

Subspace Pursuit for Compressive Sensing: Closing the Gap Between Performance and Complexity*

Wei Dai and Olgica Milenkovic
 Department of Electrical and Computer Engineering
 University of Illinois at Urbana-Champaign

Abstract—We propose a new method for reconstruction of sparse signals with and without noisy perturbations, termed the subspace pursuit algorithm. The algorithm has two important characteristics: low computational complexity, comparable to that of orthogonal matching pursuit techniques, and reconstruction accuracy of the same order as that of LP optimization methods. The presented analysis shows that in the noiseless setting, the proposed algorithm can exactly reconstruct arbitrary sparse signals provided that the sensing matrix satisfies the restricted isometry property with a constant parameter. In the noisy setting and in the case that the signal is not exactly sparse, it can be shown that the mean squared error of the reconstruction is upper bounded by constant multiples of the measurement and signal perturbation energies.

Index Terms—Compressive sensing, orthogonal matching pursuit, reconstruction algorithms, restricted isometry property, sparse signal reconstruction

I. INTRODUCTION

Compressive sensing (CS) is a method closely connected to *transform coding*, a compression technique widely used in modern communication systems involving large scale data samples. A transform code converts input signals, embedded in a high dimensional space, into signals that lie in a space of significantly smaller dimension. Examples of transform coders include the well known wavelet transforms and the ubiquitous Fourier transform.

Compressive sensing techniques perform transform coding successfully whenever applied to so-called compressible and/or K -sparse signals, i.e., signals that can be represented by $K \ll N$ significant coefficients over an N -dimensional basis. Encoding of a K -sparse, discrete-time signal \mathbf{x} of dimension N is accomplished by computing a measurement vector \mathbf{y} that consists of $m \ll N$ linear projections of the vector \mathbf{x} , compactly described via

$$\mathbf{y} = \Phi \mathbf{x}.$$

Here, Φ represents an $m \times N$ matrix, usually over the field of real numbers. Within this framework, the projection basis is assumed to be *incoherent* with the basis in which the signal has a sparse representation [2].

This work is supported by NSF Grants CCF 0644427, 0729216 and the DARPA Young Faculty Award.

*At the time of writing this manuscript, the authors became aware of the related work by J. Tropp, D. Needell and R. Vershynin [1], where similar reconstruction algorithms are designed. Our results were developed independently, and we believe that there are significant differences in these two proposed reconstruction approaches.

Although the reconstruction of the signal $\mathbf{x} \in \mathbb{R}^N$ from the possibly noisy random projections is an ill-posed problem, the strong prior knowledge of signal sparsity allows for recovering \mathbf{x} using $m \ll N$ projections only. One of the outstanding results in CS theory is that the signal \mathbf{x} can be reconstructed using optimization strategies aimed at finding the sparsest signal that matches with the m projections. In other words, the reconstruction problem can be cast as an ℓ_0 minimization problem [3]. It can be shown that to reconstruct a K -sparse signal \mathbf{x} , ℓ_0 minimization requires only $m = K + 1$ random projections when the signal and the measurements are noise-free. Unfortunately, solving the ℓ_0 optimization is known to be NP-hard. This issue has led to a large body of work in CS theory and practice centered around the design of measurement and reconstruction algorithms with tractable reconstruction complexity.

The work by Donoho and Candès et. al. [2], [4]–[6], demonstrated that CS reconstruction is, indeed, a polynomial time problem – albeit under the constraint that more than $K + 1$ measurements are used. The key observation behind these findings is that it is not necessary to resort to ℓ_0 optimization to recover \mathbf{x} from the under-determined inverse problem; a much easier ℓ_1 optimization, based on Linear Programming (LP) techniques, yields an equivalent solution, as long as the sampling matrix Φ satisfies the so called *restricted isometry property* (RIP) with a constant parameter.

While LP techniques play an important role in designing computationally tractable CS decoders, their complexity is still highly impractical for many applications. In such cases, the need for faster decoding algorithms - preferably operating in linear time - is of critical importance, even if one has to increase the number of measurements. Several classes of low-complexity reconstruction techniques were recently put forward as alternatives to linear programming (LP) based recovery, which include group testing methods [7], and algorithms based on belief propagation [8].

Recently, a family of iterative greedy algorithms received significant attention due to their low complexity and simple geometric interpretation. They include the Orthogonal Matching Pursuit (OMP), the Regularized OMP (ROMP) and the Stagewise OMP (StOMP) algorithms. The basic idea behind these methods is to find the support of the unknown signal sequentially. At each iteration of the algorithms, one or several coordinates of the vector \mathbf{x} are selected for testing based on the correlation values between the columns of Φ and the regularized measurement vector. If deemed sufficiently

reliable, the candidates are subsequently added to the current estimate of the support set of \mathbf{x} . The pursuit algorithms iterate this procedure until all the coordinates in the correct support are in the estimated support. The computational complexity of OMP strategies depends on the number of iterations needed for exact reconstruction: standard OMP always runs through K iterations, and therefore its reconstruction complexity is roughly $O(KmN)$. This complexity is significantly smaller than that of LP methods, especially when the signal sparsity level K is small. However, the pursuit algorithms do not have provable reconstruction quality of the level of LP methods. For OMP techniques to operate successfully, one requires that the correlation between all pairs of columns of Φ is at most $1/2K$ [9], which by the Gershgorin Circle Theorem [10], represents a more restrictive constraint than the RIP. The ROMP algorithm [11] can reconstruct all K -sparse signals provided that the RIP holds with parameter $\delta_{2K} \leq 0.06/\sqrt{\log K}$, which strengthens the RIP requirements for ℓ_1 -linear programming by a factor of $\sqrt{\log K}$.

The main contribution of this paper is a new algorithm, termed the *subspace pursuit* (SP) algorithm, which exhibits low reconstruction complexity of matching pursuit techniques, but has provable reconstruction capability comparable to that of LP methods. The algorithm can operate both in the noiseless and noisy regime, allowing for exact and approximate signal recovery, respectively. For any sampling matrix Φ satisfying the RIP with a constant parameter independent on K , the SP algorithm can recover arbitrary K -sparse signals exactly from its noiseless measurements. When the measurements are inaccurate and/or the signal is not sufficiently sparse, the reconstruction distortion is upper bounded by a constant multiple of the measurement and/or signal perturbation energy. The computational complexity of the SP algorithm is upper bounded by $O(mNK)$, but can be further reduced to $O(mN \log K)$ when the nonzero entries of the sparse signal decay slowly.

The basic idea behind the SP algorithm is borrowed from sequential coding theory with backtracking, more precisely, the A^* order-statistic algorithm [12]. In this decoding framework, one first selects a set of K codewords of highest reliability that span the codespace. If the distance of the received vector to this space is deemed large, the algorithm incrementally removes and adds new basis vectors according to their reliability values, until a sufficiently close candidate codeword is identified. SP employs a similar strategy, except for the fact that at each step, the *same number* of vectors is expurgated from the candidate list. This feature is mainly introduced for simplicity of analysis: one can easily extend the algorithm to include adaptive expurgation strategies that do not necessarily work with fixed-sized lists.

In compressive sensing, the major challenge associated with sparse signal reconstruction is to identify in which subspace, generated by not more than K columns of the matrix Φ , the measured signal \mathbf{y} lies in. Once the correct subspace is determined, the non-zero signal coefficients are calculated by applying the pseudoinversion process. The defining character of the SP algorithm is the method used for finding the K columns that span the correct subspace: SP tests subsets of

K columns in a group, for the purpose of refining at each stage an initially chosen estimate for the subspace. More specifically, the algorithm maintains a list of K columns of Φ , performs a simple test in the spanned space, and then refines the list. If \mathbf{y} does not lie in the current estimate for the correct spanning space, one refines the estimate by retaining reliable candidates, discarding the unreliable ones while adding the same number of new candidates. The ‘‘reliability property’’ is captured in terms of the order statistics of the inner products of the received signal with the columns of Φ , and the subspace projection coefficients.

As a consequence, the main difference between ROMP and the SP reconstruction strategy is that the former algorithm generates a list of candidates sequentially, without backtracking: it starts with an empty list, identifies one or several reliable candidates during each iteration, and adds them to the already existing list. Once a coordinate is deemed to be reliable and is added to the list, it is not removed from it until terminating the algorithm. This search strategy is overly restrictive, since candidates have to be selected with extreme caution. In contrast, the SP algorithm incorporates a simple method for re-evaluating the reliability of all candidates at each iteration of the process.

The remainder of the paper is organized as follows. Section II introduces relevant concepts and terminology for describing the proposed CS reconstruction technique. Section III contains the algorithmic description of the SP algorithm, along with a simulation-based study of its performance when compared to OMP, ROMP, and LP methods. Section IV contains the main result of the paper pertaining to the noiseless setting: a formal proof for the guaranteed reconstruction performance and the reconstruction complexity of the SP algorithm. Section V contains the main result of the paper pertaining to the noisy setting. Concluding remarks are given in Section VI, while proofs of most of the theorems are presented in the Appendix of the paper.

II. PRELIMINARIES

A. Compressive Sensing and the Restricted Isometry Property

Let $\text{supp}(\mathbf{x})$ denote the set of indices of the non-zero coordinates of an arbitrary vector $\mathbf{x} = (x_1, \dots, x_N)$, and let $|\text{supp}(\mathbf{x})| = \|\cdot\|$ denote the support size of \mathbf{x} , or equivalently, its ℓ_0 norm¹. Assume next that $\mathbf{x} \in \mathbb{R}^N$ is an unknown signal with $|\text{supp}(\mathbf{x})| \leq K$, and let $\mathbf{y} \in \mathbb{R}^m$ be an observation of \mathbf{x} via M linear measurements, i.e.,

$$\mathbf{y} = \Phi \mathbf{x},$$

where $\Phi \in \mathbb{R}^{m \times N}$ is henceforth referred to as the *sampling matrix*.

We are concerned with the problem of low-complexity recovery of the unknown signal \mathbf{x} from the measurement \mathbf{y} . A natural formulation of the recovery problem is within an ℓ_0 norm minimization framework which seeks a solution to the problem

$$\min \|\mathbf{x}\|_0 \text{ subject to } \mathbf{y} = \Phi \mathbf{x}.$$

¹We interchangeably use both notations in the paper.

Unfortunately, solving the above ℓ_0 minimization problem is NP-hard and therefore not practical [4], [5].

One way to avoid using this computationally intractable formulation is to refer to an ℓ_1 -regularized optimization settings, i.e.,

$$\min \|\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = \Phi \mathbf{x},$$

where

$$\|\mathbf{x}\|_1 = \sum_{i=1}^N |x_i|$$

denotes the ℓ_1 norm of the vector \mathbf{x} .

The main advantage of the ℓ_1 minimization approach is that it is a convex optimization problem that can be solved efficiently by linear programming (LP) techniques. This method is therefore frequently referred to as ℓ_1 -LP reconstruction, and its reconstruction complexity equals $O(N^3)$ [4], [13].

The reconstruction accuracy of the ℓ_1 -LP method is described in terms of the so called *restricted isometry property* (RIP), formally defined below.

Definition 1 (Truncation): Let $\Phi \in \mathbb{R}^{m \times N}$ and let $I \subset \{1, \dots, N\}$. The matrix Φ_I consists of the columns of Φ with indices $i \in I$. The space spanned by the columns of Φ_I is denoted by $\text{span}(\Phi_I)$.

Definition 2 (RIP): A matrix $\Phi \in \mathbb{R}^{m \times N}$ is said to satisfy the Restricted Isometry Property (RIP) with parameters (K, δ) for $K \leq m$, $0 \leq \delta \leq 1$, if for all index sets $I \subset \{1, \dots, N\}$ such that $|I| \leq K$ and for all $\mathbf{q} \in \mathbb{R}^{|I|}$, one has

$$(1 - \delta) \|\mathbf{q}\|_2^2 \leq \|\Phi_I \mathbf{q}\|_2^2 \leq (1 + \delta) \|\mathbf{q}\|_2^2.$$

We define δ_K to be the infimum of all parameters δ for which the RIP holds, i.e.

$$\delta_K := \inf \left\{ \delta : (1 - \delta) \|\mathbf{q}\|_2^2 \leq \|\Phi_I \mathbf{q}\|_2^2 \leq (1 + \delta) \|\mathbf{q}\|_2^2, \right. \\ \left. \forall |I| \leq K, \forall \mathbf{q} \in \mathbb{R}^{|I|} \right\}.$$

Remark 1 (RIP and eigenvalues): If a sampling matrix $\Phi \in \mathbb{R}^{m \times N}$ satisfies the RIP with parameters (K, δ_K) , then for all $I \subset \{1, \dots, N\}$ such that $|I| \leq K$, it holds that

$$1 - \delta_K \leq \lambda_{\min}(\Phi_I^* \Phi_I) \leq \lambda_{\max}(\Phi_I^* \Phi_I) \leq 1 + \delta_K,$$

where λ_{\min} and λ_{\max} denote the minimum and maximum eigenvalues of Φ , respectively.

Remark 2 (Matrices satisfying the RIP): Most known examples of matrices satisfying the RIP property with optimal or near-optimal performance guarantees are random. Examples include:

- 1) Random matrices with i.i.d. entries that follow either the Gaussian distribution, Bernoulli distribution with zero mean and variance $1/n$, or any other distribution that satisfies certain tail decay laws. It was shown in [13] that the RIP for a randomly chosen matrix from such ensembles holds with overwhelming probability whenever

$$K \leq C \frac{m}{\log(N/m)},$$

where C is a function of the RIP parameter.

- 2) Random matrices from the Fourier ensemble. Here, one randomly selects m rows from the $N \times N$ discrete Fourier transform matrix uniformly at random. Upon selection, the columns of the matrix are scaled to unit norm. The resulting matrix satisfies the RIP with overwhelming probability provided that

$$K \leq C \frac{m}{(\log N)^6},$$

where C depends only on the RIP parameter.

There exists an intimate relationship between the LP reconstruction accuracy and the RIP property, first described by Candés and Tao in [4]. The result in [4] shows that if the sampling matrix Φ satisfies the RIP with parameters δ_K , δ_{2K} , and δ_{3K} , such that

$$\delta_K + \delta_{2K} + \delta_{3K} < 1, \quad (1)$$

then the ℓ_1 -LP algorithm will reconstruct all K -sparse signals exactly.

For our subsequent derivations, we need two results summarized in the lemma below. The first part of the claim, as well as a related modification of the second claim also appeared in [4], [11]. For completeness, we include the proof of the lemma in Appendix A.

Lemma 1 (Consequences of RIP):

- 1) (*Monotonicity of δ_K*) For any two integers $K \leq K'$,

$$\delta_K \leq \delta_{K'}.$$

- 2) (*Near orthogonality of columns*) Let $I, J \subset \{1, \dots, N\}$ be two disjoint sets, $I \cap J = \emptyset$. Suppose that $\delta_{|I|+|J|} < 1$. For arbitrary vectors $\mathbf{a} \in \mathbb{R}^{|I|}$ and $\mathbf{b} \in \mathbb{R}^{|J|}$,

$$|\langle \Phi_I \mathbf{a}, \Phi_J \mathbf{b} \rangle| \leq \delta_{|I|+|J|} \|\mathbf{a}\|_2 \|\mathbf{b}\|_2,$$

and

$$\|\Phi_I^* \Phi_J \mathbf{b}\|_2 \leq \delta_{|I|+|J|} \|\mathbf{b}\|_2.$$

The lemma implies that $\delta_K \leq \delta_{2K} \leq \delta_{3K}$, which consequently implies that $\delta_{3K} < 1/3$ is a sufficient condition for exact reconstruction of K -sparse signals. Although this condition is weaker than the one specified in Equation (1), we henceforth focus only on characterizing the performance and complexity of the SP algorithm with respect to this parameter. Our motivation for slightly weakening this RIP parameter bound is to simplify the notation used in most of the proofs, and to provide a fair comparison between different reconstruction strategies.

In order to describe the main steps of the SP algorithm, we introduce next the notion of the projection of a vector and its residue.

Definition 3 (Projection and Residue): Let $\mathbf{y} \in \mathbb{R}^m$ and $\Phi_I \in \mathbb{R}^{m \times |I|}$. Suppose that $\Phi_I^* \Phi_I$ is invertible. The projection of \mathbf{y} onto $\text{span}(\Phi_I)$ is defined as

$$\mathbf{y}_p = \text{proj}(\mathbf{y}, \Phi_I) := \Phi_I \Phi_I^\dagger \mathbf{y},$$

where

$$\Phi_I^\dagger := (\Phi_I^* \Phi_I)^{-1} \Phi_I^*$$

denotes the pseudo-inverse of the matrix Φ_I , and $*$ stands for matrix transposition.

The *residue vector* of the projection equals

$$\mathbf{y}_r = \text{resid}(\mathbf{y}, \Phi_I) := \mathbf{y} - \mathbf{y}_p.$$

We find the following properties of projections and residues of vectors useful for our subsequent derivations.

Lemma 2 (Projection and Residue):

1) (*Orthogonality of the residue*) For an arbitrary vector $\mathbf{y} \in \mathbb{R}^m$, and a sampling matrix $\Phi_I \in \mathbb{R}^{m \times K}$ of full column rank, let $\mathbf{y}_r = \text{resid}(\mathbf{y}, \Phi_I)$. Then

$$\Phi_I^* \mathbf{y}_r = 0.$$

2) (*Approximation of the projection residue*) Consider a matrix $\Phi \in \mathbb{R}^{m \times N}$. Let $I, J \subset \{1, \dots, N\}$ be two disjoint sets, $I \cap J = \emptyset$, and suppose that $\delta_{|I|+|J|} < 1$. Furthermore, let $\mathbf{y} \in \text{span}(\Phi_I)$, $\mathbf{y}_p = \text{proj}(\mathbf{y}, \Phi_J)$ and $\mathbf{y}_r = \text{resid}(\mathbf{y}, \Phi_J)$. Then

$$\|\mathbf{y}_p\|_2 \leq \frac{\delta_{|I|+|J|}}{1 - \delta_{|I|+|J|}} \|\mathbf{y}\|_2,$$

and

$$\left(1 - \frac{\delta_{|I|+|J|}}{1 - \delta_{|I|+|J|}}\right) \|\mathbf{y}\|_2 \leq \|\mathbf{y}_r\|_2 \leq \|\mathbf{y}\|_2.$$

The proof of Lemma 2 can be found in Appendix B.

III. THE SP ALGORITHM

The main steps of the SP algorithm can be described as follows.

Algorithm 1 Subspace Pursuit Algorithm

Input: K, Φ, \mathbf{y}

Initialization:

$\hat{T} = \{K \text{ indices corresponding to the largest absolute values of } \Phi^* \mathbf{y}\}.$

$\mathbf{y}_r = \text{resid}(\mathbf{y}, \Phi_{\hat{T}}).$

Iteration:

If $\mathbf{y}_r = 0$, quit the iteration; otherwise continue.

$T' = \hat{T} \cup \{K \text{ indices corresponding to the largest magnitudes of } \Phi^* \mathbf{y}_r\}.$

Let $\mathbf{x}'_p = \Phi_{T'}^* \mathbf{y}.$

$\tilde{T} = \{K \text{ indices corresponding to the largest elements of } \mathbf{x}'_p\}.$

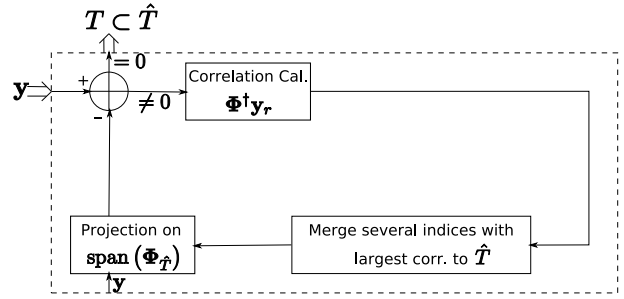
$\tilde{\mathbf{y}}_r = \text{resid}(\mathbf{y}, \Phi_{\tilde{T}}).$

If $\|\tilde{\mathbf{y}}_r\| > \|\mathbf{y}_r\|$, quit the iteration; otherwise, let $\hat{T} = \tilde{T}$ and $\mathbf{y}_r = \tilde{\mathbf{y}}_r$, and continue with a new iteration.

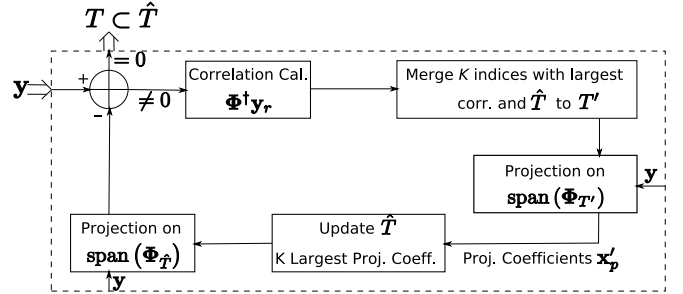
Output:

The estimated signal $\hat{\mathbf{x}}$ satisfies $\hat{\mathbf{x}}_{\{1, \dots, N\} - \hat{T}} = \mathbf{0}$ and $\mathbf{x}_{\hat{T}} = \Phi_{\hat{T}}^* \mathbf{y}.$

A schematic diagram of the SP algorithm is depicted in Fig. 1(b). For comparison, a diagram of OMP-type methods is also provided in Fig. 1(a). The subtle, but important, difference



(a) Iterations in OMP, Stagewise OMP, and Regularized OMP: in each iteration, one decides on a reliable set of candidate indices to be added into the list \hat{T} ; once a candidate is added, it remains in the list until the algorithm terminates.



(b) Iterations in the proposed Subspace Pursuit Algorithm: a list of K candidates, which is allowed to be updated during the iterations, is maintained.

Figure 1. Description of reconstruction algorithms for K -sparse signals: though both approaches look similar, the basic ideas behind are quite different.

between the two schemes lies in the approach used to generate \hat{T} , the estimate of the correct support set T . In OMP strategies, during each iteration one decides the algorithm selects one or several indices that represent good partial support set estimates and adds them to \hat{T} . Once an index is added into \hat{T} , it remains in this set throughout the remainder of the process. As a result, strict inclusion rules are needed to ensure that a significant fraction of the newly added indices belongs to the correct support T . On the other hand, in the SP algorithm, an estimate \hat{T} of size K is maintained and refined during each iteration. An index, which is considered reliable in some iteration but shown to be wrong at a later iteration, can be added into or removed from the estimated support set freely. The expectation is that the recursive refinements of the estimate of the support set will lead to subspaces with strictly decreasing distance from the measurement vector \mathbf{y} .

We performed extensive computer simulations in order to compare the accuracy of different reconstruction algorithms empirically. In the compressive sensing framework, all sparse signals are expected to be exactly reconstructed as long as the level of the sparsity is below a certain threshold. For empirical testing, we adopt the simulation strategy described in [6] for simulating the empirical frequency of exact reconstruction. The steps of the testing strategy are listed below.

- 1) For given values of the parameters m and N , choose a signal sparsity level K such that $K \leq m/2$;
- 2) Randomly generate a $m \times N$ sampling matrix Φ from the standard i.i.d. Gaussian ensemble;
- 3) Select a support set T of size $|T| = K$ uniformly at

random, and generate the sparse signal vector \mathbf{x} by either one of the following two methods:

- a) Draw the elements of the vector \mathbf{x} restricted to T from the standard Gaussian distribution; we refer to this type of signal as a *Gaussian* signal. Or,
- b) set all entries of \mathbf{x} supported on T to ones; we refer to this type of signal as *zero-one* signal.

Note that zero-one sparse signals are of spatial interest for the comparative study, since they represent a particularly challenging case for OMP-type of reconstruction strategies.

- 4) Compute the measurement $\mathbf{y} = \Phi\mathbf{x}$, apply a reconstruction algorithm to obtain an estimate of \mathbf{x} , $\hat{\mathbf{x}}$, and compare \mathbf{x} to $\hat{\mathbf{x}}$;
- 5) Repeat the process 500 times for each K , and then simulate the same algorithm for different values of m and N .

The improved reconstruction capability of the SP method, compared to that of the OMP and ROMP algorithms, is illustrated by two examples shown in Fig. 2. Here, the signals are drawn both according to the Gaussian and zero-one model, and the benchmark performance of the LP reconstruction technique is plotted as well.

Figure 2 depicts the empirical frequency of exact reconstruction. The numerical values on the x -axis denote the sparsity level K , while the numerical values on the y -axis represent the fraction of exactly recovered test signals. Of particular interest is the sparsity level at which the recovery rate drops below 100% - i.e. the *critical sparsity* - which, when exceeded, leads to errors in the reconstruction algorithm applied to some of the signals from the given class.

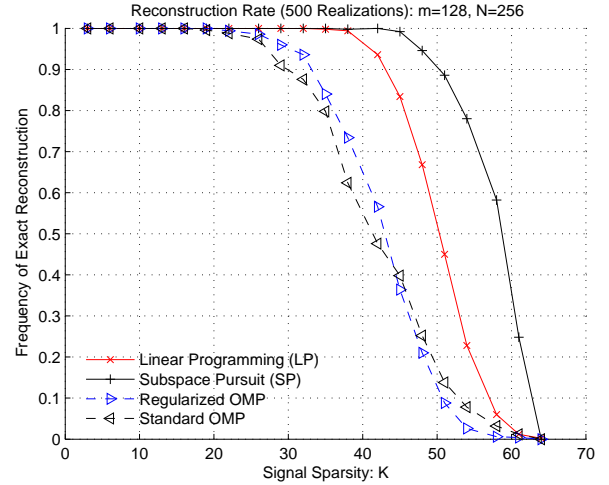
The simulation results reveal that the critical sparsity of the SP algorithm by far exceeds that of the OMP and ROMP techniques, for both Gaussian and zero-one inputs. The reconstruction capability of the SP algorithm is comparable to that of the LP based approach: the SP algorithm has a slightly higher critical sparsity for Gaussian signals, but also a slightly lower critical sparsity for zero-one signals. However, the SP algorithms significantly outperforms the LP method when it comes to reconstruction complexity. As we analytically demonstrate in the exposition to follow, the reconstruction complexity of the SP algorithm for both Gaussian and zero-one sparse signals is $O(mN \log K)$. At the same time, the complexity of LP algorithms based on interior point methods is $O(m^2 N^{3/2})$ [14].

IV. RECOVERY OF SPARSE SIGNAL

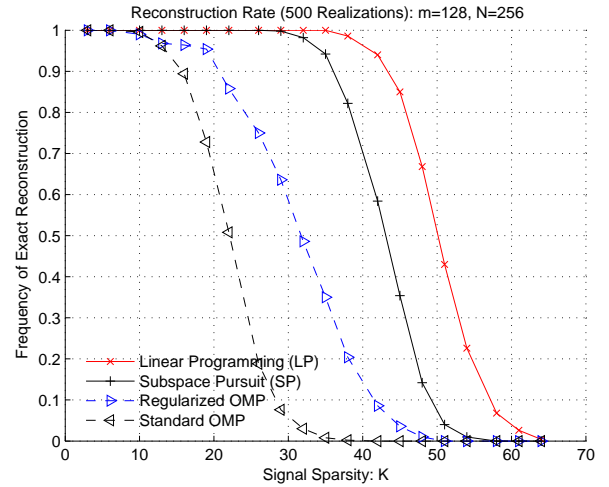
For simplicity, we start by analyzing the reconstruction performance of SP algorithms applied to sparse signals in the noiseless setting. The techniques used in this context, and the insights obtained are also applicable to the analysis of SP reconstruction schemes with signal or/and measurement perturbations.

A sufficient condition for exact reconstruction of arbitrary sparse signals is stated in the following theorem.

Theorem 1: Let $\mathbf{x} \in \mathbb{R}^N$ be a K -sparse signal, and let its corresponding measurement be $\mathbf{y} = \Phi\mathbf{x} \in \mathbb{R}^m$. If the



(a) Simulations for Gaussian sparse signals: OMP and ROMP start to fail when $K \geq 19$ and when $K \geq 22$ respectively, ℓ_1 -LP begins to fail when $K \geq 35$, and the SP algorithm fails only when $K \geq 45$.



(b) Simulations for zero-one sparse signals: both OMP and ROMP starts to fail when $K \geq 10$, ℓ_1 -LP begins to fail when $K \geq 35$, and the SP algorithm fails when $K \geq 45$.

Figure 2. Simulations of the exact recovery rate: compared to OMPs, the SP algorithm has significantly larger critical sparsity.

sampling matrix Φ satisfies the RIP with parameter

$$\delta_{3K} < 0.06, \quad (2)$$

then the SP algorithm is guaranteed to exactly recover \mathbf{x} from \mathbf{y} via a finite number of iterations.

This sufficient condition is proved by applying Theorems 2 and 6. The computational complexity is related to the number of iterations required for exact reconstruction, and discussed at the end of Section IV-C. Before we go to the details, let us sketch the main ideas behind the proof.

As before, denote the estimate of $\text{supp}(\mathbf{x})$ at the beginning of a given iteration by \hat{T} , and the estimate of the support set at the end of the iteration by \tilde{T} , which also serves as the estimate for the next iteration. Let

$$\hat{\mathbf{x}}_0 = \mathbf{x}_{T-\hat{T}} \text{ and } \tilde{\mathbf{x}}_0 = \mathbf{x}_{T-\tilde{T}}.$$

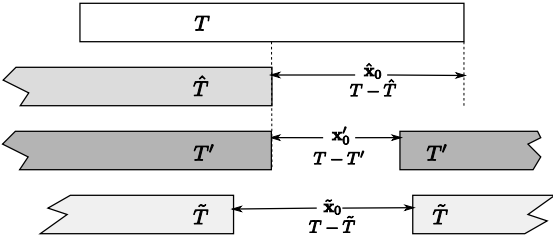


Figure 3. Illustration of sets and signal coefficient vectors

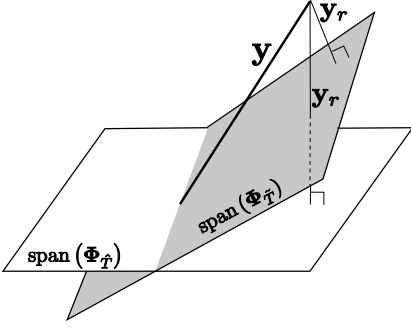


Figure 4. After each iteration, a K -dimensional hyper-plane closer to \mathbf{y} is obtained.

The vectors $\hat{\mathbf{x}}_0$ and $\tilde{\mathbf{x}}_0$ represent the residual signals based upon the estimates of $\text{supp}(\mathbf{x})$ before and after a given iteration of the SP algorithm is completed, respectively (see Fig. 3 for illustration). Provided that the sampling matrix Φ satisfies the RIP with constant (2), it holds that

$$\|\tilde{\mathbf{x}}_0\|_2 < \|\hat{\mathbf{x}}_0\|_2,$$

which implies that at each iteration, the SP algorithm identifies a K -dimensional space that reduces the reconstruction error of the vector \mathbf{x} . See Fig. 4 for an illustration. We are now ready to formally state this observation as follows.

Theorem 2: Assume that the conditions of Theorem 1 hold. For each iteration of the SP algorithm, it holds that

$$\|\tilde{\mathbf{x}}_0\|_2 \leq c_K \|\hat{\mathbf{x}}_0\|_2, \quad (3)$$

and

$$\|\tilde{\mathbf{y}}_r\|_2 \leq \frac{1 + \delta_{3K}}{1 - 2\delta_{3K}} c_K \|\hat{\mathbf{y}}_r\|_2 < \|\hat{\mathbf{y}}_r\|_2, \quad (4)$$

where

$$c_K = \frac{\sqrt{10\delta_{3K}}}{1 - \delta_{3K}}.$$

To prove Theorem 2, we need to take a closer look at the operations executed during each iteration of the SP algorithm. During one iteration, two basic sets of computations and comparisons are performed: first, given \hat{T} , K additional candidate indices for inclusion into the estimate of the support set are identified; and second, given T' , K reliable indices out of the total $2K$ indices are selected for future testing. This set of candidate indices is represented by \tilde{T} . In Subsections IV-A and IV-B, we provide the intuition for choosing to perform SP support reconstruction according to these rules. Now, let

$$\mathbf{x}'_0 = \mathbf{x}_{T-T'}$$

be the residue signal coefficient vector corresponding to the support set estimate T' .

To proceed, we need the following two theorems.

Theorem 3: It holds that

$$\|\mathbf{x}'_0\|_2 \leq \frac{\sqrt{10\delta_{2K}}}{1 + \delta_{2K}} \|\hat{\mathbf{x}}_0\|_2.$$

The proof of the theorem is postponed to Appendix D.

Theorem 4: The following inequality is valid

$$\|\tilde{\mathbf{x}}_0\|_2 \leq \frac{1 + \delta_{3K}}{1 - \delta_{3K}} \|\mathbf{x}'_0\|_2.$$

The proof of the result is deferred to Appendix E.

Based on Theorems 3 and 4, one arrives at the result claimed in Equation (3).

Furthermore, according to Lemmas 1 and 2, we have

$$\begin{aligned} \|\tilde{\mathbf{y}}_r\|_2 &= \|\text{resid}(\mathbf{y}, \Phi_{\tilde{T}})\|_2 \\ &\leq \|\Phi_{T-\tilde{T}} \tilde{\mathbf{x}}_0\|_2 \\ &\leq (1 + \delta_{3K}) c_K \|\tilde{\mathbf{x}}_0\|_2, \end{aligned}$$

and

$$\begin{aligned} \|\hat{\mathbf{y}}_r\|_2 &= \|\text{resid}(\mathbf{y}, \Phi_{\hat{T}})\|_2 \\ &\geq \frac{1 - 2\delta_{3K}}{1 - \delta_{3K}} \|\Phi_{T-\hat{T}} \hat{\mathbf{x}}_0\|_2 \\ &\geq (1 - 2\delta_{3K}) \|\hat{\mathbf{x}}_0\|_2. \end{aligned}$$

Upon combining the two inequalities described above, we obtain the following upper bound

$$\|\tilde{\mathbf{y}}_r\|_2 \leq \frac{1 + \delta_{3K}}{1 - 2\delta_{3K}} c_K \|\hat{\mathbf{y}}_r\|_2.$$

Finally, elementary calculations show that when $\delta_{3K} < 0.06$,

$$\frac{1 + \delta_{3K}}{1 - 2\delta_{3K}} c_K < 1,$$

which completes the proof of Theorem 2.

A. Why Does Correlation Maximization Work for the SP Algorithm?

Both in the initialization step and during each iteration of the SP algorithm, we select K indices that maximize the correlations between the column vectors and the residual measurement. Henceforth, this step is referred to as *correlation maximization* (CM). Consider the ideal case where all columns of Φ are orthogonal². In this scenario, the signal coefficients can be easily recovered by calculating the correlations $\langle \mathbf{v}_i, \mathbf{y} \rangle$ - i.e., all indices with non-zero magnitude are in the correct support of the sensed vector. Now assume that the sampling matrix Φ satisfies the RIP. Recall that the RIP (see Lemma 1) implies that the columns are locally near-orthogonal. Consequently, for any j not in the correct support, the magnitude of the correlation $\langle \mathbf{v}_j, \mathbf{y} \rangle$ is expected to be small, and more precisely, upper bounded by $\delta_{K+1} \|\mathbf{x}\|_2$. This seems to provide a very simple intuition why correlation maximization allows for exact reconstruction, but the correct problems in reconstruction arise

²Of course, in this case no compression is possible.

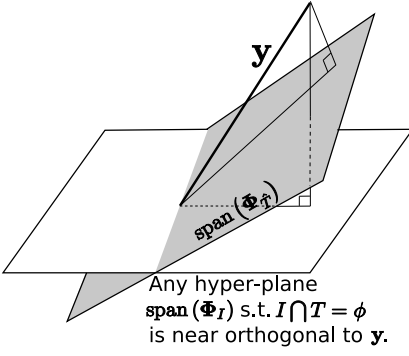


Figure 5. Correlation maximization works in the SP setting.

due to the following fact. Although it is clear that for all $j \notin T$, the values of $|\langle \mathbf{v}_j, \mathbf{y} \rangle|$ are upper bounded by $\delta_{K+1} \|\mathbf{x}\|$, it may also happen that for all $i \in T$, the values of $|\langle \mathbf{v}_i, \mathbf{y} \rangle|$ are small as well. Dealing with order statistics in this scenario cannot be immediately proved to be a good reconstruction strategy. The following example illustrates this point.

Example 1: Without loss of generality, let $T = \{1, \dots, K\}$. Let the vectors \mathbf{v}_i ($i \in T$) be orthonormal, and let the remaining columns \mathbf{v}_j , $j \notin T$, of Φ be constructed randomly, using i.i.d. Gaussian samples. Consider the following normalized zero-one sparse signal,

$$\mathbf{y} = \frac{1}{\sqrt{K}} \sum_{i \in T} \mathbf{v}_i.$$

Then, for K sufficiently large,

$$|\langle \mathbf{v}_i, \mathbf{y} \rangle| = \frac{1}{\sqrt{K}} \ll 1, \text{ for all } 1 \leq i \leq K.$$

It is straightforward to envision the existence of a $j \notin T$ such that

$$|\langle \mathbf{v}_j, \mathbf{y} \rangle| \approx \delta_{K+1} > \frac{1}{\sqrt{K}}.$$

The latter inequality is critical, because achieving very small values for the RIP parameter is a challenging task.

This example represents a particularly challenging case for the OMP algorithm. Therefore, one of the major constraints imposed on the OMP algorithm is the requirement that

$$\max_{i \in T} |\langle \mathbf{v}_i, \mathbf{y} \rangle| = \frac{1}{\sqrt{K}} > \max_{j \notin T} |\langle \mathbf{v}_j, \mathbf{y} \rangle| \approx \delta_{K+1}.$$

To meet this requirement, δ_{K+1} has to be less than $1/\sqrt{K}$, which decays fast as K increases.

In contrast, the SP algorithm allows for some $j \notin T$ to be such that

$$\max_{i \in T} |\langle \mathbf{v}_i, \mathbf{y} \rangle| < |\langle \mathbf{v}_j, \mathbf{y} \rangle|.$$

As long as Equation (2) holds, the indices in the correct support of \mathbf{x} , which account for the most significant part of the energy of the signal, are captured by the CM procedure. Detailed descriptions of how this can be achieved are provided in the proofs of the previously stated Theorems 5 and 3.

Let us first focus on the initialization step. By the definition of the set \hat{T} in the initialization stage of the algorithm, the set of the K selected columns ensures that

$$\|\Phi_{\hat{T}}^* \mathbf{y}\|_2 = \sqrt{\sum_{i \in \hat{T}} |\langle \mathbf{v}_i, \mathbf{y} \rangle|^2} \geq (1 - \delta_{2K}) \|\mathbf{x}\|_2. \quad (5)$$

This is a consequence of the result of Theorem 5. Now, if we assume that the estimate \hat{T} is disjoint from the correct support, i.e., that $\hat{T} \cap T = \phi$, then by the near orthogonality property of Lemma 1, one has

$$\|\Phi_{\hat{T}}^* \mathbf{y}\|_2 \leq \delta_{2K} \|\mathbf{x}\|_2.$$

The last inequality clearly contradicts (5) whenever $\delta_{2K} < \delta_{3K} < 1/2$. Consequently,

$$\hat{T} \cap T \neq \phi,$$

and at least one correct element of the support of \mathbf{x} is in the set \hat{T} . This phenomenon is depicted in Fig. 5 and quantitatively detailed in Theorem 5.

Theorem 5: After the initialization step, one has

$$\|\mathbf{x}_{\hat{T} \cap T}\|_2 \geq \frac{1 - 3\delta_{2K}}{1 + \delta_{2K}} \|\mathbf{x}\|_2,$$

and

$$\|\mathbf{x}_{T - \hat{T}}\|_2 \leq \frac{\sqrt{8\delta_{2K} + 4\delta_{2K}^2}}{1 + \delta_{2K}} \|\mathbf{x}\|_2.$$

The proof of the theorem is postponed to Appendix C.

To study the effect of correlation maximization during each iteration, one has to observe that correlation calculations are performed with respect to the vector

$$\mathbf{y}_r = \text{resid}(\mathbf{y}, \Phi_{\hat{T}})$$

instead of being performed with respect to the vector \mathbf{y} . As a consequence, to show that the CM process captures a significant part of residual signal energy requires an analysis including a number of technical details. These can be found in the Proof of Theorem 3.

B. Identifying Indices Outside of the Correct Support Set

Note that there are $2K$ indices in the set T' , among which at least K of them do not belong to the correct support set T . In order to expurgate those indices from T' , or equivalently, in order to find a K -dimensional subspace of the space $\text{span}(\Phi_{T'})$ closest to \mathbf{y} , we need to estimate these K incorrect indices.

Define $\Delta T = T' - \hat{T}$. This set contains the K indices which are deemed incorrect. If $\Delta T \cap T = \phi$, our estimate of incorrect indices is perfect. However, sometimes $\Delta T \cap T \neq \phi$. This means that among the estimated incorrect indices, there are some candidates that actually belong to the correct support set T . The question of interest is how often these correct indices are erroneously removed from the support estimate, and how quickly the algorithm manages to restore them back.

First, we claim that the reduction in the $\|\cdot\|_2$ norm induced by such erroneous expurgation is small. The intuitive explanation for this claim is as follows. Let us assume that all the

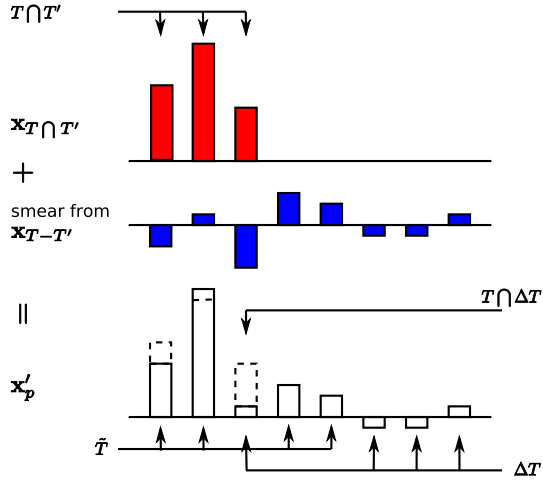


Figure 6. The projection coefficient vector \mathbf{x}'_p is a smeared version of the vector $\mathbf{x}_{T \cap T'}$.

indices in the support of \mathbf{x} have been successfully captured, or equivalently, that $T \subset T'$. When we project \mathbf{y} onto the space span $(\Phi_{T'})$, it can be shown that its corresponding vector \mathbf{x}'_p satisfies

$$\mathbf{x}'_p = \mathbf{x}_{T'},$$

and that it contains at least K zeros. Consequently, the K indices with smallest magnitude - equal to zero - are clearly not in the correct support set.

However, the situation changes when $T \not\subset T'$, or equivalently, when $T - T' \neq \emptyset$. After the projection, one has

$$\mathbf{x}'_p \neq \mathbf{x}_{T'}.$$

The projection vector \mathbf{x}'_p can be viewed as a smeared version of $\mathbf{x}_{T'}$ (see Fig. 6 for illustration): the coefficients indexed by elements outside the support of \mathbf{x} may become non-zero; the coefficients indexed by elements in the support set T may experience changes in their magnitudes. Fortunately, the level of this smear is proportional to the norm of the residual signal \mathbf{x}'_0 , which can be proved to be small according to the analysis accompanying Theorem 3. As long as the smear is not severe, the largest projection coefficients still serve as good estimates of the correct signal coefficients restricted to T' , and the correct support set T . This intuitive explanation is formalized in the previously stated Theorem 5.

C. Convergence of the SP Algorithm

In this subsection, we upper bound the number of iterations needed to reconstruct an arbitrary K -sparse signal using the SP algorithm.

Given an arbitrary K -sparse signal \mathbf{x} , we first arrange its elements in decreasing order of magnitude. Without loss of generality, assume that

$$|x_1| \geq |x_2| \geq \dots \geq |x_K| > 0,$$

and that $x_j = 0, \forall j > K$. Define

$$\rho_{\min} := \frac{|x_K|}{\|\mathbf{x}\|_2} = \frac{\min_{1 \leq i \leq K} x_i}{\sqrt{\sum_{i=1}^K x_i^2}}.$$

Let n_{it} denote the number of iterations of the SP algorithm needed for exact reconstruction of \mathbf{x} . Then the following theorem upper bounds n_{it} in terms of c_K and ρ_{\min} . It can be viewed as a bound on the complexity/performance trade-off for the SP algorithm.

Theorem 6: The number of iterations of the SP algorithm is upper bounded by

$$n_{\text{it}} \leq \min \left(\frac{-\log \rho_{\min}}{-\log c_K} + 1, \frac{1.5 \cdot K}{-\log c_K} \right).$$

This result is a combination of Theorems 7 and 8, described below.

Theorem 7: One has

$$n_{\text{it}} \leq \frac{-\log \rho_{\min}}{-\log c_K} + 1.$$

Theorem 8: It can be shown that

$$n_{\text{it}} \leq \frac{1.5 \cdot K}{-\log c_K}.$$

The proof of Theorem 7 is intuitively clear and presented below, while the proof of Theorem 8 is more technical and postponed to Appendix F.

Proof of Theorem 7: This theorem is proved by a contradiction. Let \tilde{T} be the estimate of T after

$$\frac{-\log \rho_{\min}}{-\log c_K} + 1$$

iterations. Suppose that $T \not\subset \tilde{T}$, or equivalently, $T - \tilde{T} \neq \emptyset$. Then

$$\begin{aligned} \|\mathbf{x}_{T-\tilde{T}}\|_2 &= \sqrt{\sum_{i \in T-\tilde{T}} x_i^2} \\ &\geq \min_{i \in T} |x_i| = \rho_{\min} \|\mathbf{x}\|_2. \end{aligned}$$

However, according to Theorem 2,

$$\begin{aligned} \|\mathbf{x}_{T-\tilde{T}}\|_2 &\leq (c_K)^{n_{\text{it}}} \|\mathbf{x}\|_2 \\ &= c_K \rho_{\min} \|\mathbf{x}\|_2 < \rho_{\min} \|\mathbf{x}\|_2, \end{aligned}$$

where the last inequality follows from the assumption that $c_K < 1$. This contradiction completes the proof. ■

A drawback of Theorem 7 is that it sometimes overestimates the number of iterations, especially when $\rho_{\min} \ll 1$. The example to follow illustrates this point.

Example 2: Let $K = 2, x_1 = 2^{10}, x_2 = 1, x_3 = \dots = x_N = 0$. Suppose that the sampling matrix Φ satisfies the RIP with

$$c_K = \frac{\sqrt{10\delta_{3K}}}{1 - \delta_{3K}} = \frac{1}{2}.$$

Noting that $\rho_{\min} \lesssim 2^{-10}$, Theorem 6 implies that

$$n_{\text{it}} \leq 11.$$

Indeed, if we take a close look at the steps of the SP algorithm, we can verify that

$$n_{\text{it}} \leq 1.$$

After the initialization step, by Theorem 5, it can be shown that

$$\|\hat{\mathbf{x}}_0\|_2 \leq \frac{\sqrt{4\delta_{2K} + 8\delta_{2K}^2}}{1 + \delta_{2K}} \leq c_K \|\mathbf{x}\|_2 \leq \frac{\|\mathbf{x}\|_2}{2}.$$

As a result, the estimate \hat{T} must contain the index one and $\|\hat{\mathbf{x}}_0\|_2 \leq 1$. After the first iteration, since

$$\|\tilde{\mathbf{x}}_0\|_2 \leq \frac{1}{2} \|\hat{\mathbf{x}}_0\|_2 \leq \frac{1}{2} < \min_{i \in T} |x_i|,$$

we have $T \subset \tilde{T}$.

This example suggests that the upper bound in Equation (7) can be tightened when $\rho_{\min} \ll 1$. Based on the idea behind this example, another approach to upper bounding n_{it} is described in Theorem 8 and its validity proved in Appendix F.

It is clear that the number of iterations required for exact reconstruction depends on the values of the entries of the sparse signal itself. We therefore focus our attention on the following three particular classes of sparse signals.

- 1) *Zero-one sparse signals*. As explained before, zero-one signals are in the most challenging reconstruction category for the well-known OMP algorithm. However, this class of signals has the best upper bound on the convergence rate of the SP algorithm. Elementary calculations reveal that $\rho_{\min} = 1/\sqrt{K}$ and that

$$n_{it} \leq \frac{\log K}{2 \log(1/c_K)}.$$

- 2) *Sparse signals with power-law decaying entries (also known as compressible sparse signals)*. Signals in this category are defined via the following constraint

$$|x_i| \leq c_x \cdot i^{-p},$$

for some constants $c_x > 0$ and $p > 1$. This type of signals has been widely considered in the CS literature, since most practical and naturally occurring signals belong to this class [13]. It follows from Theorem 7 that in this case

$$n_{it} \leq \frac{p \log K}{\log(1/c_K)} (1 + o(1)),$$

where $o(1) \rightarrow 0$ when $K \rightarrow \infty$.

- 3) *Sparse signals with exponentially decaying entries*. Signals in this class satisfy

$$|x_i| \leq c_x \cdot e^{-pi},$$

for some constants $c_x > 0$ and $p > 0$. Theorem 6 implies that

$$n_{it} \leq \begin{cases} \frac{pK}{\log(1/c_K)} (1 + o(1)) & \text{if } 0 < p \leq 1.5 \\ \frac{1.5K}{\log(1/c_K)} & \text{if } p > 1.5 \end{cases},$$

where again $o(1) \rightarrow 0$ as $K \rightarrow \infty$.

Simulation results, shown in Fig. 7, indicate that the above analysis gives the right order of growth in complexity with respect to the parameter K . To generate the plots of Fig. 7, we set $m = 128$, $N = 256$, and run simulations for different classes of sparse signals. For each type of sparse signal, we selected different values for the parameter K , and

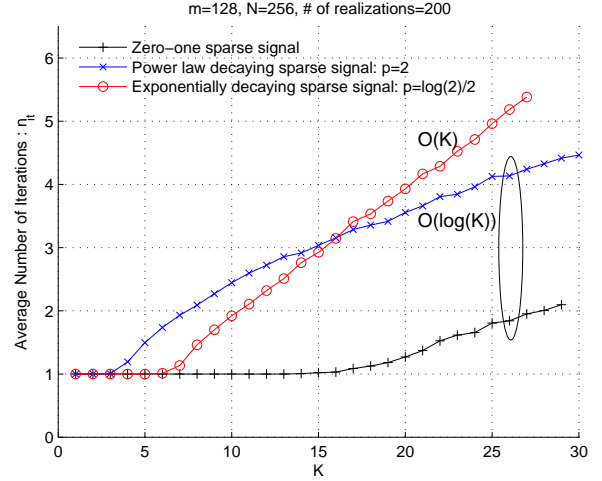


Figure 7. Convergence of the subspace pursuit algorithm for different signals.

for each K , we selected 200 different randomly generated Gaussian sampling matrices Φ and as many different support sets T . The plots depict the average number of iterations versus the signal sparsity level K , and they clearly show that $n_{it} = O(\log(K))$ for zero-one signals and sparse signals with coefficients decaying according to a power law, while $n_{it} = O(K)$ for sparse signals with exponentially decaying coefficients.

With the bound on the number of iterations required for exact reconstruction, the computational complexity of the complete SP algorithm can be easily estimated. In each iteration, CM requires mN computations, while the projections can be computed with marginal cost $O(Km)$ by the Modified Gram-Schmidt (MGS) algorithm [15]. Therefore, the total complexity of the SP algorithm is $O(mN \log K)$ for compressible sparse signals, and it is upper bounded by $O(mNK)$ for arbitrary sparse signals.

The complexity of the SP algorithm is comparable to that of OMP-type algorithms. For the standard OMP algorithm, exact reconstruction always requires K iterations. The corresponding complexity is $O(KmN)$. For the ROMP and StOMP algorithms, the challenging signals in terms of convergence rate are the sparse signals with exponentially decaying entries. When p is sufficiently large, it can be shown that both ROMP and StOMP also need $O(K)$ iterations for reconstruction, which implies computational complexity of the order of $O(KmN)$.

One advantage of the SP algorithm is that the complexity is reduced to $O(mN \log K)$ when compressible sparse signals are considered. For this class of sparse signals, to the best of the author's knowledge, there is no known formal proof that allows one to bound the complexity of the ROMP and StOMP algorithm.

V. RECOVERY OF APPROXIMATELY SPARSE SIGNALS FROM INACCURATE MEASUREMENTS

We consider first a sampling scenario in which the signal \mathbf{x} is K -sparse, but the measurement vector \mathbf{y} is subjected to an additive noise component, \mathbf{e} . The following theorem gives

a sufficient condition for convergence of the SP algorithm in terms of the RIP parameter δ_{3K} , as well as an upper bounds on the recovery distortion that depends on the energy (ℓ_2 -norm) of the error vector \mathbf{e} .

Theorem 9 (Stability under measurement perturbations):

Let $\mathbf{x} \in \mathbb{R}^N$ be such that $|\text{supp}(\mathbf{x})| \leq K$, and let its corresponding measurement be $\mathbf{y} = \Phi \mathbf{x} + \mathbf{e}$, where \mathbf{e} denotes the noise vector. Suppose that the sampling matrix satisfies the RIP with parameter

$$\delta_{3K} < 0.03. \quad (6)$$

Then the reconstruction distortion of the SP algorithm satisfies

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq c'_K \|\mathbf{e}\|_2,$$

where

$$c'_K = \frac{1 + \delta_{3K}}{\delta_{3K}(1 - \delta_{3K})}.$$

The proof of this theorem is sketched in Section V-A.

We also study the case where the signal \mathbf{x} is only *approximately* K -sparse, and the measurements \mathbf{y} is contaminated by a noise vector \mathbf{e} . To simplify the notation, we henceforth denote by \mathbf{x}_K the vector obtained from \mathbf{x} by maintaining the K entries with largest magnitude and setting all other entries in the vector to zero. In this setting, a signal \mathbf{x} is said to be approximately K -sparse if $\mathbf{x} - \mathbf{x}_K \neq \mathbf{0}$. Based on Theorem 9, we can upper bound the recovery distortion in terms of the ℓ_1 and ℓ_2 norms of $\mathbf{x} - \mathbf{x}_K$ and \mathbf{e} , respectively, as follows.

Corollary 1: (Stability under signal and measurement perturbations) Let $\mathbf{x} \in \mathbb{R}^N$ be approximately K -sparse, and let $\mathbf{y} = \Phi \mathbf{x} + \mathbf{e}$. Suppose that the sampling matrix satisfies the RIP with parameter

$$\delta_{6K} < 0.03.$$

Then

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq c'_{2K} \left(\|\mathbf{e}\|_2 + \sqrt{\frac{1 + \delta_{6K}}{K}} \|\mathbf{x} - \mathbf{x}_K\|_1 \right).$$

The proof of this corollary is given in Section V-B.

Theorem 9 and Corollary 1 provide analytical upper bounds on the reconstruction distortion of the noisy SP version of the SP algorithm. In addition to these theoretical bounds, we performed numerical simulations to empirically estimate the reconstruction distortion. In the simulations, we first select the dimension of the signal \mathbf{x} to N , and the number of measurements m . We then choose a sparsity level K such that $K \leq m/2$. Once the parameters are chosen, an $m \times N$ sampling matrix with standard i.i.d. Gaussian entries is generated. For a given K , the support set T of size $|T| = K$ is selected uniformly at random. A zero-one sparse signal is constructed as in the previous section. Finally, either signal or a measurement perturbations are added as follows:

- 1) *Signal perturbations*: the signal entries on T are kept unchanged but the signal entries out of T are perturbed by i.i.d. Gaussian $\mathcal{N}(0, \sigma_s^2)$ samples.
- 2) *Measurement perturbations*: the perturbation vector \mathbf{e} is generated from a Gaussian distribution with zero mean

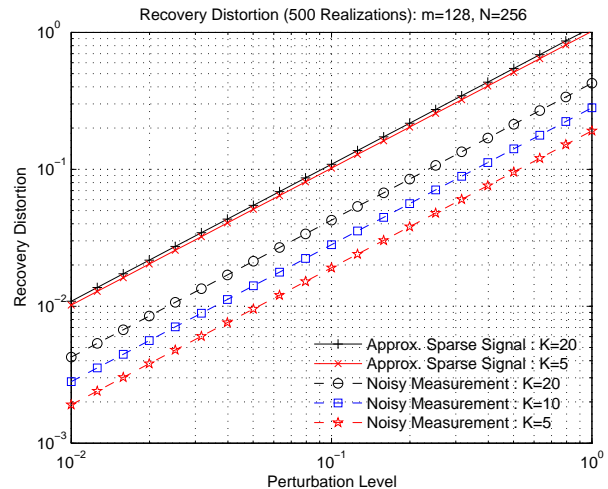


Figure 8. Reconstruction distortion under signal or measurement perturbations: both perturbation level and reconstruction distortion are described via the ℓ_2 norm.

and covariance matrix $\sigma_e^2 \mathbf{I}_m$, where \mathbf{I}_m denotes the $m \times m$ identity matrix.

We execute the SP decoding reconstruction process on \mathbf{y} , 500 times for each K , σ_s^2 and σ_e^2 . The reconstruction distortion $\|\mathbf{x} - \hat{\mathbf{x}}\|_2$ is obtained via averaging over all these instances, and the results are plotted in Fig. 8. Consistent with the findings of Theorem 9 and Corollary 1, we observe that the recovery distortion increases linearly with the ℓ_2 -norm of measurement errors. Even more encouraging is the fact that the empirical reconstruction distortion is typically much smaller than the corresponding upper bounds. This is likely due to the fact that, in order to simplify the expressions involved, many constants and parameters used in the proof were upper bounded.

A. Recovery Distortion under Measurement Perturbations

The first step towards proving Theorem 9 is to upper bound the reconstruction error for a given estimated support set \hat{T} , as succinctly described in the lemma to follow.

Lemma 3: Let $\mathbf{x} \in \mathbb{R}^N$ be a K -sparse vector, $\|\mathbf{x}\|_0 \leq K$, and let $\mathbf{y} = \Phi \mathbf{x} + \mathbf{e}$ be a measurement for which $\Phi \in \mathbb{R}^{m \times N}$ satisfies the RIP with parameter δ_K . For an arbitrary $\hat{T} \subset \{1, \dots, N\}$ such that $|\hat{T}| \leq K$, define $\hat{\mathbf{x}}$ as

$$\hat{\mathbf{x}}_{\hat{T}} = \Phi_{\hat{T}}^\dagger \mathbf{y},$$

and

$$\hat{\mathbf{x}}_{\{1, \dots, N\} - \hat{T}} = \mathbf{0}.$$

Then

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq \frac{1}{1 - \delta_{3K}} \|\hat{\mathbf{x}}_0\|_2 + \frac{1 + \delta_{3K}}{1 - \delta_{3K}} \|\mathbf{e}\|_2.$$

The proof of the lemma is given in Appendix G.

Next, we need to upper bound the norm $\|\hat{\mathbf{x}}_0\|_2$. To achieve this task, we describe in the theorem to follow how $\|\hat{\mathbf{x}}_0\|_2$ depends on the noise energy $\|\mathbf{e}\|_2$.

Theorem 10: Let $\hat{\mathbf{x}}_0 = \mathbf{x}_{T-\hat{T}}$, $\mathbf{x}'_0 = \mathbf{x}_{T-(\hat{T} \cup T')}$ and $\tilde{\mathbf{x}}_0 = \mathbf{x}_{T-\hat{T}}$. Suppose that

$$\|\mathbf{e}\|_2 \leq \frac{\delta_{2K}}{1 - \delta_{2K}^2} \|\hat{\mathbf{x}}_0\|_2. \quad (7)$$

Then

$$\|\mathbf{x}'_0\|_2 \leq \frac{4\sqrt{\delta_{2K}}}{1 + \delta_{2K}} \|\hat{\mathbf{x}}_0\|_2, \quad (8)$$

and

$$\|\tilde{\mathbf{x}}_0\|_2 \leq \left(\frac{4\sqrt{\delta_{3K}}}{1 - \delta_{3K}} + \frac{2\delta_{3K}}{(1 - \delta_{3K})^2} \right) \|\hat{\mathbf{x}}_0\|_2. \quad (9)$$

Furthermore, if

$$\delta_{3K} < 0.03,$$

one has

$$\|\tilde{\mathbf{y}}_r\|_2 < \|\mathbf{y}_r\|_2.$$

Proof: The upper bounds in Inequalities (8) and (9) are proved in Appendix H and I respectively. To complete the proof, we make use of Lemma 2 stated in Section II. According to this lemma, we have

$$\begin{aligned} \|\tilde{\mathbf{y}}_r\|_2 &= \|\text{resid}(\mathbf{y}, \Phi_{\hat{T}})\|_2 \\ &\leq \|\Phi_{T-\hat{T}} \mathbf{x}_{T-\hat{T}}\|_2 + \|\mathbf{e}\|_2 \\ &\leq (1 + \delta_{3K}) \|\tilde{\mathbf{x}}_0\|_2 + \|\mathbf{e}\|_2 \\ &\leq \left((1 + \delta_{3K}) c'_K + \frac{\delta_{3K}}{1 - \delta_{3K}^2} \right) \|\hat{\mathbf{x}}_0\|_2, \end{aligned}$$

and

$$\begin{aligned} \|\mathbf{y}_r\|_2 &= \|\text{resid}(\mathbf{y}, \Phi_{\hat{T}})\|_2 \\ &\geq \frac{1 - 2\delta_{3K}}{1 - \delta_{3K}} (\|\Phi_{\hat{T}} \hat{\mathbf{x}}_0\| - \|\mathbf{e}\|_2) \\ &\geq \frac{1 - 2\delta_{3K}}{1 - \delta_{3K}} \left((1 - \delta_{3K}) \|\hat{\mathbf{x}}_0\|_2 - \frac{\delta_{3K}}{1 - \delta_{3K}^2} \|\hat{\mathbf{x}}_0\|_2 \right) \\ &\geq \left(1 - 2\delta_{3K} - \frac{\delta_{3K}}{1 - \delta_{3K}^2} \right) \|\hat{\mathbf{x}}_0\|_2. \end{aligned}$$

Elementary calculation reveal that as long as $\delta_{3K} < 0.03$, we have $\|\tilde{\mathbf{y}}_r\| < \|\mathbf{y}_r\|$. This completes the proof of the theorem. ■

Based on Theorem 10, we conclude that when the SP algorithm terminates, the inequality (7) is violated and we must have

$$\|\mathbf{e}\|_2 > \frac{\delta_{3K}}{1 - \delta_{3K}^2} \|\hat{\mathbf{x}}_0\|_2.$$

Under this assumption, it follows from Lemma 3 that

$$\begin{aligned} \|\mathbf{x} - \hat{\mathbf{x}}\|_2 &\leq \left(\frac{1}{1 - \delta_{3K}} \frac{1 - \delta_{3K}^2}{\delta_{3K}} + \frac{1 + \delta_{3K}}{1 - \delta_{3K}} \right) \|\mathbf{e}\|_2 \\ &= \frac{1 + \delta_{3K}}{\delta_{3K} (1 - \delta_{3K})} \|\mathbf{e}\|_2, \end{aligned}$$

which completes the proof of Theorem 9.

B. Recovery Distortion under Signal and Measurement Perturbations

The proof of Corollary 1 is based on the following two lemmas, which are proved in [16] and [17], respectively.

Lemma 4: Suppose that the sampling matrix $\Phi \in \mathbb{R}^{m \times N}$ satisfies the RIP with parameter δ_K . Then, for every $\mathbf{x} \in \mathbb{R}^N$, one has

$$\|\Phi \mathbf{x}\|_2 \leq \sqrt{1 + \delta_K} \left(\|\mathbf{x}\|_2 + \frac{1}{\sqrt{K}} \|\mathbf{x}\|_1 \right).$$

Lemma 5: Let $\mathbf{x} \in \mathbb{R}^d$ be K -sparse, and let \mathbf{x}_K denote the vector obtained from \mathbf{x} by keeping its K entries of largest magnitude, and by setting all its other components to zero. Then

$$\|\mathbf{x} - \mathbf{x}_K\|_2 \leq \frac{\|\mathbf{x}\|_1}{2\sqrt{K}}.$$

To prove the corollary, consider the measurement vector

$$\begin{aligned} \mathbf{y} &= \Phi \mathbf{x} + \mathbf{e} \\ &= \Phi \mathbf{x}_{2K} + \Phi (\mathbf{x} - \mathbf{x}_{2K}) + \mathbf{e}. \end{aligned}$$

By Theorem 9, one has

$$\|\hat{\mathbf{x}} - \mathbf{x}_{2K}\|_2 \leq C_{6K} (\|\Phi (\mathbf{x} - \mathbf{x}_{2K})\|_2 + \|\mathbf{e}\|_2),$$

and invoking Lemma 4 shows that

$$\begin{aligned} \|\Phi (\mathbf{x} - \mathbf{x}_{2K})\|_2 &\leq \sqrt{1 + \delta_{6K}} \left(\|\mathbf{x} - \mathbf{x}_{2K}\|_2 + \frac{\|\mathbf{x} - \mathbf{x}_{2K}\|_1}{\sqrt{6K}} \right). \end{aligned}$$

Furthermore, Lemma 5 implies that

$$\begin{aligned} \|\mathbf{x} - \mathbf{x}_{2K}\|_2 &= \|(\mathbf{x} - \mathbf{x}_K) - (\mathbf{x} - \mathbf{x}_K)_K\|_2 \\ &\leq \frac{1}{2\sqrt{K}} \|\mathbf{x} - \mathbf{x}_K\|_1. \end{aligned}$$

Therefore,

$$\begin{aligned} \|\Phi (\mathbf{x} - \mathbf{x}_{2K})\|_2 &\leq \sqrt{1 + \delta_{6K}} \left(\frac{\|\mathbf{x} - \mathbf{x}_K\|_1}{2\sqrt{K}} + \frac{\|\mathbf{x} - \mathbf{x}_{2K}\|_1}{\sqrt{6K}} \right) \\ &\leq \sqrt{1 + \delta_{6K}} \frac{\|\mathbf{x} - \mathbf{x}_K\|_1}{\sqrt{K}}, \end{aligned}$$

and

$$\|\hat{\mathbf{x}} - \mathbf{x}_{2K}\|_2 \leq c'_{2K} \left(\|\mathbf{e}\|_2 + \sqrt{1 + \delta_{6K}} \frac{\|\mathbf{x} - \mathbf{x}_K\|_1}{\sqrt{K}} \right),$$

which completes the proof.

VI. CONCLUSION

We introduced a new algorithm, termed subspace pursuit, for low-complexity recovery of sparse signals sampled by matrices satisfying the RIP with a constant parameter δ_{3K} . Also presented were simulation results demonstrating that the recovery performance of the algorithm matches, and sometimes even exceeds, that of the LP programming technique; and, simulations showing that the number of iterations executed by the algorithm for zero-one sparse signals and compressible signals is of the order $O(\log K)$.

APPENDIX

We provide next detailed proofs for the lemmas and theorems stated in the paper.

A. Proof of Lemma 1

- 1) The first part of the lemma follows directly from the definition of δ_K . Every vector $\mathbf{q} \in \mathbb{R}^K$ can be extended to a vector $\mathbf{q}' \in \mathbb{R}^{K'}$ by attaching $K' - K$ zeros to it. From the fact that for all $J \subset \{1, \dots, N\}$ such that $|J| \leq K'$, and all $\mathbf{q}' \in \mathbb{R}^{K'}$, one has

$$(1 - \delta_{K'}) \|\mathbf{q}'\|_2^2 \leq \|\Phi_J \mathbf{q}'\|_2^2 \leq (1 + \delta_{K'}) \|\mathbf{q}'\|_2^2,$$

it follows that

$$(1 - \delta_{K'}) \|\mathbf{q}\|_2^2 \leq \|\Phi_I \mathbf{q}\|_2^2 \leq (1 + \delta_{K'}) \|\mathbf{q}\|_2^2$$

for all $|I| \leq K$ and $\mathbf{q} \in \mathbb{R}^K$. Since δ_K is defined as the infimum of all parameter δ that satisfy the above relationship, $\delta_K \leq \delta_{K'}$.

- 2) The inequality

$$|\langle \Phi_I \mathbf{a}, \Phi_J \mathbf{b} \rangle| \leq \delta_{|I|+|J|} \|\mathbf{a}\|_2 \|\mathbf{b}\|_2$$

obviously holds if either one of the norms $\|\mathbf{a}\|_2$ and $\|\mathbf{b}\|_2$ is zero. Assume therefore that neither one of them is zero, and define

$$\begin{aligned} \mathbf{a}' &= \mathbf{a} / \|\mathbf{a}\|_2, \quad \mathbf{b}' = \mathbf{b} / \|\mathbf{b}\|_2, \\ \mathbf{x}' &= \Phi_I \mathbf{a}, \quad \mathbf{y}' = \Phi_J \mathbf{b}. \end{aligned}$$

Note that the RIP implies that

$$\begin{aligned} 2(1 - \delta_{|I|+|J|}) &\leq \|\mathbf{x}' + \mathbf{y}'\|_2^2 \\ &= \left\| \begin{bmatrix} \Phi_i \Phi_j \\ \Phi_i \Phi_j \end{bmatrix} \begin{bmatrix} \mathbf{a}' \\ \mathbf{b}' \end{bmatrix} \right\|_2^2 \leq 2(1 + \delta_{|I|+|J|}), \end{aligned} \quad (10)$$

and similarly,

$$\begin{aligned} 2(1 - \delta_{|I|+|J|}) &\leq \|\mathbf{x}' - \mathbf{y}'\|_2^2 \\ &= \left\| \begin{bmatrix} \Phi_i \Phi_j \\ -\Phi_i \Phi_j \end{bmatrix} \begin{bmatrix} \mathbf{a}' \\ \mathbf{b}' \end{bmatrix} \right\|_2^2 \leq 2(1 + \delta_{|I|+|J|}). \end{aligned}$$

We thus have

$$\langle \mathbf{x}', \mathbf{y}' \rangle \leq \frac{\|\mathbf{x}' + \mathbf{y}'\|_2^2 - \|\mathbf{x}' - \mathbf{y}'\|_2^2}{4} \leq \delta_{|I|+|J|},$$

$$-\langle \mathbf{x}', \mathbf{y}' \rangle \leq \frac{\|\mathbf{x}' - \mathbf{y}'\|_2^2 - \|\mathbf{x}' + \mathbf{y}'\|_2^2}{4} \leq \delta_{|I|+|J|},$$

and therefore

$$\frac{|\langle \Phi_I \mathbf{a}, \Phi_J \mathbf{b} \rangle|}{\|\mathbf{a}\|_2 \|\mathbf{b}\|_2} = |\langle \mathbf{x}', \mathbf{y}' \rangle| \leq \delta_{|I|+|J|}.$$

Now,

$$\begin{aligned} \|\Phi_I^* \Phi_J \mathbf{b}\|_2 &= \max_{\mathbf{q}: \|\mathbf{q}\|_2=1} \|\mathbf{q}^* (\Phi_I^* \Phi_J \mathbf{b})\|_2 \\ &\leq \max_{\mathbf{q}: \|\mathbf{q}\|_1=1} \delta_{|I|+|J|} \|\mathbf{q}\|_2 \|\mathbf{b}\|_2 \\ &= \delta_{|I|+|J|} \|\mathbf{b}\|_2, \end{aligned}$$

which completes the proof.

B. Proof of Lemma 2

- 1) The first claim is proved by observing that

$$\begin{aligned} \Phi_I^* \mathbf{y}_r &= \Phi_I^* (\mathbf{y} - \Phi_I (\Phi_I^* \Phi_I)^{-1} \Phi_I^* \mathbf{y}) \\ &= \Phi_I^* \mathbf{y} - \Phi_I^* \mathbf{y} = \mathbf{0}. \end{aligned}$$

- 2) To prove the second part of the lemma, let

$$\mathbf{y}_p = \Phi_I \mathbf{x}_p, \text{ and } \mathbf{y} = \Phi_I \mathbf{x}.$$

By Lemma 1, we have

$$\begin{aligned} |\langle \mathbf{y}_p, \mathbf{y} \rangle| &\leq \delta_{|I|+|J|} \|\mathbf{x}_p\|_2 \|\mathbf{x}\|_2 \\ &\leq \delta_{|I|+|J|} \frac{\|\mathbf{y}_p\|_2}{\sqrt{1 - \delta_{|I|}}} \frac{\|\mathbf{y}\|_2}{\sqrt{1 - \delta_{|J|}}} \\ &\leq \frac{\delta_{|I|+|J|}}{1 - \delta_{|I|+|J|}} \|\mathbf{y}_p\|_2 \|\mathbf{y}\|_2. \end{aligned}$$

Since

$$\langle \mathbf{y}_p, \mathbf{y} \rangle = \langle \mathbf{y}_p, \mathbf{y}_p + \mathbf{y}_r \rangle = \|\mathbf{y}_p\|_2^2,$$

we have

$$\|\mathbf{y}_p\|_2 \leq \frac{\delta_{|I|+|J|}}{1 - \delta_{|I|+|J|}} \|\mathbf{y}\|_2.$$

Furthermore, since

$$\|\mathbf{y}_r\|_2 = \|\mathbf{y} - \mathbf{y}_p\|_2 \geq \|\mathbf{y}\|_2 - \|\mathbf{y}_p\|_2$$

and since

$$\|\mathbf{y}_r\|_2 = \|\mathbf{y} - \mathbf{y}_p\|_2 \leq \|\mathbf{y}\|_2 + \|\mathbf{y}_p\|_2,$$

one can show that

$$1 - \frac{\delta_{|I|+|J|}}{1 - \delta_{|I|+|J|}} \leq \frac{\|\mathbf{y}_r\|_2}{\|\mathbf{y}\|_2} \leq 1 + \frac{\delta_{|I|+|J|}}{1 - \delta_{|I|+|J|}}.$$

Observing that

$$\|\mathbf{y}_r\|_2^2 + \|\mathbf{y}_p\|_2^2 = \|\mathbf{y}\|_2^2,$$

we finally show that

$$\left(1 - \frac{\delta_{|I|+|J|}}{1 - \delta_{|I|+|J|}}\right) \|\mathbf{y}\|_2 \leq \|\mathbf{y}_r\|_2 \leq \|\mathbf{y}\|_2.$$

C. Proof of Theorem 5

The first step consists in proving Inequality (5), which reads as

$$\|\Phi_{\hat{T}}^* \mathbf{y}\|_2 \geq (1 - \delta_{2K}) \|\mathbf{x}\|_2.$$

By assumption, $|T| \leq K$, so that

$$\|\Phi_T^* \mathbf{y}\|_2 = \|\Phi_T^* \Phi_T \mathbf{x}\|_2 \geq (1 - \delta_{2K}) \|\mathbf{x}\|_2,$$

which provides the desired proof. According to the definition of \hat{T} ,

$$\begin{aligned} \|\Phi_{\hat{T}}^* \mathbf{y}\|_2 &= \max_{|I| \leq K} \sqrt{\sum_{i \in I} |\langle \mathbf{v}_i, \mathbf{y} \rangle|^2} \\ &\geq \|\Phi_T^* \mathbf{y}\|_2 \geq (1 - \delta_{2K}) \|\mathbf{x}\|_2. \end{aligned}$$

The second step is to partition the estimate of the support set \hat{T} into two subsets: the set $\hat{T} \cap T$, containing the indices

in the correct support set, and $\hat{T} - T$, the set of incorrectly selected indices. Then

$$\begin{aligned} \|\Phi_{\hat{T}}^* \mathbf{y}\|_2 &\leq \|\Phi_{\hat{T} \cap T}^* \mathbf{y}\|_2 + \|\Phi_{\hat{T} - T}^* \mathbf{y}\|_2 \\ &\leq \|\Phi_{\hat{T} \cap T}^* \mathbf{y}\|_2 + \delta_{2K} \|\mathbf{x}\|_2, \end{aligned}$$

where the last inequality follows from the near-orthogonality property of Lemma 1.

Furthermore,

$$\begin{aligned} \|\Phi_{\hat{T} \cap T}^* \mathbf{y}\|_2 &\leq \|\Phi_{\hat{T} \cap T}^* \Phi_{\hat{T} \cap T} \mathbf{x}_{\hat{T} \cap T}\|_2 \\ &\quad + \|\Phi_{\hat{T} \cap T}^* \Phi_{T - \hat{T}} \mathbf{x}_{T - \hat{T}}\|_2 \\ &\leq (1 + \delta_{2K}) \|\mathbf{x}_{\hat{T} \cap T}\|_2 + \delta_{2K} \|\mathbf{x}\|_2. \end{aligned}$$

Combining the two inequalities above, one can show that

$$\|\Phi_{\hat{T}}^* \mathbf{y}\|_2 \leq (1 + \delta_{2K}) \|\mathbf{x}_{\hat{T} \cap T}\|_2 + 2\delta_{2K} \|\mathbf{x}\|_2.$$

By invoking Inequality (5) it follows that

$$(1 - \delta_{2K}) \|\mathbf{x}\|_2 \leq (1 + \delta_{2K}) \|\mathbf{x}_{\hat{T} \cap T}\|_2 + 2\delta_{2K} \|\mathbf{x}\|_2.$$

Hence,

$$\|\mathbf{x}_{\hat{T} \cap T}\|_2 \geq \frac{1 - 3\delta_{2K}}{1 + \delta_{2K}} \|\mathbf{x}\|_2.$$

To complete the proof, we observe that

$$\begin{aligned} \|\mathbf{x}_{T - \hat{T}}\|_2 &= \sqrt{\|\mathbf{x}\|_2^2 - \|\mathbf{x}_{\hat{T} \cap T}\|_2^2} \\ &\leq \frac{\sqrt{8\delta_{2K} + 4\delta_{2K}^2}}{1 + \delta_{2K}} \|\mathbf{x}\|_2. \end{aligned}$$

D. Proof of Theorem 3

The proof of this theorem heavily relies on the following technical (and tedious) notation, some of which has been previously described in the paper, but is repeated in this section for completeness:

- \mathbf{y}_r = resid($\mathbf{y}, \Phi_{\hat{T}}$), denotes the residue of the projection of \mathbf{y} onto the space span($\Phi_{\hat{T}}$);
- \mathbf{x}_r is the coefficient vector corresponding to \mathbf{y}_r , i.e., $\mathbf{y}_r = \Phi_{T \cup \hat{T}} \mathbf{x}_r$;
- $\hat{\mathbf{y}}_0$ = $\Phi_{T - \hat{T}} \mathbf{x}_{T - \hat{T}}$, is the component of the measurement which has not been captured by the set \hat{T} ;
- $\hat{\mathbf{x}}_0$ = $\mathbf{x}_{T - \hat{T}}$, denotes the part of the signal not captured by \hat{T} ;
- $\hat{\mathbf{y}}_{0,p}$ = proj($\hat{\mathbf{y}}_0, \Phi_{\hat{T}}$) denotes the projection of $\hat{\mathbf{y}}_0$ onto span($\Phi_{\hat{T}}$);
- $\hat{\mathbf{x}}_{0,p}$ is used to denote the projection coefficient vector corresponding to $\hat{\mathbf{y}}_{0,p}$, i.e., $\hat{\mathbf{y}}_{0,p} = \Phi_{\hat{T}} \hat{\mathbf{x}}_{0,p}$;
- T'' denotes the set of K residual indices with maximum correlation magnitudes $|\langle \mathbf{v}_i, \mathbf{y}_r \rangle|$;
- \mathbf{y}'_c = $\Phi_{T \cap T''} \mathbf{x}_{T \cap T''}$ denotes the component of the measured vector included through the set T'' ;
- \mathbf{x}'_c = $\mathbf{x}_{T \cap T''}$, denotes part of the sample signal supported on T'' ;
- \mathbf{y}'_0 = $\Phi_{T - T'} \mathbf{x}_{T - T'}$, corresponds to the measurement vector not captured by $T' = \hat{T} \cup T''$.

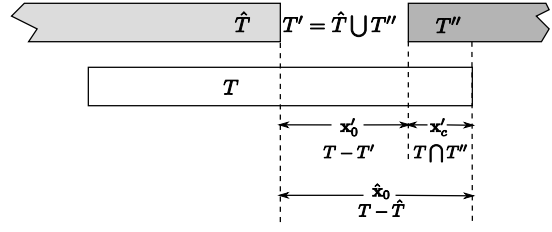


Figure 9. Illustration of sets and signal coefficient vectors for Theorem 3

$$\mathbf{x}'_0 = \mathbf{x}_{T - T'}, \text{ is the part of the signal not captured by } T'.$$

For clarity, some of the sets and vectors in the list above are depicted in Fig. 9.

With the above notation, the main step of the proof is to show that CM allows for capturing a significant part of the residual signal power, that is,

$$\|\mathbf{x}'_0\|_2 \leq c_1 \|\hat{\mathbf{x}}_0\|_2$$

for some constant c_1 . Note that $\hat{\mathbf{x}}_0$ is composed of \mathbf{x}'_0 and \mathbf{x}'_c , i.e.,

$$\hat{\mathbf{x}}_0 = [(\mathbf{x}'_0)^*, (\mathbf{x}'_c)^*]^*,$$

so that

$$\|\mathbf{x}'_0\|_2^2 = \|\hat{\mathbf{x}}_0\|_2^2 - \|\mathbf{x}'_c\|_2^2.$$

The most difficult part of our demonstration is to upper bound $\|\mathbf{x}'_c\|_2$.

The roadmap of the proof can be formed by establishing the validity of the following four claims.

1) If we write

$$\Phi_{T \cup \hat{T}} = [\Phi_{T - \hat{T}} \Phi_{\hat{T}}],$$

then

$$\mathbf{y}_r = \Phi_{T \cup \hat{T}} \mathbf{x}_r,$$

where

$$\mathbf{x}_r = [\hat{\mathbf{x}}_0^*, -\hat{\mathbf{x}}_{0,p}^*]^*.$$

We claim that

$$\|\hat{\mathbf{x}}_{0,p}\|_2 \leq \frac{\delta_{2K}}{1 - \delta_{2K}} \|\hat{\mathbf{x}}_0\|_2.$$

2) It holds that

$$\|\Phi_{T''}^* \mathbf{y}_r\|_2 \geq (1 - 2\delta_{2K}) \|\hat{\mathbf{x}}_0\|_2.$$

3) The corresponding upper bound reads as

$$\|\Phi_{T''}^* \mathbf{y}_r\|_2 \leq (1 + \delta_{2K}) \|\mathbf{x}'_c\|_2 + \frac{2\delta_{2K} - \delta_{2K}^2}{1 - \delta_{2K}} \|\hat{\mathbf{x}}_0\|_2.$$

4) Finally,

$$\|\mathbf{x}'_0\|_2 \leq \frac{\sqrt{10\delta_{2K}}}{1 + \delta_{2K}} \|\hat{\mathbf{x}}_0\|_2.$$

Proof: The claims can be established as demonstrated below.

1) It is clear that

$$\begin{aligned} \mathbf{y}_r &= \text{resid}(\mathbf{y}, \Phi_{\hat{T}}) = \text{resid}(\hat{\mathbf{y}}_0, \Phi_{\hat{T}}) \\ &= \hat{\mathbf{y}}_0 - \Phi_{\hat{T}} \left((\Phi_{\hat{T}}^* \Phi_{\hat{T}})^{-1} \Phi_{\hat{T}}^* \hat{\mathbf{y}}_0 \right) \\ &= \Phi_{T-\hat{T}} \hat{\mathbf{x}}_0 - \Phi_{\hat{T}} \hat{\mathbf{x}}_{0,p} \\ &= [\Phi_{T-\hat{T}}, \Phi_{\hat{T}}] \begin{bmatrix} \hat{\mathbf{x}}_0 \\ -\hat{\mathbf{x}}_{0,p} \end{bmatrix}. \end{aligned}$$

As a consequence of the RIP,

$$\begin{aligned} \|\hat{\mathbf{x}}_{0,p}\|_2 &= \left\| (\Phi_{\hat{T}}^* \Phi_{\hat{T}})^{-1} \Phi_{\hat{T}}^* (\Phi_{T-\hat{T}} \hat{\mathbf{x}}_0) \right\|_2 \\ &\leq \frac{1}{1-\delta_K} \delta_{2K} \|\hat{\mathbf{x}}_0\|_2 \leq \frac{\delta_{2K}}{1-\delta_{2K}} \|\hat{\mathbf{x}}_0\|_2. \end{aligned}$$

This proves the stated claim.

2) Note that

$$\mathbf{y}_r = \text{resid}(\mathbf{y}, \Phi_{\hat{T}}) \in \text{span}(\Phi_{T \cup \hat{T}}),$$

and that \mathbf{y}_r is orthogonal to $\Phi_{\hat{T}}$. We therefore have

$$\begin{aligned} \|\Phi_{T-\hat{T}}^* \mathbf{y}_r\|_2 &= \|\Phi_{T \cup \hat{T}}^* \mathbf{y}_r\|_2 = \|\Phi_{T \cup \hat{T}}^* (\Phi_{T \cup \hat{T}} \mathbf{x}_r)\|_2 \\ &\geq (1-\delta_{2K}) \|\mathbf{x}_r\|_2 \\ &\geq (1-\delta_{2K}) (\|\hat{\mathbf{x}}_0\|_2 - \|\hat{\mathbf{x}}_{0,p}\|_2) \\ &\geq (1-2\delta_{2K}) \|\hat{\mathbf{x}}_0\|_2. \end{aligned}$$

Since the set T'' is chosen so as to maximize the correlations with the residual vector, we can show that

$$\|\Phi_{T''}^* \mathbf{y}_r\|_2 \geq \|\Phi_{T-\hat{T}}^* \mathbf{y}_r\|_2 \geq (1-2\delta_{2K}) \|\hat{\mathbf{x}}_0\|_2,$$

which completes the proof.

3) Using the decomposition

$$\mathbf{y}_r = [\Phi_{T-\hat{T}}, \Phi_{\hat{T}}] [\hat{\mathbf{x}}_0^*, -\hat{\mathbf{x}}_{0,p}^*]^*,$$

we can show that

$$\begin{aligned} \|\Phi_{T''}^* \mathbf{y}_r\|_2 &\leq \|\Phi_{T''}^* \Phi_{T-\hat{T}} \mathbf{x}_{T-\hat{T}}\|_2 + \|\Phi_{T''}^* \Phi_{\hat{T}} \hat{\mathbf{x}}_{0,p}\|_2 \\ &\leq \|\Phi_{T''}^* \Phi_{T-\hat{T}} \mathbf{x}_{T-\hat{T}}\|_2 + \frac{\delta_{2K}^2}{1-\delta_{2K}} \|\hat{\mathbf{x}}_0\|_2. \end{aligned} \quad (11)$$

Since $\hat{T} \cap T'' = \emptyset$, we can partition the set $T - \hat{T}$ as

$$T - \hat{T} = (T \cap T'') \cup (T - \hat{T} - T'').$$

Then

$$\begin{aligned} &\|\Phi_{T''}^* \Phi_{T-\hat{T}} \mathbf{x}_{T-\hat{T}}\|_2 \\ &\leq \|\Phi_{T \cap T''}^* \Phi_{T-\hat{T}} \mathbf{x}_{T-\hat{T}}\|_2 + \|\Phi_{T-\hat{T}-T''}^* \Phi_{T-\hat{T}} \mathbf{x}_{T-\hat{T}}\|_2 \\ &\leq \|\Phi_{T \cap T''}^* \Phi_{T \cap T''} \mathbf{x}_{T \cap T''}\|_2 \\ &\quad + \|\Phi_{T \cap T''}^* \Phi_{T-\hat{T}-T''} \mathbf{x}_{T-\hat{T}-T''}\|_2 + \delta_{2K} \|\hat{\mathbf{x}}_0\|_2 \\ &\leq (1+\delta_{2K}) \|\mathbf{x}_{T \cap T''}\|_2 + \delta_{2K} \|\hat{\mathbf{x}}_0\|_2 + \delta_{2K} \|\hat{\mathbf{x}}_0\|_2. \end{aligned} \quad (12)$$

Upon substituting Inequality (12) into (11), we obtain

$$\|\Phi_{T''}^* \mathbf{y}_r\|_2 \leq (1+\delta_{2K}) \|\mathbf{x}_{T \cap T''}\|_2 + \frac{2\delta_{2K} - \delta_{2K}^2}{1-\delta_{2K}} \|\hat{\mathbf{x}}_0\|_2.$$

4) Combining the second and the third claims of the proof, we find that

$$\begin{aligned} \|\mathbf{x}'_c\|_2 &= \|\mathbf{x}_{T \cap T''}\|_2 \\ &\geq \frac{1}{1+\delta_{2K}} \left(1 - 2\delta_{2K} - \frac{2\delta_{2K} - \delta_{2K}^2}{1-\delta_{2K}} \right) \|\hat{\mathbf{x}}_0\|_2 \\ &= \frac{1-5\delta_{2K}+3\delta_{2K}^2}{1-\delta_{2K}^2} \|\hat{\mathbf{x}}_0\|_2. \end{aligned}$$

Based on this inequality, we can show that

$$\begin{aligned} \|\mathbf{x}'_0\|_2 &= \sqrt{\|\hat{\mathbf{x}}_0\|_2^2 - \|\mathbf{x}'_c\|_2^2} \\ &\leq \|\hat{\mathbf{x}}_0\|_2 \sqrt{1 - \left(\frac{1-5\delta_{2K}+3\delta_{2K}^2}{1-\delta_{2K}^2} \right)^2}. \end{aligned}$$

To make this result more tractable for subsequent analysis, we observe that

$$\begin{aligned} &(1-\delta_{2K}^2)^2 - (1-5\delta_{2K}+3\delta_{2K}^2)^2 \\ &\leq (1-\delta_{2K}^2)^2 - (1-5\delta_{2K}+\delta_{2K}^2)^2 \\ &= 10\delta_{2K} - 29\delta_{2K}^2 + 10\delta_{2K}^3 \\ &\leq 10\delta_{2K} (1-\delta_{2K})^2, \end{aligned}$$

so that

$$\|\mathbf{x}'_0\|_2 \leq \frac{\sqrt{10\delta_{2K}}}{1+\delta_{2K}} \|\hat{\mathbf{x}}_0\|_2,$$

as claimed. \blacksquare

E. Proof of Theorem 4

As in the previous subsection, we first introduce the notation followed in this part of the manuscript:

- \mathbf{y}'_0 = $\Phi_{T-T'} \mathbf{x}_{T-T'}$ denotes the part of the measurement vector not captured by T' ;
- \mathbf{x}'_0 = $\mathbf{x}_{T-T'}$ denotes part of the signal \mathbf{x} not captured by T' ;
- $\mathbf{y}'_{0,p}$ = $\text{proj}(\mathbf{y}'_0, \Phi_{T'})$ denotes the projection of \mathbf{y}'_0 onto $\text{span}(\Phi_{T'})$;
- $\mathbf{x}'_{0,p}$ denotes the projection coefficient vector corresponding to $\mathbf{y}'_{0,p}$, i.e., $\mathbf{y}'_{0,p} = \Phi_{T'} \mathbf{x}'_{0,p}$;
- \mathbf{y}'_p = $\text{proj}(\mathbf{y}', \Phi_{T'})$ denotes the projection of \mathbf{y} onto $\text{span}(\Phi_{T'})$;
- \mathbf{x}'_p stands for the projection coefficient vector corresponding to \mathbf{y}'_p , i.e., $\mathbf{y}'_p = \Phi_{T'} \mathbf{x}'_p$;
- \tilde{T} denotes the estimate of the K indices in T upon completion of an iteration (i.e., the set of those K indices that are deemed sufficiently reliable);
- ΔT = $T' - \tilde{T}$ consists of the set of indices estimated to be incorrect;
- $\Delta \mathbf{x}_0$ = $\mathbf{x}_{T \cap \Delta T}$ denotes the signal component erroneously removed from the list at the given iteration;
- $\tilde{\mathbf{x}}_0$ = $\mathbf{x}_{T-\tilde{T}}$, denotes the signal component not captured by \tilde{T} .

As for the previous proof, the sets and signal coefficient vectors introduced above are illustrated in Fig. 10. The previously studied concept of the *smear* of a vector is also depicted in the same figure.

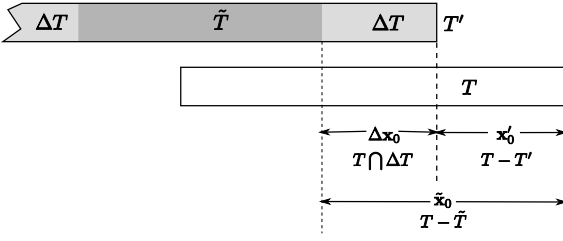


Figure 10. Illustration of sets and signal coefficient vectors for Theorem 4

To prove the theorem, we again proceed with establishing the validity of four different claims, listed below.

1) It can be shown that

$$\|\mathbf{x}'_{0,p}\|_2 \leq \frac{\delta_{3K}}{1 - \delta_{3K}} \|\mathbf{x}'_0\|_2.$$

2) For any index $i \in \hat{T} \cup T''$,

$$(\mathbf{x}'_p)_i = \begin{cases} \mathbf{x}_i + (\mathbf{x}'_{0,p})_i & \text{if } i \in T \\ (\mathbf{x}'_{0,p})_i & \text{if } i \notin T \end{cases}.$$

3) One has

$$\|\Delta \mathbf{x}_0\|_2 \leq 2 \|\mathbf{x}'_{0,p}\|_2.$$

4) And, finally,

$$\|\tilde{\mathbf{x}}_0\|_2 \leq \frac{1 + \delta_{3K}}{1 - \delta_{3K}} \|\mathbf{x}'_0\|_2.$$

Proof: The proofs proceed as follows.

1) To prove the first claim, we only need to note that

$$\begin{aligned} & \|\mathbf{x}'_{0,p}\|_2 \\ &= \left\| (\Phi_{T'}^*, \Phi_{T'})^{-1} \Phi_{T'}^* (\Phi_{T-T'} \mathbf{x}'_0) \right\|_2 \\ &\leq \frac{1}{1 - \delta_{2K}} \delta_{3K} \|\mathbf{x}'_0\|_2 \leq \frac{\delta_{3K}}{1 - \delta_{3K}} \|\mathbf{x}'_0\|_2. \end{aligned}$$

2) This claim is proved by partitioning the entries of the sampling matrix as follows. First, we write

$$\Phi_{T'} = [\Phi_{T \cap T'}, \Phi_{T-T'}].$$

Then, we observe that

$$\begin{aligned} & \Phi_{T \cap T'} \mathbf{x}_{T \cap T'} \\ &= [\Phi_{T \cap T'}, \Phi_{T-T'}] \begin{bmatrix} \mathbf{x}_{T \cap T'} \\ \mathbf{0} \end{bmatrix} \\ &= \Phi_{T'} \begin{bmatrix} \mathbf{x}_{T \cap T'} \\ \mathbf{0} \end{bmatrix}. \end{aligned}$$

Consequently,

$$\begin{aligned} \mathbf{x}'_p &= (\Phi_{T'}^*, \Phi_{T'})^{-1} \Phi_{T'}^* \mathbf{y} \\ &= (\Phi_{T'}^*, \Phi_{T'})^{-1} \Phi_{T'}^* (\Phi_{T \cap T'} \mathbf{x}_{T \cap T'}) \\ &\quad + (\Phi_{T'}^*, \Phi_{T'})^{-1} \Phi_{T'}^* (\Phi_{T-T'} \mathbf{x}_{T-T'}) \\ &= (\Phi_{T'}^*, \Phi_{T'})^{-1} \Phi_{T'}^* \Phi_{T'} \begin{bmatrix} \mathbf{x}_{T \cap T'} \\ \mathbf{0} \end{bmatrix} + \mathbf{x}'_{0,p} \\ &= \begin{bmatrix} \mathbf{x}_{T \cap T'} \\ \mathbf{0} \end{bmatrix} + \mathbf{x}'_{0,p}, \end{aligned}$$

which establishes the stated result.

3) As described before, if $T \subset T'$, then $\Delta T \cap T = \phi$ and $\|\Delta \mathbf{x}_0\|_2 = 0$. However, if $T - T' \neq \phi$, the projection coefficients \mathbf{x}'_p is a smeared version of $\mathbf{x}_{T'}$. By the second claim of this proof, the smear is simply $\mathbf{x}'_{0,p}$ and its energy equals $\|\mathbf{x}'_{0,p}\|_2$.

In what follows, we first show that the energy of the projection vector \mathbf{x}'_p restricted to ΔT is smaller than the energy of the smear, i.e.,

$$\left\| (\mathbf{x}'_p)_{\Delta T} \right\|_2 \leq \|\mathbf{x}'_{0,p}\|_2.$$

Consider an arbitrary index set $\Delta T'' \subset T'$ of cardinality K that is disjoint from T , $\Delta T'' \cap T = \phi$. Such a set $\Delta T''$ exists because $|T' - T| \geq K$. By the second claim in this proof,

$$\begin{aligned} \left\| (\mathbf{x}'_p)_{\Delta T''} \right\|_2 &= \sqrt{\sum_{i \in \Delta T''} (\mathbf{x}'_p)_i^2} \\ &= \sqrt{\sum_{i \in \Delta T''} (\mathbf{x}'_{0,p})_i^2} \leq \|\mathbf{x}'_{0,p}\|_2. \end{aligned}$$

Since ΔT is chosen to contain the K projection coefficients with the smallest magnitudes,

$$\left\| (\mathbf{x}'_p)_{\Delta T} \right\|_2 \leq \left\| (\mathbf{x}'_p)_{\Delta T''} \right\|_2 \leq \|\mathbf{x}'_{0,p}\|_2.$$

Next, we decompose the vector $(\mathbf{x}'_p)_{\Delta T}$ into a signal component and a smear component. Then

$$\begin{aligned} \left\| (\mathbf{x}'_p)_{\Delta T} \right\|_2 &= \left\| \mathbf{x}_{\Delta T} + (\mathbf{x}'_{0,p})_{\Delta T} \right\|_2 \\ &\geq \|\mathbf{x}_{\Delta T}\|_2 - \left\| (\mathbf{x}'_{0,p})_{\Delta T} \right\|_2. \end{aligned}$$

We also have

$$\begin{aligned} \|\Delta \mathbf{x}_0\|_2 &= \|\mathbf{x}_{\Delta T}\|_2 \leq \left\| (\mathbf{x}'_p)_{\Delta T} \right\|_2 + \left\| (\mathbf{x}'_{0,p})_{\Delta T} \right\|_2 \\ &\leq 2 \|\mathbf{x}'_{0,p}\|_2, \end{aligned}$$

which completes this part of the proof.

4) This claim is proved by combining the first three parts, and it results in

$$\begin{aligned} \|\tilde{\mathbf{x}}_0\|_2 &\leq \|\Delta \mathbf{x}_0\|_2 + \|\mathbf{x}'_0\|_2 \\ &\leq 2 \|\mathbf{x}'_{0,p}\|_2 + \|\mathbf{x}'_0\|_2 \\ &\leq \frac{2\delta_{3K}}{1 - \delta_{3K}} \|\mathbf{x}'_0\|_2 + \|\mathbf{x}'_0\|_2 \\ &= \frac{1 + \delta_{3K}}{1 - \delta_{3K}} \|\mathbf{x}'_0\|_2. \end{aligned}$$

F. Proof of Theorem 8

Without loss of generality, assume that

$$|x_1| \geq |x_2| \geq \dots \geq |x_K| > 0.$$

The following iterative algorithm is employed to generate a partition of the support set T that will establish the correctness of the claimed result.

Algorithm 2 Partitioning of the support set T

Initialization:

Let $T_1 = \{1\}$, $i = 1$ and $j = 1$.

Iteration:

If $i = K$, quit the iterations; otherwise, continue.

If

$$\|\mathbf{x}_{\{i+1, \dots, K\}}\|_2 \geq \frac{1}{2} |x_i|,$$

then we set $T_j = T_j \cup \{i+1\}$; otherwise, we have

$$\|\mathbf{x}_{\{i+1, \dots, K\}}\|_2 < \frac{1}{2} |x_i|,$$

and we set $j = j + 1$ and $T_j = \{i+1\}$.

Let $i = i + 1$. Continue with a new iteration.

Suppose that after the iterative partition, we have

$$T = T_1 \cup T_2 \cup \dots \cup T_J,$$

where $J \leq K$ is the number of the subsets in the partition.

Let $s_j = |T_j|$, $j = 1, \dots, J$. It is clear that

$$\sum_{j=1}^J s_j = K.$$

Then Theorem 8 is proved by invoking the following lemma.

Lemma 6:

1) For a given j , let $|T_j| = s$, and let

$$T_j = \{i, i+1, \dots, i+s-1\}.$$

Then,

$$|x_{i+s-1-k}| \leq 3^k |x_{i+s-1}|, \text{ for all } 0 \leq k \leq s-1, \quad (13)$$

and therefore

$$|x_{i+s-1}| \geq \frac{2}{3^s} \|\mathbf{x}_{\{i, \dots, K\}}\|_2. \quad (14)$$

2) Let

$$n_j = \left\lceil \frac{\log 2 - s_j \log 3}{\log c_K} \right\rceil, \quad (15)$$

where $\lceil \cdot \rceil$ denotes the ceiling function. Then for any $1 \leq j_0 \leq J$, after

$$\sum_{j=1}^{j_0} n_j$$

iterations, the SP algorithm has the property that

$$\bigcup_{j=1}^{j_0} T_j \subset \tilde{T}. \quad (16)$$

More specifically, after

$$\sum_{j=1}^J n_j \leq \frac{1.5 \cdot K}{-\log c_K} \quad (17)$$

iterations, the SP algorithm guarantees that $T \subset \tilde{T}$.

Proof: Both parts of this lemma are proved by mathematical induction as follows.

1) By the construction of T_j ,

$$\|\mathbf{x}_{\{i+s, \dots, K\}}\|_2 \leq \frac{1}{2} |x_{i+s-1}|.$$

On the other hand,

$$\begin{aligned} \frac{1}{2} |x_{i+s-2}| &\leq \|\mathbf{x}_{\{i+s-1, \dots, K\}}\|_2 \\ &\leq \|\mathbf{x}_{\{i+s, \dots, K\}}\|_2 + |x_{i+s-1}| \\ &\leq \frac{3}{2} |x_{i+s-1}|. \end{aligned}$$

It follows that

$$|x_{i+s-2}| \leq 3 |x_{i+s-1}|.$$

Now suppose that for any $1 < k_0 \leq s-1$,

$$|x_{i+s-1-k}| \leq 3^k |x_{i+s-1}| \text{ for all } 1 \leq k \leq k_0 - 1.$$

Then,

$$\begin{aligned} \frac{1}{2} |x_{i+s-1-k_0}| &\leq \|\mathbf{x}_{\{i+s-k_0, \dots, K\}}\| \\ &\leq |x_{i+s-k_0}| + \dots + |x_{i+s-1}| \\ &\quad + \|\mathbf{x}_{\{i+s, \dots, K\}}\| \\ &\leq \left(3^{k_0-1} + \dots + 1 + \frac{1}{2} \right) |x_{i+s-1}| \\ &\leq \frac{3^{k_0}}{2} |x_{i+s-1}|. \end{aligned}$$

This proves Equation (13) of the lemma. Inequality (14) then follows from the observation that

$$\begin{aligned} \|\mathbf{x}_{\{i, \dots, K\}}\|_2 &\leq |x_i| + \dots + |x_{i+s-1}| + \|\mathbf{x}_{\{i+s, \dots, K\}}\|_2 \\ &\leq \left(3^{s-1} + \dots + 1 + \frac{1}{2} \right) |x_{i+s-1}| \\ &\leq \frac{3^s}{2} |x_{i+s-1}|. \end{aligned}$$

2) From (15), it is clear that for $1 \leq j \leq J$,

$$c_K^{n_j} \leq \frac{2}{3^{s_j}}.$$

According to Theorem 2, after n_1 iterations,

$$\|\tilde{\mathbf{x}}_0\|_2 \leq \frac{2}{3^{s_1}} \|\mathbf{x}\|_2.$$

On the other hand, for any $i \in T_1$, it follows the first part of this lemma that

$$|x_i| \geq |x_{s_1}| \geq \frac{2}{3^{s_1}} \|\mathbf{x}\|.$$

Therefore,

$$T_1 \subset \tilde{T}.$$

Now, suppose that for a given $j_0 \leq J$, after $\sum_{j=1}^{j_0-1} n_j$ iterations, we have

$$\bigcup_{j=1}^{j_0-1} T_j \subset \tilde{T}.$$

Let $T_0 = \bigcup_{j=1}^{j_0-1} T_j$. Then

$$\|\tilde{\mathbf{x}}_0\|_2 = \|\mathbf{x}_{T-\tilde{T}}\|_2 \leq \|\mathbf{x}_{T-T_0}\|_2.$$

Denote the smallest coordinate in T_{j_0} by i , and the largest coordinate in T_{j_0} by k . Then

$$|x_k| \geq \frac{2}{3^{s_{j_0}}} \|\mathbf{x}_{\{i, \dots, k\}}\|_2 = \frac{2}{3^{s_{j_0}}} \|\mathbf{x}_{T-T_0}\|_2.$$

After n_{j_0} more iterations, we obtain \tilde{T}' and $\tilde{\mathbf{x}}'_0$. Consequently,

$$\|\tilde{\mathbf{x}}'_0\|_2 \leq \frac{2}{3^{s_{j_0}}} \|\tilde{\mathbf{x}}_0\|_2 \leq \frac{2}{3^{s_{j_0}}} \|\mathbf{x}_{T-T_0}\|_2 \leq |x_k|.$$

As a result, we conclude that

$$T_{j_0} \subset \tilde{T}$$

after $\sum_{j=1}^{j_0} n_j$ iterations, which proves inequality (16). Now in order to ensure that $T \subset \tilde{T}$, the SP algorithm needs at most

$$\begin{aligned} \sum_{j=1}^J n_j &\leq \sum_{j=1}^J \frac{s_j \log 3 - \log 2 + 1}{-\log c_K} \\ &\leq \frac{K \log 3 + J(1 - \log 2)}{-\log c_K} \\ &\leq \frac{K(\log 3 + 1 - \log 2)}{-\log c_K} \leq \frac{K \cdot 1.5}{-\log c_K} \end{aligned}$$

iterations. This completes the proof of the last claim. \blacksquare

G. Proof of Lemma 3

The claim in the lemma is established through the following chain of inequalities.

$$\begin{aligned} \|\mathbf{x} - \hat{\mathbf{x}}\|_2 &\leq \|\mathbf{x}_{\hat{T}} - \Phi_{\hat{T}}^\dagger \mathbf{y}\|_2 + \|\mathbf{x}_{T-\hat{T}}\|_2 \\ &= \|\mathbf{x}_{\hat{T}} - \Phi_{\hat{T}}^\dagger (\Phi_T \mathbf{x}_T + \mathbf{e})\|_2 + \|\mathbf{x}_{T-\hat{T}}\|_2 \\ &\leq \|\mathbf{x}_{\hat{T}} - \Phi_{\hat{T}}^\dagger (\Phi_T \mathbf{x}_T)\|_2 + \|\Phi_{\hat{T}}^\dagger \mathbf{e}\|_2 + \|\mathbf{x}_{T-\hat{T}}\|_2 \\ &\leq \|\mathbf{x}_{\hat{T}} - \Phi_{\hat{T}}^\dagger (\Phi_{T \cap \hat{T}} \mathbf{x}_{T \cap \hat{T}})\|_2 \\ &\quad + \|\Phi_{\hat{T}}^\dagger \Phi_{T-\hat{T}} \mathbf{x}_{T-\hat{T}}\|_2 \\ &\quad + \frac{\sqrt{1+\delta_K}}{1-\delta_K} \|\mathbf{e}\| + \|\mathbf{x}_{T-\hat{T}}\|_2 \\ &\leq 0 + \left(\frac{\delta_{2K}}{1-\delta_K} + 1 \right) \|\mathbf{x}_{T-\hat{T}}\|_2 + \frac{\sqrt{1+\delta_K}}{1-\delta_K} \|\mathbf{e}\|_2 \\ &\leq \frac{1}{1-\delta_{2K}} \|\mathbf{x}_{T-\hat{T}}\|_2 + \frac{\sqrt{1+\delta_K}}{1-\delta_K} \|\mathbf{e}\|_2. \end{aligned}$$

Note that the next to last inequality is a consequence of the fact that

$$\|\mathbf{x}_{\hat{T}} - \Phi_{\hat{T}}^\dagger (\Phi_{T \cap \hat{T}} \mathbf{x}_{T \cap \hat{T}})\|_2 = 0.$$

By relaxing the upper bound in terms of replacing δ_{2K} by δ_{3K} , we obtain

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq \frac{1}{1-\delta_{3K}} \|\mathbf{x}_{T-\hat{T}}\|_2 + \frac{1+\delta_{3K}}{1-\delta_{3K}} \|\mathbf{e}\|_2.$$

This completes the proof of the lemma.

H. Proof of Inequality (8)

Following the same notations outlined in Section D, we have

$$\begin{aligned} \|\Phi_{T' \cup T}^* \mathbf{y}_r\|_2 &\geq \left\| \Phi_{T'}^* \Phi_{T \cup T'} \hat{T} \mathbf{x}_r \right\|_2 - \|\Phi_{T'}^* \mathbf{e}\|_2 \\ &\geq \left\| \Phi_{T-\hat{T}}^* \Phi_{T \cup T'} \hat{T} \mathbf{x}_r \right\|_2 - \sqrt{1+\delta_K} \|\mathbf{e}\|_2 \\ &\geq (1-2\delta_{2K}) \|\hat{\mathbf{x}}_0\|_2 - \sqrt{1+\delta_K} \|\mathbf{e}\|_2 \\ &\geq (1-2\delta_{2K}) \|\hat{\mathbf{x}}_0\|_2 - \left(1 + \frac{1}{2}\delta_{2K}\right) \|\mathbf{e}\|_2. \end{aligned}$$

On the other hand,

$$\begin{aligned} \|\Phi_{T' \cup T}^* \mathbf{y}_r\|_2 &\leq \left\| \Phi_{T'}^* \Phi_{T \cup T'} \hat{T} \mathbf{x}_r \right\|_2 + \|\Phi_{T'}^* \mathbf{e}\|_2 \\ &\leq (1+\delta_{2K}) \|\mathbf{x}_{T \cap T'}\|_2 + \frac{2\delta_{2K} - \delta_{2K}^2}{1-\delta_{2K}} \|\hat{\mathbf{x}}_0\|_2 \\ &\quad + \left(1 + \frac{1}{2}\delta_{2K}\right) \|\mathbf{e}\|_2. \end{aligned}$$

By combining these two bounds we obtain

$$\begin{aligned} \|\mathbf{x}_{T \cap T'}\|_2 &\geq \frac{1-5\delta_{2K}+3\delta_{2K}^2}{1-\delta_{2K}^2} \|\hat{\mathbf{x}}_0\|_2 - \frac{2+\delta_{2K}}{1+\delta_{2K}} \|\mathbf{e}\|_2 \\ &\geq \frac{1-5\delta_{2K}+3\delta_{2K}^2}{1-\delta_{2K}^2} \|\hat{\mathbf{x}}_0\|_2 - 2\|\mathbf{e}\|_2. \end{aligned}$$

Recall next the result in Inequality (7), stating that

$$\|\mathbf{e}\|_2 \leq \frac{\delta_{2K}}{1-\delta_{2K}^2} \|\hat{\mathbf{x}}_0\|_2.$$

The two above inequalities imply that

$$\|\mathbf{x}_{T \cap T'}\|_2 \geq \frac{1-7\delta_{2K}+\delta_{2K}^2}{1-\delta_{2K}^2} \|\hat{\mathbf{x}}_0\|_2.$$

Therefore,

$$\begin{aligned} \|\mathbf{x}'_0\|_2 &\leq \frac{\sqrt{(1-\delta_{2K}^2)^2 - (1-7\delta_{2K}+\delta_{2K}^2)^2}}{1-\delta_{2K}^2} \|\hat{\mathbf{x}}_0\|_2 \\ &= \frac{\sqrt{14\delta_{2K}(1-(\frac{7}{2}+\frac{2}{7})\delta_{2K}+\delta_{2K}^2)}}{1-\delta_{2K}^2} \|\hat{\mathbf{x}}_0\|_2 \\ &\leq \frac{\sqrt{16\delta_{2K}(1-2\delta_{2K}+\delta_{2K}^2)}}{1-\delta_{2K}^2} \|\hat{\mathbf{x}}_0\|_2 \\ &= \frac{4\sqrt{\delta_{2K}}}{1+\delta_{2K}} \|\hat{\mathbf{x}}_0\|_2, \end{aligned}$$

as claimed.

I. Proof of Inequality (9)

We start by first upper bounding the norm $\|\mathbf{x}'_{0,p}\|_2$ as outlined below.

$$\begin{aligned} &\left\| \Phi_{\hat{T} \cup T'}^\dagger (\Phi_{T-\hat{T} \cup T'} \mathbf{x}'_0 + \mathbf{e}) \right\|_2 \\ &\leq \left\| \Phi_{\hat{T} \cup T'}^\dagger \Phi_{T-\hat{T} \cup T'} \mathbf{x}'_0 \right\|_2 + \left\| \Phi_{\hat{T} \cup T'}^\dagger \mathbf{e} \right\|_2 \\ &\leq \frac{\delta_{3K}}{1-\delta_{3K}} \|\mathbf{x}'_0\|_2 + \frac{\sqrt{1+\delta_{3K}}}{1-\delta_{3K}} \|\mathbf{e}\|_2 \\ &\leq \frac{\delta_{3K}}{1-\delta_{3K}} \|\mathbf{x}'_0\|_2 + \frac{1+\frac{1}{2}\delta_{3K}}{1-\delta_{3K}} \|\mathbf{e}\|_2. \end{aligned}$$

Then, similar type of arguments as those used in Section E, establish that

$$\begin{aligned} \|\tilde{\mathbf{x}}_0\|_2 &\leq 2\|\mathbf{x}'_{0,p}\|_2 + \|\mathbf{x}'_0\|_2 \\ &\leq \frac{1 + \delta_{3K}}{1 - \delta_{3K}} \|\mathbf{x}'_0\|_2 + \frac{2 + \delta_{3K}}{1 - \delta_{3K}} \|\mathbf{e}\|_2 \\ &\leq \frac{4\sqrt{\delta_{3K}}}{1 - \delta_{3K}} \|\hat{\mathbf{x}}_0\|_2 + \frac{2 + \delta_{3K}}{1 - \delta_{3K}} \|\mathbf{e}\|_2. \end{aligned}$$

Recalling the assumption in (7), we arrive at

$$\begin{aligned} \|\tilde{\mathbf{x}}_0\|_2 &\leq \frac{4\sqrt{\delta_{3K}}}{1 - \delta_{3K}} \|\hat{\mathbf{x}}_0\|_2 + \frac{2 + \delta_{3K}}{1 - \delta_{3K}} \frac{\delta_{3K}}{1 - \delta_{3K}^2} \|\hat{\mathbf{x}}_0\|_2 \\ &\leq \frac{4\sqrt{\delta_{3K}}}{1 - \delta_{3K}} \|\hat{\mathbf{x}}_0\|_2 + \frac{2\delta_{3K}}{(1 - \delta_{3K})^2} \|\hat{\mathbf{x}}_0\|_2, \end{aligned}$$

thereby proving the claimed result.

REFERENCES

- [1] J. Tropp, D. Needell, and R. Vershynin, "Iterative signal recovery from incomplete and inaccurate measurements," in *Information Theory and Applications*, Jan. 27 - Feb. 1 San Deigo, CA, 2008.
- [2] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [3] R. Venkataramani and Y. Bresler, "Sub-nyquist sampling of multiband signals: perfect reconstruction and bounds on aliasing error," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 3, 12-15 May Seattle, WA, 1998, pp. 1633–1636 vol.3.
- [4] E. Candès and T. Tao, "Decoding by linear programming," *Information Theory, IEEE Transactions on*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [5] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [6] E. Candès, R. Mark, T. Tao, and R. Vershynin, "Error correction via linear programming," in *IEEE Symposium on Foundations of Computer Science (FOCS)*, 2005, pp. 295 – 308.
- [7] G. Cormode and S. Muthukrishnan, "Combinatorial algorithms for compressed sensing," in *Proceedings of the 40th Annual Conference on Information Sciences and Systems*, 2006.
- [8] S. Sarvotham, D. Baron, and R. Baraniuk, "Compressed sensing reconstruction via belief propagation." *Preprint*, 2006.
- [9] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.
- [10] R. S. Varga, *Geršgorin and His Circles*. Berlin: Springer-Verlag, 2004.
- [11] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," submitted.
- [12] Y. Han, C. Hartmann, and C.-C. Chen, "Efficient priority-first search maximum-likelihood soft-decision decoding of linear block codes," *IEEE Trans. on Information Theory*, vol. 51, pp. 1514–1523, 1993.
- [13] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.
- [14] I. E. Nesterov, A. Nemirovskii, and Y. Nesterov, *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM, 1994.
- [15] Å. Björck, *Numerical Methods for Least Squares Problems*. SIAM, 1996.
- [16] A. Gilbert, M. Strauss, J. Tropp, and R. Vershynin, "One sketch for all: Fast algorithms for compressed sensing," in *Symp. on Theory of Computing (STOC)*, San Diego, CA, June, 2007.
- [17] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," preprint, 2008.