Near Oracle Performance of Signal Space Greedy Methods

R. Giryes and D. Needell

February 12, 2014

Abstract

Compressive sampling (CoSa) is a new methodology which demonstrates that sparse signals can be recovered from a small number of linear measurements. Greedy algorithms like CoSaMP have been designed for this recovery, and variants of these methods have been adapted to the case where sparsity is with respect to some arbitrary dictionary rather than an orthonormal basis. In this work we present an analysis of this so-called Signal Space CoSaMP method when the measurements are corrupted with mean-zero white Gaussian noise. We establish near-oracle performance for recovery of signals sparse in some arbitrary dictionary.

1 Introduction

We consider the compressive sensing problem which aims to recover a signal $x \in \mathbb{R}^d$ from noisy measurements

$$y = Mx + e,$$

where $M \in \mathbb{R}^{m \times d}$ is a known linear operator and $e \in \mathbb{R}^d$ is additive bounded noise, i.e. $\|e\|_2^2 \leq \varepsilon^2$. A typical assumption in this context is that the signal $x$ is sparse. There are several notions of sparsity, the simplest of which is that the signal itself has a small number of non-zero elements: $\|x\|_0 \leq k$, where $\|x\|_0 = |\text{supp}(x)|$ denotes the $\ell_0$ quasi-norm. We call such signals $k$-sparse.

A common approach to the compressive sensing problem utilizes the following optimization problem, deemed $\ell_1$-synthesis,

$$\hat{x}_{\ell_1} = \arg\min \|x\|_1 \quad \text{s.t.} \quad \|y - Mx\|_2 \leq \varepsilon. \quad (2)$$

One can guarantee accurate recovery using this approach when the measurement operator $M$ satisfies the Restricted Isometry Property (RIP) [1], which states that for some small enough constant $\delta_k < 1$,

$$(1 - \delta_k) \|x\|^2 \leq \|Mx\|^2 \leq (1 + \delta_k) \|x\|^2 \quad \text{for all } k\text{-sparse } x.$$

for some small enough constant $\delta_k < 1$.

It has been shown [2][3][4] that when the signal $x$ is $k$-sparse and $M$ satisfies the RIP with $\delta_{2k} < 0.4652$, the program (2) accurately recovers the signal,

$$\|\hat{x}_{\ell_1} - x\|_2 \leq C_{\ell_1} \varepsilon. \quad (3)$$

However, this simple notion of sparsity limits the reality of compressive sensing applications, so we instead consider signals sparse in some dictionary $D \in \mathbb{R}^{d \times n}$:

$$x = D\alpha \quad \text{for some } \|\alpha\|_0 \leq k.$$
In this setting one can utilize the same $\ell_1$-synthesis program to obtain a candidate coefficient vector $\hat{\alpha}_{\ell_1}$ and then estimate the signal $x$ by $\hat{x}_{\ell_1} = D\hat{\alpha}_{\ell_1}$. Initial work on this problem shows that under stringent requirements on the dictionary $D$, accurate recovery is possible (see e.g. [5, 6]).

Alternatively, one can solve the $\ell_1$-analysis problem which minimizes coefficients in the analysis domain,

$$\hat{x}_{\ell_1} = \text{argmin } \|D^* x\|_1 \text{ s.t. } \|y - Mx\|_2 \leq \varepsilon.$$  \hspace{1cm} (4)

In [7], the authors prove accurate recovery using this approach when the operators $M$ and $D$ satisfy the $D$-RIP:

$$(1 - \delta_k) \|D\alpha\|^2 \leq \|MD\alpha\|^2 \leq (1 + \delta_k) \|D\alpha\|^2 \text{ for all } k\text{-sparse } \alpha.$$  \hspace{1cm} (5)

Another approach to solving the compressive sensing problem is to use a greedy algorithm. Recently introduced methods that use this strategy are the CoSaMP [8], IHT [9], and HTP [10] methods. Greedy methods attempt to uncover the support of the signal iteratively, and then utilize a simple least-squares problem to estimate the entire signal.

Recently, the greedy approaches have been adapted to the setting in which signals are sparse with respect to arbitrary dictionaries. In particular, the Signal Space CoSaMP variant of the CoSaMP method [8] is shown in Algorithm 1. Here and throughout, the subscript $T$ denotes the restriction to elements (or columns) indexed by $T$. $\mathcal{S}_k(y)$ denotes the operator which returns the support of the best $k$-sparse representation of $y$ in the dictionary $D$, and $P_T$ denotes the projection onto the range of $D_T$.

This method is analyzed in [11], under the assumption of the $D$-RIP [5] and the assumption that one has access to projections $\mathcal{S}_k$ which satisfy

$$\|\mathcal{S}_k(z) - \mathcal{S}_k^*(z)\|_2 \leq \min \left( c_1 \|\mathcal{S}_k^*(z)\|_2, c_2 \|z - \mathcal{S}_k^*(z)\|_2 \right),$$  \hspace{1cm} (6)

where $\mathcal{S}_k^*$ denotes the optimal projection:

$$\mathcal{S}_k^*(z) = \text{argmin } \|z - P_T z\|_2^2.$$  \hspace{1cm} (7)

\begin{algorithm}[h]
\caption{Signal Space CoSaMP (SSCoSaMP)}
\begin{algorithmic}
\Require $k, M, D, y, a$ where $y = Mx + e$, $k$ is the sparsity of $x$ under $D$ and $e$ is the additive noise. $\mathcal{S}_{ak,1}$ and $\mathcal{S}_{k,2}$ are two near optimal support selection schemes.
\Ensure $\hat{x}$: $k$-sparse approximation of $x$.
\State Initialize the support $T^0 = \emptyset$, the residual $y^0 = y$ and set $t = 0$.
\While {halting criterion is not satisfied}
\State $t = t + 1$.
\State Find new support elements: $T_\Delta = \mathcal{S}_{ak,1}(M^*y^{t-1})$.
\State Update the support: $T^t = T^{t-1} \cup T_\Delta$.
\State Compute the representation: $x_p = D(MD_{T^t})^\dagger y = D\left\{\text{argmin}_\alpha \|y - MD\hat{\alpha}\|_2^2 \text{ s.t. } \hat{\alpha}_{(T^t)^c} = 0\right\}$.
\State Shrink support: $T^t = \mathcal{S}_{k,2}(x_p)$.
\State Calculate new representation: $y^t = P_T^* x_p$.
\State Update the residual: $y^t = y - Mx^t$.
\EndWhile
\State Form final solution $\hat{x} = x^t$.
\end{algorithmic}
\end{algorithm}
Under these requirements, the authors prove that the method accurately recovers the \( k \)-sparse signal as in [3].

Although the assumption on the approximate projections is also made for other methods [12, 13], it is unknown whether such methods can be obtained. Recently, Giryes and Needell [14] relaxed these assumptions by introducing the notion of near-optimal projections.

**Definition 1.1** A procedure \( \mathcal{S}_k \) implies a near-optimal projection \( \mathbf{P}_{\mathcal{S}_k} \) with constants \( C_k \) and \( \tilde{C}_k \) if for any \( z \in \mathbb{R}^d \), \(|\mathcal{S}_k(z)| \leq \zeta \), with \( \zeta \geq 1 \), and

\[
\|z - \mathbf{P}_{\mathcal{S}_k}(z)\|_2^2 \leq C_k \|z - \mathbf{P}_{\mathcal{S}_k}(z)\|_2^2 \quad \text{as well as} \quad \|\mathbf{P}_{\mathcal{S}_k}(z)\|_2 \geq \tilde{C}_k \|\mathbf{P}_{\mathcal{S}_k}(z)\|_2,
\]

where \( \mathbf{P}_{\mathcal{S}_k} \) denotes the optimal projection via (7).

They prove that when the dictionary \( \mathbf{D} \) is incoherent or satisfies the RIP, that many standard algorithms in compressive sensing give near-optimal projections satisfying (8). This improves upon previous results since even in this case, it is unknown whether any methods exist that satisfy the stricter requirements of (6). In particular, they prove the following result:

**Theorem 1.2** Let \( \mathbf{M} \) satisfy the D-RIP with a constant \( \delta_{(3\zeta + 1)k} (\zeta \geq 1) \). Suppose that \( \mathcal{S}_{k,1} \) and \( \mathcal{S}_{k,2} \) are near optimal projections (as in Definition 1.1) with constants \( C_k, \tilde{C}_k \) and \( C_{2k}, \tilde{C}_{2k} \) respectively. Apply SSCoSaMP (with \( a = 2 \)) and let \( \mathbf{x}^t \) denote the approximation after \( t \) iterations. If \( \delta_{(3\zeta + 1)k} < \epsilon_{C_k, \tilde{C}_{2k}, \gamma} \) and

\[
(1 + C_k) \left( 1 - \frac{\tilde{C}_{2k}}{(1 + \gamma)^2} \right) < 1,
\]

then after a constant number of iterations \( t^* \) it holds that

\[
\|\mathbf{x}^{t^*} - \mathbf{x}\|_2 \leq \eta_0 \|\mathbf{e}\|_2,
\]

where \( \gamma \) is an arbitrary constant, and \( \eta_0 \) is a constant depending on \( \delta_{(3\zeta + 1)k}, C_k, \tilde{C}_{2k}, \gamma \). The constant \( \epsilon_{C_k, \tilde{C}_{2k}, \gamma} \) is greater than zero if and only if (9) holds.

### 1.1 Our contribution

In this work we extend the results of [14] to provide near-oracle recovery guarantees when the measurement noise \( \mathbf{e} \) is mean-zero Gaussian noise. We focus on the Signal Space CoSaMP method, but analogous results can be obtained for other methods. Our main result is summarized by the following theorem.

**Theorem 1.3** Let \( \mathbf{M} \) satisfy the D-RIP with a constant \( \delta_{(3\zeta + 1)k} (\zeta \geq 1) \) and \( \mathbf{e} \) be white Gaussian noise with variance \( \sigma^2 \). Suppose that \( \mathcal{S}_{k,1} \) and \( \mathcal{S}_{k,2} \) are near optimal projections (as in Definition 1.1) with constants \( C_k, \tilde{C}_k \) and \( C_{2k}, \tilde{C}_{2k} \) respectively. Apply SSCoSaMP (with \( a = 2 \)) and let \( \mathbf{x}^t \) denote the approximation after \( t \) iterations. If \( \delta_{(3\zeta + 1)k} < \epsilon_{C_k, \tilde{C}_{2k}, \gamma} \) and

\[
(1 + C_k) \left( 1 - \frac{\tilde{C}_{2k}}{(1 + \gamma)^2} \right) < 1,
\]

then after a constant number of iterations \( t^* \) it holds with high probability that that

\[
\|\mathbf{x}^{t^*} - \mathbf{x}\|_2 = O(k \log(n)\sigma^2),
\]

where \( \gamma \) is an arbitrary constant, and the constant \( \epsilon_{C_k, \tilde{C}_{2k}, \gamma} \) is greater than zero if and only if (11) holds.
Remarks.

1. This improves upon Theorem 1.2 in general, since \( \|e\|_2 \) is expected to be on the order of \( \sqrt{n} \sigma \) when \( e \) is mean-zero Gaussian noise with variance \( \sigma^2 \). These results align with those of standard compressive sensing when the dictionary \( D \) is the identity [15][16][17].

2. This bound is, up to a constant and a \( \log(n) \) factor, the same as the one we get if we use an oracle that foreknows the true support of the original signal \( x \). The oracle estimator and its error will be defined and calculated hereafter. Note that the \( \log \) factor is inevitable for any practical estimator that does not have access to oracle information [18].

1.2 Organization

We establish some required notation and preliminary lemmas in Section 2. In Section 3 we present the oracle estimator in the signal domain and calculate its recovery error. In Section 4 we present our main results, which imply the near-oracle performance of Theorem 1.3. Our proofs are included in Section 5. We conclude our work in Section 6.

2 Notation and Consequences of D-RIP

As usual, we let \( \| \cdot \|_2 \) denote the Euclidean (\( \ell_2 \)) norm of a vector, and \( \| \cdot \| \) the spectral (\( \ell_2 \rightarrow \ell_2 \)) norm of a matrix. We write the \( d \times d \) identity matrix as \( I_d \). For an index set \( T \), we denote by \( D_T \) the sub-matrix of \( D \) whose columns are indexed by \( T \). \( P_T = D_T D_T^\dagger \) denotes the orthogonal projection onto range\( (D_T) \) and \( Q_T = I_d - P_T \) the orthogonal projection onto its orthogonal complement.

We next recall some elementary consequences of the D-RIP, whose proofs can be found in [14].

Lemma 2.1 If \( M \) satisfies the D-RIP with a constant \( \delta_k \) then
\[
\|MP_T\|_2^2 \leq 1 + \delta_k \quad \text{and} \quad \|P_T(I - M^* M)P_T\| \leq \delta_k
\]
for every \( T \) such that \( |T| \leq k \).

Lemma 2.2 If \( M \) satisfies the D-RIP [5] then
\[
\|P_{T_1}(I - M^* M)P_{T_2}\| \leq \delta_k,
\]
for any \( T_1 \) and \( T_2 \) with \( |T_1| \leq k_1, |T_2| \leq k_2 \), and \( k_1 + k_2 \leq k \).

Lemma 2.3 (Approximate projections) For any vector \( v \in \mathbb{R}^d \) that has a \( k \)-sparse representation and a support set \( T \) such that \( |T| \leq k \), and for any \( z \in \mathbb{R}^d \) we have that
\[
\|z - P_{\mathcal{F}_k(z)}z\|_2^2 \leq C_k \|v - z\|_2^2, \quad \text{and}
\]
\[
\|P_{\mathcal{F}_k(z)}z\|_2^2 \geq \tilde{C}_k \|P_T z\|_2^2.
\]

Finally, an elementary fact that we will also utilize.

Proposition 2.4 For any two given vectors \( x_1, x_2 \) and a constant \( c > 0 \) it holds that
\[
\|x_1 + x_2\|_2^2 \leq (1 + c) \|x_1\|_2^2 + (1 + \frac{1}{c}) \|x_2\|_2^2.
\]
3 The Oracle Estimator in the Signal Domain

Before we proceed to develop our main result for SSCoSaMP, we start by asking what is the error of an estimator that foreknows the support of the original signal \( x \). Let \( T \) be the true support of \( x \), then the oracle estimator is simply

\[
\hat{x}_O = D_T (MD_T)^\dagger y, \tag{17}
\]

i.e., the minimizer of

\[
\min_{\tilde{x}} \| y - M\tilde{x} \|_2^2 \quad s.t. \quad \tilde{x} = D\tilde{\alpha}, \tilde{\alpha}_T = 0. \tag{18}
\]

The oracle's error is given by the following lemma

**Lemma 3.1** Let \( M \) satisfy the D-RIP and \( x \) be a signal with a \( k \)-sparse representation \( \alpha \) under a dictionary \( D \). The oracle estimator's error is

\[
\frac{k\sigma^2}{1 + \delta_k} \leq \mathbb{E} \| x - \hat{x}_O \|_2^2 \leq \frac{k\sigma^2}{1 - \delta_k}. \tag{19}
\]

**Proof:** Since \( x \) is supported on \( T \) we can write it as \( x = D\alpha = D_T \alpha_T \). Plugging (1) in (17) we have

\[
\hat{x}_O = D_T (MD_T)^\dagger (MD_T\alpha_T + e) = x + D_T (MD_T)^\dagger e. \tag{20}
\]

Thus, the oracle's error equals

\[
\mathbb{E} \| x - \hat{x}_O \|_2^2 = \mathbb{E} \| D_T (MD_T)^\dagger e \|_2^2. \tag{21}
\]

Using the D-RIP we have

\[
\frac{1}{1 + \delta_k} \mathbb{E} \| MD_T (MD_T)^\dagger e \|_2^2 \leq \mathbb{E} \| x - \hat{x}_O \|_2^2 \leq \frac{1}{1 - \delta_k} \mathbb{E} \| MD_T (MD_T)^\dagger e \|_2^2 \tag{22}
\]

The proof ends by noticing that \( MD_T (MD_T)^\dagger \) is a projection operator. Therefore, from the properties of white Gaussian noise we have

\[
\mathbb{E} \| MD_T (MD_T)^\dagger e \|_2^2 = \text{trace} \left( MD_T (MD_T)^\dagger \right) \sigma^2 = \text{trace} \left( (MD_T)^\dagger MD_T \right) \sigma^2 = k\sigma^2. \tag{23}
\]

\[\square\]

4 Main Results

Though the oracle's error is promising, it is unattainable as we do not have the support of the original signal. We turn to analyze the SSCoSaMP method which is a feasible algorithm for signal recovery. We provide theoretical guarantees for the recovery performance of SSCoSaMP when the measurement noise is Gaussian. We assume \( a = 2 \) in the algorithm, however, analogous results for other values can be obtained similarly.
4.1 Theorem Conditions

Before we present the proof of the main result, we recall the conditions which guarantee the assumptions of Theorem 1.3. The first requirement, that \( \delta_{2(1+\varepsilon)k} \leq \varepsilon^2 c_k \tilde{C}_{2k} \gamma \) for a constant \( c_k \tilde{C}_{2k} \gamma > 0 \), holds for many families of random matrices when \( m \geq \frac{c_k}{\varepsilon^2} k \log \left( \frac{n}{k \varepsilon} \right) \). The more challenging assumption in the theorem is the condition (11), which requires \( c_k \tilde{C}_{2k} \gamma \), and \( \tilde{C}_{2k} \gamma \) to be close to 1. However, we do have an access to such projection operators in many practical settings, and these are not supported by the guarantees provided in previous results [11, 12, 13, 20]. In fact, when the dictionary \( \mathbf{D} \) is incoherent or satisfies the RIP itself, then simple thresholding or standard compressive sensing algorithms can be used for the projection. See Sec. 4 of [14] for a detailed discussion.

4.2 SCoSaMP Near-Oracle Guarantees

As in [8, 4], our proof utilizes an iteration invariant which guarantees that each iteration exponentially reduces the recovery error, down to the noise floor.

**Theorem 4.1** Let \( \mathbf{M} \) satisfy the D-RIP with constants \( \delta_{(1+\varepsilon)k} \delta_{3\varepsilon k} \delta_{(3\varepsilon+1)k} \) and let \( \mathcal{S}_{k,1} \) and \( \mathcal{S}_{2k,2} \) be near optimal projections as in Definition 1.7 with constants \( c_k, \tilde{C}_k \) and \( C_{2k}, \tilde{C}_{2k} \) respectively. Then

\[
\| \mathbf{x}^t - \mathbf{x} \|_2 \leq \rho \| \mathbf{x} - \mathbf{x}^{t-1} \|_2 + \eta \| \mathbf{P}_{\tilde{T}_e} \mathbf{M}^* \mathbf{e} \|_2,
\]

for constants \( \rho \) and \( \eta \), and where

\[
\tilde{T}_e = \arg\max_{\tilde{T} : |\tilde{T}| \leq 3\kappa} \| \mathbf{P}_{\tilde{T}} \mathbf{M}^* \mathbf{e} \|_2.
\]

The iterates converge, i.e. \( \rho < 1 \), if \( \delta_{(3\varepsilon+1)k} < \varepsilon^2 c_k \tilde{C}_{2k} \gamma \), for some positive constant \( \varepsilon^2 c_k \tilde{C}_{2k} \gamma \), and (11) holds.

An immediate corollary of the above theorem yields the following.

**Theorem 4.2** Let \( \mathbf{M} \) satisfy the D-RIP with constants \( \delta_{(1+\varepsilon)k} \delta_{3\varepsilon k} \delta_{(3\varepsilon+1)k} \) and let \( \mathcal{S}_{k,1} \) and \( \mathcal{S}_{2k,2} \) be near optimal projections as in Definition 1.7 with constants \( c_k, \tilde{C}_k \) and \( C_{2k}, \tilde{C}_{2k} \) respectively. Then after a constant number of iterations \( t^* = \left\lceil \frac{\log(\log(1/\rho))}{\log(1/\rho)} \right\rceil \) it holds that

\[
\| \mathbf{x}^{t^*} - \mathbf{x} \|_2 \leq \left( 1 + \frac{1 - \rho^{t^*}}{1 - \rho} \right) \eta \| \mathbf{P}_{\tilde{T}_e} \mathbf{M}^* \mathbf{e} \|_2,
\]

where \( \eta \) is a constant and \( \tilde{T}_e \) is defined as in (25).

**Proof:** By using (24) and recursion we have that after \( t^* \) iterations

\[
\| \mathbf{x}^{t^*} - \mathbf{x} \|_2 \leq \rho^{t^*} \| \mathbf{x} - \mathbf{x}_0 \|_2 + (1 + \rho + \rho^2 + \ldots + \rho^{t^*-1}) \eta \| \mathbf{P}_{\tilde{T}_e} \mathbf{M}^* \mathbf{e} \|_2
\]

\[
\leq \left( 1 + \frac{1 - \rho^{t^*}}{1 - \rho} \right) \eta \| \mathbf{P}_{\tilde{T}_e} \mathbf{M}^* \mathbf{e} \|_2,
\]

where the last inequality is due to the equation of the geometric series, the choice of \( t^* \), and the fact that \( \mathbf{x}_0 = 0 \). \( \square \)

To prove the near oracle bound we need the following lemma, whose proof is presented in Section 5.
Lemma 4.3 If $e$ is zero-mean white Gaussian noise with variance $\sigma^2$ then with probability exceeding $1 - \frac{2}{(3\zeta k)!}n^{-\beta}$ we have

$$\|P_TM^*e\|_2 \leq \sqrt{(1 + \delta_{3\zeta k})3\zeta k\left(1 + \sqrt{2(1 + \beta)\log(n)}\right)\sigma}. \quad (28)$$

This lemma together with Corollary 4.2 provides the following near-oracle performance theorem.

Theorem 4.4 Assume the conditions of Theorem 4.1. Then after a constant number of iterations $t^* = \left\lceil \frac{\log(\|x\|_2/\|e\|_2)}{\log(1/\rho)} \right\rceil$ it holds with probability exceeding $1 - \frac{2}{(3\zeta k)!}n^{-\beta}$ that

$$\|x^* - x\|_2 \leq \left(1 + \frac{1 - \rho^*}{1 - \rho}\right)\eta \sqrt{(1 + \delta_{3\zeta k})3\zeta k\left(1 + \sqrt{2(1 + \beta)\log(n)}\right)\sigma}. \quad (29)$$

Note that Theorem 4.4 implies our main result, Theorem 1.3. We have thus established that SS-CoSaMP provides near-oracle performance when the noise is mean-zero Gaussian.

5 Proofs

5.1 Proof of Lemma 4.3

We rely on the proof technique of Lemma 3 in [21]. Without loss of generality, we prove for the case of $\sigma = 1$. By simple scaling we get the above result for any value of $\sigma$. Using Lemma 2.1 we have that for any $e_1, e_2 \in \mathbb{R}^d$ and any support $\tilde{T}$, $|\tilde{T}| \leq 3\zeta k$,

$$\|P_{\tilde{T}}M^*(e_1 - e_2)\|_2 \leq \sqrt{1 + \delta_{3\zeta k}\|e_1 - e_2\|_2}. \quad (30)$$

Thus we can say that $\|P_{\tilde{T}}M^*\|_2^2$ is a $\sqrt{1 + \delta_{3\zeta k}}$-Lipschitz functional. Using trace and expectation properties we have

$$\mathbb{E}\|P_{\tilde{T}}M^*e\|_2^2 = \mathbb{E}\left[\text{trace}(e^*MP_{\tilde{T}}P_{\tilde{T}}M^*)\right] = \text{trace}(MP_{\tilde{T}}P_{\tilde{T}}M^*\mathbb{E}[e^*]) = \text{trace}(MP_{\tilde{T}}P_{\tilde{T}}M^*), \quad (31)$$

where the last equality is due to $\mathbb{E}[e^*] = I$. Note that $\text{trace}(MP_{\tilde{T}}P_{\tilde{T}}M^*)$ equals the sum of the singular values of $MP_{\tilde{T}}$. Since $P_{\tilde{T}}$ is a projection to a subspace of dimension $3\zeta k$, there are at most $3\zeta k$ non-zero singular values. By the D-RIP, we thus have that

$$\mathbb{E}\|P_{\tilde{T}}M^*e\|_2^2 \leq (1 + \delta_{3\zeta k})3\zeta k, \quad (32)$$

and from Jensen’s inequality it follows that

$$\mathbb{E}\|P_{\tilde{T}}M^*e\|_2 \leq \sqrt{(1 + \delta_{3\zeta k})3\zeta k}, \quad (33)$$

Using concentration of measure in Gauss space [22, 23] we have

$$\mathbb{P}\left(\|P_{\tilde{T}}M^*(e)\|_2 - \mathbb{E}\|P_{\tilde{T}}M^*e\|_2 \geq t\right) \leq 2\exp\left(-\frac{t^2}{2(1 + \delta_{3\zeta k})}\right). \quad (34)$$
Using (33) we have \( \|P_T M^* e\|_2 - \sqrt{(1 + \delta_{3\zeta k})3\zeta k} \leq \|P_T M^* e\|_2 - \mathbb{E}[\|P_T M^* e\|_2] \) and thus
\[
P\left( \|P_T M^* e\|_2 - \sqrt{(1 + \delta_{3\zeta k})3\zeta k} \geq t \right) \leq \mathbb{P}\left( \|P_T M^* e\|_2 - \mathbb{E}[\|P_T M^* e\|_2] \geq t \right)
\]
Combining (35) and (34) yields
\[
P\left( \|P_T M^* e\|_2 \geq \sqrt{(1 + \delta_{3\zeta k})3\zeta k + t} \right) \leq 2 \exp\left( \frac{-t^2}{2(1 + \delta_{3\zeta k})} \right).
\]
Selecting \( t = \sqrt{(1 + \delta_{3\zeta k})3\zeta k \sqrt{2(1 + \beta) \log(n)}} \) we have \( e^{-\frac{t^2}{2(1 + \delta_{3\zeta k})}} = n^{-3\zeta k(1 + \beta)} \). Using a union bound we have
\[
P\left( \|P_T M^* e\|_2 \geq \sqrt{(1 + \delta_{3\zeta k})3\zeta k \left( 1 + \sqrt{2(1 + \beta) \log(n)} \right)} \right)
\]
Combining (36) and (34) yields
\[
\sum_{\hat{r} : |\hat{r}| = 3\zeta k} P\left( \|P_T M^* e\|_2 \geq \sqrt{(1 + \delta_{3\zeta k})3\zeta k \left( 1 + \sqrt{2(1 + \beta) \log(n)} \right)} \right)
\]
which completes the claim.

5.2 Proof of Theorem 4.1

We turn now to prove the iteration invariant, Theorem 4.1. Instead of presenting the proof directly, we divide the proof into several lemmas. The first lemma gives a bound for \( \|x_p - x\|_2 \) as a function of \( \|P_T M^* e\|_2 \) and \( \|Q_{\hat{r}}(x_p - x)\|_2 \).

Lemma 5.1 If \( M \) has the D-RIP with a constant \( \delta_{3\zeta k} \), then with the notation of Algorithm 7 we have
\[
\|x_p - x\|_2 \leq \frac{1}{\sqrt{1 - \delta_{3\zeta k}^2 (\zeta k + 1)}} \|Q_{\hat{r}}(x_p - x)\|_2 + \frac{1}{1 - \delta_{3\zeta k}^2 (\zeta k + 1)} \|P_T M^* e\|_2 \tag{38}
\]

Proof: Since \( x_p \triangleq D\alpha_p \) is the minimizer of \( \|y - M\tilde{x}\|_2 \) with the constraints \( \tilde{x} = D\tilde{\alpha} \) and \( \tilde{\alpha}_{(\hat{r})} = 0 \), then
\[
\langle Mx_p - y, Mv \rangle = 0 \tag{39}
\]
for any vector \( v = D\tilde{\alpha} \) such that \( \tilde{\alpha}_{(\hat{r})} = 0 \). Substituting \( y = M\tilde{x} + e \) with simple arithmetic gives
\[
\langle x_p - x, M^* Mv \rangle = \langle e, Mv \rangle \tag{40}
\]
where \( v = D\tilde{\alpha} \) and \( \tilde{\alpha}_{(\hat{r})} = 0 \). To bound \( \|P_{\hat{r}}(x_p - x)\|_2^2 \), we use (40) with \( v = P_{\hat{r}}(x_p - x) \), which gives
\[
\|P_{\hat{r}}(x_p - x)\|_2^2 = \langle x_p - x, P_{\hat{r}}(x_p - x) \rangle \tag{41}
\]

where the first inequality follows from the Cauchy-Schwartz inequality, the projection property that $P_{\tilde{\mathcal{T}}_T} = P_{\tilde{\mathcal{T}}_T} P_{\tilde{\mathcal{T}}_T}$, and the fact that $x_p - x = P_{\tilde{\mathcal{T}}_T} (x_p - x)$. The last inequality is due to the $D$-RIP property, the fact that $|\tilde{\mathcal{T}}_T| \leq 3\zeta k$ and Lemma 2.2. After simplification of (41) by $\|P_{\tilde{\mathcal{T}}_T} (x_p - x)\|_2$ we have

$$\|P_{\tilde{\mathcal{T}}_T} (x_p - x)\|_2 \leq \delta_{(3\zeta+1)k} \|x_p - x\|_2 + \|P_{\tilde{\mathcal{T}}_T} M^* e\|_2.$$  

Utilizing the last inequality with the fact that $x_p - x = P_{\tilde{\mathcal{T}}_T} (x_p - x) + (x_p - x) - P_{\tilde{\mathcal{T}}_T} (x_p - x)$ gives

$$\|x_p - x\|_2 \leq \|P_{\tilde{\mathcal{T}}_T} (x_p - x)\|_2 + \delta_{(3\zeta+1)k} \|x_p - x\|_2 + \|P_{\tilde{\mathcal{T}}_T} M^* e\|_2^2.$$ (42)

The last equation is a second order polynomial of $\|x_p - x\|_2$. Thus its larger root is an upper bound for it and together with (25) this gives the inequality in (39). For more details look at the derivation of (13) in [4].

The second lemma bounds $\|x^T - x\|_2$ in terms of $\|P_{\tilde{\mathcal{T}}_T} (x_p - x)\|_2$ and $\|P_{\tilde{\mathcal{T}}_T} M^* e\|_2$ using the first lemma.

**Lemma 5.2** Given that $\mathcal{S}_{k,2}$ is a near support selection scheme with a constant $C_k$, if $M$ has the $D$-RIP with a constant $\delta_{(3\zeta+1)k}$, then

$$\|x^T - x\|_2 \leq \rho_1 \|P_{\tilde{\mathcal{T}}_T} (x_p - x)\|_2 + \eta_1 \|P_{\tilde{\mathcal{T}}_T} M^* e\|_2$$ (43)

**Proof:** Denote $w = x_p$. We start with the following observation

$$\|x - x^T\|_2^2 = \|x - w + w - x^T\|_2^2 = \|x - w\|_2^2 + \|x^T - w\|_2^2 + 2(x - w)^* (w - x^T),$$ (44)

and turn to bound the second and last terms in the right hand side. For the second term, using the fact that $x^T = P_{\mathcal{S}_{k,2}(w)} w$ with (14) gives

$$\|x^T - w\|_2^2 \leq C_k \|x - w\|_2^2.$$ (45)

For bounding the last term, we look at its absolute value and use (10) with $u = w - x^T = P_{\tilde{\mathcal{T}}_T} (w - x^T)$. This leads to

$$|(x - w)^* (w - x^T)| = |(x - w)^* (I - M^* M) (w - x^T) - e^* M (w - x^T)|.$$ 

By using the triangle and Cauchy-Schwartz inequalities with the fact that $x - w = P_{T \cup \tilde{\mathcal{T}}_T} (x - w)$ and $w - x^T = P_{\tilde{\mathcal{T}}_T} (w - x^T)$ we have

$$|(x - w)^* (w - x^T)| \leq \|x - w\|_2 \|P_{T \cup \tilde{\mathcal{T}}_T} (I - M^* M) P_{\tilde{\mathcal{T}}_T}\|_2 \|w - x^T\|_2 + \|P_{\tilde{\mathcal{T}}_T} M^* e\|_2 \|w - x^T\|_2 \leq \delta_{(3\zeta+1)k} \|x - w\|_2 \|w - x^T\|_2 + \|P_{\tilde{\mathcal{T}}_T} M^* e\|_2 \|w - x^T\|_2,$$ (46)

where the last inequality is due to the $D$-RIP definition, the fact that $|\tilde{\mathcal{T}}_T| \leq 3\zeta k$ and Lemma 2.2.
By substituting (45) and (46) into (44) we have

\[ \| x - x' \|^2 \leq (1 + C_k) \| x - w \|^2 + 2\delta(\delta_{(3\xi+1)k})^2 \sqrt{C_k} \| x - w \|^2 + 2\sqrt{C_k} \| P_{\tilde{T}} x - M^* e \|^2 \| x - w \|^2 \]

\[ \leq \left( 1 + 2\delta(\delta_{(3\xi+1)k})^2 \sqrt{C_k} + C_k \right) \| x - w \|^2 + 2\sqrt{C_k} \| P_{\tilde{T}} x - M^* e \|^2 \]  

\[ \leq \left( 1 + 2\delta(\delta_{(3\xi+1)k})^2 \sqrt{C_k} + C_k \right) \left( \frac{1 - \delta(\delta_{(3\xi+1)k})^2}{1 - \delta(\delta_{(3\xi+1)k})^2} \right) \| Q_{\tilde{T}} (x - w) \|^2 + \frac{2 + \sqrt{C_k} + C_k}{1 - \delta(\delta_{(3\xi+1)k})^2} \| Q_{\tilde{T}} (x - w) \|^2 \]

where for the second inequality we use the fact that \( \delta(\delta_{(3\xi+1)k}) \leq 1 \) combined with the inequality of Lemma 5.1 and for the last inequality we use the fact that \( 1 + (1 + \delta_{(3\xi+1)k})^2 \sqrt{C_k} + C_k \leq (1 + 2\delta_{(3\xi+1)k})^2 \sqrt{C_k} + C_k \left( \frac{2 + \sqrt{C_k} + C_k}{1 + \sqrt{C_k} + C_k} \right) \) together with a few algebraic steps. Taking square-root on both sides of (47) and using (25) provide the desired result.

The last lemma bounds \( \| Q_{\tilde{T}} (x - x_p) \| \) with \( \| x^{t-1} - x \| \) and \( \| P_{\tilde{T}} x - M^* e \| \).

**Lemma 5.3** Given that \( S_{2\xi+1,1} \) is a near optimal support selection scheme with a constant \( C_{2\xi} \), if \( M \) has the D-RIP with constants \( \delta(\delta_{(3\xi+1)k})^2 \) and \( \delta_{2\xi} \), then

\[ \| Q_{\tilde{T}} (x - x) \| \leq \eta_2 \| P_{\tilde{T}} x - M^* e \| + \rho_2 \| x - x^{t-1} \|. \]

**Proof:** Looking at the step of finding new support elements one can observe that \( P_{\Lambda_\xi} \) is a near optimal projection operator for \( M^* y' = M^* (y - Mx^{t-1}) \). Noticing that \( Q_{\tilde{T}} \subseteq \tilde{T} \) and then using (15) with \( P_{T^{t-1} \cup T} \) gives

\[ \| P_{\tilde{T}} x - M^* (y - Mx^{t-1}) \| \leq \| P_{\tilde{T}} x - M^* (y - Mx^{t-1}) \|^2 \leq C_{2\xi} \| P_{T^{t-1} \cup T} M^* (y - Mx^{t-1}) \|^2 \]  

We start by bounding the left hand side of (49) from above. Using Proposition 2.2 with \( \gamma > 0 \) and \( a > 0 \) we have

\[ \| P_{\tilde{T}} x - M^* (y - Mx^{t-1}) \|^2 \leq \left( 1 + \frac{1}{\gamma} \right) \| P_{\tilde{T}} x - M^* e \|^2 + (1 + \gamma_1) \| P_{\tilde{T}} x - M^* M (y - Mx^{t-1}) \|^2 \]

\[ \leq \left( 1 + \frac{1 + \gamma_1}{\gamma} \right) \| P_{\tilde{T}} x - M^* e \|^2 + (1 + \alpha + \gamma_1) \| P_{\tilde{T}} x - M^* M (y - Mx^{t-1}) \|^2 \]

where the last inequality is due to Lemma 2.1 and (14).
We continue with bounding the right hand side of (49) from below. For the first element we use Proposition 2.4 with constants \( \gamma_2 > 0 \) and \( \beta > 0 \), and (14) to achieve

\[
\|P_{T^t \cup T}M^*(y - Mx^{t-1})\|_2^2 \geq \frac{1}{\gamma_2} \|P_{T^t \cup T}M^*(x - x^t)\|_2^2 - \frac{1}{\gamma_2} \|P_{T^t \cup T}M^*e\|_2^2
\]

(51)

\[
\geq \frac{1}{1 + \gamma_2} \frac{1}{1 + \gamma_2} \|x - x^{t-1}\|_2^2 - \frac{1}{\gamma_2} \|P_{T^t \cup T}M^*e\|_2^2
\]

\[
- \frac{1}{\beta} \frac{1}{1 + \gamma_2} \|P_{T^t \cup T}(M^*M - I_d)(x - x^{t-1})\|_2^2
\]

\[
\geq \frac{1}{1 + \beta} - \frac{\delta(\xi+1)k}{\beta} \frac{1}{1 + \gamma_2} \|x - x^{t-1}\|_2^2.
\]

By combining (50) and (51) with (49) and then using (25) we have

\[
(1 + \alpha)(1 + \gamma_1) \|Q_{\tilde{T}}(x - x^{t-1})\|_2^2 \leq \frac{1}{\gamma_1(1 + \alpha)} \|P_{\tilde{T}}M^*e\|_2^2 + \tilde{C}_2k \frac{1}{\gamma_2} \|P_{\tilde{T}}M^*e\|_2^2
\]

\[
+ \left(1 + \alpha + \delta(\xi+1)k + \frac{\delta(\xi+1)k}{\alpha}\right)(1 + \gamma_1) \|x - x^{t-1}\|_2^2
\]

\[
- \tilde{C}_k(1 + \gamma_1) \frac{1}{1 + \beta} \|x - x^{t-1}\|_2^2.
\]

Division of both sides by \((1 + \alpha)(1 + \gamma_1)\) yields

\[
\|Q_{\tilde{T}}(x - x^{t-1})\|_2^2 \leq \left(\frac{1}{\gamma_1(1 + \alpha)} + \frac{\tilde{C}_2k}{\gamma_2(1 + \alpha)(1 + \gamma_1)}\right) \|P_{\tilde{T}}M^*e\|_2^2
\]

\[
+ \left(1 + \frac{\delta(\xi+1)k}{\alpha} - \frac{\tilde{C}_2k}{(1 + \alpha)(1 + \gamma_1)(1 + \gamma_2)} \left(1 + \frac{\delta(\xi+1)k}{\beta}\right)\right) \|x - x^{t-1}\|_2^2.
\]

Substituting \( \beta = \frac{\sqrt{\delta(\xi+1)k}}{1 - \sqrt{\delta(\xi+1)k}} \) gives

\[
\|Q_{\tilde{T}}(x - x^{t-1})\|_2^2 \leq \left(\frac{1}{\gamma_1(1 + \alpha)} + \frac{\tilde{C}_2k}{\gamma_2(1 + \alpha)(1 + \gamma_1)}\right) \|P_{\tilde{T}}M^*e\|_2^2
\]

\[
+ \left(1 + \frac{\delta(\xi+1)k}{\alpha} - \frac{\tilde{C}_2k}{(1 + \alpha)(1 + \gamma_1)(1 + \gamma_2)} \left(1 + \sqrt{\delta(\xi+1)k}\right)\right) \|x - x^{t-1}\|_2^2,
\]

Using \( \alpha = \frac{\sqrt{\delta(\xi+1)k}}{1 - \sqrt{\delta(\xi+1)k}} \) yields

\[
\|Q_{\tilde{T}}(x - x^{t-1})\|_2^2 \leq \left(\frac{1}{\gamma_1(1 + \alpha)} + \frac{\tilde{C}_2k}{\gamma_2(1 + \alpha)(1 + \gamma_1)}\right) \|P_{\tilde{T}}M^*e\|_2^2
\]

\[
+ \left(-\sqrt{\delta(\xi+1)k} - \frac{\tilde{C}_2k}{(1 + \gamma_1)(1 + \gamma_2)} \left(1 - \sqrt{\delta(\xi+1)k}\right)\right) \|x - x^{t-1}\|_2^2,
\]

The values of \( \gamma_1, \gamma_2 \) give a tradeoff between the convergence rate and the size of the noise coefficient. For smaller values we get better convergence rate but higher amplification of the noise. We make no
optimization on them and choose them to be \( \gamma_1 = \gamma_2 = \gamma \) where \( \gamma \) is an arbitrary number greater than 0. Thus we have

\[
\|Q_{\tilde{T}}(x - x^{t-1})\|_2^2 \leq \left( \frac{1}{\gamma(1 + \alpha)} + \frac{\tilde{C}_{2k}}{\gamma(1 + \alpha)(1 + \gamma)} \right) \|P_{T^*}M^* e\|_2^2 + \left( - \frac{\sqrt{\delta_{(3\xi+1)k}} - \frac{\sqrt{C_{2k}}}{1 + \gamma} \left( 1 - \sqrt{\delta_{(3\xi+1)k}} \right)^2 + 1 \right) \|x - x^{t-1}\|_2^2.
\]

Using the triangle inequality and the fact that \( Q_{\tilde{T}}x_p = Q_{\tilde{T}}x^{t-1} = 0 \) gives the desired result. \( \square \)

With the aid of the above three lemmas we turn to the proof of the iteration invariant, Theorem 4.1.

**Proof of Theorem 4.1**: Substituting the inequality of Lemma 5.3 into the inequality of Lemma 5.2 gives (24) with \( \rho = \rho_1 \rho_2 \) and \( \eta = \eta_1 + \rho_1 \eta_2 \). The iterates converge if \( \rho_1^2 \rho_2^2 < 1 \). Since \( \delta_{(3\xi+1)k} \leq \delta_{3k} \leq \delta_{(3\xi+1)k} \) this holds if

\[
\left( 1 + 2\delta_{(3\xi+1)k} \sqrt{C_k} + C_k \right) \left( \frac{1}{1 - \delta_{(3\xi+1)k}^2} \right) \left( 1 - \left( \frac{\sqrt{C_{2k}}}{1 + \gamma} + 1 \right) \sqrt{\delta_{(3\xi+1)k}} \right)^2 < 1.
\]

Since \( \delta_{(3\xi+1)k} < 1 \), we have \( \delta_{(3\xi+1)k} < \delta_{(3\xi+1)k} < \sqrt{\delta_{(3\xi+1)k}} \). Using this fact and expanding (57) yields the stricter condition

\[
\left( (1 + C_k) \left( 1 - \left( \frac{\sqrt{C_{2k}}}{1 + \gamma} \right)^2 \right) - 1 \right) + 2(1 + C_l) \left( \frac{\sqrt{C_{2k}}}{1 + \gamma} + 1 \right) \sqrt{\delta_{(3\xi+1)k}}
\]

\[
+ \left( 2 \sqrt{C_k} \left( 1 - \left( \frac{\sqrt{C_{2l}}}{1 + \gamma} \right)^2 \right) \right) - (1 + C_l) \left( 1 + \frac{\sqrt{C_{2k}}}{1 + \gamma} \right)^2 + 4 \sqrt{C_k} \left( \frac{\sqrt{C_{2k}}}{1 + \gamma} + 1 \right) + 2 \delta_{(3\xi+1)k} < 0.
\]

The above equation has a positive solution if and only if (11) holds. Denoting its positive solution by \( \epsilon_{\xi+1}, \tilde{C}_{2k}, \gamma \) we have that the expression holds when \( \delta_{(3\xi+1)k} \leq \epsilon_{C_{\xi+1}} \), which completes the proof. Note that in the proof we have

\[
\eta_1 = \frac{\sqrt{\frac{1 + C_k}{1 + \gamma} + 2 \sqrt{C_k} + C_k}}{1 - \delta_{(3\xi+1)k}}, \quad \eta_2 = \frac{1 + \delta_{3k}}{\gamma(1 + \alpha)} \frac{(1 + \delta_{(3\xi+1)k}) \tilde{C}_{2k}}{\gamma(1 + \alpha)(1 + \gamma)} - \frac{1}{\gamma(1 + \alpha)} + \frac{\tilde{C}_{2k}}{\gamma(1 + \alpha)(1 + \gamma)}
\]

\[
\rho_1^2 = \left( \frac{1}{\gamma(1 + \alpha)} + \frac{\tilde{C}_{2k}}{\gamma(1 + \alpha)(1 + \gamma)} \right), \quad \rho_2^2 = 1 - \left( \sqrt{\frac{\delta_{(3\xi+1)k}}{1 + \gamma}} \left( 1 - \sqrt{\delta_{(3\xi+1)k}} \right)^2 \right),
\]

\[
\alpha = \frac{\tilde{C}_{2k}}{\sqrt{\frac{1 + \gamma}{1 + \gamma}} (1 - \sqrt{\delta_{(3\xi+1)k}})^2 - \delta_{(3\xi+1)k}} = \frac{\tilde{C}_{2k}}{\sqrt{\frac{1 + \gamma}{1 + \gamma}} (1 - \sqrt{\delta_{(3\xi+1)k}})^2 - \delta_{(3\xi+1)k}}
\]

and \( \gamma > 0 \) is an arbitrary constant. \( \square \)
6 Conclusion

The Signal Space CoSaMP method was studied in the case of arbitrary noise \cite{24, 11, 14} under the assumptions of the D-RIP and approximate projections. As in \cite{14}, the assumptions in this work on the approximate projections allow for standard compressed sensing algorithms to be used when the dictionary D satisfies properties like the RIP or incoherence. In this correspondence, we have presented performance guarantees for this method in the presence of white Gaussian noise, which are comparable to those obtained from an oracle which provides the support of the signal. Our bounds are also of the same order as those for standard greedy algorithms like IHT and CoSaMP \cite{17}, but ours hold also for signals sparse with respect to an arbitrary dictionary.

Acknowledgment

R. Giryes is grateful to the Azrieli Foundation for the award of an Azrieli Fellowship. D. Needell was partially supported by the Simons Foundation Collaboration grant.

References

[1] E. Candès, T. Tao, Decoding by linear programming, IEEE Trans. Inf. Theory 51 (12) (2005) 4203 – 4215.

[2] E. J. Candès, J. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Communications on Pure and Applied Mathematics 59 (8) (2006) 1207–1223.

[3] E. J. Candès, T. Tao, Near-optimal signal recovery from random projections: Universal encoding strategies?, IEEE Trans. Inf. Theory. 52 (12) (2006) 5406 –5425.

[4] S. Foucart, Sparse recovery algorithms: sufficient conditions in terms of restricted isometry constants, in: Approximation Theory XIII, Springer Proceedings in Mathematics, 2010, pp. 65–77.

[5] H. Rauhut, K. Schnass, P. Vandergheynst, Compressed sensing and redundant dictionaries, IEEE Trans. Inf. Theory. 54 (5) (2008) 2210 –2219.

[6] M. Elad, P. Milanfar, R. Rubinstein, Analysis versus synthesis in signal priors, Inverse Problems 23 (3) (2007) 947–968.

[7] E. J. Candès, Y. C. Eldar, D. Needell, P. Randall, Compressed sensing with coherent and redundant dictionaries, Appl. Comput. Harmon. Anal 31 (1) (2011) 59 – 73.

[8] D. Needell, J. Tropp, CoSaMP: Iterative signal recovery from incomplete and inaccurate samples, Appl. Comput. Harmon. Anal 26 (3) (2009) 301 –321.

[9] T. Blumensath, M. Davies, Iterative hard thresholding for compressed sensing, Appl. Comput. Harmon. Anal 27 (3) (2009) 265 –274.

[10] S. Foucart, Hard thresholding pursuit: an algorithm for compressive sensing, SIAM J. Numer. Anal. 49 (6) (2011) 2543–2563.
[11] M. Davenport, D. Needell, M. Wakin, Signal space cosamp for sparse recovery with redundant diction- 
aries, IEEE Trans. Inf. Theory. 59 (10) (2013) 6820–6829.
[12] T. Blumensath, Sampling and reconstructing signals from a union of linear subspaces, IEEE Trans. 
Inf. Theory. 57 (7) (2011) 4660–4671.
[13] R. Giryes, S. Nam, M. Elad, R. Gribonval, M. Davies, Greedy-like algorithms for the cosparse analysis 
model, Linear Algebra and its Applications 441 (0) (2014) 22 – 60, special Issue on Sparse Approximate 
Solution of Linear Systems.
[14] R. Giryes, D. Needell, Greedy signal space methods for incoherence and beyond, submitted (2013).
[15] E. Candès, T. Tao, The Dantzig selector: Statistical estimation when p is much larger than n, Annals 
Of Statistics 35 (2007) 2313.
[16] P. Bickel, Y. Ritov, A. Tsybakov, Simultaneous analysis of lasso and dantzig selector, Annals of Statis-
tics 37 (4) (2009) 1705–1732.
[17] R. Giryes, M. Elad, RIP-based near-oracle performance guarantees for SP, CoSaMP, and IHT, IEEE 
Trans. Signal Process. 60 (3) (2012) 1465–1468.
[18] E. Candès, Modern statistical estimation via oracle inequalities, Acta Numerica 15 (2006) 257–325.
[19] S. Mendelson, A. Pajor, N. Tomczak-Jaegermann, Uniform uncertainty principle for bernoulli and 
subgaussian ensembles, Constructive Approximation 28 (2008) 277–289.
[20] R. Giryes, M. Elad, Iterative hard thresholding for signal recovery using near optimal projections, in: 
Proceedings of the 10th Int. Conf. on Sampling Theory Appl. (SAMPTA), 2013, pp. 212–215.
[21] D. L. Donoho, I. M. Johnstone, Ideal denoising in an orthonormal basis chosen from a library of 
bases, Comptes Rendus Acad. Sci., Ser. I 319 (1994) 1317–1322.
[22] G. Pisier, Probabilistic methods in the geometry of banach spaces, in: G. Letta, M. Pratelli (Eds.), 
Probability and Analysis, Vol. 1206 of Lecture Notes in Mathematics, Springer Berlin Heidelberg, 
1986, pp. 167–241.
[23] V. D. Milman, G. Schechtman, Asymptotic Theory of Finite Dimensional Normed Spaces, Springer-
Verlag New York, Inc., New York, NY, USA, 1986.
[24] M. Davenport, M. Wakin, Compressive sensing of analog signals using discrete prolate spheroidal 
sequences, Appl. Comput. Harmon. Anal. 33 (3) (2012) 438–472.