, instead of a fixed matrix, we have a family of matrices. □

Notation: We assume a signal $\boldsymbol{y}$ is k-subspace sparse w.r.t. $\mathcal{R}(A_1), \ldots, \mathcal{R}(A_m)$. Without loss of generality, $\boldsymbol{y}$ can be written as $\boldsymbol{y} = \sum_{i=1}^{k} w_i \boldsymbol{z}_i$, where $\boldsymbol{z}_j \in S_j$ and $\|\boldsymbol{z}_j\|_2 = 1$ for $j = 1, \ldots, k$, and $|w_1| \geq |w_2| \geq \cdots \geq |w_k| > 0$. Let $\boldsymbol{x}_A \doteq [\, \|A_1 \boldsymbol{x}_1\|_2 \;\; \|A_2 \boldsymbol{x}_2\|_2 \ldots \|A_m \boldsymbol{x}_m\|_2 \,]^T$. Since the A_i's are assumed to be full column rank matrices, $\|\boldsymbol{x}_A\|_0$ is the number of nonzero $\boldsymbol{x}_i$'s. We use $\pi_i(\cdot)$ to denote the projection operator onto subspace $S_i = \mathcal{R}(A_i)$, $i = 1, \ldots, m$.

3. SUBSPACE MATCHING PURSUIT

Analogous to the Orthogonal Matching Pursuit (OMP) algorithm [5, 6], we propose a greedy algorithm for recovering block-sparse solutions that is guaranteed to recover the correct subspace components when the subspaces are sufficiently incoherent.

Algorithm 1 Subspace Matching Pursuit (SMP).

Input: Bases $A_1, A_2, \ldots, A_m$, a vector $\boldsymbol{y}$, and its sparsity k.

Initialize: $\boldsymbol{r}_0 = \boldsymbol{y}, t = 1, \Lambda_0 = \emptyset$, and Φ_0 a null matrix.
while $\boldsymbol{r}_{t-1} \neq \boldsymbol{0}$ and $t \leq k$ **do**
    Find the index λ_t as follows

$$\lambda_t = \operatorname*{arg\,min}_{j=1,\ldots,m} \left\| \boldsymbol{r}_{t-1} - \pi_j(\boldsymbol{r}_{t-1}) \right\|_2.$$

    $\Lambda_t \leftarrow \Lambda_{t-1} \cup \{\lambda_t\}$ and $\Phi_t \leftarrow [\, \Phi_{t-1} \; A_{\lambda_t} \,]$.
    $\hat{\boldsymbol{z}}_t = \arg\min_{\boldsymbol{z}} \|\boldsymbol{y} - \Phi_t \, \boldsymbol{z}\|_2$,
    $\boldsymbol{x}_{\lambda_j} = \hat{\boldsymbol{z}}_{t \,|\, A_{\lambda_j}}$ for $j = 1, \ldots, t$ and $\boldsymbol{x}_i = 0$ for $i \notin \Lambda_t$.
    $\boldsymbol{r}_t \leftarrow \boldsymbol{y} - \Phi_t \, \hat{\boldsymbol{z}}_t$ and $t \leftarrow t + 1$.
end while

Output: Coefficients $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$.

Theorem 1. Suppose that a signal $\boldsymbol{y}$ is k-subspace sparse w.r.t. $\mathcal{R}(A_1), \ldots, \mathcal{R}(A_m)$, and

$$k < \frac{1}{2}\left(1 + \frac{1}{\mu}\right), \tag{3}$$

where μ is the incoherence between the subspaces $\mathcal{R}(A_1)$, $\mathcal{R}(A_2), \ldots, \mathcal{R}(A_m)$. Then, SMP is guaranteed to find the k subspaces exactly.

Proof. The proof is an extension of the proof of Theorem 3 in [8]. One of the first k subspaces is chosen at the first step if

$$\|\pi_1(\boldsymbol{y})\|_2 > \|\pi_t(\boldsymbol{y})\|_2, \quad t > k. \tag{4}$$

We now find a lower bound to the left hand side:

$$\begin{aligned} \|\pi_1(\boldsymbol{y})\|_2 &= \left\| w_1 \boldsymbol{z}_1 + \sum_{i=2}^{k} w_i \pi_1(\boldsymbol{z}_i) \right\|_2 \\ &\geq |w_1| - \sum_{i=2}^{k} |w_i| \, \|\pi_1(\boldsymbol{z}_i)\|_2 \geq |w_1| - \sum_{i=2}^{k} |w_1| \mu \\ &= |w_1| \, (1 - \mu(k-1)), \end{aligned}$$

and an upper bound to the right hand side:

$$\|\pi_t(\boldsymbol{y})\|_2 = \Big\| \sum_{i=1}^{k} w_i \, \pi_t(\boldsymbol{z}_i) \Big\|_2 \leq |w_1| \, \mu \, k.$$

Then, Equation (4) holds if $|w_1| \, (1 - \mu(k-1)) > |w_1| \, \mu \, k$, which leads to equation (3).

From the above inequality, we infer that if equation (3) is satisfied, the first iteration chooses a correct subspace. The residual $\boldsymbol{r}_1$ is orthogonal to the range space of A_{λ_1} due to the least squares step and can be expressed as

$$\boldsymbol{r}_1 = \boldsymbol{y} - A_{\lambda_1} \hat{\boldsymbol{z}}_1 = \sum_{i=1}^{k} A_i \tilde{\boldsymbol{x}}_i = \sum_{i=1}^{k} \tilde{w}_i \tilde{\boldsymbol{z}}_i,$$

where $\tilde{\boldsymbol{z}}_j \in S_j$, $\|\tilde{\boldsymbol{z}}_j\|_2 = 1 \, \forall j = 1, \ldots, k$, $|\tilde{w}_1| \geq |\tilde{w}_2| \geq \cdots \geq |\tilde{w}_k| > 0$. Repeating the same arguments as above, we are guaranteed to find a subspace from the true support set. The orthogonality of $\boldsymbol{r}_t$ to Φ_t at each iteration ensures that a subspace is not chosen a second time. Thus, after at most k iterations, the algorithm finds the smallest set of subspaces whose direct sum contains $\boldsymbol{y}$. □

The above proof implicitly assumes that the sparsest subspace representation of a signal is necessarily unique. The following theorem, along with Lemma 1, establishes that if equation (3) holds, then uniqueness of representation is guaranteed.

Theorem 2. Given m subspaces, if $\boldsymbol{y}$ can be represented as a linear sum of elements from k distinct subspaces such that $k < \frac{1}{2} \operatorname{spark}(S_1, \ldots, S_m)$, then the sparse decomposition is necessarily unique *i.e.,* $\boldsymbol{y}$ cannot be represented as the linear sum of elements from a different set of k subspaces.

Proof. Assume the contrary. Then, we have $\boldsymbol{y} = \sum_{i=1}^{k} \boldsymbol{z}_{\lambda_i} w_i = \sum_{j=1}^{l} \boldsymbol{z}_{\eta_j} v_j$ for some $l \leq k$, where $\boldsymbol{z}_i \in S_i$. Thus, $\sum_{i=1}^{k} \boldsymbol{z}_{\lambda_i} w_i - \sum_{j=1}^{l} \boldsymbol{z}_{\eta_j} v_j = 0$, a linear sum of at most $k + l < \operatorname{spark}(S_1, \ldots, S_m)$ subspaces, which is a contradiction to the definition of spark. □

Now suppose we are given a noisy observation $\boldsymbol{y} = \boldsymbol{y}_0 + \boldsymbol{n}$ of a k-subspace sparse signal $\boldsymbol{y}_0 = \sum_{i=1}^{k} A_i \, \boldsymbol{x}_i$. Assuming that the noise $\boldsymbol{n}$ is bounded $\|\boldsymbol{n}\|_2 \leq \epsilon$, we have:

Theorem 3. Let $\hat{\boldsymbol{x}}$ be the output of applying SMP to $\boldsymbol{y}$, which stops when the residual $\|\boldsymbol{r}_t\| \leq \epsilon$. If

$$k < \frac{1+\mu}{2\mu} - \frac{1}{\mu} \cdot \frac{\epsilon}{|w_k|}, \tag{5}$$

then the support is correctly recovered: $\operatorname{supp}(\hat{\boldsymbol{x}}_A) = \operatorname{supp}(\boldsymbol{x}_A)$, and the estimation error is bounded:

$$\|\hat{\boldsymbol{x}}_A - \boldsymbol{x}_A\|_2^2 \leq \frac{\epsilon^2}{1 - \mu(k-1)}. \tag{6}$$

Proof. We first show the support is correctly recovered, which can be established by similar reasoning as the proof of Theorem 1. The SMP operates by projecting $\boldsymbol{y} = \boldsymbol{y}_0 + \boldsymbol{n}$ onto each subspace S_i in turn. Again, it will select one of the first k subspaces at the first step if equation (4) is satisfied. The left hand side of equation (4) can be bounded below by $\|\pi_1(\boldsymbol{y})\|_2 \geq \|\pi_1(\boldsymbol{y}_0)\|_2 - \|\pi_1(\boldsymbol{n})\|_2 \geq |w_1|\,(1-\mu(k-1))-\epsilon$, and similarly a bound for the right hand side is

$$\|\pi_t(\boldsymbol{y})\|_2 \leq |w_1|\,\mu\,k+\epsilon.$$

Together they lead to the following sufficient condition

$$k < \frac{1+\mu}{2\mu} - \frac{1}{\mu}\cdot\frac{\epsilon}{|w_1|} \tag{7}$$

for (4) to hold. Repeating the argument above for each iteration, we have the following conditions that guarantee SMP to succeed after k itereations

$$k < \frac{1+\mu}{2\mu} - \frac{1}{\mu}\cdot\frac{\epsilon}{|w_i|}, \quad i=1,2,\ldots,k. \tag{8}$$

Since $|w_i|$ is ordered $|w_1| \geq |w_2| \geq \cdots \geq |w_k| > 0$, it is sufficient to only require equation (5).

Now we turn to the claim about the estimation error. Define $\boldsymbol{e} = \hat{\boldsymbol{x}} - \boldsymbol{x}$. Let $\boldsymbol{u} = \boldsymbol{e}_A$, $\psi_i = \frac{A_i\boldsymbol{e}_i}{\|A_i\boldsymbol{e}_i\|_2}$ for $i = 1,2,\ldots,m$. We use the triangle inequality to get

$$\begin{aligned}\|\hat{\boldsymbol{x}}_A - \boldsymbol{x}_A\|_2^2 &= \sum_{i=1}^m \left(\|A_i(\boldsymbol{x}_i+\boldsymbol{e}_i)\|_2 - \|A_i\boldsymbol{x}_i\|_2\right)^2 \\ &\geq \sum_{i=1}^m \|A_i\boldsymbol{e}_i\|_2^2 = \|\boldsymbol{e}_A\|_2^2 = \|\boldsymbol{u}\|_2^2.\end{aligned}$$

Define $\Psi \doteq [\,\psi_1\,\psi_2\,\ldots\,\psi_m\,]$. Each column in Ψ is from a different subspace and the subspaces have a mutual incoherence factor of μ. It follows that the kth singular value of Ψ is bounded from below by $(1-\mu(k-1))^{1/2}$ (Lemma 2.2 in [9].) That is, $\|\Psi\boldsymbol{u}\|_2^2 \geq \|\boldsymbol{u}\|_2^2(1-\mu(k-1))$. Therefore

$$\|\hat{\boldsymbol{x}}_A - \boldsymbol{x}_A\|_2^2 \leq \|\boldsymbol{u}\|_2^2 \leq \frac{\|\Psi\boldsymbol{u}\|_2^2}{1-\mu(k-1)} \leq \frac{\epsilon^2}{1-\mu(k-1)}.$$

□

4. SUBSPACE BASE PURSUIT

Ultimately, the separation problem for a subspace-sparse signal is to solve the following problem:

$$(P_0) \quad \min \|\boldsymbol{x}_A\|_0 \quad \text{subject to} \quad \boldsymbol{y} = \sum_{i=1}^m A_i\boldsymbol{x}_i. \tag{9}$$

Since (P_0) is in general an NP-hard problem, we seek to solve its relaxed convex version, by replacing $\|\boldsymbol{x}_A\|_0$ with $\|\boldsymbol{x}_A\|_1$. We call this approach Subspace Base Pursuit (SBP) [1], analogous to Basis Pursuit (BP) for the vector case [7]:

$$(P_1) \quad \min \sum_{i=1}^m \|A_i\boldsymbol{x}_i\|_2 \quad \text{subject to} \quad \boldsymbol{y} = \sum_{i=1}^m A_i\boldsymbol{x}_i. \tag{10}$$

The following theorem states that under the same condition as the SMP on the sparsity of the signal, the SBP finds the correct sparse solution.

[1] This coincides with the scheme proposed in [3, 4] if $A_1,\ldots,A_m$ are orthogonal matrices.

Theorem 4. Suppose that $\boldsymbol{y}$ is k-subspace sparse w.r.t. $\mathcal{R}(A_1),\ldots,\mathcal{R}(A_m)$, and

$$k < \frac{1}{2}\left(1+\frac{1}{\mu}\right), \tag{11}$$

where μ is the incoherence between the subspaces $\mathcal{R}(A_1)$, $\mathcal{R}(A_2)$, …, $\mathcal{R}(A_m)$. Then, SBP is guaranteed to find the k subspaces exactly.

Proof. We have already proved that if equation (11) holds, then the solution is necessarily unique. To prove that SBP finds this solution, we prove that the set of alternative solutions is empty.

Let $\mathcal{C}_s$ be the set of $\boldsymbol{e}$ such that $\boldsymbol{x}+\boldsymbol{e}$ is an alternative solution: $\mathcal{C}_s \doteq \{\boldsymbol{e}|\boldsymbol{e}\neq 0, \|(\boldsymbol{e}+\boldsymbol{x})_A\|_1 \leq \|\boldsymbol{x}_A\|_1, A\boldsymbol{e}=0\}$. It can be shown that

$$\mathcal{C}_s \subseteq \mathcal{C}_1 \doteq \Big\{\boldsymbol{e}\;\Big|\;\boldsymbol{e}\neq 0, \|\boldsymbol{e}_A\|_1 \leq 2\sum_{i=1}^k \|A_i\boldsymbol{e}_i\|_2, A\boldsymbol{e}=0\Big\}.$$

Let $\boldsymbol{u} = \boldsymbol{e}_A$, $\psi_i = \frac{A_i\boldsymbol{e}_i}{\|A_i\boldsymbol{e}_i\|_2}$ for $i = 1,2,\ldots,m$. Define $\Psi \doteq [\,\psi_1\,\ldots\,\psi_m\,]$. Then, $\mathcal{C}_1$ is empty iff $\mathcal{C}_2$ is empty, where

$$\mathcal{C}_2 \doteq \Big\{\boldsymbol{u}\;\Big|\;\boldsymbol{u}\neq 0, \|\boldsymbol{u}\|_1 \leq 2\sum_{i=1}^k u_i, \Psi\boldsymbol{u}=0\Big\}, \tag{12}$$

and $u_i = \|A_i\boldsymbol{e}_i\|_2$. We now prove that $\mathcal{C}_2$ is empty.

Each column in Ψ is from a different subspace and the subspaces have a mutual incoherence factor of μ. Therefore,

$$u_i \leq \frac{\mu}{1+\mu}\|\boldsymbol{u}\|_1, \quad i=1,2,\ldots,m. \tag{13}$$

We observe that

$$2\sum_{i=1}^k u_i \leq \frac{2k\mu}{1+\mu}\|\boldsymbol{u}\|_1 < \|\boldsymbol{u}\|_1. \tag{14}$$

Thus, by equation (12), we see that $\mathcal{C}_2$ is empty. Hence, SBP finds the unique sparsest solution. □

In Theorem 4, the observation $\boldsymbol{y}$ contains no noise. Consider the following modified version of the SBP:

$$(P_{1,\delta}) \quad \min \sum_{i=1}^m \|A_i\boldsymbol{x}_i\|_2 \;\; \text{subject to} \;\; \Big\|\boldsymbol{y} - \sum_{i=1}^m A_i\boldsymbol{x}_i\Big\|_2 \leq \delta. \tag{15}$$

The following theorem shows that $P_{1,\delta}$ is robust to small noise in the observation.

Theorem 5. Suppose that we are given a noisy observation $\boldsymbol{y} = \boldsymbol{y}_0 + \boldsymbol{n}$ of a k-subspace sparse signal $\boldsymbol{y}_0 = \sum_{i=1}^k A_i\,\boldsymbol{x}_i$, and

$$k < \frac{1}{4}\left(1+\frac{1}{\mu}\right), \tag{16}$$

where μ is the incoherence between the subspaces $\mathcal{R}(A_1)$, $\mathcal{R}(A_2)$, …, $\mathcal{R}(A_m)$. Then, provided $\epsilon \leq \delta$, the solution $\hat{\boldsymbol{x}}_1,\ldots,\hat{\boldsymbol{x}}_m$ returned by $P_{1,\delta}$ has the following property.

$$\|\hat{\boldsymbol{x}}_A - \boldsymbol{x}_A\|_2^2 \leq \frac{(\epsilon+\delta)^2}{1-\mu(4k-1)}. \tag{17}$$

Proof. The proof follows along similar lines to the proof of Theorem 3.1 in [9], and hence we borrow notation from there. Let $\boldsymbol{e} = \hat{\boldsymbol{x}} - \boldsymbol{x}$. Since $\hat{\boldsymbol{x}}$ is the optimal solution to $P_{1,\delta}$, we have that $\|\hat{\boldsymbol{x}}_A\|_1 \leq \|\boldsymbol{x}_A\|_1$. It follows that $\|\boldsymbol{e}_A\|_1 \leq 2\sum_{i=1}^{k} \|A_i \boldsymbol{e}_i\|_2$.

Let $\boldsymbol{u} = \boldsymbol{e}_A$, $\psi_i = \frac{A_i \boldsymbol{e}_i}{\|A_i \boldsymbol{e}_i\|_2}$, and $\Psi = [\psi_1 \ \psi_2 \ldots \psi_m]$. Then, applying the triangle inequality, we have that $\|\Psi \boldsymbol{u}\|_2 \leq \epsilon + \delta$. Define $\Delta = \epsilon + \delta$. Note that each off-diagonal entry in the Gram matrix $G = \Psi^T \Psi$ has absolute value $\leq \mu$. Therefore, we have

$$\begin{aligned} \Delta^2 &\geq \boldsymbol{u}^T G \boldsymbol{u} = \|\boldsymbol{u}\|_2^2 + \boldsymbol{u}^T (G - I) \boldsymbol{u} \\ &\geq \|\boldsymbol{u}\|_2^2 - |\boldsymbol{u}|^T |G - I| |\boldsymbol{u}| \geq \|\boldsymbol{u}\|_2^2 - \mu |\boldsymbol{u}|^T |\mathbf{1} - I| |\boldsymbol{u}| \\ &= (1 + \mu)\|\boldsymbol{u}\|_2^2 - \mu \|\boldsymbol{u}\|_1^2. \end{aligned}$$

Since $\|\hat{\boldsymbol{x}}_A - \boldsymbol{x}_A\|_2 \leq \|\boldsymbol{u}\|_2$, instead of directly looking for a bound of $\|\hat{\boldsymbol{x}}_A - \boldsymbol{x}_A\|_2$, we try to find a bound for $\|\boldsymbol{u}\|_2$. This new problem can be posed as the solution to an optimization problem of the form

$$\max \|\boldsymbol{u}\|_2 \text{ s.t.} \left\{ \|\boldsymbol{u}\|_1 \leq 2 \sum_{i=1}^{k} u_i, (1+\mu)\|\boldsymbol{u}\|_2^2 - \mu \|\boldsymbol{u}\|_1^2 \leq \Delta^2 \right\}.$$

Now we are in the exactly same situation as equation (3.11) in [9]. Therefore by the same reasoning, one can show that if $k < (1/\mu + 1)/4$,

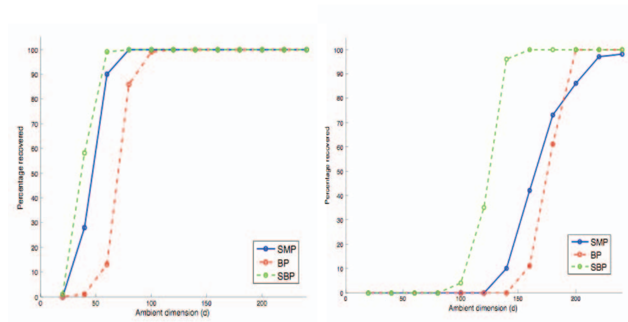
$$\|\boldsymbol{u}\|_2^2 \leq \frac{(\epsilon + \delta)^2}{1 - \mu(4k - 1)}.$$

Hence equation (17) follows. □

5. SIMULATIONS

We test the proposed algorithms with random Gaussian matrices[2] $A_1, \ldots, A_m$ and generate observation vectors $\boldsymbol{y}$ with different levels of subspace sparsity. SMP and SBP are used to separate the signal $\boldsymbol{y}$. Instead of recovering $\boldsymbol{x}_i$'s, we focus on recovering their supports on the subspaces. We assume that the subspace-sparsity level k is known a priori. For the SBP, we choose the k subspaces corresponding to the k largest values of $\{\|A_i \boldsymbol{x}_i\|_2\}_{i=1}^m$. The recovery is considered a success if the k chosen subspaces match the ground truth. We compare our results to BP [7], where subspace structure is ignored while solving for the $\boldsymbol{x}_i$'s. The k subspaces chosen correspond to the k largest values of $\{\|\boldsymbol{x}_i\|_2\}_{i=1}^m$. The simulations are done using the CVX package.

Figure 1 shows the percentage of successful recoveries as a function of ambient dimension d. Among the two proposed algorithms, SBP performs better than SMP at the expense of increased computational cost – convex optimization v.s. a greedy iterative algorithm.



(a) $k = 5$ (b) $k = 20$

Figure 1: Support recovery rates (over 100 trials) vs ambient dimension d. In each trial, $m = 50$, and each subspace was 5-dimensional. In general, SBP (green) has a higher recovery rate than SMP (blue), which in turn is higher than the original BP (red) algorithm, specifically when the signal is less subspace-sparse.

6. CONCLUSION

We have extended classical sparse recovery algorithms to separate subspace-sparse signals, and provided easily verifiable sufficient conditions for correct separation. Simulations on synthetic data show that the proposed subspace matching or base pursuit algorithms perform better than the conventional basis pursuit which ignores the subspace structures, particularly when the ambient dimension is small. The proposed algorithms can potentially be used in speech recognition [10] or other source separation problems.

7. REFERENCES

[1] 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, no. 6, pp. 797–829, Jun 2006.

[2] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, Dec 2005.

[3] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signalswith an optimal number of measurements," *preprint*, Mar 2008.

[4] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a union of subspaces," *preprint*, Jul 2008.

[5] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, Oct 2004.

[6] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, Dec 2007.

[7] S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1998.

[8] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *preprint*, 2007.

[9] D. L. Donoho, M. Elad, and V. 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, Jan 2006.

[10] J. F. Gemmeke and B. Cranen, "Noise reduction through compressed sensing," in *Interspeech 2008*, Sep 2008.

[2]For large linear systems, it can be shown that such matrices have low mutual subspace incoherence μ with overwhelming probability.