<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper published in *IEEE Transactions on Signal Processing*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Blasco-Serrano, R., Zachariah, D., Sundman, D., Thobaben, R., Skoglund, M. (2014)
A Measurement Rate-MSE Tradeoff for Compressive Sensing Through Partial Support Recovery.
IEEE Transactions on Signal Processing, 62(18): 4643-4658
<http://dx.doi.org/10.1109/TSP.2014.2321739>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

© 2014 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-144414>

A Measurement Rate-MSE Tradeoff for Compressive Sensing Through Partial Support Recovery

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

Abstract—We consider the problem of estimating sparse vectors from noisy linear measurements in the high dimensionality regime. For a fixed number k of non-zero entries, we study the fundamental relationship between two relevant quantities: the measurement rate, which characterizes the asymptotic behavior of the dimensions of the measurement matrix in terms of the ratio $m/\log n$ (m being the number of measurements and n the dimension of the sparse vector), and the estimation mean square error. First, we use an information-theoretic approach to derive sufficient conditions on the measurement rate to reliably recover a part of the support set that represents a certain fraction of the total vector power. Second, we characterize the mean square error of an estimator that uses partial support set information. Using these two parts, we derive a tradeoff between the measurement rate and the mean square error. This tradeoff is achievable using a two-step approach: first support set recovery, then estimation of the active components. Finally, for both deterministic and random vectors, we perform a numerical evaluation to verify the advantages of the methods based on partial support set recovery.

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

I. INTRODUCTION

Compressive sensing deals with the recovery and estimation of sparse vectors from underdetermined sets of measurements [1]–[4]. For example, given the noise measurements $\mathbf{y} \in \mathbb{R}^m$ of the k -sparse vector $\mathbf{x} \in \mathbb{R}^n$ (i.e., all but $k \ll n$ entries of $\mathbf{x}$ are zero), obtained using the measurement matrix $\phi \in \mathbb{R}^{m \times n}$:

$$\mathbf{y} = \phi \mathbf{x} + \mathbf{z}, \quad (1)$$

a common goal is to obtain an approximation $\hat{\mathbf{x}}$ that is close to $\mathbf{x}$ in some sense, for example in mean square error (MSE). This problem arises often in practice as many signals of interest are sparse [1], [2], [4]–[7]. Several algorithms with affordable complexity have been proposed to implement the necessary operations for sparse signal processing: LASSO [8], Orthogonal Matching Pursuit (OMP) and variants thereof [9]–[12], Subspace Pursuit (SP) [13], Compressive Sampling Matching Pursuit (CoSaMP) [14], or Iterative Hard Thresholding (IHT) [15], among others.

The support set of a sparse vector (i.e., the set of non-zero entries) plays a fundamental role in the inference of sparse signals from an underdetermined relation of linear measurements in noise. In fact in some applications detecting this set is the ultimate goal [16]–[18]. In other cases, the knowledge of the support set is used as a tool to estimate some further parameters

about the vector or directly, to estimate the vector itself. Several works have tried to unveil the fundamental tradeoffs between the sparsity level k , the number of measurements m , and the signal dimension n for reliable detection of the support set. Necessary and sufficient conditions for complete support recovery were established in [19]–[22] for different regimes of (k, m, n) using different recovery methods. In [23]–[25], Jin et al. exploited the connections of complete support recovery to the problem of communicating over multiple access channels (MAC) to derive such conditions on the ratio $m/\log n$. Among others, they studied the case of fixed k , which is the only one considered here. For partial support recovery, which is a central component in our study, Reeves and Gastpar obtained necessary and sufficient conditions in [26] and [27], respectively. In their work, they observed that in the large system setting the conditions are less stringent than those for complete support recovery. In [28] necessary and sufficient conditions were derived for different support recovery criteria (including partial and complete recovery). In one or another way, most of these works also highlighted the difficulty in detecting entries with small magnitudes.

The problem of reconstructing or estimating sparse vectors has also attracted substantial attention. For the measurement model in (1), which is the one considered here, many works have studied stable recovery conditions that guarantee that the error between the estimate and the true vector is, at most, proportional to the noise norm. Wu and Verdú derived such conditions in terms of the ratio m/n for a stochastic model of the sparse signal and showed that Gaussian sensing matrices have optimal thresholds. For Basis Pursuit Denoising (BPDN), Candès et al. derived sufficient conditions for stability [29] and showed that it can provide a k -sparse approximation of any arbitrary vector [29, Theorem 2]. Their conditions are expressed in terms of restrictive isometry constants and can be applied to both deterministic and random matrices, including the Gaussian ensemble [30]. Conditions for stable recovery for the SP, CoSaMP, and IHT algorithms were derived in [13], [14], and [15], respectively (see also [31]). In [32], Donoho et al. derived sufficient conditions for stable recovery and partial support recovery of sparse signals using a combinatorial formulation (based on the L_0 “norm”) as well as for convex relaxations (see also [33]). As we will shortly see, all these works implicitly showed the existence of a tradeoff between the triple (k, m, n) and the MSE, and characterized it.

Another line of work has pursued to determine bounds on the estimation performance. In [34] the authors derived the Cramér-Rao bound of the the oracle estimator that knows the support set [35], [36], averaged over the ensemble measurement matrices. Moreover, [35] established the asymptotic achievability of the bound with estimators that lack knowledge about the support set by using the information-theoretic methods developed in [28]. Non-asymptotic bounds on the estimation performance for fixed measurement matrices were derived in [37] using the framework of the constrained Cramér-Rao bound. Finally, the performance of the linear minimum mean

This paper was partly presented at ICASSP 2013. Part of this work has been performed in the framework of the Network of Excellence ACROPOLIS and funded by the European Union under FP7 ICT Objective 1.1 - The Network of the Future.

All authors were affiliated at the time of submission with the KTH Royal Institute of Technology and the ACCESS Linnaeus Centre, Stockholm, Sweden. R. Blasco-Serrano and D. Sundman are now with Ericsson Research, Stockholm, Sweden. D. Zachariah is now with the Uppsala University, Uppsala, Sweden. (e-mail: ricardo.blasco@ieee.org, dave.zachariah@it.uu.se, dennis.sundman@ericsson.com, ragnart@kth.se, skoglund@kth.se).

square error (LMMSE) reconstruction method with partially correct support set information was studied in [38].

A. Contributions and Outline

In this paper, we study the fundamental relationship between two relevant quantities in compressive sensing: the measurement rate, which describes the asymptotic behavior of the relation $m/\log n$ (Definition 1), and the estimation mean square error. More specifically, we derive sufficient conditions in terms of the measurement rate to achieve a certain mean square error when performing estimation of a deterministic but unknown vector with a fixed level of sparsity through partial support set recovery. Our approach is the following.

First, we establish a condition on the number of measurements that are asymptotically sufficient for reliable recovery of parts of the support set corresponding to a certain fraction of the total power (Section III-A). Our results show that it is possible to reduce the measurement rate at the expense of recovering smaller parts of the support set, still with high reliability. Then, we characterize the estimation mean square error performance averaged over the ensemble of Gaussian measurement matrices when the sparse estimator relies only on partial knowledge about the support set (Section III-B).

We use these two results to derive a tradeoff between the measurement rate and the estimation mean square error (Section III-C). Our results show that it is often possible to significantly reduce the measurement rate at a minimal cost in terms of MSE. Indeed, we derive conditions on the signal of interest that guarantee that partial support recovery yields such reduction in measurement rate (Section III-D).

We conclude our study by analyzing the advantages of partial support set recovery methods in the scenario where the vector of interest is drawn from a random distribution (Section IV). We observe that neglecting the elements in the support set with little power relaxes the requirements in terms of measurement rate and, thus, reduces the number of measurement outages. In this way the MSE performance averaged over the distribution of the vectors is significantly improved.

B. Relation to Previous Work

The results on partial support recovery reported in this paper (Proposition 1 and Corollary 1) are related to the work in [28], although our notion of recovery is more stringent as it penalizes incorrect detection. For the sublinear sparsity regime $k = o(n)$, which is the relevant case here, [28] shows that reliable recovery is possible if and only if $C_1 < m/(k \log(n - k)) < C_2$ asymptotically for some $C_1, C_2 > 0$ that depend on the total power of the sparse signal $\|x\|^2$, the noise variance, and the fraction of power to be detected. In contrast, by using the tools introduced in [25] we characterize the constants appearing in our sufficient conditions in a way that allows us to study the role of the non-zero entries of the sparse signal (cf. (4) and (5)).

Our measurement rate-MSE tradeoff (Proposition 3) is related to the results in [13]–[15], [29], [32], [33]. However, we observe the following significant differences. Regarding the assumptions:

- We consider a statistical characterization of the noise (in particular, Gaussian) instead of the worst-case approach obtained by using bounds on the L_2 norm (e.g., $\|z\| \leq \epsilon$). As discussed in [31], for worst-case noise characterizations one cannot expect a denoising effect.

- Most of these other works consider the case of arbitrary (k, m, n) . In contrast, we keep the sparsity level k fixed and let $m, n \rightarrow \infty$. This allows us to derive conditions that explicitly show the role of the non-zero entries of the sparse vector.

Regarding the results:

- As opposed to the stability results in [13]–[15], [29], [32], [33], our approach guarantees reliable recovery of (parts of) the support set, which is a highly desirable property.
- We derive tighter bounds on the MSE performance by considering the asymptotic regime of (k, m, n) described previously. To be more specific, our bound on the MSE (cf. (11)–(12)) has the form $\|x_{und}\|_{L_2}^2 + O(1/m)$. The first term in the bound corresponds to the L_2 norm of the components in the signal that remain undetected. This contrasts with the results derived in [31, Theorem 3.1] for the expected performance for SP [13], CoSaMP [14], and IHT [15]. In this case, the bounds include an additive term $C(\|x_{und}\|_{L_2} + 1/\sqrt{k}\|x_{und}\|_{L_1})^2$ where $C > 1$ is a (reportedly loose) constant that depends on the algorithm.

Finally, regarding the methodology:

- Part of our results rely on computationally prohibitive tools, in particular, those for support recovery. In contrast, many of the other works use convex optimization tools, which can be implemented in practice.

Our work is also related to previous studies on the MSE bounds in [34], [36]–[38]. However, our approach is different since we draw an explicit connection between measurement rate and MSE. Moreover, we consider a different regime of the parameters (fixed k and growing m and n). The papers [34], [36] consider the linear sparsity regime (i.e., k growing linearly with n) and rely on complete support recovery to achieve the Cramér-Rao bound. With their method, complete support requires $\|x\|^2 \rightarrow \infty$ (for a discussion on this, see [28, Section I-C]), which is quite different from our model with fixed $\|x\|^2$. In addition, we are more conservative when analyzing the effect on the MSE of the errors of the support recovery map (see Section III-C in this paper). Nonetheless, in spite of the differences of regimes, for the particular case of complete support recovery our conclusions are similar. Observe also that for the case of strictly partial support recovery, the direct comparison with the Cramér-Rao bound (as defined in [34], [36], [37]) is not meaningful because our estimators are necessarily biased.

Finally, in contrast to many previous works our results do not rely on explicit knowledge of the sparsity level k ; at most they require an arbitrary but fixed upper bound.

II. PRELIMINARIES

Notation: Upper-case letters denote random variables or vectors and lower-case letters denote their realizations. The statistical expectation over the distribution of X is denoted by $\mathbb{E}_X \{\cdot\}$. Vectors and matrices are represented with bold face letters x, ϕ . The j^{th} column of a matrix ϕ and the entry in row i , column j are denoted by ϕ_j and $\phi_{i,j}$, respectively. Similarly, x_i refers to the i^{th} entry in vector x . A vector $x \in \mathbb{R}^n$ is k -sparse if only k of its entries are non-zero. The operators $\|\cdot\|$ and $\text{tr}\{\cdot\}$ denote the L_2 norm of a vector/matrix and the trace of a square matrix, respectively. The size- l identity matrix is denoted by $\mathbf{I}_l$. Sets are collections of unique objects and are denoted using calligraphic letters (e.g., $\mathcal{S}$ or $\mathcal{S}$). The cardinality of a set is denoted by $|\mathcal{S}|$. Given a vector $x = [x_1, \dots, x_n]^T$ and two integers i, j , with $1 \leq i \leq j \leq n$, $x_i^j = [x_i, \dots, x_j]^T$.

Similarly, given a set $\mathcal{S} = \{s_1, \dots, s_l\}$, $\mathbf{x}_{\mathcal{S}}$ is the subvector $[x_{s_1}, \dots, x_{s_l}]^T$. If ϕ is a matrix, $\phi_{\mathcal{S}}$ denotes the submatrix obtained by taking only the columns specified by the set $\mathcal{S}$. Probability events are also denoted by calligraphic letters. The complement of a set or event $\mathcal{E}$ is denoted by $\mathcal{E}^c$. $o(\cdot)$ and $O(\cdot)$ denote the standard Ordo notation.

A. System Model and Motivation

Consider the following model for sparse signals. The k -sparse vector $\mathbf{x} \in \mathbb{R}^n$ is defined component-wise as

$$x_i = \begin{cases} w_j & \text{if } i = s_j, \\ 0 & \text{if } i \notin \mathcal{S} \end{cases}$$

for $i = \{1, \dots, n\}$, where $\mathcal{S} \subset \{1, \dots, n\}$ is the support set, $[s_1, \dots, s_k]^T$ is an arbitrary ordering of the k elements in $\mathcal{S}$, and $\mathbf{w} \in \mathbb{R}^k$ is a deterministic but unknown vector with k non-zero entries sorted by decreasing magnitude. To emphasize the dependence of $\mathbf{x}$ on $\mathbf{w}$ and $\mathcal{S}$, we will often write $\mathbf{x}(\mathbf{w}, \mathcal{S})$. This vector is observed using the random measurement matrix $\Phi \in \mathbb{R}^{m \times n}$ in the presence of independent random noise $\mathbf{Z}$:

$$\mathbf{Y} = \Phi \mathbf{x}(\mathbf{w}, \mathcal{S}) + \mathbf{Z}. \quad (2)$$

The entries of the measurement matrix are independent and identically distributed (i.i.d.) according a Gaussian distribution $\mathcal{N}(0, P_{\Phi})$. Similarly, we have i.i.d. $Z_i \sim \mathcal{N}(0, P_Z)$. We define the measurement signal-to-noise ratio as $\text{SNR} \triangleq P_{\Phi}/P_Z$.

We are interested in estimating $\mathbf{x}(\mathbf{w}, \mathcal{S})$. In particular, we construct an estimator $\hat{\mathbf{X}}$ that first attempts to detect the support set and then estimates the values of the corresponding entries. We study sufficient conditions on the number of measurements that ensure that the MSE

$$\text{mse}(\mathbf{x}(\mathbf{w}, \mathcal{S})) \triangleq \mathbb{E}_{\Phi, \mathbf{Z}} \left\{ \|\mathbf{x}(\mathbf{w}, \mathcal{S}) - \hat{\mathbf{X}}\|^2 \right\}$$

is arbitrarily close to a target value.

In our analysis, we allow both the vector dimension n and the number of measurements m to go to infinity while keeping k , $\mathcal{S}$, and $\mathbf{w}$ fixed. For this purpose, we consider a sequence of support recovery problems $\{\mathbf{x}^n(\mathbf{w}, \mathcal{S}), \Phi^{(n)}, \mathbf{Y}^{(n)}\}$, indexed by $n \in \mathbb{N}_+$. Here $\mathbf{x}^n(\mathbf{w}, \mathcal{S})$ denotes a sequence of k -sparse vectors in $\mathbb{R}^n$ with fixed support set $\mathcal{S}$ and non-zero entries $\mathbf{w} \in \mathbb{R}^k$. Since n indexes the sequence, we will denote the number of measurements by m_n (e.g., $\Phi^{(n)} \in \mathbb{R}^{m_n \times n}$). The measurement rate of a sequence of support recovery problems is a quantity of especial relevance in our analysis.

Definition 1 (Measurement rate).

$$r \triangleq \liminf_{n \rightarrow \infty} \frac{m_n}{\log_2 n}.$$

◇

We divide the problem into two parts: first, support recovery and then, signal estimation. We will formulate both sub-problems separately in Section II-B but first we motivate the necessity for partial support recovery.

For a variation of the model in (2) that considers random support set (but fixed $\mathbf{w}$), [25, Theorems 1 and 2] determines necessary and sufficient conditions on r for complete support recovery (i.e., reliable detection of $\mathcal{S}$). The following example illustrates some of the challenges implied by this result (the conditions can be obtained by setting $\gamma = 1$ in (5)):

Example 1. Let $\frac{P_{\Phi}}{P_Z} = 10$ and $\mathbf{w} \in \mathbb{R}^2$.

- Case 1: $w_1^2 = w_2^2 = 0.5$ requires $r > 1.16$.

- Case 2: $w_1^2 = 0.95$ and $w_2^2 = 0.05$ requires $r > 3.42$.
- Case 3: $w_1^2 = 0.99$ and $w_2^2 = 0.01$ requires $r > 14.54$.

In all three cases, the sparse vector has $\|\mathbf{w}\|^2 = 1$. This means that it is the uneven distribution of the power between w_1 and w_2 that varies significantly the requirements for complete support recovery. ◇

Motivated by this example, we seek to develop methods for recovering parts of the support set with less stringent conditions on r than those in [25, Theorems 1 and 2].

B. Problem Formulation

We are interested in finding sufficient conditions on the measurement rate r that ensure that is possible to estimate a k -sparse vector $\mathbf{x}^n(\mathbf{w}, \mathcal{S})$ with a certain MSE performance. Our approach is to divide the problem into two parts: first, partial support recovery and then, signal estimation. In the following, we formulate each of these two sub-problems separately.

1) *Partial Support Set Recovery*: To study the problem of partial support recovery, we introduce the following definitions that extend the framework for complete support recovery from [25]. The central element is concept of γ -support set $\mathcal{S}_{\gamma}$.

Definition 2 (γ -support set). Let $\mathbf{x} \in \mathbb{R}^n$ and $\gamma \in (0, 1]$. A γ -support set $\mathcal{S}_{\gamma}$ of $\mathbf{x}$ is any subset of $\{1, \dots, n\}$ such that

$$\sum_{i \in \mathcal{S}_{\gamma}} x_i^2 \geq \gamma \|\mathbf{x}\|^2$$

and that has the smallest possible size $\lambda(\mathbf{w}, \gamma) = |\mathcal{S}_{\gamma}|$. ◇

Observe that this definition precludes $x_i = 0$ for any $i \in \mathcal{S}_{\gamma}$. Note also that there might exist more than one γ -support set but they all have the same size $\lambda(\mathbf{w}, \gamma)$. This size clearly depends on $\mathbf{w}$ and γ , but to keep the notation simpler in the rest of the paper we will simply denote it by λ . In addition, we define

$$\mathcal{S}_{\gamma} \triangleq \{\mathcal{T} \in \mathcal{P}(\{1, \dots, n\}) : \mathcal{T} \text{ is a } \gamma\text{-support set of } \mathbf{x}\},$$

$$\underline{\mathcal{S}}_{\gamma} \triangleq \bigcup_{\mathcal{T} \in \mathcal{S}_{\gamma}} \mathcal{T},$$

where $\mathcal{P}(\{1, \dots, n\})$ is the power set of $\{1, \dots, n\}$ (i.e., the set of all possible subsets of $\{1, \dots, n\}$). Unlike $\mathcal{S}_{\gamma}$, the sets $\mathcal{S}_{\gamma}$ and $\underline{\mathcal{S}}_{\gamma}$ are always unique given a fixed γ . For the case $\gamma = 1$, we have $\mathcal{S}_{\gamma} = \underline{\mathcal{S}}_{\gamma} = \mathcal{S}$ and $\mathcal{S}_{\gamma} = \{\mathcal{S}\}$.

To estimate a γ -support set of a sparse vector, we use a *support recovery map* d_{γ} , which maps every pair of vector of observations and measurement matrix $(\mathbf{y}, \phi)$ to a subset of $\{1, \dots, n\}$:

$$d_{\gamma} : \mathbb{R}^{m_n} \times \mathbb{R}^{m_n \times n} \rightarrow \mathcal{P}(\{1, \dots, n\}).$$

For given vector $\mathbf{x}^n(\mathbf{w}, \mathcal{S})$ and support recovery map d_{γ} , we define the error probability as

$$P_e(\mathbf{x}^n(\mathbf{w}, \mathcal{S}), d_{\gamma}) \triangleq \Pr(d_{\gamma}(\mathbf{Y}, \Phi) \notin \mathcal{S}_{\gamma}). \quad (3)$$

Here the probability is taken over the distributions of the noise and the measurement matrix.

Our first goal is to characterize a large class of sparse vectors in terms of a condition on the measurement rate r for which there exist support recovery maps with arbitrarily low $P_e(\mathbf{x}^n(\mathbf{w}, \mathcal{S}), d_{\gamma})$ for every vector in the class. As opposed to most previous works, we do not endow the support recovery map with explicit knowledge on the sizes of $\mathcal{S}_{\gamma}$ or $\mathcal{S}$, (i.e., λ or k , respectively). We will only assume that the support recovery map knows a fixed upper bound $k_{\max}$ on k (i.e., independent of n and m_n) and restrict accordingly the class of detectable vectors. We defer the analysis to Section III-A.

2) *Signal Estimation Using Partial Support Set Knowledge:* In the second part of our study, we determine the MSE performance of an estimator that uses a γ -support set. That is, we are interested in the MSE performance that is achievable by an estimator $\hat{\mathbf{x}}(\mathcal{S}_\gamma)$,

$$\text{mse}(\mathbf{x}, \mathcal{S}_\gamma) \triangleq \mathbb{E}_{\Phi, \mathbf{Z}} \left\{ \|\mathbf{x} - \hat{\mathbf{x}}(\mathcal{S}_\gamma)\|^2 \right\}$$

that knows some $\mathcal{S}_\gamma \in \mathcal{S}_\gamma$. We consider this in Section III-B.

III. MAIN RESULTS

In this section, we present our main contributions.

A. Partial Support Set Recovery

Given $P_\Phi > 0$, $P_Z > 0$, $\mathbf{w} \in \mathbb{R}^k$, and $\gamma \in (0, 1]$, we define

$$c_i(\mathbf{w}, \gamma) \triangleq \frac{1}{2i} \log_2 \left(\frac{P_\Phi \sum_{j=\lambda-i+1}^k w_j^2 + P_Z}{(1-\gamma) \|\mathbf{w}\|^2 P_\Phi + P_Z} \right), \quad (4)$$

for $i \in \{1, \dots, \lambda\}$, where λ is the size of the γ -support set (recall that λ depends on $\mathbf{w}$ and γ , see Definition 2), and

$$r^*(\mathbf{w}, \gamma) \triangleq \max_{i \in \{1, \dots, \lambda\}} \frac{1}{c_i(\mathbf{w}, \gamma)}. \quad (5)$$

We use (5) to characterize a class of sequences of sparse vectors that are suitable for partial support recovery.

Definition 3. Given $\rho \geq 0$, $\gamma \in (0, 1]$, and $k \in \mathbb{N}_+$, we define

$$\mathcal{C}^n(\rho, \gamma, k) \triangleq \{\mathbf{x}^n(\mathbf{w}, \mathcal{S}) : |\mathcal{S}| \leq k, r^*(\mathbf{w}, \gamma) < \rho\}.$$

□

Our main result regarding partial support set recovery for the measurement model in (2) is the following proposition.

Proposition 1. Given any $\gamma \in (0, 1]$ and $k_{\max} \in \mathbb{N}_+$, there exists a fixed sequence of support recovery maps $d_\gamma^{(n)}$ with measurement rate r and such that

$$\lim_{n \rightarrow \infty} P_e \left(\mathbf{x}^n(\mathbf{w}, \mathcal{S}), d_\gamma^{(n)} \right) = 0$$

for every $\mathbf{x}^n(\mathbf{w}, \mathcal{S}) \in \mathcal{C}^n(r, \gamma, k_{\max})$. In particular, there exists one such sequence $d_\gamma^{(n)}$ with

$$P_e \left(\mathbf{x}^n(\mathbf{w}, \mathcal{S}), d_\gamma^{(n)} \right) \leq o(1/m_n).$$

□

Proof: The proof is provided in Appendix A-A. ■

For a given $\mathbf{x}^n(\mathbf{w}, \mathcal{S})$, Proposition 1 implies that it suffices that the number of measurements m_n grows with the vector dimension n so that $r > r^*(\mathbf{w}, \gamma)$ in order to detect a γ -support set with vanishing error probability.

Remark 1. By setting $\gamma = 1$ in (5) (i.e., detecting all k components in the support set), we recover [25, Theorem 1]. Proposition 1 implies that this is possible even without knowledge of the support set size k , as long as it is fixed. □

Remark 2. As [25, Theorem 1], our result only depends on the sparse vector $\mathbf{x}$ through the non-zero components $\mathbf{w}$. □

To illustrate the advantages of partial support recovery, we revisit Example 1.

Example 2. Let $\frac{P_\Phi}{P_Z} = 10$ and $\mathbf{w} \in \mathbb{R}^2$.

- For Case 2 in Example 1, the choice $\gamma = 0.95$ yields $r^*(\mathbf{w}, \gamma) = 0.71$. In this case, partial support recovery

(only the position of w_1 is detected) requires roughly one fifth of the measurements of complete support recovery.

- For Case 3 in Example 1, the choice $\gamma = 0.99$ yields $r^*(\mathbf{w}, \gamma) = 0.60$. In this case, partial support recovery (only the position of w_1 is detected) requires roughly 4% of the number of measurements of complete support recovery. □

This shows that large reductions of measurement rate are possible. We emphasize that the savings depend on the choice of γ , as shown in the following example.

Example 3. Let $\mathbf{w} \in \mathbb{R}^2$ 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 $\mathbf{x}$. However, $r^*(\mathbf{w}, \gamma_1) > r^*(\mathbf{w}, \gamma_2)$ for all SNR. □

In Lemma 2 in Appendix F we prove that for a given λ , $r^*(\mathbf{w}, \gamma)$ is minimized by choosing $\gamma = \gamma_\lambda^{\text{opt}}$, where

$$\gamma_\lambda^{\text{opt}} \triangleq \frac{\sum_{j=1}^\lambda w_j^2}{\|\mathbf{w}\|^2}. \quad (6)$$

It is necessary to emphasize that, although partial support recovery is often possible at lower measurements rates than complete support recovery, this is not always the case. We illustrate this with a final example.

Example 4. Consider Case 1 in Example 1. For any SNR, the sufficient condition in terms of the measurement rate for complete support recovery (i.e., $\gamma = 1$) is less stringent than that for partial support recovery, regardless of the choice of $\gamma \in (0, 1)$. For example, for $\frac{P_\Phi}{P_Z} = 10$ we have $r^*(\mathbf{w}, 1) = 1.16$; with $\gamma_1 = 0.95$ (both entries detected) we have $r^*(\mathbf{w}, \gamma_1) = 1.39$; and with $\gamma_2 = 0.45$ (one entry detected) we have $r^*(\mathbf{w}, \gamma_2) = 2.63$. □

These last two examples show that the choice of γ should be influenced by our prior knowledge of $\mathbf{w}$, if any. We will discuss in Section III-D under which conditions partial support recovery is possible at lower measurement rate than complete support recovery.

The sufficient conditions in Proposition 1 suffer from an SNR saturation effect caused by the small but strictly non-zero entries of a sparse vector. We have that

$$\lim_{\text{SNR} \rightarrow \infty} c_i(\mathbf{w}, \gamma) = \frac{1}{2i} \log_2 \left(\frac{\sum_{j=\lambda-i+1}^k w_j^2}{(1-\gamma) \|\mathbf{w}\|^2} \right) \quad (7)$$

for any $\gamma \in (0, 1)$. That is, by letting $\text{SNR} \rightarrow \infty$, we are increasing the power contribution of $\mathbf{x}$. This includes also the power of the smaller entries in $\mathbf{x}$ that are treated as noise when performing partial support recovery (i.e., the denominator in (4)). In this situation, the reasonable approach is to increase the power to be detected (i.e., increase γ) or even perform complete support recovery. From our results on the MSE performance it will be clear that given a $\mathcal{S}_{\gamma_1}$, we can identify a $\mathcal{S}_{\gamma_2}$ for $0 < \gamma_2 \leq \gamma_1 \leq 1$ with high probability.

One small problem with the formulation of Proposition 1 is that there is not always a unique γ -support set (see Section II-B1). To avoid the ambiguity regarding the output of the support recovery map, we extend the result to the detection of $\underline{\mathcal{S}}_\gamma$, which is unique.

Corollary 1. Under the conditions of Proposition 1, there exists a sequence of partial support recovery maps $d_\gamma^{(n)}$ such that

$$\Pr \left(d_\gamma^{(n)}(\mathbf{Y}^{(n)}, \Phi^{(n)}) \neq \underline{\mathcal{S}}_\gamma \right) \leq o(1/m_n).$$

Proof: The proof is provided in Appendix A-D. ■

Finally, observe that given the previous results, standard concentration arguments based on the Markov inequality show that with high probability, any randomly chosen ϕ will have small error probability (e.g., see [28, Theorem 1.2]).

B. Estimation Mean Square Error

Consider now the measurement model given in (2) for fixed (m, n) and an arbitrary k -sparse vector $\mathbf{x}(\mathbf{w}, \mathcal{S})$.

Proposition 2. *Given some $\mathcal{T} \subseteq \mathcal{S}$, it is possible to estimate any $\mathbf{x}(\mathbf{w}, \mathcal{S})$ with MSE*

$$\text{mse}(\mathbf{x}(\mathbf{w}, \mathcal{S}), \mathcal{T}) = \|\mathbf{x}_{\mathcal{T}^c}\|^2 + \frac{\xi l}{m - l - 1}, \quad (8)$$

where $l \triangleq |\mathcal{T}|$, $\xi \triangleq P_Z/P_\Phi + \|\mathbf{x}_{\mathcal{T}^c}\|^2$, and $\mathbf{x}_{\mathcal{T}^c}$ is the subvector of $\mathbf{x}(\mathbf{w}, \mathcal{S})$ that contains the non-zero entries of $\mathbf{x}(\mathbf{w}, \mathcal{S})$ not included in $\mathcal{T}$ (i.e., in $\mathcal{S} \setminus \mathcal{T}$). □

Proof: The proof is provided in Appendix B. ■

The first part of the MSE expression in (8) corresponds to the residual entries in $\mathbf{x}(\mathbf{w}, \mathcal{S})$; that is, those whose index was not part of $\mathcal{T}$. This term has fixed magnitude, independent of m and n . The second part of the MSE corresponds to the estimation noise and consists of two contributions (cf. ξ): the measurement noise $\mathbf{Z}$ and the residual entries. This term depends on the matrix dimensions and decays roughly as $1/m$ with the number of measurements.

The behavior in (8) is not unexpected, since given (a fraction of) the support set, the estimation of $\mathbf{x}$ amounts to finding the solution of a noisy overdetermined system of equations. In this sense, the MSE expression in (8) has an oracular interpretation similar to those in [34], [35], and [37], for the case where the oracle provides only partial information about the support set. Note also that for $\mathcal{T} \neq \mathcal{S}$, the estimator constructed in the proof is biased.

Remark 3. *For the case $\mathcal{T} \in \mathcal{S}_\gamma$, we have*

$$\text{mse}(\mathbf{x}(\mathbf{w}, \mathcal{S}), \mathcal{S}_\gamma) = \|\mathbf{x}_{\mathcal{S}_\gamma^c}\|^2 + \frac{\xi \lambda}{m - \lambda - 1}. \quad (9)$$

Similarly, if $\mathcal{T} = \mathcal{S}_\gamma$, then

$$\text{mse}(\mathbf{x}(\mathbf{w}, \mathcal{S}), \mathcal{S}_\gamma) = \|\mathbf{x}_{\mathcal{S}_\gamma^c}\|^2 + \frac{\xi \lambda}{m - \lambda - 1}, \quad (10)$$

where $\lambda \triangleq |\mathcal{S}_\gamma|$. ◇

Note that in general, (9) depends on $\mathbf{x}$ and the choice of $\mathcal{S}_\gamma$ but (10) depends only on $\mathbf{x}$ and γ . Before drawing a connection between Corollary 1 and Proposition 2, we introduce the following definition:

$$\text{mse}^*(\mathbf{w}, \gamma) \triangleq \|\mathbf{x}_{\mathcal{S}_\gamma^c}\|^2 + O(1/m_n). \quad (11)$$

C. Measurement Rate-MSE Tradeoff

Consider the measurement model in (2). The concatenation of the results in Sections III-A and III-B yields the following proposition.

Proposition 3. *Given any $\gamma \in (0, 1]$ and $k_{\max} \in \mathbb{N}_+$, there exists a fixed sequence of estimators $\hat{\mathbf{x}}^n$ with measurement rate τ and mean square error*

$$\text{mse}(\mathbf{x}^n(\mathbf{w}, \mathcal{S})) = \text{mse}^*(\mathbf{w}, \gamma)$$

Fig. 1. Measurement rate vs. normalized MSE: all pairs (τ, mse) above the solid line are asymptotically achievable. The markers identify the (τ^*, mse^*) pairs achievable using several different values of γ .

for every $\mathbf{x}^n(\mathbf{w}, \mathcal{S}) \in \mathcal{C}^n(\tau, \gamma, k_{\max})$. □

Proof: The proof is provided in Appendix C. ■

Remark 4. *This proposition characterizes a tradeoff between the measurement rate τ and the mean square error mse that is achievable for a given $\mathbf{x}^n(\mathbf{w}, \mathcal{S}) \in \mathcal{C}^n(\rho, \gamma, k_{\max})$ using the sequence of estimators. Moreover, the pairs (τ, mse) are achievable by first performing (partial) support set recovery followed by the estimation of the detected active components.* ◇

The bound on the $\text{mse}(\mathbf{x}^n(\mathbf{w}, \mathcal{S}))$ obtained in the proof, up to a positive term $o(1/m_n)$, is given by (see (50))

$$\|\mathbf{x}_{\mathcal{S}_\gamma^c}\|^2 + \frac{k_{\max}(P_Z + (\xi + k_{\max}\|\mathbf{x}\|^2)P_\Phi)}{P_\Phi(m_n - k_{\max} - 1)}. \quad (12)$$

Observe that the MSE behavior in (12) is worse than that characterized in (10). That is, although the tradeoff describes the performance that is achievable in two steps, first recovering $\mathcal{S}_\gamma$ and then estimating the values, the resulting MSE is not equal to that of the second step. In fact, the main difference between both characterizations comes from the fact that (10) is obtained under the assumption that a *correct* $\mathcal{S}_\gamma$ is available. In contrast, (12) is derived using an estimate $\hat{\mathcal{S}}_\gamma$, which is correct with high probability but not always. Since the support recovery map has no means of telling with absolute certainty whether a specific $\hat{\mathcal{S}}_\gamma$ is correct or not, it cannot avoid making bad estimates. In some works, this issue is neglected and it is decided that $\hat{\mathbf{x}} = \mathbf{0}$ whenever $\hat{\mathcal{S}}_\gamma$ is incorrect. This simplifies the analysis and yields a more optimistic expression for the MSE. We believe that it is necessary to be more conservative when bounding the MSE in this cases. Nonetheless, asymptotically, the difference between (12) and (10) is negligible, as the gap vanishes with m_n . Moreover, with very high probability (depending on the recovery map), the MSE behavior is indeed given by (10).

In Fig. 1, we show typical examples of pairs (τ, mse) . This corresponds to a realization of $\mathbf{w}$ with $k = 10$ and i.i.d. $w_j \sim \mathcal{N}(0, 1/\sqrt{k})$. The MSE is normalized by $\|\mathbf{w}\|^2$ so that the values range from 0 to 1. The solid line represents the boundary of the region of pairs (τ, mse) in Proposition 3. All pairs above this curve are asymptotically achievable by selecting γ appropriately. The staircase-like behavior is due

to the discrete nature of the support set: once a certain measurement rate is achieved it is possible to recover a larger fraction of the support set and thus abruptly reduce the MSE. The outer corner points of the region correspond to pairs with

$$\begin{aligned} r &= r^*(\mathbf{w}, \gamma_i^{opt}), \\ mse &= mse^*(\mathbf{w}, \gamma_i^{opt}) \end{aligned}$$

for some $i \in \{1, \dots, k\}$ with γ_i^{opt} as defined in (6). In practice one usually has no knowledge about the structure of $\mathbf{w}$ and thus γ needs to be chosen arbitrarily. To illustrate the performance in this case, we have included the (r^*, mse^*) pair for several arbitrary choices of γ . This figure shows that it is often possible to drastically reduce 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 $\gamma = 1$, complete support recovery [25]) to $r \approx 7$ only incurs in a relative MSE of 0.0028 if γ is chosen carefully.

D. Region of Interest

We observed in Example 4 that partial support recovery does not always yield a measurement rate requirement that is lower than that for complete support recovery. In some situations, this is the case even if the parameter γ for partial support recovery is optimally chosen, as we also saw in the example.

Intuitively, whenever one of the entries is significantly smaller than the others, partial support recovery has a lower measurement rate requirement (cf. (4) and (5)). The following proposition formalizes this observation by establishing a sufficient condition in terms of the values of the entries of $\mathbf{w}$. Consider the following definitions:

$$c_i^{ex} \triangleq c_i(\mathbf{w}, 1) \quad (13)$$

$$r^{ex} \triangleq r^*(\mathbf{w}, 1) \quad (14)$$

for $i \in \{1, \dots, k\}$, and

$$c_i^{par, \lambda} \triangleq c_i(\mathbf{w}, \gamma_\lambda^{opt}) \quad (15)$$

$$r^{par, \lambda} \triangleq r^*(\mathbf{w}, \gamma_\lambda^{opt}) \quad (16)$$

for $i \in \{1, \dots, \lambda\}$ with $\lambda < k$ and γ_λ^{opt} as in (6). Equations (13)-(14) characterize the requirement for complete support recovery, whereas (15)-(16) characterize the recovery of the support corresponding to the largest λ entries when γ is optimally chosen (see Lemma 2 in Appendix F).

Proposition 4. *If $c_i^{ex} \leq c_t^{ex}$ for all $i \in \{1, \dots, k - \lambda\}$ and $t \in \{k - \lambda + 1, \dots, k\}$ then $r^{ex} \geq r^{par, \lambda}$.* $\square$

Proof: The proof is provided in Appendix D-A. $\blacksquare$

In fact, partial support recovery strictly reduces the required measurement rate if the inequality in the condition is strict.

Corollary 2. *If $c_i^{ex} < c_t^{ex}$ for all $i \in \{1, \dots, k - \lambda\}$ and $t \in \{k - \lambda + 1, \dots, k\}$ then $r^{ex} > r^{par, \lambda}$.* $\square$

Proof: The proof is provided in Appendix D-B. $\blacksquare$

IV. RANDOM SIGNALS

The derivation of the measurement rate-mean square error tradeoff in the previous section relied on support recovery with a vanishing error probability. This was possible thanks to the deterministic nature of $\mathbf{x}$ (or more precisely of $\mathbf{w}$, see [25]), by increasing (m_n, n) at a measurement rate above $r^*(\mathbf{w}, \gamma)$.

Fig. 2. Measurement outage probability P_o as a function of the measurement SNR for partial support recovery of 90%, 95%, 99%, and 99.9% of the signal power, and for complete support recovery of a Gaussian $\mathbf{W}$.

Let $k \in \mathbb{N}_+$ be fixed and consider the following variation of the measurement model in (2)

$$\mathbf{Y} = \Phi \mathbf{X}(\mathbf{W}, \mathfrak{S}) + \mathbf{Z}$$

where both $\mathbf{W} \in \mathbb{R}^k$ and the support set $\mathfrak{S}$ are randomly generated.¹ The measurement matrix $\Phi \in \mathbb{R}^{m_n \times n}$ and the noise $\mathbf{Z} \in \mathbb{R}^{m_n}$ consist of Gaussian i.i.d. entries, as before. We assume that Φ , $\mathbf{Z}$, $\mathbf{W}$, and $\mathfrak{S}$ are mutually independent.

Now, consider a sequence of support recovery problems $\{\mathbf{X}^n(\mathbf{W}, \mathfrak{S}), \Phi^{(n)}, \mathbf{Y}^{(n)}\}$. Given any fixed measurement rate r , for most distributions of interest for $\mathbf{W}$ there is a non-vanishing probability of error for complete support detection [25]. The reason is that it is always possible that some realizations of $\mathbf{W}$ contain small entries that push the measurement requirements beyond r . We refer to this type of errors as measurement outages in analogy with the multiple access fading channel. These outages can be of two types: (i) failing to detect some of the entries of the support set and (ii) producing false estimates. In this section, we study the advantages of *partial but reliable* support set estimation for the case of random $\mathbf{X}(\mathbf{W}, \mathfrak{S})$ in terms of the measurement outage probability and the average mean square error.

A. Measurement Outage Probability

For any fixed $\gamma \in (0, 1]$, we define the measurement outage probability as

$$P_o \triangleq \Pr(\mathbf{X}^n(\mathbf{W}, \mathfrak{S}) \notin \mathcal{C}^n(r, \gamma, k_{max})).$$

This is the probability that a realization $\mathbf{x}$ does not belong to our class of detectable vectors in Definition 3. Observe that given our model and the definition of the class, this only depends on $\mathbf{X}^n(\mathbf{W}, \mathfrak{S})$ through $r^*(\mathbf{W}, \gamma)$. It can be shown that P_o is an upper bound on the probability that the output of a good sequence of support recovery maps does not yield a correct γ -support set (see [25, Theorem 5]) or, more generally, all γ -support sets, as in Corollary 1.

Let $\mathbf{W}$ have independent Gaussian distributed entries with zero mean and unit variance (i.e., $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_k)$) and let the

¹We use $\mathfrak{S}$ to denote a random support set. A deterministic support sets is denoted by $\mathcal{S}$, as before.

Fig. 3. Critical situation: the size of the γ -support set is λ . However, this choice of γ is just above the fraction of power contained in the $\lambda - 1$ largest components. The support recovery map treats the fraction $(1 - \gamma)$ of power, which increases with the SNR, as noise.

support set be randomly distributed over all size- k subsets of $\{1, \dots, n\}$. In Fig. 2, we illustrate the measurement outage probability P_o as a function of the SNR for $k = 10$ and different values of γ . We observe that for fixed SNR, partial support recovery with a properly chosen γ yields significantly lower measurement outage probabilities. Conversely, given a target measurement outage probability P_o it is possible to significantly reduce the SNR if the components in $\mathbf{W}$ with less power are of minor concern.

Observe that the optimal choice of γ depends on the measurement SNR. Moreover, note that the measurement outage probability curves flatten off at large SNR for all $\gamma < 1$. This has two causes. First, the SNR saturation effect in our sufficient conditions due to the small entries described in Section III-A (cf., (7)). Second, since γ is set independently of the realization of $\mathbf{W}$, the following effect takes place. Consider $c_i(\mathbf{w}, \gamma)$ for $i = 1$ and observe that although we always have that

$$\sum_{j=\lambda}^k w_j^2 > (1 - \gamma) \|\mathbf{w}\|^2, \quad (17)$$

there exist realizations of $\mathbf{W}$ that yield an arbitrarily small gap between both sides in (17), given that γ is chosen independently. That is, if λ is the size of the γ -support sets, the parameter γ is too close (from above) to the fraction of power encompassed by the largest $\lambda - 1$ components. This effect, illustrated in Fig. 3, renders support set detection extremely difficult in the presence of noise disturbances. In terms of required measurement rate, $c_1(\mathbf{w}, \gamma)$ vanishes and thus $r^*(\mathbf{w}, \gamma)$ becomes arbitrarily large. However, this only becomes noticeable in terms of P_o at high SNR. For low and moderate values of the SNR, it is the presence of small entries in $\mathbf{w}$ that is the dominating source of measurement outages.

B. Mean Square Error

We have seen that partial support recovery increases the reliability of detection by neglecting some of the entries in $\mathbf{X}$. In this section, we evaluate the implications that this has in terms of the estimation MSE when $\mathbf{X}$ is random. For any $\gamma \in (0, 1]$, consider the expectation of the MSE

$$\text{mse}_\gamma \triangleq \mathbb{E}_{\mathbf{X}} \{\text{mse}(\mathbf{X})\} \quad (18)$$

taken over the distribution of $\mathbf{X}$. The results in Section III are a valid characterization for those realizations $\mathbf{x} \in \mathcal{C}^n(\mathbf{r}, \gamma, k_{\max})$. Establishing an MSE description for those $\mathbf{x} \notin \mathcal{C}^n(\mathbf{r}, \gamma, k_{\max})$ would require a complete description of the support recovery map, even when it errs. Such description is beyond the scope of this work. Instead, we establish a bound on the MSE and study its behavior.

As discussed before, a (partial) support recovery map has no means of detecting whether the estimate of the support set is correct or not. Thus, in the event of a support recovery error,

Fig. 4. MSE bound $\overline{\text{mse}}_\gamma$ as a function of the measurement SNR for estimation using partial support recovery of 90%, 95%, 99%, and 99.9% of the signal power, and for complete support recovery of a Gaussian $\mathbf{W}$.

the squared error can be arbitrarily large for a given realization of $\mathbf{x}$. Nevertheless, the following proposition establishes that the MSE averaged over the distribution of $\mathbf{X}$ is bounded. Again, we make the additional assumption that the size of the support set is bounded by some arbitrary but fixed and known $k_{\max}$.

Proposition 5. *Let $\mathbf{X}$ be a k -sparse random vector with a distribution that has bounded support. For any measurement rate $\mathbf{r}$ there exists a sequence of estimates of $\mathbf{X}$ with average MSE that satisfies $\limsup_{n \rightarrow \infty} \text{mse}_\gamma \leq \overline{\text{mse}}_\gamma$ where*

$$\overline{\text{mse}}_\gamma \triangleq \int_{\mathcal{C}^c} \|\mathbf{x}_{\mathcal{S}_\gamma}\|^2 f(\mathbf{x}) d\mathbf{x} + \int_{\mathcal{C}} \|\mathbf{x}\|^2 f(\mathbf{x}) d\mathbf{x}, \quad (19)$$

with $\mathcal{C} \triangleq \mathcal{C}^n(\mathbf{r}, \gamma, k_{\max})$ and its complement $\mathcal{C}^c$. $\square$

Proof: The proof is provided in Appendix E. $\blacksquare$

The first integral in (19) corresponds to the mean square error incurred in the estimation of those vectors whose γ -support sets are correctly detected. The second integral in (19) corresponds to the estimation error otherwise and amounts to the whole contribution of $\mathbf{x}$. That is, even though the estimator cannot know that $\mathcal{S}_\gamma$ was not correctly detected, the resulting MSE contribution does not exceed $\|\mathbf{x}\|^2$ on average.

In Fig. 4, we show the behavior of $\overline{\text{mse}}_\gamma$ for the setting introduced in Section IV-A (i.e., entries $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_k)$ for $k = 10$) for different values of the measurement SNR. We observe the following remarkable facts. First, if $\overline{\text{mse}}_\gamma$ is the metric of interest, insisting on complete support recovery (i.e., $\gamma = 1$) is detrimental; the same performance can be attained at a much lower SNR through partial support recovery by selecting the parameter γ properly. Second, observe again that for $\gamma < 1$, the curves flatten off at high SNR. Two factors contribute to this: the residual error corresponding to the discarded terms (i.e., the first integral in (19)) and the measurement outages (i.e., the second integral in (19)), which are insensitive to the growth in SNR, as discussed before.

Finally, we note that the performance in terms of P_o and $\overline{\text{mse}}_\gamma$ does not vary substantially if other distributions for $\mathbf{W}$ are considered. As long as there is a non-negligible probability of having small, non-zero entries, the description here applies.

V. SUMMARY AND CONCLUDING REMARKS

In this paper, we have studied the fundamental tradeoff between the measurement rate and the estimation mean square error in compressive sensing. First, we have derived guarantees for asymptotically reliable recovery of a fraction of the support set. Then, we have derived an expression for the mean square error that is also achievable in the asymptotic regime. The combination of these two results allowed us to determine an achievable measurement rate-MSE region.

Our results show that in general, methods based on partial support recovery can significantly reduce the measurement rate whenever there are some signal components that are distinctly smaller than the rest. Furthermore, the degradation in terms of relative MSE is negligible in these cases. When applied to the estimation of random signals, partial support recovery methods can drastically reduce the occurrence of measurement outages at a minimal cost in terms of MSE. Conversely, large savings in SNR are possible by using partial support recovery-based estimation.

ACKNOWLEDGEMENT

We would like to thank the associate editor and the reviewers for their insightful comments that have greatly improved this paper. We would also like to thank Mattias Andersson for many discussions regarding the contents of the paper.

APPENDIX A PROOFS FOR PARTIAL SUPPORT RECOVERY

A. Proof of Proposition 1

The proof of Proposition 1 uses the random coding method of the proof of [25, Theorem 1]. However, we encounter the following differences:

- 1) The size of the γ -support set depends on the value of γ . Thus, the support recovery map needs to estimate this parameter as well. Our approach is to attempt to detect the γ -support set in an iterative fashion, increasing the size of the candidate set in each iteration.
- 2) Complete support recovery (i.e., [25, Theorem 1]) uses a detection method that only depends on the statistics of the noise and of the entries in the measurement matrix, in particular their second moments. In contrast, partial support recovery needs some knowledge about the power of the signal to be estimated.

Proof of Proposition 1: Let $\gamma \in (0, 1]$ and let $\epsilon_n, \epsilon_{1,n}$ be two sequences of positive numbers such that $\epsilon_n \rightarrow 0$ and $\epsilon_{1,n} \rightarrow 0$ as $n \rightarrow \infty$. These two sequences cannot be arbitrary; we will characterize them in the following. A summary of their properties is included at the end of this section.

Support recovery map: Our support recovery map $d_\gamma^{(n)}$ is the following variation of that described in [25]. Given the vector of measurements $\mathbf{Y}^{(n)}$ and the matrix $\Phi^{(n)}$:

- 1) Form an estimate of $\|\mathbf{w}\|$ as

$$\hat{R} = \sqrt{\frac{\frac{1}{m_n} \|\mathbf{Y}\|^2 - P_Z}{P_\Phi}}.$$

Note that since $\mathbf{x}$ is k -sparse we have that $\|\mathbf{x}\| = \|\mathbf{w}\|$ and thus $\hat{R}$ is also an estimate of $\|\mathbf{x}\|$.

- 2) Repeat for $l = 1, \dots, k_{max}$, in increasing order:

- a) Let $\mathcal{B}_l(\hat{R} + \epsilon_{1,n})$ be the l -dimensional ball of radius $\hat{R} + \epsilon_{1,n}$ with respect to the L_2 norm:

$$\mathcal{B}_l(\hat{R} + \epsilon_{1,n}) \triangleq \{\mathbf{b} \in \mathbb{R}^l : \|\mathbf{b}\| \leq \hat{R} + \epsilon_{1,n}\}.$$

For the given $\epsilon_{1,n}$, consider the (non-unique) sets of points in $\mathcal{B}_l(\hat{R} + \epsilon_{1,n})$ such that l -dimensional balls of radius $\epsilon_{1,n}$ centered on these points cover the whole ball $\mathcal{B}_l(\hat{R} + \epsilon_{1,n})$. Let $\mathcal{Q}_l(\hat{R} + \epsilon_{1,n}, \epsilon_{1,n})$ be one such set that has the smallest number of points. That is, for every $\mathbf{b} \in \mathcal{B}_l(\hat{R} + \epsilon_{1,n})$, there exists at least one $\hat{\mathbf{w}} \in \mathcal{Q}_l(\hat{R} + \epsilon_{1,n}, \epsilon_{1,n})$ such that $\|\mathbf{b} - \hat{\mathbf{w}}\|^2 \leq \epsilon_{1,n}$ and $|\mathcal{Q}_l(\hat{R} + \epsilon_{1,n}, \epsilon_{1,n})|$ is minimal. Observe that $\mathcal{B}_l$ is compact and, thus, a minimal set $\mathcal{Q}_l$ exists [39].

- b) Find a set $\mathcal{T} \subseteq \{1, \dots, n\}$ of size l such that

$$\frac{1}{m_n} \left\| \mathbf{Y} - \sum_{i=1}^l \hat{W}_i \Phi_{t_i}^{(n)} \right\|^2 \leq (1 - \gamma) \hat{R}^2 P_\Phi + P_Z + \epsilon_n \quad (20)$$

for some $\hat{\mathbf{W}} = [\hat{W}_1, \dots, \hat{W}_l]^T \in \mathcal{Q}_l(\hat{R} + \epsilon_{1,n}, \epsilon_{1,n})$, where $[t_1, \dots, t_l]$ is the natural ordering of the elements in $\mathcal{T}$ and $\Phi_{t_i}^{(n)}$ is the column of $\Phi^{(n)}$ in position t_i . The process stops when the first set $\mathcal{T}$ that satisfies (20) is found. This set is the desired estimate (i.e., $\hat{\mathcal{S}}_\gamma = \mathcal{T}$). If no such set of size l is found, increase l and start again.

For each l , $\mathcal{Q}_l(\hat{R} + \epsilon_{1,n}, \epsilon_{1,n})$ is a set of length- l vectors that are estimates of the first l components of $\mathbf{w}$ (i.e., $\mathbf{w}_l^T$) in arbitrary order. In Appendix G we have included some basic properties of the sets $\mathcal{Q}_l$ that are used in this proof. In particular, note that the size of the sets may grow as $\epsilon_{1,n} \rightarrow 0$. As we will see, $\epsilon_{1,n}$ has to decay slowly enough to ensure that this growth is asymptotically negligible (in terms of error probability).

Error analysis: Let $\mathbf{x}^n(\mathbf{w}, \mathcal{S}) \in \mathcal{C}^n(\mathbf{r}, \gamma, k_{max})$. Given the measurement model in (2), we define

$$\mathcal{E} \triangleq \{d_\gamma^{(n)}(\mathbf{Y}, \Phi^{(n)}) \notin \mathcal{S}_\gamma\}.$$

Thus, $\Pr(\mathcal{E})$ corresponds to the error probability in (3). We want to show that the probability of error of the support recovery map described above can be made arbitrarily small by choosing a sufficiently large n . Let $\mathcal{I}_{\epsilon_{1,n}}$ denote the interval $(-\epsilon_{1,n}, \epsilon_{1,n})$ and consider the following events:

$$\mathcal{E}_\mathcal{T} \triangleq \left\{ \exists \hat{\mathbf{W}} \in \mathcal{Q}_l(\hat{R} + \epsilon_{1,n}, \epsilon_{1,n}) \text{ such that (20) holds} \right\},$$

defined for any $\mathcal{T} \subseteq \{1, \dots, n\}$, and

$$\begin{aligned} \mathcal{E}_{aux} \triangleq & \left\{ \hat{R}^2 - \|\mathbf{w}\|^2 \in \mathcal{I}_{\epsilon_{1,n}} \right\} \cap \left\{ \bigcap_{i \in \mathcal{S}} \left\{ \frac{\mathbf{Z}^{(n)T} \Phi_i^{(n)}}{m_n} \in \mathcal{I}_{\epsilon_{1,n}} \right\} \right\} \\ & \cap \left\{ \left\{ \frac{\|\mathbf{Z}^{(n)}\|^2}{m_n} - P_Z \in \mathcal{I}_{\epsilon_{1,n}} \right\} \cap \left\{ \frac{\|\Phi^{(n)}\|^2}{nm_n} - P_\Phi \in \mathcal{I}_{\epsilon_{1,n}} \right\} \right\} \\ & \cap \left\{ \bigcap_{i \in \mathcal{S}} \bigcap_{\substack{j \in \mathcal{S} \\ j \neq i}} \left\{ \frac{\Phi_i^{(n)T} \Phi_j^{(n)}}{m_n} \in \mathcal{I}_{\epsilon_{1,n}} \right\} \right\}. \end{aligned}$$

For convenience, in the following we drop the superindex (n) in the random vectors. Using these events, we obtain

$$\Pr(\mathcal{E}) \leq \Pr\left(\bigcup_{l=1}^{\lambda} \left(\bigcup_{\substack{\mathcal{T}: |\mathcal{T}|=l \\ \mathcal{T} \notin \mathcal{S}_\gamma}} \mathcal{E}_\mathcal{T}\right) \cup \left(\bigcup_{\mathcal{T} \in \mathcal{S}_\gamma} \mathcal{E}_\mathcal{T}\right)^c\right).$$

Using basic set operations and bounds, it is easy to show that

$$\begin{aligned} \Pr(\mathcal{E}) &\leq \Pr(\mathcal{E}_{aux}^c) \\ &+ \underbrace{\sum_{\mathcal{T} \in \mathcal{S}_\gamma} \Pr(\mathcal{E}_\mathcal{T}^c \cap \mathcal{E}_{aux})}_{P_2} + \underbrace{\sum_{l=1}^{\lambda} \sum_{\substack{\mathcal{T}: |\mathcal{T}|=l \\ \mathcal{T} \notin \mathcal{S}_\gamma}} \Pr(\mathcal{E}_\mathcal{T} \cap \mathcal{E}_{aux})}_{P_3}. \end{aligned} \quad (21)$$

The first term in (21) is an upper bound on the probability that the realization of the measurement matrix or the noise significantly deviate from their expected behavior. Using Chernoff bounds, it is easy to show that if $\epsilon_{1,n}$ decays slowly enough, then $\Pr(\mathcal{E}_{aux}^c) \leq o(1/m_n)$. Each of the summands in P_2 is the probability that a γ -support set (i.e., $\mathcal{T} \in \mathcal{S}_\gamma$ and thus $|\mathcal{T}| = \lambda$) does not satisfy (20). We have that

$$\Pr(\mathcal{E}_\mathcal{T}^c \cap \mathcal{E}_{aux}) \leq 1 - \Pr(\mathcal{E}_\mathcal{T} | \mathcal{E}_{aux}).$$

Note that in event $\mathcal{E}_\mathcal{T}$, the right-hand side of (20) depends on the random estimate $\hat{R}$. Our first step is to obtain a lower bound on $\Pr(\mathcal{E}_\mathcal{T} | \mathcal{E}_{aux})$ that depends instead on the deterministic quantity $\|\mathbf{w}\|$. Consider the following inequality

$$\frac{1}{m_n} \left\| \mathbf{Y} - \sum_{i=1}^{\lambda} \hat{W}_i \Phi_{t_i} \right\|^2 \leq (1-\gamma)(\|\mathbf{w}\|^2 - \epsilon_{1,n})P_\Phi + P_Z + \epsilon_n \quad (22)$$

where the right-hand side is now deterministic. We have that

$$\Pr(\mathcal{E}_\mathcal{T} | \mathcal{E}_{aux}) \geq \Pr\left(\exists \hat{\mathbf{W}} \in \mathcal{Q}_\lambda(\hat{R} + \epsilon_{1,n}, \epsilon_{1,n}) \text{ s.t. (22)} \middle| \mathcal{E}_{aux}\right). \quad (23)$$

To establish (23) we have used that $\|\mathbf{w}\|^2 - \epsilon_{1,n} < \hat{R}^2$ by the condition $\mathcal{E}_{aux}$. In Appendix A-B, we show that

$$\Pr\left(\exists \hat{\mathbf{W}} \in \mathcal{Q}_\lambda(\hat{R} + \epsilon_{1,n}, \epsilon_{1,n}) \text{ s.t. (22) holds} \middle| \mathcal{E}_{aux}\right) = 1 \quad (24)$$

if $\epsilon_n > \delta^{(n)}(\epsilon_{1,n})$, where $\delta^{(n)}(\epsilon_{1,n})$ is a positive function of $\epsilon_{1,n}$ given in Appendix A-B. Moreover, $\delta^{(n)}(\epsilon_{1,n}) \rightarrow 0$ as $\epsilon_{1,n} \rightarrow 0$. Consequently, $\Pr(\mathcal{E}_\mathcal{T}^c \cap \mathcal{E}_{aux}) = 0$ for $\mathcal{T} \in \mathcal{S}_\gamma$ and, thus, $P_2 = 0$ if $\epsilon_n > \delta^{(n)}(\epsilon_{1,n})$.

We encounter a similar problem with the random threshold when upper bounding each of the terms in the double summation defining P_3 . Consider the inequality

$$\frac{1}{m_n} \left\| \mathbf{Y} - \sum_{i=1}^l \hat{W}_i \Phi_{t_i} \right\|^2 \leq (1-\gamma)(\|\mathbf{w}\|^2 + \epsilon_{1,n})P_\Phi + P_Z + \epsilon_n \quad (25)$$

We have that

$$\begin{aligned} \Pr(\mathcal{E}_\mathcal{T} \cap \mathcal{E}_{aux}) &\leq \Pr(\mathcal{E}_\mathcal{T} | \mathcal{E}_{aux}) \\ &\leq \Pr\left(\exists \hat{\mathbf{W}} \in \mathcal{Q}_l(\hat{R} + \epsilon_{1,n}, \epsilon_{1,n}) \text{ s.t. (25) holds} \middle| \mathcal{E}_{aux}\right). \end{aligned} \quad (26)$$

To establish (26) we have used that $\hat{R}^2 < \|\mathbf{w}\|^2 + \epsilon_{1,n}$ by the condition $\mathcal{E}_{aux}$. In Appendix A-C, we show that for sufficiently large n ,

$$\Pr(\mathcal{E}_\mathcal{T} | \mathcal{E}_{aux}) \leq q_l(\epsilon_{1,n}) 2^{-\frac{m}{2} \log_2 \left(\frac{P_\Phi \sum_{j=d+1}^k w_j^2 + P_Z - \delta_1(\epsilon_{1,n})}{(1-\gamma)(\|\mathbf{w}\|^2 + \epsilon_{1,n})P_\Phi + P_Z + \epsilon_n} \right)} \quad (27)$$

where $l = |\mathcal{T}|$, d is the number of correct estimates in the set $\mathcal{T}$ (i.e., $d = |\mathcal{T} \cap \mathcal{S}|$), $q_l(\epsilon_{1,n}) \triangleq |\mathcal{Q}_l(\|\mathbf{w}\| + \delta_2(\epsilon_{1,n}), \epsilon_{1,n})|$, $l \in \{1, \dots, \lambda\}$, and $\delta_1(\epsilon_{1,n})$ and $\delta_2(\epsilon_{1,n})$ are positive functions of $\epsilon_{1,n}$ that tend to 0 as $\epsilon_{1,n} \rightarrow 0$.

We emphasize that (27) is only valid for sufficiently large n . This is a consequence of the fact that, without knowledge on the structure of $\mathbf{w}$, we cannot choose a fixed threshold ϵ_n that discriminates between correct and incorrect estimates of γ -support sets for all possible $\mathbf{w}$ and γ . Therefore, the rest of the proof assumes implicitly that n is sufficiently large.

The exponential bound in (27) only depends on $\mathcal{T}$ through d . Thus, we can replace the second summation in the definition of P_3 that runs over sets with a fixed cardinality by a summation that runs over the parameter d . For given k and n , the support recovery maps tries not more than

$$d! \binom{k}{d} \binom{n-k}{l-d} (l-d)!$$

different sets $\mathcal{T}$ with $|\mathcal{T}| = l$ and $|\mathcal{T} \cap \mathcal{S}| = d$. Thus,

$$P_3 \leq \sum_{l=1}^{\lambda} \sum_{d=0}^l d! \binom{k}{d} \binom{n-k}{l-d} (l-d)! \Pr(\mathcal{E}_\mathcal{T} \cap \mathcal{E}_{aux}).$$

Note that by Lemma 3-2 in Appendix G we have that $q_l(\epsilon_{1,n}) \leq q_\lambda(\epsilon_{1,n})$ for all $l \in \{1, \dots, \lambda\}$. For the sake of compactness we define

$$c(\epsilon_{1,n}) \triangleq (\lambda!)^2 \binom{k}{\lfloor \lambda/2 \rfloor} q_\lambda(\epsilon_{1,n})$$

that groups and upper bounds all terms that are independent of m_n and n . We obtain the following upper bound on P_3

$$c(\epsilon_{1,n}) \sum_{l=1}^{\lambda} \sum_{d=0}^l \binom{n}{l-d} 2^{-\frac{m_n}{2} \log_2 \left(\frac{P_\Phi \sum_{j=d+1}^k w_j^2 + P_Z - \delta_1(\epsilon_{1,n})}{(1-\gamma)(\|\mathbf{w}\|^2 + \epsilon_{1,n})P_\Phi + P_Z + \epsilon_n} \right)}.$$

Now, using the change of variables $i = l - d$, and the bound $\binom{n}{i} \leq 2^{i \log_2 n}$ we upper bound P_3 further by

$$c(\epsilon_{1,n}) \sum_{l=1}^{\lambda} \sum_{i=0}^l 2^{i \log_2 n - \frac{m_n}{2} \log_2 \left(\frac{P_\Phi \sum_{j=l-i+1}^k w_j^2 + P_Z - \delta_1(\epsilon_{1,n})}{(1-\gamma)(\|\mathbf{w}\|^2 + \epsilon_{1,n})P_\Phi + P_Z + \epsilon_n} \right)}.$$

It is easy to see that in order to enforce $P_3 \rightarrow 0$ as m_n increases, it suffices to consider the terms in the sum corresponding to $l = \lambda$ only. In addition, note that $c(\epsilon_{1,n})$ has to grow slowly enough so that P_3 can be made arbitrarily small. From Lemma 3-3 in Appendix G, we know that $c(\epsilon_{1,n})$ grows with $\epsilon_{1,n}$, as $\epsilon_{1,n}^{-\lambda}$, at most. Observe that if $\epsilon_{1,n}^{-1}$ grows at most polynomially with $\frac{m_n}{\log_2 n}$, then

$$\lim_{n \rightarrow \infty} \left(\frac{1}{\epsilon_{1,n}} \right)^\lambda 2^{-\frac{m_n}{\log_2 n} a} = 0$$

for every $a > 0$. The support recovery map does not know the exact value of λ when choosing $\epsilon_{1,n}$ but it can use the conservative upper bound given by k_{max} .

Using this result, together with the fact that $\epsilon_n \rightarrow 0$ and $\epsilon_{1,n} \rightarrow 0$ as $n \rightarrow \infty$, we see that

$$\liminf_{n \rightarrow \infty} \frac{m_n}{\log_2 n} > \left(\frac{1}{2i} \log_2 \left(\frac{P_\Phi \sum_{j=\lambda-i+1}^k w_j^2 + P_Z}{(1-\gamma) \|\mathbf{w}\|^2 P_\Phi + P_Z} \right) \right)^{-1}$$

for all $i \in \{1, \dots, \lambda\}$ is a sufficient condition for recovery of a γ -support set with $P_3 \rightarrow 0$ exponentially as $m_n \rightarrow \infty$. This is satisfied for every $\mathbf{x}^n(\mathbf{w}, \mathcal{S}) \in \mathcal{C}^n(\mathbf{r}, \gamma, k_{max})$. Collecting the bounds for $\Pr(\mathcal{E}_{aux}^c)$, P_2 , and P_3 , we conclude that $\Pr(\mathcal{E}) \leq o(1/m_n)$. This completes the proof. ■

To summarize the properties of $\epsilon_{1,n}$ and ϵ_n : the sequence $\epsilon_{1,n} > 0$ is chosen to decay to 0 slow enough with n so that: i) $\Pr(\mathcal{E}_{aux}^c) \leq o(1/m_n)$ and ii) $(\epsilon_{1,n})^{-k_{max}}$ grows at most polynomially with $\frac{m_n}{\log_2 n}$. The sequence $\epsilon_n > 0$ also decays to 0 and must satisfy $\epsilon_n > \delta^{(n)}(\epsilon_{1,n})$ for each n , where $\delta^{(n)}(\epsilon_{1,n})$ is given in Appendix A-B.

B. Proof of (24)

Consider the following sets:

$$\mathcal{D} \triangleq \mathcal{T} \cap \mathcal{S}, \quad (28)$$

$$\mathcal{F} \triangleq \mathcal{T} \cap \mathcal{S}^c, \quad (29)$$

$$\mathcal{U} \triangleq \mathcal{S} \setminus \mathcal{D} = \mathcal{S} \setminus \mathcal{T}. \quad (30)$$

That is, $\mathcal{D}$ contains the elements in $\mathcal{T}$ that are correct guesses (i.e., they belong to the support set $\mathcal{S}$), $\mathcal{F}$ contains the rest of elements of $\mathcal{T}$ (i.e., wrong guesses), and $\mathcal{U}$ contains the elements in the support set that are not in $\mathcal{T}$ (i.e., undetected). Observe that for $\mathcal{T} \in \mathcal{S}_\gamma$, we have $\mathcal{F} = \emptyset$ and $|\mathcal{T}| = \lambda$.

Given a set $\mathcal{T}$ and a vector $\hat{\mathbf{w}} \in \mathcal{Q}_\lambda$, we construct $\hat{\mathbf{x}}$ component-wise as follows

$$\hat{x}_i = \begin{cases} \hat{w}_j & \text{if } i = t_j, \\ 0 & \text{otherwise,} \end{cases} \quad (31)$$

where $t_1 < t_2 < \dots$ is the natural ordering of the elements in $\mathcal{T}$. Let

$$C_1 \triangleq \frac{1}{m_n} \left\| \sum_{j \in \mathcal{T}} (x_j - \hat{X}_j) \Phi_j + \sum_{j \in \mathcal{U}} x_j \Phi_j + \mathbf{Z} \right\|^2$$

where $\hat{\mathbf{W}}$ and $\hat{\mathbf{X}}$ are related as in (31). Given $\mathcal{E}_{aux}$, we know that $\|\mathbf{w}\|^2 < \hat{R}^2 + \epsilon_{1,n}$. Thus, by construction, there will be at least one $\hat{\mathbf{w}} \in \mathcal{Q}_\lambda(\hat{R} + \epsilon_{1,n}, \epsilon_{1,n})$ such that

$$\sum_{j \in \mathcal{T}} (x_j - \hat{x}_j)^2 \leq \epsilon_{1,n},$$

where $\hat{\mathbf{x}}$ is constructed from $\hat{\mathbf{w}}$ and $\mathcal{T}$ as in (31). For this $\hat{\mathbf{x}}$, if we expand the squared norm in C_1 and use the condition $\mathcal{E}_{aux}$, we see that C_1 is in the interval

$$\left(P_\Phi \sum_{j \in \mathcal{U}} x_j^2 + P_Z - \delta_3(\epsilon_{1,n}), P_\Phi \sum_{j \in \mathcal{U}} x_j^2 + P_Z + \delta_3(\epsilon_{1,n}) \right)$$

where the function $\delta_3(\epsilon_{1,n})$ is given by

$$t \left(1 + P_\Phi + 4k_{max}^2 \left(\hat{R}^2 + 2t + \sqrt{t} + (1 + \sqrt{t}) \sqrt{\hat{R}^2 + t} \right) \right)$$

for $t = \epsilon_{1,n}$. Observe that $\delta_3(\epsilon_{1,n})$ is a positive function of $\epsilon_{1,n} > 0$ such that $\delta_3(\epsilon_{1,n}) \rightarrow 0$ as $\epsilon_{1,n} \rightarrow 0$. Moreover, it only depends on parameters that are known by the support recovery map and, thus, can be constructed.

Now, let $\delta^{(n)}(\epsilon_{1,n}) \triangleq \delta_3(\epsilon_{1,n}) + (1 - \gamma)P_\Phi \epsilon_{1,n}$. Observe that if $\epsilon_n > \delta^{(n)}(\epsilon_{1,n})$, then

$$\begin{aligned} C_1 &\leq P_\Phi \sum_{j \in \mathcal{U}} x_j^2 + P_Z + \delta^{(n)}(\epsilon_{1,n}) \\ &\leq (1 - \gamma)(\|\mathbf{w}\|^2 - \epsilon_{1,n})P_\Phi + P_Z + \epsilon_n \end{aligned}$$

always, because

$$P_\Phi \sum_{j \in \mathcal{U}} x_j^2 < (1 - \gamma) \|\mathbf{w}\|^2 P_\Phi$$

if $\mathcal{T} \in \mathcal{S}_\gamma$. The condition $\epsilon_n > \delta^{(n)}(\epsilon_{1,n})$ expresses the fact that the sensitivity threshold of the support recovery map has to be large enough to allow for the small deviations in the power of the measurement matrices, noise, etc.

C. Proof of (27)

The proof of (27) follows closely the steps in [25, Theorem 1]. We are interested in bounding

$$\Pr \left(\exists \hat{\mathbf{W}} \in \mathcal{Q}_l(\hat{R} + \epsilon_{1,n}, \epsilon_{1,n}) \text{ s.t. (25) holds} \middle| \mathcal{E}_{aux} \right) \quad (32)$$

for $\mathcal{T} \notin \mathcal{S}_\gamma$ with $|\mathcal{T}| \leq \lambda$. Consider the sets $\mathcal{D}, \mathcal{F}$, and $\mathcal{U}$, defined in (28), (29), (30), respectively. Consider also the event

$$\mathcal{E}_{cond} \triangleq \left(\bigcap_{i \in \mathcal{S}} \{\Phi_i = \phi_i\} \right) \cap \{\mathbf{Z} = \mathbf{z}\} \cap \mathcal{E}_{aux}$$

and note that we can write (32) as

$$\begin{aligned} &\int \overbrace{\Pr \left(\exists \hat{\mathbf{W}} \in \mathcal{Q}_l(\hat{R} + \epsilon_{1,n}, \epsilon_{1,n}) \text{ s.t. (25) holds} \middle| \mathcal{E}_{cond} \right)}^{P_4} \\ &\quad \times f(\phi_1, \dots, \phi_k, \mathbf{z} | \mathcal{E}_{aux}) d\phi_1 \dots d\phi_k d\mathbf{z}, \end{aligned} \quad (33)$$

where $f(\cdot)$ is the relevant probability density function. We concentrate on P_4 and distinguish two cases: i) $\mathcal{D} = \mathcal{T}$ and ii) $\mathcal{D} \neq \mathcal{T}$, which we analyze in the following. Recall that for given $\mathcal{T}$, we construct $\hat{\mathbf{x}}$ from $\hat{\mathbf{w}}$ as in (31) and similarly for $\hat{\mathbf{W}}$ and $\hat{\mathbf{X}}$.

Case 1: $\mathcal{D} = \mathcal{T}$: In this case, all elements are correct estimates (i.e., $\mathcal{T} \subseteq \mathcal{S}$) but they do not encompass enough power: $\sum_{i \in \mathcal{T}} x_i^2 < \gamma \|\mathbf{w}\|^2$. First, let

$$C_3 \triangleq \frac{1}{m_n} \left\| \sum_{j \in \mathcal{D}} (x_j - \hat{X}_j) \phi_j + \sum_{j \in \mathcal{U}} x_j \phi_j + \mathbf{z} \right\|^2.$$

We have that

$$\begin{aligned}
\Pr(\mathcal{E}_T | \mathcal{E}_{cond}) &= \Pr(\exists \hat{\mathbf{W}} \in \mathcal{Q}_l(\hat{R} + \epsilon_{1,n}, \epsilon_{1,n}) \text{ s.t.} \\
&\quad C_3 \leq (1 - \gamma) \hat{R}^2 P_\Phi + P_Z + \epsilon_n | \mathcal{E}_{cond}) \\
&\leq \Pr(\exists \hat{\mathbf{W}} \in \mathcal{Q}_l(\|\mathbf{w}\| + \delta_2(\epsilon_{1,n}), \epsilon_{1,n}) \text{ s.t.} \\
&\quad C_3 \leq (1 - \gamma)(\|\mathbf{w}\|^2 + \epsilon_{1,n}) P_\Phi + P_Z + \epsilon_n | \mathcal{E}_{cond}) \\
&\leq \sum_{\hat{\mathbf{w}}} \Pr(C_4 \leq (1 - \gamma)(\|\mathbf{w}\|^2 + \epsilon_{1,n}) P_\Phi + P_Z + \epsilon_n | \mathcal{E}_{cond})
\end{aligned} \tag{34}$$

with the summation over $\hat{\mathbf{w}} \in \mathcal{Q}_l(\|\mathbf{w}\| + \delta_2(\epsilon_{1,n}), \epsilon_{1,n})$ and

$$\begin{aligned}
\delta_2(\epsilon_{1,n}) &\triangleq \epsilon_{1,n} + \sqrt{\|\mathbf{w}\|^2 + \epsilon_{1,n} - \|\mathbf{w}\|}, \\
C_4 &\triangleq \frac{1}{m_n} \left\| \sum_{j \in \mathcal{D}} (x_j - \hat{x}_j) \phi_j + \sum_{j \in \mathcal{U}} x_j \phi_j + \mathbf{z} \right\|^2.
\end{aligned}$$

Thus, all terms inside the probability expression for P_4 are deterministic quantities. Moreover, if we express the norm in terms of all cross products and use the properties of the event $\mathcal{E}_{aux}$, we see that C_4 is in the interval

$$\left(P_\Phi \sum_{j \in \mathcal{U}} x_j^2 + P_Z - \delta_4(\epsilon_{1,n}), P_\Phi \sum_{j \in \mathcal{U}} x_j^2 + P_Z + \delta_4(\epsilon_{1,n}) \right)$$

where $\delta_4(\epsilon_{1,n})$ is a positive function of $\epsilon_{1,n}$ such that $\delta_4(\epsilon_{1,n}) \rightarrow 0$ as $\epsilon_{1,n} \rightarrow 0$. Now, write

$$\begin{aligned}
P_\Phi \sum_{j \in \mathcal{U}} x_j^2 + P_Z - \delta_4(\epsilon_{1,n}) &\leq C_4 \\
&\leq (1 - \gamma) \|\mathbf{w}\|^2 P_\Phi + P_Z + \epsilon_n
\end{aligned}$$

and note that the second inequality will not be satisfied for sufficiently large n because $\epsilon_{1,n} \rightarrow 0$ and $\epsilon_n \rightarrow 0$ and

$$P_\Phi \sum_{j \in \mathcal{U}} x_j^2 > (1 - \gamma) \|\mathbf{w}\|^2 P_\Phi.$$

Thus, for sufficiently large n , we have that $\Pr(\mathcal{E}_T | \mathcal{E}_{cond}) = 0$ for Case 1.

Case 2: $\mathcal{D} \neq \mathcal{T}$: In this case, some of the elements in $\mathcal{T}$ are incorrect estimates. The proof relies on the following Lemma from [24].

Lemma 1. *Let $0 < \beta < \alpha$. Let $\mathbf{u} \in \mathbb{R}^m$ be a vector such that $\frac{1}{m} \|\mathbf{u}\|^2 \in (\alpha - \beta, \alpha + \beta)$. Let $\mathbf{V} \in \mathbb{R}^m$ be a random vector with i.i.d. $V_i \sim \mathcal{N}(0, \sigma_v^2)$. Then, for any $\varpi \in (0, \alpha - \beta)$,*

$$\Pr\left(\frac{1}{m} \|\mathbf{u} - \mathbf{V}\|^2 \leq \varpi\right) \leq 2^{-\frac{m}{2} \log\left(\frac{\alpha - \beta}{\varpi}\right)}.$$

□

Let

$$C_5 \triangleq \frac{1}{m_n} \left\| \sum_{j \in \mathcal{D}} (x_j - \hat{X}_j) \phi_j + \sum_{j \in \mathcal{U}} x_j \phi_j - \sum_{j \in \mathcal{F}} \hat{X}_j \Phi_j + \mathbf{z} \right\|^2.$$

Following similar steps to (34)-(35), we obtain

$$\begin{aligned}
&\Pr(\mathcal{E}_T | \mathcal{E}_{cond}) \\
&\leq \sum_{\hat{\mathbf{w}}} \Pr\left(C_6 \leq (1 - \gamma)(\|\mathbf{w}\|^2 + \epsilon_{1,n}) P_\Phi + P_Z + \epsilon_n | \mathcal{E}_{cond}\right)
\end{aligned}$$

with the summation over $\hat{\mathbf{w}} \in \mathcal{Q}_l(\|\mathbf{w}\| + \delta_2(\epsilon_{1,n}), \epsilon_{1,n})$ and

$$C_6 \triangleq \frac{1}{m_n} \left\| \sum_{j \in \mathcal{D}} (x_j - \hat{x}_j) \phi_j + \sum_{j \in \mathcal{U}} x_j \phi_j - \sum_{j \in \mathcal{F}} \hat{x}_j \Phi_j + \mathbf{z} \right\|^2.$$

We now use Lemma 1 with

$$\begin{aligned}
\mathbf{u} &= \sum_{j \in \mathcal{D}} (x_j - \hat{x}_j) \phi_j + \sum_{j \in \mathcal{U}} x_j \phi_j + \mathbf{z}, \\
\mathbf{V} &= - \sum_{j \in \mathcal{F}} \hat{x}_j \Phi_j.
\end{aligned}$$

For this choice of $\mathbf{u}$, we have $\frac{1}{m_n} \|\mathbf{u}\|^2 \in (\alpha - \beta, \alpha + \beta)$ with

$$\begin{aligned}
\alpha &= \left(\sum_{j \in \mathcal{D}} (x_j - \hat{x}_j)^2 + \sum_{j \in \mathcal{U}} x_j^2 \right) P_\Phi + P_Z, \\
\beta &= \delta_1(\epsilon_{1,n}), \\
\varpi &= (1 - \gamma)(\|\mathbf{w}\|^2 + \epsilon_{1,n}) P_\Phi + P_Z + \epsilon_n,
\end{aligned}$$

where $\delta_1(\epsilon_{1,n})$ is a positive function of $\epsilon_{1,n}$ such that $\delta_1(\epsilon_{1,n}) \rightarrow 0$ as $\epsilon_{1,n} \rightarrow 0$. Observe also that for sufficiently large n ,

$$\varpi < \sum_{j \in \mathcal{U}} x_j^2 P_\Phi + P_Z - \delta_1(\epsilon_{1,n}) \leq \alpha - \beta,$$

as required by Lemma 1. We further see that

$$\alpha \geq P_\Phi \sum_{j \in \mathcal{U}} x_j^2 + P_Z \geq P_\Phi \sum_{j=|\mathcal{D}|+1}^k w_j^2 + P_Z.$$

We have used the fact that the sum of the magnitudes of any set of $k - |\mathcal{D}|$ non-zero entries in $\mathbf{x}$ must be no less than the sum of the $k - |\mathcal{D}|$ smallest-magnitude entries. These correspond to the last $k - |\mathcal{D}|$ entries in $\mathbf{w}$. Thus,

$$\begin{aligned}
P_4 &\leq \sum_{\hat{\mathbf{w}}} 2^{-\frac{m_n}{2} \log\left(\frac{\alpha - \beta}{\varpi}\right)} \\
&\leq \sum_{\hat{\mathbf{w}}} 2^{-\frac{m_n}{2} \log\left(\frac{P_\Phi \sum_{j=|\mathcal{D}|+1}^k w_j^2 + P_Z - \delta_1(\epsilon_{1,n})}{(1 - \gamma)(\|\mathbf{w}\|^2 + \epsilon_{1,n}) P_\Phi + P_Z + \epsilon_n}\right)}
\end{aligned}$$

where the summations are take over all $\hat{\mathbf{w}} \in \mathcal{Q}_l(\|\mathbf{w}\| + \delta_2(\epsilon_{1,n}), \epsilon_{1,n})$. Since the exponential term does not depend on $\hat{\mathbf{w}}$, we just need to multiply it by the number of terms in the sum, which is $|\mathcal{Q}_l(\|\mathbf{w}\| + \delta_2(\epsilon_{1,n}), \epsilon_{1,n})|$. Finally, note that this bound does not depend on any of the integration variables in (33). Combining Cases 1 and 2, (27) follows immediately.

D. Proof of Corollary 1

Proof: Consider a variation of the support recovery method used in the proof of Proposition 1 that, instead of stopping after finding the first set $\hat{\mathcal{S}}_\gamma$, it checks all the set with the same size and puts out the union of all those satisfying (20) (i.e., $\hat{\mathcal{S}} = \cup_i \hat{\mathcal{S}}_{\gamma,i}$). It is easy to see that the proof of Proposition 1 applies to this method as well. ■

APPENDIX B PROOF OF PROPOSITION 2

We now study the performance in terms of the MSE of an estimator $\hat{\mathbf{X}}$ for $\mathbf{x}$ that uses a set $\mathcal{T} \subseteq \mathcal{S}$. To make explicit the dependency of $\hat{\mathbf{X}}$ on $\mathcal{T}$ we write $\hat{\mathbf{X}}(\mathcal{T})$. Recall that in our model both the noise and the measurement matrix are random. We restrict our attention to the set Ω consisting of all ϕ with full column rank and all $\mathbf{z}$. This restriction entails no loss of generality because the set of column-rank deficient ϕ has Lebesgue measure zero.

Proof of Proposition 2: Let $\mathcal{T} \subseteq \mathcal{S}$ and let $\mathcal{T}^c$ identify the rest of the support set (i.e., $\mathcal{T}^c = \mathcal{S} \setminus \mathcal{T}$). Given $\mathcal{T}$, with elements $t_1 < t_2 < \dots$, we construct the estimate $\hat{\mathbf{X}}(\mathcal{T})$ of $\mathbf{x}$ component-wise as

$$\hat{X}_i = \begin{cases} \hat{X}_{\mathcal{T},j} & \text{for } i = t_j, \\ 0 & \text{for } i \notin \mathcal{T} \end{cases}$$

for $i \in \{1, \dots, n\}$ and some estimator $\hat{\mathbf{X}}_{\mathcal{T}}$ that will be specified later. The MSE of $\hat{\mathbf{X}}(\mathcal{T})$ is defined as

$$\text{mse}(\mathbf{x}, \mathcal{T}) \triangleq \mathbb{E}_{\mathbf{Z}, \Phi} \left\{ \|\mathbf{x} - \hat{\mathbf{X}}(\mathcal{T})\|^2 \right\}$$

with the implicit assumption that the expectation, over noise and measurement matrices, is restricted to Ω . We have that

$$\text{mse}(\mathbf{x}, \mathcal{T}) = \underbrace{\mathbb{E}_{\mathbf{Z}, \Phi} \left\{ \|\mathbf{x}_{\mathcal{T}} - \hat{\mathbf{X}}_{\mathcal{T}}\|^2 \right\}}_{\text{mse}(\mathbf{x}_{\mathcal{T}})} + \|\mathbf{x}_{\mathcal{T}^c}\|^2. \quad (36)$$

Let $\text{mse}(\mathbf{x}_{\mathcal{T}})$ denote the first term in (36). We see that

$$\text{mse}(\mathbf{x}_{\mathcal{T}}) = \mathbb{E}_{\Phi_{\mathcal{T}}} \left\{ \underbrace{\mathbb{E}_{\mathbf{Z}, \Phi_{\mathcal{T}^c}} \left\{ \|\mathbf{x}_{\mathcal{T}} - \hat{\mathbf{X}}_{\mathcal{T}}\|^2 \right\}}_{\text{mse}(\mathbf{x}_{\mathcal{T}} | \Phi_{\mathcal{T}} = \phi_{\mathcal{T}})} \right\}. \quad (37)$$

To obtain (37) we have used the fact that the columns of the measurement matrix are independently generated. For the inner expectation expression, we see that conditioned on $\Phi_{\mathcal{T}} = \phi_{\mathcal{T}}$ the vector of measurements is

$$\mathbf{Y} = \phi_{\mathcal{T}} \mathbf{x}_{\mathcal{T}} + \sum_{i \in \mathcal{T}^c} x_i \Phi_i + \mathbf{Z}.$$

That is, both the noise and the residual terms (i.e., those in $\Phi_{\mathcal{T}^c}$) are independent Gaussian random processes. Regarding the residual terms, we see that

$$\mathbb{E}_{\Phi_{\mathcal{T}^c}} \left\{ \sum_{i \in \mathcal{T}^c} \sum_{j \in \mathcal{T}^c} x_i x_j \Phi_i \Phi_j^T \right\} = P_{\Phi} \|\mathbf{x}_{\mathcal{T}^c}\|^2 \mathbf{I}_m$$

and thus, the covariance matrix of the residual terms plus noise is $\tilde{\xi} \mathbf{I}_m$ with $\tilde{\xi} \triangleq P_Z + P_{\Phi} \|\mathbf{x}_{\mathcal{T}^c}\|^2$. That is, conditioned on $\Phi_{\mathcal{T}} = \phi_{\mathcal{T}}$, the estimation of $\mathbf{x}_{\mathcal{T}}$ corresponds to a linear estimation problem in white Gaussian noise. For the class of unbiased estimators the MSE satisfies:

$$\text{mse}(\mathbf{x}_{\mathcal{T}} | \Phi_{\mathcal{T}} = \phi_{\mathcal{T}}) \geq \tilde{\xi} \text{tr} \left\{ (\phi_{\mathcal{T}}^T \phi_{\mathcal{T}})^{-1} \right\}$$

with equality for $\hat{\mathbf{X}}_{\mathcal{T}} = \phi_{\mathcal{T}}^{\dagger} \mathbf{Y}$, where

$$\phi_{\mathcal{T}}^{\dagger} \triangleq (\phi_{\mathcal{T}}^T \phi_{\mathcal{T}})^{-1} \phi_{\mathcal{T}}^T. \quad (38)$$

Thus, for this choice of estimate,

$$\text{mse}(\mathbf{x}_{\mathcal{T}}) = \tilde{\xi} \mathbb{E}_{\Phi_{\mathcal{T}}} \left\{ \text{tr} \left\{ (\Phi_{\mathcal{T}}^T \Phi_{\mathcal{T}})^{-1} \right\} \right\} \quad (39)$$

$$= \frac{\xi l}{m - l - 1} \quad (40)$$

where $\xi \triangleq \tilde{\xi}/P_{\Phi}$ and $l \triangleq |\mathcal{T}|$. To calculate the expectation in (40) we have used the statistical characterization of the moments of the inverted Wishart matrix (e.g., see [40]). This completes the asymptotic characterization given in (8). ■

APPENDIX C PROOF OF PROPOSITION 3

Proof: Consider an estimator that given the vector of measurements $\mathbf{Y}$, first obtains an estimate $\hat{\underline{\mathcal{S}}}_{\gamma}$ of the union of all support sets $\underline{\mathcal{S}}_{\gamma}$ (cf. Corollary 1) and then uses it to produce an estimate $\hat{\mathbf{X}}$ of $\mathbf{x}$. We have that:

$$\text{mse}(\mathbf{x}, \gamma) = \int_{\Omega} \|\mathbf{x} - \hat{\mathbf{x}}\|^2 f(\phi, \mathbf{z}) d\phi d\mathbf{z} \quad (41)$$

where $f(\phi, \mathbf{z})$ is the joint probability density of the noise and the measurement matrix, and Ω is the set of all ϕ with full column rank and all $\mathbf{z}$. We split the integral in (41) into

$$I_1 \triangleq \int_{\mathcal{E}^c} \|\mathbf{x} - \hat{\mathbf{x}}\|^2 f(\phi, \mathbf{z}) d\phi d\mathbf{z},$$

$$I_2 \triangleq \int_{\mathcal{E}} \|\mathbf{x} - \hat{\mathbf{x}}\|^2 f(\phi, \mathbf{z}) d\phi d\mathbf{z}$$

where $\mathcal{E} \triangleq \{(\mathbf{z}, \phi) \in \Omega : \hat{\underline{\mathcal{S}}}_{\gamma} \neq \underline{\mathcal{S}}_{\gamma}\}$, and $\mathcal{E}^c$ is the complement of $\mathcal{E}$ with respect to Ω .

The term I_1 corresponds to the MSE for those realizations of the noise $\mathbf{z}$ and the measurement matrix ϕ for which the support recovery map correctly estimates all γ -support sets, that is, with $\hat{\underline{\mathcal{S}}}_{\gamma} = \underline{\mathcal{S}}_{\gamma}$. In this case, the estimate $\hat{\mathbf{x}}$ is precisely $\hat{\mathbf{x}}(\underline{\mathcal{S}}_{\gamma})$ and thus, we can write

$$I_1 = \int_{\mathcal{E}^c} \|\mathbf{x} - \hat{\mathbf{x}}(\underline{\mathcal{S}}_{\gamma})\|^2 f(\phi, \mathbf{z}) d\phi d\mathbf{z} \quad (42)$$

$$\leq \int_{\Omega} \|\mathbf{x} - \hat{\mathbf{x}}(\underline{\mathcal{S}}_{\gamma})\|^2 f(\phi, \mathbf{z}) d\phi d\mathbf{z} \quad (43)$$

$$= \text{mse}(\mathbf{w}, \underline{\mathcal{S}}_{\gamma}) \quad (44)$$

(cf. (10)), where we have used Proposition 2 to establish (44). Similarly, the term I_2 corresponds to the MSE for those $\mathbf{z}$ and ϕ such that $\hat{\underline{\mathcal{S}}}_{\gamma} \neq \underline{\mathcal{S}}_{\gamma}$. Note that for a given $\hat{\underline{\mathcal{S}}}_{\gamma}$, the estimate $\hat{\mathbf{x}}$ sets $\hat{x}_i = 0$ for all $i \notin \hat{\underline{\mathcal{S}}}_{\gamma}$. Using this we write

$$I_2 = \int_{\mathcal{E}} \left(\|\mathbf{x}_{\hat{\underline{\mathcal{S}}}_{\gamma}^c}\|^2 + \|\mathbf{x}_{\hat{\underline{\mathcal{S}}}_{\gamma}} - \hat{\mathbf{x}}_{\hat{\underline{\mathcal{S}}}_{\gamma}}\|^2 \right) f(\phi, \mathbf{z}) d\phi d\mathbf{z} \quad (45)$$

$$= I_3 + I_4 \quad (46)$$

where

$$I_3 \triangleq \int_{\mathcal{E}} \|\mathbf{x}_{\hat{\underline{\mathcal{S}}}_{\gamma}^c}\|^2 f(\phi, \mathbf{z}) d\phi d\mathbf{z}, \quad (47)$$

$$\leq \|\mathbf{x}\|^2 \int_{\mathcal{E}} f(\phi, \mathbf{z}) d\phi d\mathbf{z}. \quad (48)$$

Now, for $r > r^*(\mathbf{w}, \text{mse})$, Corollary 1 yields $I_3 \leq o(1/m_n)$. Regarding I_4 , using the expressions for the estimate $\hat{\mathbf{x}}_{\hat{\underline{\mathcal{S}}}_{\gamma}} = \phi_{\hat{\underline{\mathcal{S}}}_{\gamma}}^{\dagger} \mathbf{y}$, where $\phi_{\hat{\underline{\mathcal{S}}}_{\gamma}}^{\dagger}$ is defined analogous to (38), and for the measurement vector $\mathbf{y} = \phi \mathbf{x} + \mathbf{z}$, we write

$$\begin{aligned} I_4 &\triangleq \int_{\mathcal{E}} \|\mathbf{x}_{\hat{\underline{\mathcal{S}}}_{\gamma}} - \hat{\mathbf{x}}_{\hat{\underline{\mathcal{S}}}_{\gamma}}\|^2 f(\phi, \mathbf{z}) d\phi d\mathbf{z} \\ &= \int_{\mathcal{E}} \|\mathbf{x}_{\hat{\underline{\mathcal{S}}}_{\gamma}} - \phi_{\hat{\underline{\mathcal{S}}}_{\gamma}}^{\dagger} \phi \mathbf{x} - \phi_{\hat{\underline{\mathcal{S}}}_{\gamma}}^{\dagger} \mathbf{z}\|^2 f(\phi, \mathbf{z}) d\phi d\mathbf{z}. \end{aligned}$$

Further, using $\phi\mathbf{x} = \phi_{\underline{\mathcal{S}}_\gamma} \mathbf{x}_{\underline{\mathcal{S}}_\gamma} + \phi_{\underline{\mathcal{S}}_\gamma^c} \mathbf{x}_{\underline{\mathcal{S}}_\gamma^c}$ and the fact that $\phi_{\underline{\mathcal{S}}_\gamma}^\dagger \phi_{\underline{\mathcal{S}}_\gamma} = \mathbf{I}$ for full column-rank matrices, we obtain

$$\begin{aligned} I_4 &= \int_{\mathcal{E}} \|\phi_{\underline{\mathcal{S}}_\gamma}^\dagger \phi_{\underline{\mathcal{S}}_\gamma^c} \mathbf{x}_{\underline{\mathcal{S}}_\gamma^c} + \phi_{\underline{\mathcal{S}}_\gamma}^\dagger \mathbf{z}\|^2 f(\phi, \mathbf{z}) d\phi d\mathbf{z} \\ &\leq \int_{\Omega} \|\phi_{\underline{\mathcal{S}}_\gamma}^\dagger \phi_{\underline{\mathcal{S}}_\gamma^c} \mathbf{x}_{\underline{\mathcal{S}}_\gamma^c} + \phi_{\underline{\mathcal{S}}_\gamma}^\dagger \mathbf{z}\|^2 f(\phi, \mathbf{z}) d\phi d\mathbf{z}. \end{aligned}$$

Thus, $I_4 \leq I_5 + I_6 + 2I_7$ where

$$\begin{aligned} I_5 &\triangleq \mathbb{E}_{\mathbf{Z}, \Phi} \left\{ \|\Phi_{\underline{\mathcal{S}}_\gamma}^\dagger \Phi_{\underline{\mathcal{S}}_\gamma^c} \mathbf{x}_{\underline{\mathcal{S}}_\gamma^c}\|^2 \right\}, \\ I_6 &\triangleq \mathbb{E}_{\mathbf{Z}, \Phi} \left\{ \|\Phi_{\underline{\mathcal{S}}_\gamma}^\dagger \mathbf{Z}\|^2 \right\}, \\ I_7 &\triangleq \mathbb{E}_{\mathbf{Z}, \Phi} \left\{ \text{tr} \left\{ (\Phi_{\underline{\mathcal{S}}_\gamma}^\dagger \Phi_{\underline{\mathcal{S}}_\gamma^c} \mathbf{x}_{\underline{\mathcal{S}}_\gamma^c})^T \Phi_{\underline{\mathcal{S}}_\gamma}^\dagger \mathbf{Z} \right\} \right\}, \end{aligned}$$

with the implicit assumption that the expectations are restricted to Ω . Noting that I_5 is independent of $\mathbf{Z}$ and using the submultiplicativity property of the L_2 norm we see that

$$\begin{aligned} I_5 &\leq \mathbb{E}_{\Phi} \left\{ \|\Phi_{\underline{\mathcal{S}}_\gamma}^\dagger \Phi_{\underline{\mathcal{S}}_\gamma^c}\|^2 \|\mathbf{x}_{\underline{\mathcal{S}}_\gamma^c}\|^2 \right\} \\ &\leq \|\mathbf{x}\|^2 \mathbb{E}_{\Phi, \underline{\mathcal{S}}_\gamma} \left\{ \text{tr} \left\{ \Phi_{\underline{\mathcal{S}}_\gamma}^\dagger \mathbb{E}_{\Phi, \underline{\mathcal{S}}_\gamma} \left\{ \Phi_{\underline{\mathcal{S}}_\gamma^c} \Phi_{\underline{\mathcal{S}}_\gamma^c}^T \right\} (\Phi_{\underline{\mathcal{S}}_\gamma}^\dagger)^T \right\} \right\} \\ &= l \|\mathbf{x}\|^2 P_{\Phi} \mathbb{E}_{\Phi, \underline{\mathcal{S}}_\gamma} \left\{ \text{tr} \left\{ (\Phi_{\underline{\mathcal{S}}_\gamma}^T \Phi_{\underline{\mathcal{S}}_\gamma})^{-1} \right\} \right\} \end{aligned} \quad (49)$$

where $l = |\underline{\mathcal{S}}_\gamma^c|$. In obtaining (49) we have used the bound $\|\mathbf{x}_{\underline{\mathcal{S}}_\gamma^c}\| \leq \|\mathbf{x}\|$ for all subsets, and the fact that the submatrices $\Phi_{\underline{\mathcal{S}}_\gamma}$ and $\Phi_{\underline{\mathcal{S}}_\gamma^c}$ are independently generated in the Gaussian ensemble. Similarly, using the moments of the inverted Wishart distribution (e.g., [40]), we obtain

$$I_6 = \frac{l P_Z}{P_{\Phi}(m_n - l - 1)}.$$

Since the size of the support set is upper bounded we have that $l \leq k_{\max}$ and thus we obtain $I_5, I_6 \leq O(1/m_n)$. Finally, using the independence of the zero-mean noise, we readily see that $I_7 = 0$. Grouping all terms, we conclude that $\text{mse}(\mathbf{x}, \gamma)$ is upper bounded by

$$\|\mathbf{x}_{\underline{\mathcal{S}}_\gamma^c}\|^2 + \frac{k_{\max}(P_Z + (\xi + k_{\max}\|\mathbf{x}\|^2)P_{\Phi})}{P_{\Phi}(m_n - k_{\max} - 1)} + o(1/m_n). \quad (50)$$

Thus, in short, we have $\text{mse}(\mathbf{x}, \gamma) = \text{mse}^*(\mathbf{w}, \gamma)$. ■

APPENDIX D

PROOFS FOR THE REGION OF INTEREST

A. Proof of Proposition 4

Proof: Note that by assumption we have that

$$\frac{1}{r^{ex}} = \min_{i \in \{1, \dots, k-\lambda\}} \{c_i^{ex}\} \leq c_t^{ex}$$

for any $t \in \{k-\lambda+1, \dots, k\}$. Note also that $c_{k-\lambda}^{ex} \leq c_t^{ex}$ implies

$$\left(1 + \text{SNR} \sum_{j=\lambda+1}^k w_j^2\right)^{\frac{t}{t+\lambda-k}} \leq \left(1 + \text{SNR} \sum_{j=k-t+1}^k w_j^2\right)^{\frac{k-\lambda}{t+\lambda-k}}. \quad (51)$$

We use this to write

$$\begin{aligned} c_{k-\lambda}^{ex} &\leq c_t^{ex} \\ &= \frac{1}{2t} \log_2 \left(\frac{\left(1 + \text{SNR} \sum_{j=k-t+1}^k w_j^2\right)^{\frac{t}{t+\lambda-k}}}{\left(1 + \text{SNR} \sum_{j=k-t+1}^k w_j^2\right)^{\frac{k-\lambda}{t+\lambda-k}}} \right) \\ &\leq \frac{1}{2t} \log_2 \left(\frac{1 + \text{SNR} \sum_{j=k-t+1}^k w_j^2}{1 + \text{SNR} \sum_{j=\lambda+1}^k w_j^2} \right)^{\frac{t}{t+\lambda-k}} \\ &= c_i^{par, \lambda}. \end{aligned} \quad (52)$$

To obtain (52) we have used inequality (51). In (53) we have used the change of variables $i = t + \lambda - k$. Thus, for $i \in \{1, \dots, \lambda\}$, we have established that $c_{k-\lambda}^{ex} \leq c_i^{par, \lambda}$. This implies that $1/r^{ex} \leq 1/r^{par, \lambda}$ or, equivalently, $r^{ex} \geq r^{par, \lambda}$, as we wanted to prove. ■

B. Proof of Corollary 2

Proof: The proof of the corollary follows by noting that with the stronger assumption, the inequality in (51) is strict. This in turn implies that (52) is also a strict inequality. This establishes the claim. ■

APPENDIX E

PROOF OF PROPOSITION 5

Before proving Proposition 5, note that the results of Proposition 1 and Corollary 1 are easily extended to the case of deterministic $\mathbf{w}$ and random support set $\mathcal{S}$. In particular, this holds if the support set is chosen uniformly at random from the set of all possible size- k subsets of $\{1, \dots, n\}$. To prove this, observe that by the symmetry of the generation of the measurement matrices and the support recovery map, we have

$$\begin{aligned} P_e(\mathbf{X}^n(\mathbf{w}, \mathcal{S}), d_\gamma) &= \mathbb{E}_{\mathcal{S}} \{P_e(\mathbf{x}^n(\mathbf{w}, \mathcal{S}), d_\gamma)\} \\ &= P_e(\mathbf{x}^n(\mathbf{w}, \mathcal{S}), d_\gamma) \end{aligned} \quad (54)$$

for any deterministic $\mathcal{S}$. The probability in (54) was already analyzed in Appendix A-A. Since the same holds for random $\mathbf{W}$ independent of $\mathcal{S}$, we consider a deterministic $\mathcal{S}$.

Proof of Proposition 5: Consider any fixed $\gamma \in (0, 1]$ and $r > 0$ and let Ω be the set of all ϕ with full column rank and all $\mathbf{z}$. Let $\hat{\mathbf{X}}$ be the sequence of estimates constructed in the proof of Proposition 3 and let $\hat{\underline{\mathcal{S}}}_\gamma$ be the associated sequence of estimates of $\underline{\mathcal{S}}_\gamma$ (cf. Corollary 1). For compactness, define

$$\begin{aligned} \mathcal{C} &\triangleq \mathcal{C}^n(r, \gamma, k_{\max}), \\ \mathcal{A}_1 &\triangleq \{(\mathbf{x}, \phi, \mathbf{z}) : \mathbf{x} \in \mathcal{C}^c, (\phi, \mathbf{z}) \in \Omega, \hat{\underline{\mathcal{S}}}_\gamma = \underline{\mathcal{S}}_\gamma\}, \\ \mathcal{A}_2 &\triangleq \{(\mathbf{x}, \phi, \mathbf{z}) : \mathbf{x} \in \mathcal{C}^c, (\phi, \mathbf{z}) \in \Omega, \hat{\underline{\mathcal{S}}}_\gamma \neq \underline{\mathcal{S}}_\gamma\}, \end{aligned}$$

where $\mathcal{C}^c$ is the complement of $\mathcal{C}$. Let $\mathcal{X}$ be the domain of $\mathbf{X}$. Then, the averaged MSE in (18) can be written as

$$\text{mse}_\gamma = \int_{\mathcal{X}} \int_{\Omega} \|\mathbf{x} - \hat{\mathbf{x}}\|^2 f(\mathbf{x}, \phi, \mathbf{z}) d\mathbf{x} d\phi d\mathbf{z} = T_1 + T_2 + T_3$$

where $f(\mathbf{x}, \phi, \mathbf{z})$ is the joint probability density of the signal, the measurement matrix, and the noise, and

$$\begin{aligned} T_1 &\triangleq \int_{\mathcal{A}_1} \|\mathbf{x} - \hat{\mathbf{x}}\|^2 f(\mathbf{x}, \phi, \mathbf{z}) d\mathbf{x} d\phi d\mathbf{z}, \\ T_2 &\triangleq \int_{\mathcal{A}_2} \|\mathbf{x} - \hat{\mathbf{x}}\|^2 f(\mathbf{x}, \phi, \mathbf{z}) d\mathbf{x} d\phi d\mathbf{z}, \\ T_3 &\triangleq \int_{\mathcal{C}} \int_{\Omega} \|\mathbf{x} - \hat{\mathbf{x}}\|^2 f(\mathbf{x}, \phi, \mathbf{z}) d\mathbf{x} d\phi d\mathbf{z}. \end{aligned}$$

For T_1 , following similar steps to (42)-(44) and noting that the support of $\mathbf{X}$ is bounded, we obtain

$$\begin{aligned} T_1 &\leq \int_{C^c} \text{mse}(\mathbf{x}, \underline{\mathcal{S}}_\gamma) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{C^c} \|\mathbf{x}_{\underline{\mathcal{S}}_\gamma}\|^2 f(\mathbf{x}) d\mathbf{x} + O(1/m_n). \end{aligned}$$

Similarly to (45)-(48), for T_2 we obtain

$$\begin{aligned} T_2 &\leq \int_{C^c} \|\mathbf{x}\|^2 \Pr(\hat{\underline{\mathcal{S}}}_\gamma \neq \underline{\mathcal{S}}_\gamma | \mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &\quad + \underbrace{\int_{\mathcal{A}_2} \|\mathbf{x}_{\hat{\underline{\mathcal{S}}}_\gamma} - \hat{\mathbf{x}}_{\hat{\underline{\mathcal{S}}}_\gamma}\|^2 f(\mathbf{x}, \phi, z) d\mathbf{x} d\phi dz}_{T_4}. \end{aligned} \quad (55)$$

Using the reverse Fatou lemma [41] and applying the proof of Corollary 1 to the integrand (recall that the support of $\mathbf{X}$ is bounded) we see that

$$\limsup_{n \rightarrow \infty} \int_{C^c} \|\mathbf{x}\|^2 \Pr(\hat{\underline{\mathcal{S}}}_\gamma \neq \underline{\mathcal{S}}_\gamma | \mathbf{x}) f(\mathbf{x}) d\mathbf{x} = 0.$$

Similarly,

$$\begin{aligned} T_3 &\leq \int_C \|\mathbf{x}\|^2 f(\mathbf{x}) d\mathbf{x} \\ &\quad + \underbrace{\int_C \int_\Omega \|\mathbf{x}_{\hat{\underline{\mathcal{S}}}_\gamma} - \hat{\mathbf{x}}_{\hat{\underline{\mathcal{S}}}_\gamma}\|^2 f(\mathbf{x}, \phi, z) d\mathbf{x} d\phi dz}_{T_5}. \end{aligned} \quad (56)$$

Finally, proceeding as for the term I_4 in the proof of Proposition 3, we bound the sum of T_4 (in (55)) and T_5 (in (56))

$$\begin{aligned} T_4 + T_5 &\leq \int_{\mathcal{X}} \int_\Omega \|\mathbf{x}_{\hat{\underline{\mathcal{S}}}_\gamma} - \hat{\mathbf{x}}_{\hat{\underline{\mathcal{S}}}_\gamma}\|^2 f(\mathbf{x}, \phi, z) d\mathbf{x} d\phi dz \\ &= O(1/m_n). \end{aligned}$$

Thus, collecting the terms and taking the limit (recall that m_n grows with n) we establish the desired result

$$\limsup_{n \rightarrow \infty} \text{mse}_\gamma \leq \int_{C^c} \|\mathbf{x}_{\underline{\mathcal{S}}_\gamma}\|^2 f(\mathbf{x}) d\mathbf{x} + \int_C \|\mathbf{x}\|^2 f(\mathbf{x}) d\mathbf{x}. \quad \blacksquare$$

APPENDIX F

OPTIMAL PARAMETER FOR PARTIAL SUPPORT RECOVERY

Lemma 2. For $i \in \{1, \dots, k\}$ let $\gamma_i \triangleq \frac{\sum_{j=1}^i w_j^2}{\|\mathbf{w}\|^2}$. Given the size of the γ -support set λ , the value of γ that results in the lowest measurement rate requirement is γ_λ . $\square$

Proof: Given λ , γ must lie in the interval $[\gamma_\lambda, \gamma_{\lambda+1})$. Note that $c_i(\mathbf{w}, \gamma)$ is strictly decreasing in γ for all $i \in \{1, \dots, \lambda\}$. Thus, $\gamma = \gamma_\lambda$ results in the lowest $\mathbf{r}^*(\mathbf{w}, \gamma)$. $\blacksquare$

APPENDIX G

PROPERTIES OF THE COVERING SETS

Lemma 3. Consider the sets $\mathcal{Q}_l(r, \zeta)$ introduced in Appendix A-A. These sets satisfy the following properties:

- 1) $\mathcal{Q}_l(r, \zeta)$ is non-decreasing in r for fixed l and ζ .
- 2) $\mathcal{Q}_l(r, \zeta)$ is non-decreasing in l for fixed r and ζ .
- 3) $|\mathcal{Q}_l(r, \zeta)|$ behaves as $O(\zeta^{-l})$ with ζ for fixed l and r . $\square$

Proof: Property 1 was proved in [25]. The proof of Property 2 is straightforward. The proof of Property 3 follows from basic results on coverings [42]. $\blacksquare$

REFERENCES

- [1] D. Donoho, "Compressed sensing," *IEEE Trans. on Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [2] M. Elad, *Sparse and Redundant Representations*. New York, USA: Springer, 2010.
- [3] Y. Eldar and G. Kutyniok, Eds., *Compressed Sensing: Theory and Applications*. Cambridge, UK: Cambridge University Press, 2012.
- [4] E. Candès and M. Wakin, "An introduction to compressive sampling," *IEEE Sig. Proc. Magazine*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
- [5] M. Davenport, P. Boufounos, M. Wakin, and R. Baraniuk, "Signal processing with compressive measurements," *IEEE Journal on Selected Topics in Sig. Proc.*, vol. 4, no. 2, pp. 445–460, Apr. 2010.
- [6] N. Wagner, Y. Eldar, and Z. Friedman, "Compressed beamforming in ultrasound imaging," *IEEE Trans. on Sig. Proc.*, vol. 60, no. 9, pp. 4643–4657, Sep. 2012.
- [7] M. Mishali and Y. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *IEEE Journal on Selected Topics in Sig. Proc.*, vol. 4, no. 2, pp. 375–391, Apr. 2010.
- [8] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- [9] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1998.
- [10] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. on Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
- [11] D. Zachariah, S. Chatterjee, and M. Jansson, "Dynamic iterative pursuit," *IEEE Trans. on Sig. Proc.*, vol. 60, no. 9, pp. 4967–4972, Sep. 2012.
- [12] S. Chatterjee, D. Sundman, M. Vehkaperä, and M. Skoglund, "Projection-based and look ahead strategies for atom selection," *IEEE Trans. on Sig. Proc.*, vol. 60, no. 2, pp. 634–647, Feb. 2012.
- [13] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. on Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.
- [14] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, May 2009.
- [15] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, Nov. 2009.
- [16] G. Tang and A. Nehorai, "Performance analysis for sparse support recovery," *IEEE Trans. on Inf. Theory*, vol. 56, no. 3, pp. 1383–1399, Mar. 2010.
- [17] D. Malioutov, M. Çetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. on Sig. Proc.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.
- [18] D. Model and M. Zibulevsky, "Signal reconstruction in sensor arrays using sparse representations," *Sig. Proc.*, vol. 86, no. 3, pp. 624–638, Mar. 2006.
- [19] M. Wainwright, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," *IEEE Trans. on Inf. Theory*, vol. 55, no. 12, pp. 5728–5741, Dec. 2009.
- [20] —, "Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1 -constrained quadratic programming (Lasso)," *IEEE Trans. on Inf. Theory*, vol. 55, no. 5, pp. 2183–2202, May 2009.
- [21] K. Rahnama Rad, "Sharp upper bound on error probability of exact sparsity recovery," in *Proc. Conf. on Inf. Sciences and Systems (CISS)*, Mar. 2009, pp. 14–17.
- [22] W. Wang, M. Wainwright, and K. Ramchandran, "Information-theoretic limits on sparse signal recovery: Dense versus sparse measurement matrices," *IEEE Trans. on Inf. Theory*, vol. 56, no. 6, pp. 2967–2979, Jun. 2010.
- [23] Y. Jin and B. 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 Sig. Proc. (ICASSP)*, Apr. 2008, pp. 3921–3924.
- [24] Y. Jin, Y.-H. Kim, and B. Rao, "Performance tradeoffs for exact support recovery of sparse signals," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, Jun. 2010, pp. 1558–1562.

- [25] —, “Limits on support recovery of sparse signals via multiple-access communication techniques,” *IEEE Trans. on Inf. Theory*, vol. 57, no. 12, pp. 7877–7892, Dec. 2011.
- [26] G. Reeves and M. Gastpar, “Approximate sparsity pattern recovery: Information-theoretic lower bounds,” *IEEE Trans. on Inf. Theory*, vol. 59, no. 6, pp. 3451–3465, Jun. 2013.
- [27] —, “The sampling rate-distortion tradeoff for sparsity pattern recovery in compressed sensing,” *IEEE Trans. on Inf. Theory*, vol. 58, no. 5, pp. 3065–3092, May 2012.
- [28] M. Akçakaya and V. Tarokh, “Shannon-theoretic limits on noisy compressive sampling,” *IEEE Trans. on Inf. Theory*, vol. 56, no. 1, pp. 492–504, Jan. 2010.
- [29] E. Candès, J. K. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Communications on pure and applied mathematics*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.
- [30] E. Candès and T. Tao, “Decoding by linear programming,” *IEEE Trans. on Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [31] R. Giryes and M. Elad, “RIP-based near-oracle performance guarantees for SP, CoSaMP, and IHT,” *IEEE Trans. on Sig. Proc.*, vol. 60, no. 3, pp. 1465–1468, Mar. 2012.
- [32] D. Donoho, M. Elad, and V. Temlyakov, “Stable recovery of sparse overcomplete representations in the presence of noise,” *IEEE Trans. on Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.
- [33] J.-J. Fuchs, “Recovery of exact sparse representations in the presence of bounded noise,” *IEEE Trans. on Inf. Theory*, vol. 51, no. 10, pp. 3601–3608, 2005.
- [34] B. Babadi, N. Kalouptsidis, and V. Tarokh, “Asymptotic achievability of the Cramér-Rao bound for noisy compressive sampling,” *IEEE Trans. on Sig. Proc.*, vol. 57, no. 3, pp. 1233–1236, Mar. 2009.
- [35] E. Candès and T. Tao, “The Dantzig selector: Statistical estimation when p is much larger than n ,” *The Annals of Statistics*, vol. 35, no. 6, pp. 2313–2351, Dec. 2007.
- [36] R. Niazadeh, M. Babaie-Zadeh, and C. Jutten, “On the achievability of Cramér-Rao bound in noisy compressed sensing,” *IEEE Trans. on Sig. Proc.*, vol. 60, no. 1, pp. 518–526, Jan. 2012.
- [37] Z. Ben-Haim and Y. Eldar, “The Cramér-Rao bound for estimating a sparse parameter vector,” *IEEE Trans. on Sig. Proc.*, vol. 58, no. 6, pp. 3384–3389, Jun. 2010.
- [38] W. Xu, J. Lin, K. Niu, and Z. He, “On the performance of compressed sensing with partially correct support,” *Wireless Personal Comm.*, pp. 1–8, May 2012.
- [39] W. Rudin, *Principles of mathematical analysis*, 3rd ed. New York, USA: McGraw-Hill, 1976.
- [40] S. Press, *Applied multivariate analysis: using Bayesian and frequentist methods of inference*, 2nd ed. New York, USA: Dover Publications, 2005.
- [41] D. Williams, *Probability with martingales*. Cambridge, UK: Cambridge University Press, 1991.
- [42] N. Sloane and J. Conway, *Sphere packings, lattices and groups*, 3rd ed. Springer, 1999.

Ricardo Blasco-Serrano (S'04-M'14) received the M.Sc. degree in Telecommunications Engineering from the Technical University of Catalonia, Spain in 2007 and was awarded the Ph.D. degree in Telecommunications from the KTH Royal Institute of Technology, Stockholm, Sweden in 2013. In 2014 he joined Ericsson Research at Kista, Stockholm, Sweden. His research interests include information theory and communication theory.

Dave Zachariah (S'08-M'14) is a post-doctoral researcher at Uppsala University in Sweden. He received the M.S. degree in electrical engineering from Royal Institute of Technology (KTH), Stockholm, Sweden, in 2007. During 2007–8 he worked as a research engineer at Global IP Solutions, Stockholm. He received the Tech. Lic. and Ph.D. degrees in signal processing from KTH in 2011 and 2013, respectively.

His research interests include statistical signal processing, sensor fusion, sparse signal processing, positioning and localization, and spectral analysis.

Dennis Sundman (S'08) is a PhD student at the Communication Theory Lab at KTH Royal Institute of Technology, Sweden, which he joined shortly after receiving his M.Sc. degree in 2008. His research interest is greedy algorithms for distributed compressed sensing.

Ragnar Thobaben (M'07) received the Dipl.-Ing. degree (M.Sc.) in electrical engineering in 2001 and the Dr.-Ing. degree (Ph.D.) in electrical engineering in 2007 from the University of Kiel, Germany. In December 2006, Dr. Thobaben joined the Communication Theory Department at the KTH Royal Institute of Technology, Stockholm, Sweden, as a post-doctoral researcher, where he now serves as an Associate Professor.

Dr. Thobaben's current research activities are dedicated to the design and analysis of coding and transmission schemes for communication networks with a special focus on cognitive radio, cooperative communication, coordination, as well as physical-layer and network security.

Dr. Thobaben has served on the technical program committees for the IEEE GLOBECOM, the IEEE PIMRC, and the ISWCS and as publicity chair for the 2012 International Symposium on Turbo Codes & Iterative Information Processing. He also served as publicity chair for the 2011 IEEE Swedish Communication Technologies Workshop.

Mikael Skoglund (S'93-M'97-SM'04) received the Ph.D. degree in 1997 from Chalmers University of Technology, Sweden. In 1997, he joined the Royal Institute of Technology (KTH), Stockholm, Sweden, where he was appointed to the Chair in Communication Theory in 2003. At KTH, he heads the Communication Theory Division and he is the Assistant Dean for Electrical Engineering. He is also a founding faculty member of the ACCESS Linnaeus Center and director for the Center Graduate School.

Dr. Skoglund has worked on problems in source-channel coding, coding and transmission for wireless communications, Shannon theory and statistical signal processing. He has authored and co-authored more than 125 journal and roughly 300 conference papers, and he holds six patents.

Dr. Skoglund has served on numerous technical program committees for IEEE sponsored conferences. During 2003–08 he was an associate editor with the IEEE Transactions on Communications and during 2008–12 he was on the editorial board for the IEEE Transactions on Information Theory.