KTH Electrical Engineering

An Achievable Measurement Rate-MSE Tradeoff in Compressive Sensing Through Partial Support Recovery

RICARDO BLASCO-SERRANO, DAVE ZACHARIAH, DENNIS SUNDMAN,
RAGNAR THOBABEN, AND MIKAEL SKOGLUND

Stockholm 2013

Communication Theory Department
School of Electrical Engineering
KTH Royal Institute of Technology

AN ACHIEVABLE MEASUREMENT RATE-MSE TRADEOFF IN COMPRESSIVE SENSING THROUGH PARTIAL SUPPORT RECOVERY

Ricardo Blasco-Serrano, Dave Zachariah, Dennis Sundman, Ragnar Thobaben, and Mikael Skoglund

School of Electrical Engineering and ACCESS Linnaeus Centre
KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden
rbs@kth.se, denniss@kth.se, davez@kth.se, ragnart@kth.se, skoglund@kth.se

ABSTRACT

For compressive sensing, we derive achievable performance guarantees for recovering partial support sets of sparse vectors. The guarantees are determined in terms of the fraction of signal power to be detected and the measurement rate, defined as a relation between the dimensions of the measurement matrix. Based on this result we derive a tradeoff between the measurement rate and the mean square error, and illustrate it by a numerical example.

Index Terms— Compressive sensing, sparse signal, support recovery, MSE, performance tradeoff.

1. INTRODUCTION

Sparse signal recovery through compressive sensing is a growing field within signal processing with a wide range of applications [1, 2, 3, 4, 5, 6]. A sparse signal can be described as a vector with a large number of zero components. The 'support set' of the signal denotes the unknown set of indices of its nonzero components. This set is a central component for inference of sparse signals from an underdetermined relation of linear measurements in noise.

There exist tradeoffs between the dimensions of the sparse signal vector and the measurement vector for recovering the support set with a given sparsity level [1, 4]. The asymptotic tradeoffs for exact support set recovery in a noisy setting were studied in [7, 8, 9]. Further, asymptotically achievable Cramér-Rao bounds on the mean-square estimation error (MSE) were given in [10, 11, 12].

In this paper, we adopt the approach of [7, 8, 9] and derive achievable performance guarantees for partial support set recovery. We use the result to derive an achievable tradeoff between the mean square error and the measurement rate, defined as a relation between the dimensions of the measurement matrix. The tradeoff is illustrated by a numerical example, showing a significant potential reduction of the measurement rate at minimal increase of MSE.

Notation: Upper-case letters denote random variables or vectors and lower-case denote their realizations, e.g. $x \sim X$. The statistical expectation is denoted by $\mathbb{E}\{\cdot\}$. Vectors are represented with bold face letters $\boldsymbol{x}$. The ith entry of $\boldsymbol{x}$ is denoted by x_i. The operators $\|\cdot\|$ and $\mathrm{tr}\{\cdot\}$ denote the Frobenius norm of a vector/matrix and the trace of a square matrix, respectively. $\boldsymbol{x} \in \mathbb{R}^n$ is k-sparse if only $k \ll n$ of its entries are non-zero. Here, sets are collections of unique objects and are denoted using calligraphic letters, e.g. $\mathcal{S}$ or $\mathcal{S}$. Given a vector $\boldsymbol{x}$ and a set $\mathcal{S} = \{s_1, \ldots, s_{|\mathcal{S}|}\}$, $\boldsymbol{x}_{\mathcal{S}}$ is the subvector $(x_{s_1}, \ldots, x_{s_{|\mathcal{S}|}})$. $\mathcal{O}(\cdot)$ denotes the standard big-O notation.

2. PROBLEM FORMULATION

Let $\boldsymbol{X} \in \mathbb{R}^n$ be a k-sparse random vector and let $\boldsymbol{w} \in \mathbb{R}^k$ be a deterministic but unknown vector with the non-zero entries of $\boldsymbol{X}$ sorted in decreasing order of magnitude. The positions of these entries are the only source of randomness of $\boldsymbol{X}$ and are selected as follows. Let $\mathcal{S} = \{S_1, \ldots, S_k\}$ be chosen uniformly at random over all size-k subsets of $\{1, \ldots, n\}$, then

$$X_i = \begin{cases} w_j & \text{if } i = S_j, \\ 0 & \text{if } i \notin \{S_1, \ldots, S_k\} \end{cases}$$

for $i = \{1, \ldots, n\}$. Clearly, the size of the support set of $\boldsymbol{X}$ equals k for all possible $\mathcal{S}$. Consider the length-m vector of real-valued measurements

$$\boldsymbol{Y} = \boldsymbol{\phi} \boldsymbol{X} + \boldsymbol{Z}$$

where $\boldsymbol{\phi} \in \mathbb{R}^{m \times n}$ is a measurement matrix with average power $P_\phi = \frac{1}{nm} \|\boldsymbol{\phi}\|^2$ and $\boldsymbol{Z} \in \mathbb{R}^m$ is a noise vector with each of its entries independently and identically distributed (i.i.d.) according a Gaussian distribution $\mathcal{N}(0, P_z)$.

We consider the problem of estimating $\boldsymbol{X}$ for fixed k and varying m and n. In particular, we study the number of measurements m that suffices asymptotically to ensure estimation of $\boldsymbol{X}$ with a certain MSE as the size n of $\boldsymbol{X}$ increases. Our approach is to divide the problem into two parts: first, partial support recovery and then, signal estimation.

The first part consists of determining a relationship between the number of measurements m and the length n of

the vector X such that it is possible to recover a part of its support set that encompasses at least a fraction γ of the total power. A formal statement of the first part of the problem is the following.

For any $\gamma \in (0, 1]$, a γ-support set of X identifies a (possibly non-unique) smallest subset of the entries of X that contain at least a fraction γ of the power of X. Let ℓ be the size of the γ-support set. Note that ℓ depends on both γ and w but is, by definition, equal for all γ-support sets of X. Given the vector of measurements Y, the γ-support recovery map

$$d_\gamma : \mathbb{R}^m \mapsto \{1, \ldots, n\}^{\hat{\ell}}$$

produces an estimate $\hat{S}_\gamma$ of a γ-support set of X. The size $\hat{\ell}$ of $\hat{S}_\gamma$ is itself an estimate of ℓ. Let S_γ denote the set of *all* γ-support sets of X. For a given w and measurement matrix ϕ we define the average error probability as

$$P_e(w, \phi, \gamma) \triangleq \Pr(d_\gamma(Y) \notin S_\gamma).$$

The average is over S (i.e. the positions of the non-zero entries of X) and the noise Z. We want to determine a relationship between the number of measurements m and the length n of the vector X such that it is possible to recover a γ-support set with arbitrarily low average error probability.

In the second part of our study, we quantify the achievable MSE given that the support recovery map produces a correct estimate of a γ-support set. Namely, for a given realization x of X and conditioned on $\hat{S}_\gamma \in S_\gamma$, we establish an achievable MSE performance in estimating x.

3. MAIN RESULTS

Consider w, X, and Y as introduced in Section 2. Let $P_\Phi > 0$ be the largest allowed measurement matrix average power.

Proposition 1. *If the number of measurements m grows with the length n of the vector X so that*

$$\lim_{n \to \infty} \frac{m}{\log n} > R^\star(w, \gamma) \tag{1}$$

where

$$R^\star(w, \gamma) \triangleq \left(\min_{i \in \{1, \ldots, \ell\}} \frac{1}{2i} \log \left(\frac{P_\Phi \sum_{j=\ell-i+1}^{k} w_j^2 + P_z}{(1 - \gamma) \|w\|^2 P_\Phi + P_z} \right) \right)^{-1},$$

then there exists a sequence of measurement matrices $\phi^{(n)}$ with $P_{\phi^{(n)}} \leq P_\Phi$ and support recovery maps that detect a γ-support set S_γ with arbitrarily low average error probability.

Thus, to detect a γ-support set reliably it suffices to let the number of measurements m grow with n so that $\frac{m}{\log n} > R^\star(w, \gamma)$. Therefore we will refer to the ratio $R \triangleq \frac{m}{\log n}$ as the measurement rate.

We make the following two observations. First, note that some choices of γ are better than other ones. For example, let $w \in \mathbb{R}^2$ with $w_1^2 = 0.7, w_2^2 = 0.3$. For both $\gamma_1 = 0.4$ and $\gamma_2 = 0.6$ the γ_i-support set is just the position of w_1 in X. However, $R^\star(w, \gamma_1) > R^\star(w, \gamma_2)$. In fact it is easy to show that for a given ℓ, $R^\star(w, \gamma)$ is maximized by choosing γ to be equal to the fraction of the power of the ℓ largest entries of w.

Second, it is sometimes simpler to detect larger γ-support sets. For example, let $w \in \mathbb{R}^3$ with $w_1^2 = w_2^2 = 0.45, w_3^2 = 0.1$. Let $\gamma_1 = 0.4$ and $\gamma_2 = 0.8$. The size of the γ_1 and γ_2-support sets are 1 and 2, respectively. However, $R^\star(w, \gamma_1) > R^\star(w, \gamma_2)$. Thus, the choice of γ should be influenced by our prior knowledge of w, if any.

Now, let x be a realization of X and let $\hat{S}_\gamma$ be an estimate of a γ-support set of x. If $\hat{S}_\gamma$ is a correct estimate of a γ-support set, then we can estimate x with MSE that only depends on x through its non-zero entries, i.e. through w, and $\hat{S}_\gamma$. The MSE is characterized as follows.

Proposition 2. *Conditioned on $\hat{S}_\gamma \in S_\gamma$, it is possible to estimate x with MSE given by*

$$\mathtt{mse}^\star(w, \hat{S}_\gamma) = \|w_{\hat{S}_\gamma^c}\|^2 + \mathcal{O}(1/m) \tag{2}$$

where $w_{\hat{S}_\gamma^c}$ is the subvector of w that contains the non-zero entries of x not included in $\hat{S}_\gamma$.

Consider the pair $\left(R^\star(w, \gamma), \mathtt{mse}^\star(w, \hat{S}_\gamma) \right)$. The concatenation of the two propositions implies that it is possible to estimate x with MSE arbitrarily close to $\mathtt{mse}^\star(w, \hat{S}_\gamma)$ as long as the measurement rate is above $R^\star(w, \gamma)$. We emphasize that the MSE is an average performance characterization; that is, there is no guarantee that the estimation error for a particular realization of Y will be below the given MSE value.

In Fig. 1 we show a typical example of pairs $(R^\star, \mathtt{mse}^\star)$. This corresponds to a random realization of w with $k = 10$ and i.i.d. $w_j \sim \mathcal{N}(0, 1/\sqrt{k})$. The MSE is normalized by $\|w\|^2$ so that the values range from 0 to 1. The solid line represents the boundary of the region of pairs $(R^\star, \mathtt{mse}^\star)$ achievable by combining Propositions 1 and 2. All pairs above this curve are asymptotically achievable by selecting γ appropriately. However, in practice one usually does not have any knowledge of the structure of w and thus γ needs to be chosen arbitrarily. To illustrate the performance in this case, we have included the $(R^\star, \mathtt{mse}^\star)$ pair for several arbitrary choices of γ.

The figure also shows that it is often possible to reduce drastically the measurement rate at a very small loss in terms of MSE. For example, a reduction of the measurement rate from $R \approx 38$ (corresponding to perfect recovery [9], i.e. $\gamma = 1$) to $R \approx 7$ only incurs in a relative MSE of 0.0028 if γ is chosen carefully. Even a blind choice of $\gamma = 0.99$ yields a reduction to $R \approx 12$ for the same increase in relative MSE.
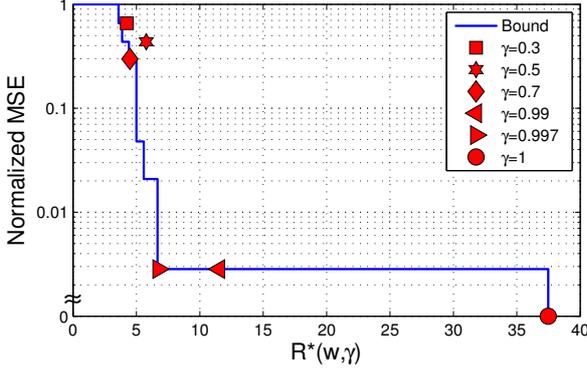
Fig. 1. Measurement rate vs. normalized MSE: all pairs $(R, \mathtt{mse})$ above the solid line are asymptotically achievable.

4. PROOFS

In this section we provide the proofs to Propositions 1 and 2.

4.1. Partial Support Recovery

Proposition 1 is based on random coding arguments and follows the lines of [9, Theorem 1]. However, as opposed to [9], we do not assume any knowledge on the size of the support set and we only detect part of it. We circumvent the first difference by applying the support recovery map for increasingly larger support set sizes until we obtain an estimate of a γ-support set. To circumvent the second difference we define the recovery threshold based on γ. This adds a difficulty due to the non-uniqueness of the γ-support for some values of γ.

Proof of Proposition 1. We sketch only the basic differences to [9, Theorem 1]. Let $\gamma \in (0, 1]$ and fix $\epsilon > 0$ and $\zeta > 0$. Consider the expectation

$$\Pr(\mathcal{E}) \triangleq \mathbb{E}_{\boldsymbol{\Phi}} \left\{ P_e(\boldsymbol{w}, \boldsymbol{\Phi}^{(n)}, \gamma) \right\}$$

taken over the random ensemble of measurement matrices $\boldsymbol{\phi}^{(n)} \sim \boldsymbol{\Phi}^{(n)}$ with i.i.d. Gaussian entries $\phi_{ij}^{(n)} \sim \mathcal{N}(0, \tilde{P}_\Phi)$ with $\tilde{P}_\Phi = P_\Phi - \epsilon$, and using the following variation of the support recovery map described in [9]. Given the vector of measurements $\boldsymbol{Y}$:

1.- Form an estimate of $\|\boldsymbol{w}\|$ (note that $\|\boldsymbol{w}\| = \|\boldsymbol{X}\|$) as

$$\hat{W} = \sqrt{\frac{\left| \frac{1}{m} \|\boldsymbol{Y}\|^2 - P_z \right|}{\tilde{P}_\Phi}}.$$

2.- For $l = 1, \ldots, n$, in increasing order:

(a) Consider the (non-unique) sets of points in $\mathcal{B}_l(\hat{W})$ (l-dimensional hypersphere of radius $\hat{W}$) such that l-dimensional hyperspheres of radius $\frac{\zeta}{2}$ centered on the points cover the whole hypersphere $\mathcal{B}_l(\hat{W})$. Let $\mathcal{Q}_l(\hat{W}, \zeta)$ be one such set that has the smallest number of points.

(b) Find a set $\mathcal{T} = \{t_1, \ldots, t_l\} \subseteq \{1, \ldots, n\}$ such that

$$\frac{1}{m} \left\| \boldsymbol{Y} - \sum_{i=1}^{l} \hat{W}_i \boldsymbol{\Phi}_{t_i}^{(n)} \right\|^2 \leq (1 - \gamma) \hat{W}^2 \tilde{P}_\Phi + \epsilon^2 \tilde{P}_\Phi + P_z \tag{3}$$

for some $\hat{\boldsymbol{W}} = [\hat{W}_1, \ldots, \hat{W}_l]^T \in \mathcal{Q}_l(\hat{W}, \zeta)$, where $\boldsymbol{\Phi}_{t_i}^{(n)}$ is the column of $\boldsymbol{\Phi}^{(n)}$ in position t_i. The process stops when the first set that satisfies (3) is found. This set is the desired estimate.

We now show that this random choice of measurement matrices and support recovery map has $\Pr(\mathcal{E}) \to 0$ as $m \to \infty$ if (1) is satisfied. To see this, consider the event

$$\mathcal{E}_\mathcal{T} \triangleq \left\{ \exists \hat{\boldsymbol{W}} \in \mathcal{Q}_l(\hat{W}, \zeta) \text{ s.t. (3) holds} \right\}$$

given a set $\mathcal{T}$. Let $\mathcal{E}_\mathcal{T}^c$ be the complement of $\mathcal{E}_\mathcal{T}$. We have that

$$\Pr(\mathcal{E}) \leq \sum_{i=1}^{\ell} \Pr \left(\bigcup_{\substack{\mathcal{T}:|\mathcal{T}|=i \\ \mathcal{T} \notin \mathcal{S}_\gamma}} \mathcal{E}_\mathcal{T} \right) + \Pr \left(\bigcap_{\mathcal{T} \in \mathcal{S}_\gamma} \mathcal{E}_\mathcal{T}^c \right)$$

The first sum upper bounds the probability that any set that is not a γ-support set satisfies (3). The second term upper bounds the probability that none of the γ-support sets satisfies (3). Following similar steps as in [9] we can show that both terms tend to 0 with increasing n if (1) is satisfied. A consequence of this is that there must exist a sequence $\phi^{(n)}$ of deterministic measurement matrices with $P_e(\boldsymbol{w}, \phi^{(n)}, \gamma) \to 0$ under the same conditions, as we wanted to prove. $\square$

4.2. MSE Performance

We now study the performance in terms of the MSE of an estimator that uses an estimate of a γ-support set. Our analysis considers the MSE averaged over the random choice of measurement matrices introduced in Section 4.1.

Proof of Proposition 2. Let $\boldsymbol{x}$ be a realization of $\boldsymbol{X}$ and let $\hat{\mathcal{S}}_\gamma$ be the output of the support recovery map. In addition, let $\hat{\mathcal{S}}_\gamma^c$ be the undetected part of the support set of $\boldsymbol{x}$. We start by introducing the event

$$\mathcal{E}_{id} \triangleq \left\{ \left\| \frac{1}{m} \boldsymbol{\Phi}_{\hat{\mathcal{S}}_\gamma}^T \boldsymbol{\Phi}_{\hat{\mathcal{S}}_\gamma} - P_\Phi \boldsymbol{I}_\ell \right\| > \delta \right\}$$

defined for arbitrary $\delta > 0$. Note that, for any such δ, by the vector Chebyshev inequality we have $\Pr(\mathcal{E}_{id}) \leq \mathcal{O}(1/m)$.

Given the output $\hat{\mathcal{S}}_\gamma$ of the support recovery map, we construct the following estimate $\hat{\boldsymbol{X}}$ of $\boldsymbol{x}$. If the event $\mathcal{E}_{id}$ happens then $\hat{\boldsymbol{X}} = \boldsymbol{0}$. Otherwise set

$$\hat{X}_i = \begin{cases} \hat{X}_{\hat{\mathcal{S}}_\gamma, i} & \text{for } i \in \hat{\mathcal{S}}_\gamma \\ 0 & \text{for } i \notin \hat{\mathcal{S}}_\gamma. \end{cases}$$

for $i \in \{1, \ldots, n\}$ and some estimator $\hat{\boldsymbol{X}}_{\hat{\mathcal{S}}_\gamma}$.

Conditioned on $\hat{S}_\gamma \in \mathcal{S}_\gamma$, the MSE of $\hat{X}$ averaged over the ensemble of measurement matrices is

$$\text{mse}(\boldsymbol{x}, \hat{S}_\gamma) \triangleq \mathbb{E}_{\boldsymbol{Y}, \boldsymbol{\Phi}} \left\{ \|\boldsymbol{x} - \hat{\boldsymbol{X}}\|^2 \right\}$$
$$= \mathbb{E}_{\boldsymbol{Y}, \boldsymbol{\Phi}} \left\{ \|\boldsymbol{x}_{\hat{S}_\gamma} - \hat{\boldsymbol{X}}_{\hat{S}_\gamma}\|^2 \right\} + \|\boldsymbol{x}_{\hat{S}_\gamma^c}\|^2. \quad (4)$$

Let $\text{mse}(\boldsymbol{x}_{\hat{S}_\gamma})$ denote the first term in (4). We have that

$$\text{mse}(\boldsymbol{x}_{\hat{S}_\gamma}) = \text{mse}(\boldsymbol{x}_{\hat{S}_\gamma}|\mathcal{E}_{id})\Pr(\mathcal{E}_{id}) + \text{mse}(\boldsymbol{x}_{\hat{S}_\gamma}|\mathcal{E}_{id}^c)\Pr(\mathcal{E}_{id}^c)$$
$$\leq \|\boldsymbol{x}_{\hat{S}_\gamma}\|^2 \mathcal{O}(1/m) + \text{mse}(\boldsymbol{x}_{\hat{S}_\gamma}|\mathcal{E}_{id}^c) \quad (5)$$

To analyse the second term in (5) we make explicit that $\boldsymbol{\Phi}$ has two independently generated parts, one part that contains the columns corresponding to $\hat{S}_\gamma$, namely $\boldsymbol{\Phi}_{\hat{S}_\gamma}$, and another part that contains the rest of the columns. In addition, note that the MSE only depends on the latter part through the columns corresponding to $\hat{S}_\gamma^c$, i.e. $\boldsymbol{\Phi}_{\hat{S}_\gamma^c}$. Using this we rewrite

$$\text{mse}(\boldsymbol{x}_{\hat{S}_\gamma}|\mathcal{E}_{id}^c) = \mathbb{E}_{\boldsymbol{\Phi}_{\hat{S}_\gamma}|\mathcal{E}_{id}^c} \left\{ \text{mse}(\boldsymbol{x}_{\hat{S}_\gamma}|\boldsymbol{\Phi}_{\hat{S}_\gamma} = \phi_{\hat{S}_\gamma}) \right\} \quad (6)$$

where

$$\text{mse}(\boldsymbol{x}_{\hat{S}_\gamma}|\boldsymbol{\Phi}_{\hat{S}_\gamma} = \phi_{\hat{S}_\gamma}) \triangleq \mathbb{E}_{\boldsymbol{\Phi}_{\hat{S}_\gamma^c}} \left\{ \mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\Phi}} \left\{ \|\boldsymbol{x}_{\hat{S}_\gamma} - \hat{\boldsymbol{X}}_{\hat{S}_\gamma}\|^2 \right\} \right\}$$

Note that $\text{mse}(\boldsymbol{x}_{\hat{S}_\gamma}|\boldsymbol{\Phi}_{\hat{S}_\gamma} = \phi_{\hat{S}_\gamma})$ is conditionally independent of $\mathcal{E}_{id}^c$ given $\phi_{\hat{S}_\gamma}$. It corresponds to the MSE incurred in obtaining $\hat{\boldsymbol{X}}_{\hat{S}_\gamma}$ when both the noise and the residual terms (i.e. those in $\hat{S}_\gamma$) are random processes. That is, for

$$\boldsymbol{Y} = \phi_{\hat{S}_\gamma} \boldsymbol{x}_{\hat{S}_\gamma} + \sum_{i \in \hat{S}_\gamma^c} x_i \boldsymbol{\Phi}_i + \boldsymbol{Z}.$$

The covariance matrix of the residual terms is given by

$$\mathbb{E} \left\{ \sum_{i \in \hat{S}_\gamma^c} \sum_{j \in \hat{S}_\gamma^c} x_i x_j \boldsymbol{\Phi}_i \boldsymbol{\Phi}_j^T \right\} = P_\Phi \|\boldsymbol{x}_{\hat{S}_\gamma^c}\|^2 \boldsymbol{I}_m.$$

Thus, the covariance matrix of the residual terms plus noise is $\boldsymbol{C} = \kappa \boldsymbol{I}_m$ with $\kappa \triangleq P_z + P_\Phi \|\boldsymbol{x}_{\hat{S}_\gamma^c}\|^2$. The estimation of $\boldsymbol{x}_{\hat{S}_\gamma}$ corresponds to a linear estimation problem in Gaussian noise. For the class of unbiased estimators the MSE satisfies:

$$\text{mse}(\boldsymbol{x}_{\hat{S}_\gamma}|\boldsymbol{\Phi}_{\hat{S}_\gamma} = \phi_{\hat{S}_\gamma}) = \kappa \text{tr}\{(\phi_{\hat{S}_\gamma}^T \phi_{\hat{S}_\gamma})^{-1}\}$$

which holds when $\hat{\boldsymbol{X}}_{\hat{S}_\gamma} = (\phi_{\hat{S}_\gamma}^T \phi_{\hat{S}_\gamma})^{-1} \phi_{\hat{S}_\gamma}^T \boldsymbol{Y}$. Thus, for this choice of estimate

$$\text{mse}(\boldsymbol{x}_{\hat{S}_\gamma}|\mathcal{E}_{id}^c) = \kappa \mathbb{E}_{\boldsymbol{\Phi}_{\hat{S}_\gamma}|\mathcal{E}_{id}^c} \left\{ \text{tr}\{(\boldsymbol{\Phi}_{\hat{S}_\gamma}^T \boldsymbol{\Phi}_{\hat{S}_\gamma})^{-1}\} \right\}.$$

Conditioned on $\mathcal{E}_{id}^c$, for any $\phi_{\hat{S}_\gamma}$ we can write

$$(\phi_{\hat{S}_\gamma}^T \phi_{\hat{S}_\gamma})^{-1} = \frac{1}{m} (\frac{1}{m} \phi_{\hat{S}_\gamma}^T \phi_{\hat{S}_\gamma})^{-1}$$
$$= \frac{1}{m P_\Phi} (\boldsymbol{I}_\ell + \boldsymbol{\psi})^{-1}$$

for some $\boldsymbol{\psi} \in \mathbb{R}^{\ell \times \ell}$ with $\|\boldsymbol{\psi}\| \leq \delta P_\Phi \triangleq \delta'$, and use the Taylor expansion of the matrix inverse to write

$$\text{tr}\{(\phi_{\hat{S}_\gamma}^T \phi_{\hat{S}_\gamma})^{-1}\} = \frac{1}{m P_\Phi} \text{tr} \left\{ \left(\boldsymbol{I}_\ell + \sum_{i=1}^\infty (-\boldsymbol{\psi})^i \right) \right\}$$
$$\leq \frac{1}{m P_\Phi} \left(\ell + \sum_{i=1}^\infty \sqrt{\ell} \|\boldsymbol{\psi}\|^i \right) \quad (7)$$
$$\leq \frac{\ell}{m P_\Phi} \left(1 + \frac{1}{\sqrt{\ell}} \frac{\delta'}{1 - \delta'} \right). \quad (8)$$

To obtain (7) we have used the bounds $\text{tr}\{\boldsymbol{\psi}\} \leq \sqrt{\ell} \|\boldsymbol{\psi}\|$, which is easily proved using the Cauchy-Schwartz inequality, and $\|\boldsymbol{\psi}^i\| \leq \|\boldsymbol{\psi}\|^i$ for $i \in \mathbb{N}$. To obtain (8) we have used that $\|\boldsymbol{\psi}\| \leq \delta'$ and calculated the geometric series (assuming $\delta' < 1$). Note that this bound is independent of ϕ_{S_γ}. Using (8) in (6) we obtain

$$\text{mse}(\boldsymbol{x}_{\hat{S}_\gamma}|\mathcal{E}_{id}^c) = \mathcal{O}(1/m).$$

This completes the asymptotic characterization of the MSE averaged over the ensemble of measurement matrices:

$$\text{mse}(\boldsymbol{x}, \hat{S}_\gamma) = \|\boldsymbol{x}_{\hat{S}_\gamma^c}\|^2 + \mathcal{O}(1/m). \quad (9)$$

We obtain (2) by noting that the preceding result only depends on $\boldsymbol{w}$ and $\hat{S}_\gamma$ because $\|\boldsymbol{x}_{\hat{S}_\gamma^c}\| = \|\boldsymbol{w}_{\hat{S}_\gamma^c}\|$. $\qquad \square$

The first term in (9) corresponds to the error incurred by not detecting the whole support set. The ordo term, which vanishes with m, includes the errors in estimating the components in $\hat{S}_\gamma$, as well as the effect of $\mathcal{E}_{id}$.

5. RELATED PRIOR WORK AND CONCLUSION

In this paper, we have derived an achievable tradeoff between the measurement rate, defined as a relation between the dimensions of the measurement matrix, and the estimation MSE. We have divided the problem into two parts.

First we have considered recovering parts of support set of the sparse signal. We have established sufficient conditions on the measurement rate to ensure partial detection of the support set based on the relative power of the non-zero entries in the sparse signal. This builds on and extends the results in [7, 8, 9], which considered only perfect recovery of the support set.

In the second part we have derived the MSE performance in estimating the entries of part of the support set. Prior work in the field considered the MSE for both ensembles of measurement matrices [10] (as we do here) and for deterministic measurement matrices [11]. However, our approach is more general in the sense that it covers the estimation of both partial and complete support sets. This was key to establishing the measurement rate-MSE tradeoff.

6. REFERENCES

[1] D.L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, Apr. 2006.

[2] E.J. Candes and M.B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[3] M.A. Davenport, P.T. Boufounos, M.B. Wakin, and R.G. Baraniuk, "Signal processing with compressive measurements," *IEEE Journal on Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 445–460, Apr. 2010.

[4] M. Elad, *Sparse and Redundant Representations*, Springer, 2010.

[5] N. Wagner, Y.C. Eldar, and Z. Friedman, "Compressed beamforming in ultrasound imaging," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4643–4657, Sept. 2012.

[6] M. Mishali and Y.C. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *IEEE Journal on Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 375–391, Apr. 2010.

[7] Y Jin and B.D. Rao, "Insights into the stable recovery of sparse solutions in overcomplete representations using network information theory," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2008, pp. 3921–3924.

[8] Y. Jin, Y.-H. Kim, and B.D. Rao, "Performance tradeoffs for exact support recovery of sparse signals," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, June 2010, pp. 1558–1562.

[9] Y. Jin, Y.-H. Kim, and B.D. Rao, "Limits on support recovery of sparse signals via multiple-access communication techniques," *IEEE Transactions on Information Theory*, vol. 57, no. 12, pp. 7877–7892, Dec. 2011.

[10] B. Babadi, N. Kalouptsidis, and V. Tarokh, "Asymptotic achievability of the Cramér-Rao bound for noisy compressive sampling," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1233–1236, Mar. 2009.

[11] Z. Ben-Haim and Y.C. Eldar, "The Cramér-Rao bound for estimating a sparse parameter vector," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3384–3389, June 2010.

[12] R. Niazadeh, M. Babaie-Zadeh, and C. Jutten, "On the achievability of Cramér-Rao bound in noisy compressed sensing," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 518–526, Jan. 2012.