Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 11–13

ON PARTIALLY SPARSE RECOVERY

A. BANDEIRA, K. SCHEINBERG AND L. N. VICENTE

ABSTRACT: In this paper we consider the problem of recovering a partially sparse solution of an underdetermined system of linear equations by minimizing the ℓ_1-norm of the part of the solution vector which is known to be sparse. Such a problem is closely related to the classical problem in Compressed Sensing where the ℓ_1-norm of the whole solution vector is minimized. We introduce analogues of restricted isometry and null space properties for the recovery of partially sparse vectors and show that these new properties are implied by their original counterparts. We show also how to extend recovery under noisy measurements to the partially sparse case.

KEYWORDS: Sparse recovery, partially sparse recovery, compressed sensing, ℓ_1-minimization, applications in sparse quadratic polynomial interpolation.
AMS SUBJECT CLASSIFICATION (1991): 65D05, 90C90.

1. Introduction

In Compressed Sensing one is interested in recovering a sparse solution $\bar{x} \in \mathbb{R}^N$ of an underdetermined system of the form $y = A\bar{x}$, given a vector $y \in \mathbb{R}^k$ and a matrix $A \in \mathbb{R}^{k \times N}$ with far fewer rows than columns ($k \ll N$). This can be accomplished by minimizing the number of non-zero components of x, i.e., the ℓ_0-*norm* of x (the ℓ_0-*norm* is defined by $\|u\|_0 = |\{i : u_i \neq 0\}|$ but, strictly speaking, it is not a norm)

$$\min \|x\|_0 \quad \text{s.t.} \quad Ax = y, \tag{1}$$

which is an NP-Hard problem. It is commonly considered sufficient to solve a more tractable approximation to this problem obtained by substituting the non-convex ℓ_0-*norm* by a relatively close convex approximation.

Recent results indicate that the ℓ_1-norm can serve as such an approximation (see [5] for a survey on some of this material) and, in fact, $\ell_1(x) = \|x\|_1$ is the tightest convex relaxation of the real-extended function $g(x)$: $g(x) = \|x\|_0$ if $\|x\|_\infty \leq 1$ and $g(x) = +\infty$ otherwise [12]. Hence problem (1) is replaced by the following optimization problem

$$\min \|x\|_1 \quad \text{s.t.} \quad Ax = y. \tag{2}$$

Date: April 15, 2011.

Support for this work was provided by FCT under the grant PTDC/MAT/098214/2008.

Note that (2) is equivalent to a linear program and thus is much easier to solve than (1).

In this paper we consider the case when it is known a priori that the solution vector consists of two parts, one of which is expected to be dense, in other words we have $x = (x_1, x_2)$, where $x_1 \in \mathbb{R}^{N-r}$ is sparse and $x_2 \in \mathbb{R}^r$ is possibly dense. A natural generalization of problem (2) to this setting of partially sparse recovery is given by

$$\min \|x_1\|_1 \quad \text{s. t.} \quad A_1 x_1 + A_2 x_2 = y, \tag{3}$$

where $A = (A_1, A_2)$, $A_1 \in \mathbb{R}^{k\times(N-r)}$, and $A_2 \in \mathbb{R}^{k\times r}$. We will refer to this setting as *partially sparse recovery of size* $N - r$.

Such type of problems arise naturally in sparse Hessian recovery (Bandeira, Scheinberg, and Vicente [2]). In particular, suppose we want to determine a model $m(z) = \alpha_1^\top \phi_1(z) + \alpha_2^\top \phi_2(z)$, written on some function basis $\phi = \{\phi_1, \phi_2\}$, by interpolating a function $f : \mathbb{R}^n \to \mathbb{R}$ on a sample set Y, where the number $k = |Y|$ of sample points is lower than the number $N = |\phi_1| + |\phi_2|$ of basis components. The interpolation conditions are represented by $M(\phi_1, Y)\alpha_1 + M(\phi_2, Y)\alpha_2 = f(Y)$, where $M(\phi, Y)$ is the Vandermonde matrix associated with the basis ϕ and the sample set Y. Now suppose that $m(z)$ has a representation in terms of ϕ_1 and ϕ_2 with α_1 a sparse vector (and α_2, possibly, dense). Then, such an interpolation model can be constructed by solving a problem of the form (3) where $(x_1, x_2) = (\alpha_1, \alpha_2)$, $(A_1, A_2) = (M(\phi_1, Y), M(\phi_2, Y))$, $y = f(Y)$, and $r = |\phi_2|$. For instance for quadratic interpolation of a function whose Hessian is sparse, but the gradient is expected to be dense, ϕ_1 contains the quadratic basis components $z_i^2/2$, $z_i z_j$, with $j > i$, $i, j \in \{1, \ldots, n\}$, whereas ϕ_2 contains the linear basis elements $1, z_1, \ldots, z_n$. The authors have used partially sparse recovery approach in [2] for building sparse quadratic interpolation models of functions with sparse Hessian. It is shown in [2] that by using randomly selected interpolation sets with $\mathcal{O}(n(\log n)^4)$ points one can achieve model accuracy similar to that of second-order Taylor polynomials if there are $\mathcal{O}(n)$ non-zero elements in the Hessian. Such a number is in comparison to $\mathcal{O}(n^2)$ interpolation points which are required by other existing results that achieved similar accuracy. The authors have also successfully applied the ℓ_1–sparse reconstruction of quadratic models [2] in interpolation-based trust-region methods for Derivative-Free Optimization [8].

One of the key applications of partially sparse recovery is image reconstruction (Vaswani and Lu [15]). In the paper [15] a sufficient condition is given under which partially sparse recovery is achieved. Their condition is weaker than the known restricted isometry property for general sparse recovery. This is natural, since the case of partial sparsity can be considered as a case of general sparsity where part of the support of the solution is known in advance. The contribution of our paper (a subset of our results appeared in [1]) is to introduce the corresponding analogues of restricted isometry and null space properties for the case of partial sparsity. We will show that these new properties are sufficient for partially sparse recovery (including the noisy case) and are implied by the original conditions of fully sparse recovery.

1.1. Notation. We will use the following notation in this paper. $[N]$ denotes the set of integers $\{1, \dots, N\}$, and $[N]^{(s)}$ denotes the set of all subsets of $[N]$ of cardinality $s \leq N$. If A is a matrix, then by $\mathcal{N}(A)$ and $\mathcal{R}(A)$ we denote the null and range spaces of A, respectively. We say that a vector x is $s-$sparse if at most s components of x are non-zero. This is also denoted by $\|x\|_0 \leq s$. Given $v \in \mathbb{R}^N$ and $S \in [N]$, $v_S \in \mathbb{R}^N$ denotes a vector defined by $(v_S)_i = v_i$, $i \in S$ and $(v_S)_i = 0$, $i \notin S$.

2. Sparse recovery in compressed sensing

The main question addressed by Compressed Sensing is under what conditions on the matrix A can a sparse vector $\bar{x}$ be recovered by solving problem (2) given A and the right hand side $y = A\bar{x}$. The next definition is a well known characterization of such matrices.

Definition 2.1 (Null Space Property). *The matrix $A \in \mathbb{R}^{k \times N}$ is said to satisfy the Null Space Property (NSP) of order s if, for every $v \in \mathcal{N}(A) \setminus \{0\}$ and for every $S \in [N]^{(s)}$, one has*

$$\|v_S\|_1 < \frac{1}{2}\|v\|_1. \tag{4}$$

The term Null Space Property was introduced in [7]. However, we note that the characterization mentioned above and formalized in the following theorem had been implicitly used in [9]. It is well known that NSP is a necessary and sufficient condition for the recovery of an s-sparse vector $\bar{x}$.

Theorem 2.1. *The matrix A satisfies the Null Space Property of order s if and only if, for every $s-$sparse vector $\bar{x}$, problem (2) with $y = A\bar{x}$ has an unique solution and it is given by $x = \bar{x}$.*

Proof: For a proof see, e.g., [13]. ∎

It is difficult to verify if a matrix satisfies the NSP. On the other hand, the *Restricted Isometry Property* (RIP), introduced in [4], is considerably more useful and insightful, although it provides only sufficient conditions for every $s-$sparse vector $\bar{x}$ to be the unique solution of (2) when $y = A\bar{x}$. We present below the definition of the RIP and the *Restricted Isometry Property Constant.*

Definition 2.2 (Restricted Isometry Property)**.** *One says that $\delta_s > 0$ is the Restricted Isometry Property Constant, or* RIP *constant, of order s of the matrix $A \in \mathbb{R}^{k \times N}$ if δ_s is the smallest positive real such that:*

$$(1 - \delta_s) \, \|x\|_2^2 \;\leq\; \|Ax\|_2^2 \;\leq\; (1 + \delta_s) \, \|x\|_2^2 \tag{5}$$

for every $s-$sparse vector x.

The following theorem (see, e.g., [13]) provides a useful sufficient condition for successful recovery by (2) with $y = A\bar{x}$.

Theorem 2.2. *Let $A \in \mathbb{R}^{k \times N}$ and $2s < N$. If*

$$2\delta_{2s} + \delta_s \;<\; 1, \tag{6}$$

where δ_s and δ_{2s} are the RIP constants of A of order s and $2s$, respectively, then, for every $s-$sparse vector $\bar{x}$, problem (2) with $y = A\bar{x}$ has an unique solution and it is given by $x = \bar{x}$.

It is known that RIP is satisfied with some probability if the entries of the matrix are randomly generated (see, e.g., [3]), but it is still an open problem to find deterministic matrices which satisfy such property when the underlying system is highly underdetermined (see [14]).

3. Partial sparse recovery

In this section we consider the following extension of the NSP to the case of partially sparse recovery.

Definition 3.1 (Null Space Property for Partially Sparse Recovery)**.** *We say that $A = (A_1, A_2)$ satisfies the Null Space Property (NSP) of order $s - r$ for* *partially sparse recovery of size* $N-r$ *with* $r \leq s$ *if* A_2 *is full column rank* ($\mathcal{N}(A_2) = \{0\}$) *and for every* $v_1 \in \mathbb{R}^{N-r} \setminus \{0\}$ *such that* $A_1 v_1 \in \mathcal{R}(A_2)$ *and every* $S \in [N-r]^{(s-r)}$*, we have*

$$\|(v_1)_S\|_1 \;<\; \frac{1}{2}\|v_1\|_1. \tag{7}$$

Note that when $r = 0$, i.e., when x does not have a known dense part, the NSP for partially sparse recovery reduces to the NSP in Definition 2.1. The new property is a necessary and sufficient condition for any solution of (3) with $y = A\bar{x}$ to satisfy $x_1 = \bar{x}_1$ if $\bar{x}_1$ is appropriately sparse.

Theorem 3.1. *The matrix* $A = (A_1, A_2)$ *satisfies the Null Space Property of order* $s-r$ *for Partially Sparse Recovery of size* $N-r$ *if and only if for every* $\bar{x} = (\bar{x}_1, \bar{x}_2)$ *such that* $\bar{x}_1 \in \mathbb{R}^{N-r}$ *is* $(s-r)-$*sparse and* $\bar{x}_2 \in \mathbb{R}^r$*, problem (3) with* $y = A\bar{x}$ *has an unique solution and it is given by* $(x_1, x_2) = (\bar{x}_1, \bar{x}_2)$*.*

Proof: The proof follows the steps of the proof of [13, Theorem 2.3] with appropriate modifications. Let us assume first that for any vector $(\bar{x}_1, \bar{x}_2) \in \mathbb{R}^N$, where $\bar{x}_1$ is an $(s-r)-$sparse vector and $\bar{x}_2 \in \mathbb{R}^r$, the minimizer (x_1, x_2) of $\|x_1\|_1$ subject to $A_1x_1 + A_2x_2 = A\bar{x}$ satisfies $x_1 = \bar{x}_1$. Consider any $v_1 \neq 0$ such that $A_1 v_1 \in \mathcal{R}(A_2)$. Then consider minimizing $\|x_1\|_1$ subject to $A_1x_1 + A_2x_2 = A_1(v_1)_S + A_2v_2$ for any $v_2 \in \mathbb{R}^r$ and for any $S \in [N-r]^{(s-r)}$. By the assumption, the corresponding minimizer (x_1, x_2) satisfies $x_1 = (v_1)_S$. Since $A_1v_1 \in \mathcal{R}(A_2)$, there exists u_2 such that $A_1(-(v_1)_{S^c}) + A_2u_2 = A_1(v_1)_S + A_2v_2$. As $-(v_1)_{S^c} \neq (v_1)_S$, $(-(v_1)_{S^c}, u_2)$ is not the minimizer of $\|x_1\|_1$ subject to $A_1x_1 + A_2x_2 = A_1(v_1)_S + A_2v_2$, hence, $\|(v_1)_{S^c}\|_1 > \|(v_1)_S\|_1$ and (7) holds.

Let us now assume that A satisfies the NSP of order $s-r$ for partially sparse recovery of size $N-r$ (Definition 3.1). Then, given a vector $(\bar{x}_1, \bar{x}_2) \in \mathbb{R}^N$, where $\bar{x}_1$ is $(s-r)-$sparse and $\bar{x}_2 \in \mathbb{R}^r$, and a vector $(u_1, u_2) \in \mathbb{R}^N$ with $u_1 \neq \bar{x}_1$ and satisfying $A_1u_1 + A_2u_2 = A_1\bar{x}_1 + A_2\bar{x}_2$, consider $(v_1, v_2) = ((\bar{x}_1 - u_1), (\bar{x}_2 - u_2)) \in \mathcal{N}(A)$, which implies $A_1v_1 \in \mathcal{R}(A_2)$ and $v_1 \neq 0$.

Thus, setting S to be the support of $\bar{x}$, one has that

$$\begin{aligned}\|\bar{x}_1\|_1 &\leq \|\bar{x}_1 - (u_1)_S\|_1 + \|(u_1)_S\|_1 \\ &= \|(v_1)_S\|_1 + \|(u_1)_S\|_1 \\ &< \|(v_1)_{S^c}\|_1 + \|(u_1)_S\|_1 \\ &= \|-(u_1)_{S^c}\|_1 + \|(u_1)_S\|_1 \\ &= \|u_1\|_1,\end{aligned}$$

(the strict inequality coming from (7)), guaranteeing that all solutions (x_1, x_2) of problem (3) with $y = A\bar{x}$ satisfy $x_1 = \bar{x}_1$.

It remains to point out that $x_2 = \bar{x}_2$ is uniquely determined by solving $A_2 x_2 = y - A_1\bar{x}_1$ if and only if A_2 is full column rank. ■

The rank assumption on A_2 is reasonable in the Derivative-Free Optimization setting, for instance, where r is smaller than the number of rows in A.

We now define an extension of the RIP to the partially sparse recovery setting. For this purpose, let $A = (A_1, A_2)$ be as considered above, under the assumption that A_2 has full column rank. Let

$$\mathcal{P} \;=\; I - A_2\left(A_2^\top A_2\right)^{-1} A_2^\top \tag{8}$$

be the matrix of the orthogonal projection from $\mathbb{R}^N$ onto $\mathcal{R}\left(A_2\right)^\perp$. Then, the problem of recovering $(\bar{x}_1, \bar{x}_2)$, where $\bar{x}_1$ is an $(s-r)-$sparse vector satisfying $A_1\bar{x}_1 + A_2\bar{x}_2 = y$, can be stated as the problem of recovering an $(s-r)-$sparse vector $x_1 = \bar{x}_1$ satisfying $(\mathcal{P}A_1)\, x_1 = \mathcal{P}y$ and then recovering $x_2 = \bar{x}_2$ satisfying $A_2 x_2 = y - A_1\bar{x}_1$. The solution of the resulting linear system in the second step exists and is unique given that A_2 has full column rank and $(\mathcal{P}A_1)\bar{x}_1 = \mathcal{P}y$. Note that the first step is now reduced to the classical setting of Compressed Sensing. This motivates the following definition of RIP for partially sparse recovery.

Definition 3.2 (Partial RIP). *We say that $\delta_{s-r}^r > 0$ is the Partial Restricted Isometry Property Constant of order $s-r$ of the matrix $A = (A_1, A_2) \in \mathbb{R}^{k\times N}$, for recovery of size $N-r$ with $r \leq s$, if A_2 is full column rank and δ_{s-r}^r is the RIP constant of order $s-r$ (see Definition 2.2) of the matrix $\mathcal{P}A_1$, where $\mathcal{P}$ is given by (8).*

Again, when $r = 0$ the Partial RIP reduces to the RIP of Definition 2.2. We also note that, given a matrix $A = (A_1, A_2) \in \mathbb{R}^{k\times N}$ with Partial RIP constants $\delta^r_{2(s-r)}$ and δ^r_{s-r} of order $2(s-r)$ and $s-r$, respectively, for recovery of size $N - r$, satisfying $2\delta^r_{2(s-r)} + \delta^r_{s-r} < 1$, then, by Theorems 2.1 and 2.2, we have that $\mathcal{P}A_1$ satisfies the NSP of order $s - r$. Thus, given $\bar{x} = (\bar{x}_1, \bar{x}_2)$ such that $\bar{x}_1 \in \mathbb{R}^{N-r}$ is $(s - r)-$sparse and $\bar{x}_2 \in \mathbb{R}^r$, $\bar{x}_1$ can be recovered by minimizing the ℓ_1-norm of x_1 subject to $(\mathcal{P}A_1)x_1 = \mathcal{P}A\bar{x}$ and, recalling that A_2 is full-column rank, $x_2 = \bar{x}_2$ is uniquely determined by $A_2x_2 = y - A_1\bar{x}_1$. (In particular, this implies that A satisfies the NSP of order $s-r$ for partially sparse recovery of size $N - r$.)

4. Partially sparse recovery implied by fully sparse recovery conditions

We are now interested in showing that partially sparse recovery is achievable under the conditions which guarantee fully sparse recovery. In particular we will show that the NSP and RIP imply, respectively, the NSP for partially sparse recovery and the partial RIP. We first establish the relationship between the corresponding null space properties.

Theorem 4.1. *If a given matrix A satisfies the NSP of order s then it satisfies the NSP for partially sparse recovery of order $s - r$ for any $r \leq s$.*

Proof: Let $A = (A_1, A_2)$ satisfy the NSP of order s. First we note that since $r \leq s$, the NSP implies that A_2 is full column rank. Let $v_1 \in \mathbb{R}^{N-r}$ be a non-zero vector such that $A_1v_1 \in \mathcal{R}(A_2)$ and let $T \in [N - r]^{(s-r)}$. Define $W = [N] \setminus [N - r]$.

Since there exists v_2 such that $A_1v_1 + A_2v_2 = 0$, we have that $v = (v_1, v_2) \in \mathcal{N}(A) \setminus \{0\}$, and therefore by setting $S = T \cup W$ and by using the NSP,

$$\|(v_1)_T\|_1 + \|v_2\|_1 \;=\; \|v_S\|_1 \;<\; \frac{1}{2}\|v\|_1 \;=\; \frac{1}{2}\|v_1\|_1 + \frac{1}{2}\|v_2\|_1.$$

Thus,

$$\|(v_1)_T\|_1 \;\leq\; \|(v_1)_T\|_1 + \frac{1}{2}\|v_2\|_1 \;\leq\; \frac{1}{2}\|v_1\|_1,$$

and A satisfies the NSP of order $s - r$ for partially sparse recovery of size $N - r$. ∎

Partial RIP is also implied by RIP without the change in the RIP constant value.

Theorem 4.2. *Let $A = (A_1, A_2)$ satisfy the RIP of order s with the RIP constant δ_s. Then A satisfies partial RIP of order $s - r$ with $\delta_{s-r}^r = \delta_s$ for partially sparse recovery of size $N - r$, for any $r \leq s$.*

Proof: First we note that since $r \leq s$, the RIP implies that A_2 is full column rank. Consider now any given $(s-r)-$sparse vector $x_1 \in \mathbb{R}^{N-r}$. First note that from Definition 2.2, RIP of order s implies

$$(1-\delta_s)\left(\|x_1\|_2^2 + \|x_2\|_2^2\right) \;\leq\; \|A_1x_1 + A_2x_2\|_2^2 \;\leq\; (1+\delta_s)\left(\|x_1\|_2^2 + \|x_2\|_2^2\right), \tag{9}$$

for the given $x_1 \in \mathbb{R}^{N-r}$ and any (possibly dense) vector $x_2 \in \mathbb{R}^r$.

Now, by setting $x_2 = -\left(A_2^\top A_2\right)^{-1} A_2^\top A_1 x_1$, one obtains

$$(1-\delta_s)\|x_1\|_2^2 \;\leq\; (1-\delta_s)\left(\|x_1\|_2^2 + \|x_2\|_2^2\right) \;\leq\; \|A_1x_1 + A_2x_2\|_2^2 \;=\; \|\mathcal{P}A_1x_1\|^2.$$

On the other hand, the choice $x_2 = 0$ provides

$$\|\mathcal{P}A_1x_1\|_2^2 \;\leq\; \|A_1x_1\|_2^2 \;\leq\; (1+\delta_s)\,\|x_1\|_2^2.$$

We have thus arrived at the conditions of Definition 3.2. ■

Remark 4.1. *Notice that Theorems 4.1 and 4.2 apply to any partition of A into A_1 and A_2 of appropriate size, hence full NSP and RIP conditions are clearly stronger than partial NSP and RIP conditions for particular A_1 and A_2.*

Remark 4.2. *Theorems 2.1, 3.1, and 4.1 guarantee that if any sparse vector is recovered by full ℓ_1-minimization, then it is recovered by partial ℓ_1-minimization if some part of the support is known. However, a particular partially sparse vector may be recovered by full ℓ_1-minimization and not by the partial one. For example, let A and $\bar{x}$ be defined as*

$$A = \begin{bmatrix} 2 & -2 & -1 \\ 0 & -1 & 1 \end{bmatrix}, \quad \bar{x} = \begin{bmatrix} -4 \\ 0 \\ 3 \end{bmatrix},$$

with $y = A\bar{x} = (-11, 3)^\top$, $s = 2$ and $r = 1$. Vector $\bar{x}$ can be reconstructed by the full ℓ_1-minimization problem

$$\min \ \|x\|_1 \quad s.t. \quad \begin{bmatrix} 2 & -2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -11 \\ 3 \end{bmatrix}.$$

However, the solution of partial ℓ_1-minimization problem

$$\min\ |x_1| + |x_2| \quad s.t. \quad \begin{bmatrix} 2 & -2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -11 \\ 3 \end{bmatrix}$$

is $(0, 8/3, 17/3)^\top$ and hence the partially sparse vector $x = (-4, 0, 3)^\top$ is not recovered. This example does contradict the theory since one can easily show that the matrix A does not satisfy the NSP.

5. Partial (and total) compressibility recovery with noisy measurements

In most realistic applications the observed measurement vector y often contains noise and the true signal vector $\bar{x}$ is not sparse but rather compressible, meaning that most components are very small but not necessarily zero. It is known, however, that Compressed Sensing is robust to noise and can approximately recover compressible vectors. This statement is formalized in the following theorem taken from [6] (see also [10, 11] and references therein for results involving smaller values of the constant in the bound (10) below).

Theorem 5.1. *Assume that the matrix $A \in \mathbb{R}^{k\times N}$ satisfies RIP with the RIP constant δ_{2s} such that*

$$\delta_{2s} \ < \ \sqrt{2} - 1. \tag{10}$$

For any $\bar{x} \in \mathbb{R}^N$, let noisy measurements $y = A\bar{x} + \epsilon$ be given satisfying $\|\epsilon\|_2 \le \eta$. Let $x^\#$ be a solution of

$$\min_{x\in\mathbb{R}^N} \|x\|_1 \quad s.t. \quad \|Ax - y\|_2 \le \eta. \tag{11}$$

Then

$$\|x^\# - \bar{x}\|_2 \ \le \ c\eta + d\frac{\sigma_s(\bar{x})_1}{\sqrt{s}}, \tag{12}$$

for constants c, d only depending on the RIP constant, and where $\sigma_s(\bar{x})_1 = \min_{x:\, \|x\|_0 \le s} \|x - \bar{x}\|_1$.

The following theorem provides an analogous result for the partially sparse recovery setting introduced in Section 3.

Theorem 5.2. *Assume that the matrix $A = (A_1, A_2) \in \mathbb{R}^{k\times N}$ satisfies partial RIP of order $2(s-r)$ for recovery of size $N - r$ with the RIP constant $\delta^r_{2(s-r)}$*

for which (10) holds. For any $\bar{x} = (\bar{x}_1, \bar{x}_2) \in \mathbb{R}^N$*, let noisy measurements* $y = A\bar{x} + \epsilon$ *be given satisfying* $\|\epsilon\|_2 \leq \eta$*. Let* $x^* = (x_1^*, x_2^*)$ *be a solution of*

$$\min_{x=(x_1,x_2)\in\mathbb{R}^N} \|x_1\|_1 \quad s.t. \quad \|Ax - y\|_2 \leq \eta. \tag{13}$$

Then

$$\|x_1^* - \bar{x}_1\|_2 \;\leq\; c\eta + d\frac{\sigma_{s-r}(\bar{x}_1)_1}{\sqrt{s-r}}, \tag{14}$$

and

$$\|x_2^* - \bar{x}_2\|_2 \;\leq\; C_2\left(2\eta + C_1\left(c\eta + d\frac{\sigma_{s-r}(\bar{x}_1)_1}{\sqrt{s-r}}\right)\right), \tag{15}$$

for constants c, d *only depending on the RIP constant* $\delta^r_{2(s-r)}$*, and where* C_1 *and* C_2 *are given by*

$$C_1 \;=\; \|A_1\|_2 \;=\; \max_{x_1\neq 0}\frac{\|A_1x_1\|_2}{\|x_1\|_2} \quad and \quad C_2 \;=\; \|A_2^\dagger\| \;=\; \frac{1}{\min_{x_2\neq 0}\frac{\|A_2x\|_2}{\|x\|_2}}.$$

(Since A_2 *is full column rank recall that* $A_2^\dagger = (A_2^\top A_2)^{-1}A_2^\top$ *and* $C_2 > 0$*.)*

Proof: From Theorem 4.2, the matrix $\mathcal{P}A_1$, where $\mathcal{P}$ is given by (8), satisfies the condition of Theorem 5.1. Thus, since $\mathcal{P}$ is a projection matrix,

$$\|\mathcal{P}A_1\bar{x}_1 - \mathcal{P}y\| \;=\; \|\mathcal{P}A\bar{x} - \mathcal{P}y\| \;\leq\; \|A\bar{x} - y\| \;\leq\; \eta,$$

and a solution $x_1^\#$ of

$$\min_{x_1\in\mathbb{R}^{N-r}} \|x_1\|_1 \quad \text{s.t.} \quad \|\mathcal{P}A_1x_1 - \mathcal{P}y\|_2 \leq \eta, \tag{16}$$

satisfies

$$\|x_1^\# - \bar{x}_1\|_2 \;\leq\; c\eta + d\frac{\sigma_{s-r}(x_1)_1}{\sqrt{s-r}}. \tag{17}$$

Now, we will prove that the solutions of problems (13) and (16) coincide in their x_1 parts, completing thus the proof of (14). Let (x_1^*, x_2^*) be a feasible point of (13). Again, since $\mathcal{P}$ is a projection matrix, we obtain that

$$\|\mathcal{P}A_1x_1^* - \mathcal{P}y\|_2 \;=\; \|\mathcal{P}(A_1x_1^* + A_2x_2^* - y)\|_2 \;\leq\; \|A_1x_1^* + A_2x_2^* - y\|_2 \;\leq\; \eta,$$

which proves that x_1^* is a feasible point of (16). Now let $x_1^\#$ be a feasible point of (16). Since $I - \mathcal{P}$ projects (orthogonally) onto the column space of A_2 there must exist an $x_2^\#$ such that $A_2x_2^\# = (I-\mathcal{P})(y - A_1x_1^\#)$, and then

$$\|A_1x_1^\# + A_2x_2^\# - y\|_2 \;=\; \|\mathcal{P}A_1x_1^\# - \mathcal{P}y\|_2 \;\leq\; \eta.$$

Therefore $(x_1^{\#}, x_2^{\#})$ is a feasible point of (13). Hence we have proved that, any solution of problem (13) is also a solution of problem (16), and the inequality (14) results directly from (17).

We now use this inequality to bound the error on the reconstruction of $\bar{x}_2$. Since both $\bar{x}$ and x^* satisfy the measurements constraints $\|Ax - y\|_2 \leq \eta$ we have that

$$\|A_1(\bar{x}_1^* - x_1) + A_2(\bar{x}_2^* - x_2)\|_2 \;\leq\; 2\eta,$$

and thus

$$\|A_2(x_2^* - \bar{x}_2)\|_2 \;\leq\; 2\eta + \|A_1(x_1^* - \bar{x}_1)\|_2.$$

Using the definitions of C_1 and C_2 we have

$$\|x_2^* - \bar{x}_2\|_2 \;\leq\; C_2\left(2\eta + C_1\|x_1^* - \bar{x}_1\|_2\right),$$

and the result (15) follows from bounding $\|x_1^* - \bar{x}_1\|_2$ by (14) in the above inequality. ■

The condition on the matrix A imposed in the previous theorem involved only its partial RIP constant. In the next theorem we describe how one can bound the constants C_1 and C_2 using the RIP constant of A, which is known to be small in many applications.

Theorem 5.3. *The constants C_1 and C_2 of Theorem 5.2 can be bounded using the RIP constant δ_s of order s of $A = (A_1, A_2) \in \mathbb{R}^{k\times N}$ in the following way:*

$$C_1 \;\leq\; \sqrt{1+\delta_s} \quad \textit{and} \quad C_2 \;\leq\; \frac{1}{\sqrt{1-\delta_s}}.$$

Proof: For every $x_1 \in \mathbb{R}^{N-r}$ there exist $m = \lceil \frac{N-r}{s} \rceil$ vectors $a_1, ..., a_m \in \mathbb{R}^N$ such that $(x_1^\top, 0)^\top = \sum_{i=1}^m a_i$, each a_i is $s-$sparse, and no two vectors share non-zero components. As a result, the bound on C_1 can be derived from

$$\|A_1x_1\|^2 \;\leq\; \sum_{i=1}^m \|Aa_i\|^2 \;\leq\; (1+\delta_s)\sum_{i=1}^m \|a_i\|^2 \;=\; (1+\delta_s)\|x_1\|^2.$$

For the bound on C_2 we note first that $A_2x_2 = A(0, x_2^\top)^\top$ and that $(0, x_2^\top)^\top$ is $s-$sparse (since $r \leq s$). By the definition of the RIP constant,

$$\|A_2x_2\|_2 \;=\; \|A(0, x_2^\top)^\top\|_2 \;\geq\; \sqrt{1-\delta_s}\|(0, x_2^\top)^\top\|_2 \;=\; \sqrt{1-\delta_s}\|x_2\|_2,$$

proving that $1/C_2 \geq \sqrt{1-\delta_s}$. ■

6. Concluding Remarks

In some applications of Compressed Sensing one may be interested in a sparse (or compressible) vector whose support is partially known in advance. In such a setting we show that one can consider the ℓ_1-minimization of the part of the vector for which the support is not known. We have shown that sparse recovery can be then ensured under conditions that are potentially weaker than those assumed for the full approach. Intuitively, the bound on the number k of measurements needed for a matrix to satisfy Partial RIP (with high probability) should decrease as the number r of known dense components increases. However, to the best of our knowledge, there is no good understanding of this dependency, and we plan to address this in our future research.

Acknowledgments. We would like to thank Rachel Ward (Courant Institute of Mathematical Sciences, NYU) for interesting discussions on the topic of this paper.

References

[1] A. Bandeira. Computation of sparse low degree interpolating polynomials and their application to derivative-free optimization. Master's thesis, Dept. of Mathematics, Univ. of Coimbra, 2010.

[2] A. Bandeira, K. Scheinberg, and L. N. Vicente. Computation of sparse low degree interpolating polynomials and their application to derivative-free optimization. 2011.

[3] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constr. Approx.*, 28:253–263, 2008.

[4] E. Candès and T. Tao. Near optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52:5406–5425, 2006.

[5] E. J. Candès. Compressive sampling. *Proceedings of the International Congress of Mathematicians Madrid 2006*, Vol. III, 2006.

[6] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *Compte Rendus de l'Academie des Sciences, Paris, Serie I*, 589–592, 2009.

[7] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *J. Amer. Math. Soc.*, 22:211–231, 2009.

[8] A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to Derivative-Free Optimization.* MPS-SIAM Series on Optimization. SIAM, Philadelphia, 2009.

[9] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decompositions. *IEEE Trans. Inform. Theory*, 47:2845–2862, 2001.

[10] S. Foucart. Sparse recovery algorithms: Sufficient conditions in terms of restricted isometry constants. In *Proceedings of the 13th International Conference on Approximation Theory*, 2011.

[11] S. Foucart and M.-J. Lai. Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$. *Appl. Comput. Harmon. Anal.*, 26:395–407, 2009.

[12] J.-B. Hiriart-Urruty. Private communication, May, 2010.

[13] H. Rauhut. Compressed sensing and structured random matrices. In M. Fornasier, editor, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, Radon Series Comp. Appl. Math., pages 1–92. 2010.

[14] T. Tao. Open question: Deterministic UUP matrices: http://terrytao.wordpress.com/2007/07/02/open-question-deterministic-uu%p-matrices, 2007.

[15] N. Vaswani and W. Lu. Modified-CS: Modifying compressive sensing for problems with partially known support. *IEEE Trans. Signal Process.*, 58:4595–4607, 2010.

A. Bandeira
Program on Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544, USA (ajsb@math.princeton.edu).

K. Scheinberg
Department of Industrial and Systems Engineering, Lehigh University, Harold S. Mohler Laboratory, 200 West Packer Avenue, Bethlehem, PA 18015-1582, USA (katyas@lehigh.edu).

L. N. Vicente
CMUC, Department of Mathematics, University of Coimbra, 3001-454 Coimbra, Portugal (lnv@mat.uc.pt).