Shannon-Theoretic Limits on Noisy Compressive Sampling

Mehmet Akçakaya, *Student Member, IEEE*, and Vahid Tarokh, *Fellow, IEEE*

***Abstract*—In this paper, we study the number of measurements required to recover a sparse signal in $\mathbb{C}^M$ with L nonzero coefficients from compressed samples in the presence of noise. We consider a number of different recovery criteria, including the exact recovery of the support of the signal, which was previously considered in the literature, as well as new criteria for the recovery of a large fraction of the support of the signal, and the recovery of a large fraction of the energy of the signal. For these recovery criteria, we prove that $O(L)$ (an asymptotically linear multiple of L) measurements are necessary and sufficient for signal recovery, whenever L grows linearly as a function of M. This improves on the existing literature that is mostly focused on variants of a specific recovery algorithm based on convex programming, for which $O(L\log(M-L))$ measurements are required. In contrast, the implementation of our proof method would have a higher complexity. We also show that $O(L\log(M-L))$ measurements are required in the sublinear regime ($L=o(M)$). For our sufficiency proofs, we introduce a Shannon-theoretic decoder based on joint typicality, which allows error events to be defined in terms of a single random variable in contrast to previous information-theoretic work, where comparison of random variables are required. We also prove concentration results for our error bounds implying that a randomly selected Gaussian matrix will suffice with high probability. For our necessity proofs, we rely on results from channel coding and rate–distortion theory.**

***Index Terms*—Compressed sensing, compressive sampling, estimation error, Fano's inequality, joint typicality, linear regime, Shannon theory, sublinear regime, support recovery.**

I. Introduction

LET $\mathbb{C}$ denote the complex field and $\mathbb{C}^M$ the M-dimensional complex space. It is well-known that $\boldsymbol{x}\in\mathbb{C}^M$ can be reconstructed from M samples. However if the number of nonzero coefficients of $\boldsymbol{x}$, denoted $||\boldsymbol{x}||_0$, is $L \ll M$ then it is natural to ask if the number of samples could be reduced while still guaranteeing faithful reconstruction of $\boldsymbol{x}$. The answer to this question is affirmative. In fact, whenever $||\boldsymbol{x}||_0 = L \ll M$, one can measure a linear combination of the components of $\boldsymbol{x}$ as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$

where $\boldsymbol{A}$ is a properly designed $N\times M$ measurement matrix with $L < N \ll M$, and use nonlinear techniques to recover $\boldsymbol{x}$. This data acquisition technique that allows one to sample sparse signals below the Nyquist rate is referred to as compressive sampling or compressed sensing [5], [12].

Manuscript received November 02, 2007; revised November 21, 2008. Current version published December 23, 2009.

The authors are with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA (e-mail: akcakaya@seas.harvard.edu; vahid@seas.harvard.edu).

Communicated by H. Bölcskei, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2009.2034796

Although the data acquisition stage is straightforward, reconstruction of $\boldsymbol{x}$ from its samples $\boldsymbol{y}$ is a nontrivial task. It can be shown [5], [17] that if every set of N columns of $\boldsymbol{A}$ are linearly independent, then a decoder can recover $\boldsymbol{x}$ uniquely from $N=2L$ samples by solving the $\mathcal{L}_0$ minimization problem

$$\min ||\boldsymbol{x}||_0 \quad \text{s.t. } \boldsymbol{y}=\boldsymbol{A}\boldsymbol{x}.$$

However, solving this $\mathcal{L}_0$ minimization problem for recovery is NP-hard [21]. In this light, alternative solution methods have been studied in the literature. One such approach is the $\mathcal{L}_1$ regularization approach, where one solves

$$\min ||\boldsymbol{x}||_1 \quad \text{s.t. } \boldsymbol{y}=\boldsymbol{A}\boldsymbol{x}$$

and then establishes criteria under which the solution to this problem is also that of the $\mathcal{L}_0$ minimization problem. In an important contribution, by considering a class of measurement matrices satisfying an eigenvalue concentration property called the restricted isometry principle, Candès and Tao showed in [5] that for $L=O(M)$ with $N=O(L)$, the solution to this $\mathcal{L}_1$ recovery problem is the sparsest (minimum $\mathcal{L}_0$) solution as long as the observations are not contaminated with (additive) noise. They also showed that matrices from a properly normalized Gaussian ensemble satisfy this property with high probability. We discuss current literature on $\mathcal{L}_1$ regularization, as well as on other reconstruction approaches in more detail in Section I-B. Another strand of work considers solving the $\mathcal{L}_0$ recovery problem for a specific class of measurement matrices, such as the Vandermonde frames [1].

In practice, however, all the measurements are noisy, i.e.,

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n} \tag{1}$$

for some additive noise $\boldsymbol{n}\in\mathbb{C}^N$. This motivates our work, where we study Shannon-theoretic limits on the recovery of sparse signals in the presence of noise. More specifically, we are interested in the order of the number of measurements required, N in terms of L, M. We consider two regimes of sparsity: The linear sparsity regime where $L=\frac{1}{\beta}M$ for $\beta>2$, and the sublinear sparsity regime where $L=o(M)$.

A. Notation and Problem Statement

We consider the noisy compressive sampling of an unknown vector, $\boldsymbol{x}\in\mathbb{C}^M$. Let $\boldsymbol{x}$ have support $\mathcal{I}=\operatorname{supp}(\boldsymbol{x})$, where

$$\operatorname{supp}(\boldsymbol{x}) = \{i \mid x_i \neq 0\}$$

with $||\boldsymbol{x}||_0 = |\mathcal{I}| = L$. We also define

$$\mu(\boldsymbol{x}) = \min_{i\in\mathcal{I}} |x_i| \tag{2}$$

and the total power of the signal

$$P = \|\boldsymbol{x}\|_2^2.$$

We consider the noisy model given in (1), where $\boldsymbol{n}$ is an additive noise vector with a complex circularly symmetric Gaussian distribution with zero mean and covariance matrix $\nu^2 I_N$, i.e., $\boldsymbol{n} \sim \mathcal{N}_C\left(0, \nu^2 I_N\right)$. Due to the presence of noise, $\boldsymbol{x}$ cannot be recovered exactly. However, a sparse recovery algorithm outputs an estimate $\hat{\boldsymbol{x}}$. For the purposes of our analysis, we assume that this estimate is exactly sparse with $\|\hat{\boldsymbol{x}}\|_0 = L$. We consider three performance metrics for the estimate

$$\text{Error Metric 1:} \quad p_1(\hat{\boldsymbol{x}}, \boldsymbol{x}) = 1 - \mathbb{I}\left(\{\hat{x}_i \neq 0 \,\forall\, i \in \mathcal{I}\}\right) \times \mathbb{I}\left(\{\hat{x}_j = 0 \,\forall\, j \notin \mathcal{I}\}\right) \quad (3)$$

$$\text{Error Metric 2:} \quad p_2(\hat{\boldsymbol{x}}, \boldsymbol{x}, \alpha) = 1 - \mathbb{I}\left(\frac{|\{i \mid \hat{x}_i \neq 0\} \cap \mathcal{I}|}{|\mathcal{I}|} > 1 - \alpha\right) \quad (4)$$

$$\text{Error Metric 3:} \quad p_3(\hat{\boldsymbol{x}}, \boldsymbol{x}, \gamma) = 1 - \mathbb{I}\left(\sum_{k \in \{i \mid \hat{x}_i \neq 0\} \cap \mathcal{I}} |x_k|^2 > (1-\gamma)P\right) \quad (5)$$

where $\mathbb{I}(\cdot)$ is the indicator function and $\alpha, \gamma \in (0,1)$. We note that the error metrics depend on $\mathrm{supp}(\hat{\boldsymbol{x}})$ but do not explicitly depend on the values of $\{\hat{x}_k\}$. Once $\mathrm{supp}(\hat{\boldsymbol{x}})$ is determined, it is simple to find the estimate $\hat{\boldsymbol{x}}$ by solving the appropriate least squares problem, $\min_{\boldsymbol{b}} \|\boldsymbol{y} - \boldsymbol{A}_{\mathrm{supp}(\hat{\boldsymbol{x}})}\boldsymbol{b}\|_2$, where $\boldsymbol{A}_{\mathrm{supp}(\hat{\boldsymbol{x}})}$ is the $N \times L$ submatrix of $\boldsymbol{A}$ limited to columns specified by $\mathrm{supp}(\hat{\boldsymbol{x}})$.

Error Metric 1 is referred to as the 0–1 loss metric, and it is the one considered by Wainwright [24]. Error Metric 2 is a statistical extension of Error Metric 1, and considers the recovery of most of the subspace information of $\boldsymbol{x}$. Error Metric 3 characterizes the recovery of most of the energy of $\boldsymbol{x}$.

Consider a sequence of vectors, $\{\boldsymbol{x}^{(M)}\}_M$ such that $\boldsymbol{x}^{(M)} \in \mathbb{C}^M$ with $\mathcal{I}^{(M)} = \mathrm{supp}(\boldsymbol{x}^{(M)})$, where $|\mathcal{I}^{(M)}| = L^{(M)}$. For $\boldsymbol{x}^{(M)}$, we will consider $N \times M$ random Gaussian measurement matrices, $\boldsymbol{A}^{(M)}$, where N is a function of M. Since the dependence of $L^{(M)}$, $\mathcal{I}^{(M)}$, and $\boldsymbol{A}^{(M)}$ on M is implied by the vector $\boldsymbol{x}^{(M)}$, we will omit the superscript for brevity, and denote the support of $\boldsymbol{x}^{(M)}$ by $\mathcal{I}$, its size by L, and any measurement matrix from the ensemble by $\boldsymbol{A}$, whenever there is no ambiguity.

A decoder $\mathcal{D}(\cdot)$ outputs a set of indices $\mathcal{D}(\boldsymbol{y})$. For a specific decoder, we consider the average probability of error, averaged over all Gaussian measurement matrices, $\boldsymbol{A}$ with the (i,j)th term $a_{i,j} \sim \mathcal{N}_C(0,1)$

$$p_{\mathrm{err}}(\mathcal{D}) = \mathbb{E}_{\boldsymbol{A}}\left(p_{\mathrm{err}}(\boldsymbol{A})\right) \quad (6)$$

where $p_{\mathrm{err}}(\boldsymbol{A}) = \mathbb{P}(p_k(\boldsymbol{x}, \hat{\boldsymbol{x}}) \neq 0)$ for $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}^{(M)} + \boldsymbol{n}$, $k = 1,2,3$ specifying the appropriate error metric[1] and $\mathbb{P}(\cdot)$ is the probability measure. We note that $p_{\mathrm{err}}(\cdot)$ is a function of a matrix, and in (6) it is a function of the random matrix $\boldsymbol{A}$.[2]

We say a decoder achieves *asymptotically reliable* sparse recovery if $p_{\mathrm{err}}(\mathcal{D}) \to 0$ as $M \to \infty$. This implies the existence of a sequence of measurement matrices $\{\boldsymbol{A}^{(M)}\}$ such that $p_{\mathrm{err}}(\boldsymbol{A}^{(M)}) \to 0$ as $M \to \infty$. Similarly, we say asymptotically reliable sparse recovery is not possible if $p_{\mathrm{err}}(\mathcal{D})$ stays bounded away from 0 as $M \to \infty$.

We also use the notation

$$f(x) \succ g(x)$$

for either $f(x) = g(x) = 0$ or for nondecreasing nonnegative functions $f(x)$ and $g(x)$, if $\exists\, x_0$ such that for all $x > x_0$

$$\frac{f(x)}{g(x)} > 1.$$

Similarly, we say $f(x) \prec g(x)$ if $g(x) \succ f(x)$.

B. Previous Work

There is a large body of literature on tractable recovery algorithms for compressive sampling. Most notably, a series of excellent papers [5], [12] on $\mathcal{L}_1$ relaxation showed that recovery is possible in the noiseless setting for $N = O(L \log(M/L))$. This translates to $N = O(L)$ and $N = O(L \log(M-L))$ measurements in the linear and sublinear sparsity regimes, respectively. The behavior of $\mathcal{L}_1$ relaxation in the presence of noise was studied in [8], [22], and bounds were derived on the distortion of the estimate $\hat{\boldsymbol{x}}$. There is also another strand of work that characterizes the performance of various recovery algorithms based on the matching pursuit algorithm, and offers similar guarantees to $\mathcal{L}_1$ relaxation [11], [17].

$\mathcal{L}_1$ relaxation in the presence of Gaussian noise $\boldsymbol{n}$ has also been studied in the literature [6], [25]. In [6], it was shown that the distortion of the estimate $\hat{\boldsymbol{x}}$ obtained via this method (the Dantzig selector) is within a factor $\log M$ of the distortion of the estimate obtained when the locations of the nonzero elements of $\boldsymbol{x}$ are known at the decoder. The problem of support recovery with respect to Error Metric 1 was first considered in [25] for this setting. Wainwright showed that the number of measurements required is $N = O(L \log(M-L))$ both in the linear and sublinear sparsity regimes, when the $\mathcal{L}_1$ constrained quadratic programming algorithm (LASSO) is used for recovery.

Recovery with respect to Error Metric 1 in an information-theoretic setting was also first studied by Wainwright in another excellent paper [24]. Using an optimal decoder that decodes to the closest subspace, it was shown that in the presence of additive Gaussian noise, $N = O(L)$ and $N = O(L \log(M-L))$ measurements were necessary and sufficient for the linear and sublinear sparsity regimes, respectively. For the linear regime, it was also required that $L\mu^2(\boldsymbol{x})/\log L \to \infty$ as $L \to \infty$, leading to $P \to \infty$. The reason for this requirement is that at high dimensions, Error Metric 1 is too stringent for an average case analysis. This is one of the reasons why we have considered other performance metrics.

Since the submission of this work, there has been more work done on information-theoretic limits of sparse recovery. In [13], recovery with respect to Error Metric 1 was considered for the fixed P regime. It was shown that in this regime, $N = O(L \log(M-L))$ measurements are necessary, which improves on previous results. Error Metric 2 was later considered independently in [19], where methods developed in [24] were used. Sparse measurement matrix ensembles instead of

[1]For Error Metric 2 and 3, appropriate values of α and γ also need to be specified. These variables are not included in the definition to ease the notation.

[2]The dependence on noise $\boldsymbol{n}$ is implicit in this notation.

Gaussian measurement ensembles were considered in [18], [26]. Necessary conditions for recovery with respect to Error Metric 1 were derived in [26]. Sufficient conditions for LASSO to asymptotically recover the whole support were obtained in [18]. We also note that there is other work that characterizes the average distortion associated with compressive sampling [20], as well as that which is associated with the broader problem of sparsely representing a given vector $\boldsymbol{y}$ [1], [14].

C. Main Results

Before we state our results, we briefly talk about our proof methodology. Our achievability proof technique is largely derived from Shannon theory [10]. We define a decoder that characterizes events based on their typicality. We call such a decoder a "joint typicality decoder." A formal definition is given in Section II-B. We note that while in Shannon theory most typicality definitions explicitly involve the entropy of a random variable or mutual information between two random variables, our definition is characterized by the noise variance ν^2. This is not surprising, since for a Gaussian vector $\boldsymbol{n} \sim \mathcal{N}_C\left(0, \nu^2 I_N\right)$, its entropy is closely related to its variance. Error events are defined based on atypicality, and the probability of these events are small as a consequence of the law of large numbers. Use of joint typicality also allows us to extend our results to various error metrics, which was not previously done in the literature. To prove the converses, we utilize Fano's inequality [15], and the rate–distortion characterization of certain sources [3].

Theorem 1.1: (Achievability for the Linear Regime) Let a sequence of sparse vectors, $\{\boldsymbol{x}^{(M)} \in \mathbb{C}^M\}_M$ with $\|\boldsymbol{x}^{(M)}\|_0 = L = \left\lfloor \frac{1}{\beta} M \right\rfloor$, where $\beta > 2$ be given. Then asymptotically reliable recovery is possible for $\left\{\boldsymbol{x}^{(M)}\right\}$ if

$$N \succ C_k L \quad k = 1, 2, 3 \tag{7}$$

for different constants $C_1, C_2, C_3 > 1$ corresponding to Error Metric 1, 2, and 3, respectively. Additionally for Error Metric 1, we require that $\frac{L\mu^4(\boldsymbol{x}^{(M)})}{\log L} \to \infty$ as $L \to \infty$. For Error Metric 2, we only require that $L\mu^2(\boldsymbol{x}^{(M)})$ and P be constants. For Error Metric 3, we only require that P be a constant. Furthermore, the constant C_1 depends only on β; C_2 depends only on α, β, P, and ν; C_3 depends only on β, γ, P, and ν.

Proof: The proof is given in Section II-C. □

As we noted previously, asymptotically reliable recovery implies the existence of a sequence of measurement matrices $\boldsymbol{A}^{(M)}$ such that $p_{\mathrm{err}}(\boldsymbol{A}^{(M)}) \to 0$ as $M \to \infty$. One can use a concentration result to show that *any* measurement matrix chosen from the Gaussian ensemble will have arbitrarily small $p_{\mathrm{err}}(\boldsymbol{A})$ with high probability.

Theorem 1.2: (Concentration of Achievability for the Linear Regime) Let the conditions of Theorem 1.1 be satisfied and consider the decoder used in its proof. Then for any measurement matrix, $\boldsymbol{A}$, chosen from the Gaussian ensemble, $\mathbb{P}(p_{\mathrm{err}}(\boldsymbol{A}) \geq \xi)$, decays to zero as a function of M for any $\xi \in (0, 1]$. For Error Metric 1, the decay is faster than $1/M^k$ for any positive k, whereas for Error Metrics 2 and 3 the decay is exponential in M.

Proof: Markov's Inequality implies

$$\mathbb{P}(p_{\mathrm{err}}(\boldsymbol{A}) \geq \xi) \leq \frac{\mathbb{E}_{\boldsymbol{A}}(p_{\mathrm{err}}(\boldsymbol{A}))}{\xi} = \frac{p_{\mathrm{err}}(\mathcal{D})}{\xi}.$$

As shown in the proof of Theorem 1.1, for Error Metric 1, $-\log p_{\mathrm{err}}(\mathcal{D}) / \log M \to \infty$ as $M \to \infty$, and for Error Metrics 2 and 3, $p_{\mathrm{err}}(\mathcal{D})$ decays exponentially fast in M, yielding the desired result for constant ξ. The results can be strengthened by letting ξ decay as well. For Error Metrics 2 and 3, $p_{\mathrm{err}}(\mathcal{D}) \leq \exp(-\kappa_0 M)$ for some constant κ_0. This implies that for any measurement matrix $\boldsymbol{A}$ from the Gaussian ensemble,

$$\mathbb{P}\left(p_{\mathrm{err}}(\boldsymbol{A}) \geq \exp\left(-\frac{\kappa_0}{2} M\right)\right) \leq \exp\left(-\frac{\kappa_0}{2} M\right)$$

i.e., any matrix $\boldsymbol{A}$ from the Gaussian ensemble has $p_{\mathrm{err}}(\boldsymbol{A}) \leq \exp\left(-\frac{\kappa_0}{2} M\right)$ with high probability. Simlarly, for Error Metric 1, we have $\mathbb{P}\left(p_{\mathrm{err}}(\boldsymbol{A}) \geq \frac{1}{M^{k_1}}\right) \leq \frac{1}{M^{k_2}}$ for any positive k_1 and k_2. We also note that an exponential decay is possible for Error Metric 1, if one further assumes that $\mu(\boldsymbol{x}^{(M)})$ is a constant, which is a stronger assumption than that required for Theorem 1.1. □

We also derive necessary conditions for asymptotically reliable recovery.

Theorem 1.3: (Converse for the Linear Regime) Let a sequence of sparse vectors, $\{\boldsymbol{x}^{(M)} \in \mathbb{C}^M\}_M$ with $\|\boldsymbol{x}^{(M)}\|_0 = L = \left\lfloor \frac{1}{\beta} M \right\rfloor$, where $\beta > 2$ be given. Then asymptotically reliable recovery is not possible for $\left\{\boldsymbol{x}^{(M)}\right\}$

$$N \prec C_k L, \quad k = 4, 5, 6 \tag{8}$$

for different constants C_4, C_5, C_6 corresponding to Error Metric 1, 2, and 3 respectively. $C_4 > 0$ depends only on β, P, and ν; $C_5 \geq 0$ depends only on α, β, P, and ν; $C_6 \geq 0$ depends only on β, γ, P, and ν. Additionally, for Error Metric 3, we require that the nonzero terms decay to zero at the same rate.[3]

Proof: The proof is given in Section III. □

A result similar to Corollary 1.2 can be obtained for the converse case. This result states that if N is less than a constant multiple of L, then $p_{\mathrm{err}}(\boldsymbol{A}) \to 1$ with overwhelming probability (i.e., the probability goes to 1 exponentially fast as a function of N).

Corollary 1.4: Let a sequence of sparse vectors, $\{\boldsymbol{x}^{(M)} \in \mathbb{C}^M\}_M$ with $\|\boldsymbol{x}^{(M)}\|_0 = L = \left\lfloor \frac{1}{\beta} M \right\rfloor$, where $\beta > 2$ be given. Then for $\xi > 0$, for any measurement matrix $\boldsymbol{A}$ from the Gaussian ensemble, $\mathbb{P}\left(p_{\mathrm{err}}(\boldsymbol{A}) \to 1\right)$ goes to 1 exponentially fast as a function of M if $N \prec \hat{C}_k L$, where $\hat{C}_k \leq C_k$ for $k = 4, 5, 6$ are positive constants for Error Metrics 1, 2, and 3, respectively. $\hat{C}_4$ depends only on β, P, ν, ξ. $\hat{C}_5$, and $\hat{C}_6$ additionally depend on α and γ, respectively.

Proof: The proof for Error Metric 1 is given in Section III-A. The proofs for Error Metrics 2 and 3 are analogous. □

Thus, for the linear sparsity regime, Theorems 1.1 and 1.3 show that $O(L)$ measurements are necessary and sufficient for

[3]That is, $c_t \leq \frac{\max_i |x_i|}{\min_i |x_i|} \leq C_t$ for constants c_t and C_t.

asymptotically reliable recovery. For Error Metric 1, Theorem 1.1 shows that there is a clear gap between the $O(L)$ measurements required for a joint typicality decoder and $O(L\log(M-L))$ measurements required by $\mathcal{L}_1$ constrained quadratic programming [25]. In our proof, it is required that $\frac{L\mu^4(\boldsymbol{x}^{(M)})}{\log L} \to \infty$ as $L \to \infty$, which implies that P grows without bound as a function of N. We note that a similar result was derived in [24]. For Error Metrics 2 and 3, Theorem 1.1 implies $O(L)$ measurements are sufficient and the power of the signal P remains a constant. This is a much less stringent requirement than that for Error Metric 1, and is due to the more statistical nature of the error metrics considered.

The converse to this theorem is established in Theorem 1.3, which demonstrates that $O(L)$ measurements are necessary. We note that although the converse theorem is stated for a Gaussian measurement matrix $\boldsymbol{A}$, the proof extends to any matrix with the property that any group of L entries in every row has $\mathcal{L}_2$ norm less than or equal to L.

Finally, as stated previously, Corollary 1.2 implies that with overwhelming probability any given $N \times M$ Gaussian measurement matrix $\boldsymbol{A}$ can be used for asymptotically reliable sparse recovery for Error Metrics 2 and 3 as long as N is greater than a specified constant times L. A weaker concentration result is readily obtained for Error Metric 1, and this can be strengthened if $\mu(\boldsymbol{x})$ is constant. Corollary 1.4 states that if the number of measurements is less than specified constant multiples of L, then with overwhelming probability, any matrix from the Gaussian ensemble will have error probability approaching 1.

We next state the analogous results for the sublinear sparsity regime.

Theorem 1.5: (Achievability for the Sublinear Regime) Let a sequence of sparse vectors, $\{\boldsymbol{x}^{(M)} \in \mathbb{C}^M\}_M$ with $\|\boldsymbol{x}^{(M)}\|_0 = L = o(M)$ be given. Then asymptotically reliable recovery is possible for $\{\boldsymbol{x}^{(M)}\}$ if

$$N \succ C_k' L \log(M-L), \quad k = 1,2,3 \tag{9}$$

for different constants C_1', C_2', $C_3' > 0$ corresponding to Error Metric 1, 2, and 3, respectively. Additionally, for Error Metric 1, we require that $L\mu^4(\boldsymbol{x}^{(M)}) \to \infty$ as $L \to \infty$. For Error Metric 2, we only require that $L\mu^2(\boldsymbol{x}^{(M)})$ and P be constants. For Error Metric 3, we only require that P be a contant. Furthermore, the constant C_1' depends only on ν; C_2' depends only on α, P, and ν; C_3' depends only on γ, P and ν.

Proof: The proof largely follows that of Theorem 1.1. An outline of the proof highlighting the differences is given in Section IV. □

Theorem 1.6: (Converse for the Sublinear Regime) Let a sequence of sparse vectors, $\{\boldsymbol{x}^{(M)} \in \mathbb{C}^M\}_M$ with $\|\boldsymbol{x}^{(M)}\|_0 = L = o(M)$ be given. Then asymptotically reliable recovery is not possible for $\{\boldsymbol{x}^{(M)}\}$

$$N \prec C_k' L \log(M-L), \quad k = 4,5,6 \tag{10}$$

for different constants C_4', C_5', C_6' corresponding to Error Metric 1, 2, and 3, respectively. $C_4' > 0$ depends only on P and ν; $C_5' > 0$ depends only on α, P and ν; $C_6 \geq 0$ depends only on γ, P, and ν. Additionally, for Error Metric 3, we require that the nonzero terms decay to zero at the same rate.

Proof: The proof largely follows that of Theorem 1.3. An outline of the proof highlighting the differences is given in Section IV. □

For the sublinear regime, we have that $O(L\log(M-L))$ measurements are necessary and sufficient for asymptotically reliable recovery. This matches the number of measurements required by the tractable $\mathcal{L}_1$ regularization algorithm [25].

We finish our results with a theorem that shows the importance of recovery with respect to Error Metric 3 in terms of the estimation error. We first consider the idealized case when a genie provides $\text{supp}(\boldsymbol{x})$ to the decoder. The decoder outputs an estimate $\hat{\boldsymbol{x}}_{\text{genie}}$. Let $\Delta_{\text{genie}} = \|\boldsymbol{x} - \hat{\boldsymbol{x}}_{\text{genie}}\|_2$. We will show that any matrix $\boldsymbol{A}$ chosen from the Gaussian ensemble satisfies certain eigenvalue concentration inequalities with overwhelming probability and $p_{\text{err}}(\boldsymbol{A})$ with respect to Error Metric 3 is exponentially small in N for a given signal $\boldsymbol{x}$, and that these imply the mean-square distortion of the estimate is within a constant factor of Δ^2_{genie} for sufficiently large N. We next state our results characterizing the distortion of the estimate in the linear sparsity regime, when a joint typicality decoder is used for recovery with respect to Error Metric 3.

Theorem 1.7: (Average Distortion of The Estimate For Error Metric 3): Suppose the conditions of Theorem 1.1 are satisfied for recovery with respect to Error Metric 3. Let Δ_{genie} be the mean-square distortion of the estimate when a genie provides $\text{supp}(\boldsymbol{x}) = \mathcal{I}$ to the decoder. Let $N > \hat{C}L$ for a constant $\hat{C}$ as specified in Section V-A. Then with overwhelming probability, for any measurement matrix $\boldsymbol{A}$ from the Gaussian ensemble and for sufficiently large M, the joint typicality decoder outputs an estimate $\hat{\boldsymbol{x}}$ such that

$$\mathbb{E}_{\boldsymbol{n}}\left(\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2^2\right) < \tilde{C}\mathbb{E}_{\boldsymbol{n}}\left(\Delta^2_{\text{genie}}\right)$$

for a constant $\tilde{C}$ that depends only on β, γ, P, ν, and ϵ_L.

Proof: The proof is given in Section V-A. □

One can compare this result to that of [6], where the estimation error, $\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2^2$ is within $\Delta^2_{\text{genie}} \log M$ when a tractable $\mathcal{L}_1$ regularization algorithm is used for recovery. Thus, if a joint typicality decoder (with respect to Error Metric 3) is used, the $\log M$ factor could be improved in the linear sparsity regime, while maintaining constant P. We also note that if the conditions of Theorem $I.1$ are satisfied for Error Metric 1, it can be shown that the estimation error of the joint typicality estimate converges to Δ^2_{genie} in the linear sparsity regime, as one would intuitively expect [2].

D. Paper Organization

The outline of the rest of the paper is given next. In Section II, we formally define the concept of joint typicality in the setting of compressive sampling and prove Theorem 1.1. In Section III, we provide the proofs for Theorem 1.3 and Corollary 1.4. In Section IV, we prove the analogous theorems for the sublinear sparsity regime, $L = o(M)$. Finally, in Section V, we provide the proof of Theorem 1.7.

II. Achievability Proofs for the Linear Regime

A. Notation

Let $\boldsymbol{a}_i$ denote the ith column of $\boldsymbol{A}$. For the measurement matrix $\boldsymbol{A}$, we define $\boldsymbol{A}_{\mathcal{J}}$ to be the matrix whose columns are $\{\boldsymbol{a}_j : j \in \mathcal{J}\}$. For any given matrix $\boldsymbol{B}$, we define $\boldsymbol{\Pi_B}$ to be the orthogonal projection matrix onto the subspace spanned by the columns of $\boldsymbol{B}$, i.e., $\boldsymbol{\Pi_B} = \boldsymbol{B}(\boldsymbol{B}^*\boldsymbol{B})^{-1}\boldsymbol{B}^*$. Similarly, we define $\boldsymbol{\Pi}_{\boldsymbol{B}}^{\perp}$ to be the projection matrix onto the orthogonal complement of this subspace, i.e., $\boldsymbol{\Pi}_{\boldsymbol{B}}^{\perp} = \boldsymbol{I} - \boldsymbol{\Pi_B}$.

B. Joint Typicality

In our analysis, we will use Gaussian measurement matrices and a suboptimal decoder based on joint typicality, as defined below.

Definition 2.1: (Joint Typicality) We say an $N \times 1$ noisy observation vector $\boldsymbol{y} = \boldsymbol{Ax} + \boldsymbol{n}$ and a set of indices $\mathcal{J} \subset \{1, 2, \dots, M\}$, with $|\mathcal{J}| = L$ are δ-jointly typical if $\operatorname{rank}(\boldsymbol{A}_{\mathcal{J}}) = L$ and

$$\left| \frac{1}{N} \left\| \boldsymbol{\Pi}_{\boldsymbol{A}_{\mathcal{J}}}^{\perp} \boldsymbol{y} \right\|^2 - \frac{N-L}{N}\nu^2 \right| < \delta \tag{11}$$

where $\boldsymbol{n} \sim \mathcal{N}_C(0, \nu^2 I_N)$, the (i,j)th entry of $\boldsymbol{A}$, $a_{ij} \sim \mathcal{N}_C(0,1)$, and $\|\boldsymbol{x}\|_0 = L$.

Lemma 2.2: For an index set $\mathcal{I} \subset \{1, 2, \dots, M\}$ with $|\mathcal{I}| = L$

$$\mathbb{P}(\operatorname{rank}(\boldsymbol{A}_{\mathcal{I}}) < L) = 0.$$

Lemma 2.3:

- Let $\mathcal{I} = \operatorname{supp}(\boldsymbol{x})$ and assume (without loss of generality) that $\operatorname{rank}(\boldsymbol{A}_{\mathcal{I}}) = L$. Then for $\delta > 0$

$$\mathbb{P}\left(\left| \frac{1}{N} \left\| \boldsymbol{\Pi}_{\boldsymbol{A}_{\mathcal{I}}}^{\perp} \boldsymbol{y} \right\|^2 - \frac{N-L}{N}\nu^2 \right| > \delta \right) \le 2\exp\left(-\frac{\delta^2}{4\nu^4} \frac{N^2}{N - L + \frac{2\delta}{\nu^2}N} \right). \tag{12}$$

- Let $\mathcal{J}$ be an index set such that $|\mathcal{J}| = L$ and $|\mathcal{I} \cap \mathcal{J}| = K < L$, where $\mathcal{I} = \operatorname{supp}(\boldsymbol{x})$ and assume that $\operatorname{rank}(\boldsymbol{A}_{\mathcal{J}}) = L$. Then $\boldsymbol{y}$ and $\mathcal{J}$ are δ-jointly typical with probability

$$\mathbb{P}\left(\left| \frac{1}{N} \left\| \boldsymbol{\Pi}_{\boldsymbol{A}_{\mathcal{J}}}^{\perp} \boldsymbol{y} \right\|^2 - \frac{N-L}{N}\nu^2 \right| < \delta \right) \le \exp\left(-\frac{N-L}{4} \left(\frac{\sum_{k \in \mathcal{I}\setminus\mathcal{J}} |x_k|^2 - \delta'}{\sum_{k \in \mathcal{I}\setminus\mathcal{J}} |x_k|^2 + \nu^2} \right)^2 \right) \tag{13}$$

where

$$\delta' = \delta \frac{N}{N-L}.$$

Proof: We first note that for

$$\boldsymbol{y} = \boldsymbol{Ax} + \boldsymbol{n} = \sum_{i \in \mathcal{I}} x_i \boldsymbol{a}_i + \boldsymbol{n}$$

we have

$$\boldsymbol{\Pi}_{\boldsymbol{A}_{\mathcal{I}}}^{\perp} \boldsymbol{y} = \boldsymbol{\Pi}_{\boldsymbol{A}_{\mathcal{I}}}^{\perp} \boldsymbol{n}$$

and

$$\boldsymbol{\Pi}_{\boldsymbol{A}_{\mathcal{J}}}^{\perp} \boldsymbol{y} = \boldsymbol{\Pi}_{\boldsymbol{A}_{\mathcal{J}}}^{\perp} \left(\textstyle\sum_{i \in \mathcal{I}\setminus\mathcal{J}} x_i \boldsymbol{a}_i + \boldsymbol{n} \right).$$

Furthermore, $\boldsymbol{\Pi}_{\boldsymbol{A}_{\mathcal{I}}}^{\perp} = \boldsymbol{U}_{\mathcal{I}} \boldsymbol{D} \boldsymbol{U}_{\mathcal{I}}^{\dagger}$, where $\boldsymbol{U}_{\mathcal{I}}$ is a unitary matrix that is a function of $\{\boldsymbol{a}_i : i \in \mathcal{I}\}$ (and independent of $\boldsymbol{n}$). $\boldsymbol{D}$ is a diagonal matrix with $N - L$ diagonal entries equal to 1, and the rest equal to 0. It is easy to see that

$$\left\| \boldsymbol{\Pi}_{\boldsymbol{A}_{\mathcal{I}}}^{\perp} \boldsymbol{y} \right\|^2 = \|\boldsymbol{D}n'\|^2$$

where $\boldsymbol{n}'$ has independent and identically distributed (i.i.d.) entries with distribution $\mathcal{N}_C(0, \nu^2)$. Without loss of generality, we may assume the nonzero entries of $\boldsymbol{D}$ are on the first $N - L$ diagonals, thus

$$\|\boldsymbol{D}n'\|^2 = |n_1'|^2 + \cdots + |n_{N-L}'|^2.$$

Similarly, $\boldsymbol{\Pi}_{\boldsymbol{A}_{\mathcal{J}}}^{\perp} = \boldsymbol{U}_{\mathcal{J}} \boldsymbol{D} \boldsymbol{U}_{\mathcal{J}}^{\dagger}$, where $\boldsymbol{U}_{\mathcal{J}}$ is a unitary matrix that is a function of $\{\boldsymbol{a}_j : j \in \mathcal{J}\}$ (and independent of $\boldsymbol{n}$ and $\{\boldsymbol{a}_i : i \in \mathcal{I}\setminus\mathcal{J}\}$) and $\boldsymbol{D}$ is as discussed above. Thus, $\boldsymbol{a}_i' = \boldsymbol{U}_{\mathcal{J}}^{\dagger} \boldsymbol{a}_i$ has i.i.d. entries with distribution $\mathcal{N}_C(0, 1)$ for all $i \in \mathcal{I}\setminus\mathcal{J}$. It is easy to see that $\boldsymbol{n}'' = \boldsymbol{U}_{\mathcal{J}}^{\dagger} \boldsymbol{n}$ also has i.i.d. entries with $\mathcal{N}_C(0, \nu^2)$. Thus

$$\left\| \boldsymbol{\Pi}_{\boldsymbol{A}_{\mathcal{J}}}^{\perp} \boldsymbol{y} \right\|^2 = \|\boldsymbol{D}w\|^2 = |w_1|^2 + \cdots + |w_{N-L}|^2$$

where w_i are i.i.d. with distribution $\mathcal{N}_C\left(0, \sigma_{\mathcal{J}}^2\right)$, where

$$\sigma_{\mathcal{J}}^2 = \sum_{k \in \mathcal{I}\setminus\mathcal{J}} |x_k|^2 + \nu^2.$$

Let $\Omega_1 = \frac{\|\boldsymbol{D}n'\|^2}{\nu^2}$ and $\Omega_2 = \frac{\|\boldsymbol{D}w\|^2}{\sigma_{\mathcal{J}}^2}$. We note that both Ω_1 and Ω_2 are chi-square random variables with $(N - L)$ degrees of freedom. Thus, to bound these probabilities, we must bound the tail of a chi-square random variable. We have

$$\begin{aligned} \mathbb{P}&\left(\left| \frac{1}{N} \left\| \boldsymbol{\Pi}_{\boldsymbol{A}_{\mathcal{I}}}^{\perp} \boldsymbol{y} \right\|^2 - \frac{N-L}{N}\nu^2 \right| > \delta \right) \\ &= \mathbb{P}\left(|\Omega_1 - (N-L)| > \frac{\delta}{\nu^2}N \right) \\ &= \mathbb{P}\left(\Omega_1 - (N-L) < -\frac{\delta}{\nu^2}N \right) \\ &\quad + \mathbb{P}\left(\Omega_1 - (N-L) > \frac{\delta}{\nu^2}N \right) \end{aligned} \tag{14}$$

and

$$\begin{aligned} \mathbb{P}&\left(\left| \frac{1}{N} \left\| \boldsymbol{\Pi}_{\boldsymbol{A}_{\mathcal{J}}}^{\perp} \boldsymbol{y} \right\|^2 - \frac{N-L}{N}\nu^2 \right| < \delta \right) \\ &= \mathbb{P}\left(\left| \Omega_2 - (N-L)\frac{\nu^2}{\sigma_{\mathcal{J}}^2} \right| < \frac{\delta}{\sigma_{\mathcal{J}}^2}N \right) \\ &\le \mathbb{P}\left(\Omega_2 - (N-L) < -(N-L)\left(1 - \frac{\nu^2}{\sigma_{\mathcal{J}}^2}\right) + \frac{\delta}{\sigma_{\mathcal{J}}^2}N \right). \end{aligned} \tag{15}$$

For a chi-square random variable, Ω with $(N - L)$ degrees of freedom [4], [16]

$$\mathbb{P}\left(\Omega - (N-L) \le -2\sqrt{(N-L)\lambda} \right) \le e^{-\lambda} \tag{16}$$

and

$$\mathbb{P}\left(\Omega - (N-L) \geq 2\sqrt{(N-L)\lambda} + 2\lambda\right) \leq e^{-\lambda}. \quad (17)$$

By replacing $\Omega = \Omega_1$ and

$$\lambda = \left(\frac{\delta N}{2\nu^2\sqrt{N-L}}\right)^2$$

in (16) and

$$\lambda = \frac{1}{4}\left(\sqrt{N-L+\frac{2\delta}{\nu^2}N} - \sqrt{N-L}\right)^2 \geq \frac{\delta^2}{4\nu^4}\frac{N^2}{N-L+\frac{2\delta}{\nu^2}N}$$

in (17), we obtain using (14)

$$\mathbb{P}\left(\left|\frac{1}{N}\left\|\mathbf{\Pi}^{\perp}_{\boldsymbol{A}_{\mathcal{I}}}\boldsymbol{y}\right\|^2 - \frac{N-L}{N}\nu^2\right| > \delta\right) \leq \exp\left(-\frac{\delta^2}{4\nu^4}\frac{N^2}{N-L}\right) + \exp\left(-\frac{\delta^2}{4\nu^4}\frac{N^2}{N-L+\frac{2\delta}{\nu^2}N}\right) \leq 2\exp\left(-\frac{\delta^2}{4\nu^4}\frac{N^2}{N-L+\frac{2\delta}{\nu^2}N}\right).$$

Similarly, by replacing $\Omega = \Omega_2$ and

$$\lambda = \left(\frac{\sqrt{N-L}}{2}\left(1-\frac{\nu^2}{\sigma^2_{\mathcal{J}}}\right) - \frac{\delta}{\sigma^2_{\mathcal{J}}}\frac{N}{2\sqrt{(N-L)}}\right)^2 = \left(\frac{\sqrt{N-L}}{2}\left(1-\frac{\nu^2}{\sigma^2_{\mathcal{J}}} - \frac{\delta}{\sigma^2_{\mathcal{J}}}\frac{N}{N-L}\right)\right)^2$$

in (16), we obtain using (15)

$$\mathbb{P}\left(\left|\frac{1}{N}\left\|\mathbf{\Pi}^{\perp}_{\boldsymbol{A}_{\mathcal{J}}}\boldsymbol{y}\right\|^2 - \frac{N-L}{N}\nu^2\right| < \delta\right) \leq \exp\left(-\frac{N-L}{4}\left(\frac{\sigma^2_{\mathcal{J}} - \nu^2 - \delta'}{\sigma^2_{\mathcal{J}}}\right)^2\right) = \exp\left(-\frac{N-L}{4}\left(\frac{\sum_{k\in\mathcal{I}\setminus\mathcal{J}}|x_k|^2 - \delta'}{\sum_{k\in\mathcal{I}\setminus\mathcal{J}}|x_k|^2 + \nu^2}\right)^2\right). \quad \square$$

C. *Proof of Theorem 1.1*

We define the event

$$E_{\mathcal{J}} = \{\boldsymbol{y} \text{ and } \mathcal{J} \text{ are } \delta\text{- jointly typical}\}$$

for all $\mathcal{J} \subset \{1,\ldots,M\}, |\mathcal{J}| = L$.

We also define the error event

$$E_0 = \{\mathrm{rank}(\boldsymbol{A}_{\mathcal{I}}) < L\}$$

which results in an order reduction in the model, and implies that the decoder is looking through subspaces of incorrect dimension. By Lemma 2.2, we have $\mathbb{P}(E_0) = 0$.

Since the relationship between M and $\boldsymbol{x}^{(M)}$ is implicit in the following proofs, we will suppress the superscript and just write $\boldsymbol{x}$ for brevity.

1) Proof of Achievability for Error Metric 1: Clearly, the decoder fails if E_0 or $E^C_{\mathcal{I}}$ occur or when one of $E_{\mathcal{J}}$ occurs for $\mathcal{J} \neq \mathcal{I}$. Thus

$$p_{\mathrm{err}}(\mathcal{D}) = \mathbb{P}\left(E_0 \cup E^C_{\mathcal{I}} \bigcup_{\mathcal{J},\mathcal{J}\neq\mathcal{I},|\mathcal{J}|=L} E_{\mathcal{J}}\right) \leq \mathbb{P}\left(E^C_{\mathcal{I}}\right) + \sum_{\mathcal{J},\mathcal{J}\neq\mathcal{I},|\mathcal{J}|=L}\mathbb{P}(E_{\mathcal{J}}).$$

We let $N = (4C_0+1)L$ where $C_0 > 2 + \log(\beta - 1)$ is a constant. Thus, $\delta' = \frac{4C_0+1}{4C_0}\delta = C'_0\delta$ with $C'_0 > 1$. Also by the statement of Theorem 1.1, we have $L\mu^4(\boldsymbol{x})$ grows faster than $\log L$. We note that this requirement is milder than that of [24], where the growth requirement is on $\mu^2(\boldsymbol{x})$ rather than $\mu^4(\boldsymbol{x})$. Since the decoder needs to distinguish between even the smallest nonoverlapping coordinates, we let $\delta' = \zeta\mu^2(\boldsymbol{x})$ for $0 < \zeta < 1$. For computational convenience, we will only consider $2/3 < \zeta < 1$.

By Lemma 2.3

$$\mathbb{P}\left(E^C_{\mathcal{I}}\right) \leq 2\exp\left(-\frac{\zeta^2 C_0}{\nu^2}\frac{L\mu^4(\boldsymbol{x})}{\nu^2 + 2\zeta\mu^2(\boldsymbol{x})}\right)$$

and by the condition on the growth of $\mu(\boldsymbol{x})$, the term in the exponent grows faster than $\log L$. Thus, $\mathbb{P}\left(E^C_{\mathcal{I}}\right)$ goes to 0 faster than $\exp(-c\log L)$.

Again, by Lemma 2.3, for $\mathcal{J}$ with $|\mathcal{I}\cap\mathcal{J}| = K$

$$\mathbb{P}(E_{\mathcal{J}}) \leq \exp\left(-\frac{N-L}{4}\left(\frac{\sum_{k\in\mathcal{I}\setminus\mathcal{J}}|x_k|^2 - \delta'}{\sum_{k\in\mathcal{I}\setminus\mathcal{J}}|x_k|^2 + \nu^2}\right)^2\right).$$

Since $\sum_{k\in\mathcal{I}\setminus\mathcal{J}}|x_k|^2 \geq (L-K)\mu^2(\boldsymbol{x})$, we have

$$\mathbb{P}(E_{\mathcal{J}}) \leq \exp\left(-\frac{N-L}{4}\left(\frac{(L-K)\mu^2(\boldsymbol{x}) - \delta'}{(L-K)\mu^2(\boldsymbol{x}) + \nu^2}\right)^2\right) \quad (18)$$

where $\mu(\boldsymbol{x})$ is defined in (2).

The condition of Theorem 1.1 on $\mu(\boldsymbol{x})$ implies that $\mathbb{P}(E_{\mathcal{J}}) \to 0$ for all K. We note that this condition also implies $P \to \infty$ as N grows without bound. This is due to the stringent requirements imposed by Error Metric 1 in high dimensions.

By a simple counting argument, the number of subsets $\mathcal{J}$ that overlaps $\mathcal{I}$ in K indices (and such that $\mathrm{rank}(\boldsymbol{A}_{\mathcal{J}}) = L$) is upper-bounded by

$$\binom{L}{K}\binom{M-L}{L-K}.$$

Thus, we get the equation at the top of the following page.

We will now show that the summation goes to 0 as $M \to \infty$. We use the following bound:

$$\exp\left(K'\log\left(\frac{L}{K'}\right)\right) \leq \binom{L}{K'} \leq \exp\left(K'\log\left(\frac{Le}{K'}\right)\right) \quad (19)$$

$$p_{\mathrm{err}}(\mathcal{D}) \leq 2\exp\left(-\frac{\zeta^2 C_0}{\nu^2}\frac{L\mu^4(\boldsymbol{x})}{\nu^2+2\zeta\mu^2(\boldsymbol{x})}\right) + \sum_{K=0}^{L-1}\binom{L}{L-K}\binom{M-L}{L-K}\exp\left(-\frac{N-L}{4}\left(\frac{(L-K)\mu^2(\boldsymbol{x})-\delta'}{(L-K)\mu^2(\boldsymbol{x})+\nu^2}\right)^2\right)$$
$$= 2\exp\left(-\frac{\zeta^2 C_0}{\nu^2}\frac{L\mu^4(\boldsymbol{x})}{\nu^2+2\zeta\mu^2(\boldsymbol{x})}\right) + \sum_{K'=1}^{L}\binom{L}{K'}\binom{M-L}{K'}\exp\left(-\frac{N-L}{4}\left(\frac{(K')\mu^2(\boldsymbol{x})-\delta'}{(K')\mu^2(\boldsymbol{x})+\nu^2}\right)^2\right).$$

to upper-bound each term of summation $s_{K'}$ by

$$s_{K'} \leq \exp\Bigg(K'\log\left(\frac{Le}{K'}\right) + K'\log\left(\frac{(M-L)e}{K'}\right) - \frac{N-L}{4}\left(\frac{K'\mu^2(\boldsymbol{x})-\delta'}{K'\mu^2(\boldsymbol{x})+\nu^2}\right)^2\Bigg)$$
$$= \exp\Bigg(L\frac{K'}{L}\log\frac{e}{\frac{K'}{L}} + L\frac{K'}{L}\log\frac{(\beta-1)e}{\frac{K'}{L}} - C_0 L\left(\frac{L\frac{K'}{L}\mu^2(\boldsymbol{x})-\delta'}{L\frac{K'}{L}\mu^2(\boldsymbol{x})+\nu^2}\right)^2\Bigg).$$

We upper-bound the whole summation by maximizing the function

$$\begin{aligned} f(z) &= Lz\log\frac{e}{z} + Lz\log\frac{(\beta-1)e}{z} - C_0 L\left(\frac{Lz\mu^2(\boldsymbol{x})-\delta'}{Lz\mu^2(\boldsymbol{x})+\nu^2}\right)^2 \\ &= -2Lz\log z + Lz(2+\log(\beta-1)) - C_0 L\left(\frac{Lz\mu^2(\boldsymbol{x})-\zeta\mu^2(\boldsymbol{x})}{Lz\mu^2(\boldsymbol{x})+\nu^2}\right)^2 \end{aligned} \tag{20}$$

for $z \in \left[\frac{1}{L}, 1\right]$. If $f(z)$ attains its maximum at z_0, we then have

$$\sum_{K'=1}^{L} s_{K'} \leq L\exp(f(z_0)).$$

For clarity of presentation, we will now state two technical lemmas.

Lemma 2.4: Let $g(z)$ be a twice differentiable function on $[a,b]$ that has a continuous second derivative. If $g(a) < 0$, $g(b) < 0$; $g'(a) < 0$, $g'(b) > 0$, and $g''(a) < 0$, $g''(b) < 0$, then $g''(x)$ is equal to 0 for at least two points in $[a,b]$.

Proof: Since $g'(a) < 0$ and $g'(b) > 0$, $g'(x)$ has to be increasing in a subset $E \subset [a,b]$. Then $g''(x) > 0$ for some $x_0 \in E$. Since $g''(a) < 0$, $g''(x_0) > 0$, and $g''(x)$ is continuous, there exists $x_1 \in [a, x_0]$ such that $g''(x_1) = 0$. Similarly, since $g''(b) < 0$, there exists $x_2 \in [x_0, b]$ such that $g''(x_2) = 0$. □

Lemma 2.5: Let $p(z) = a_4 z^4 + a_3 z^3 + a_2 z^2 + a_1 z + a_0$ be a polynomial over $\mathbb{R}$ such that a_4, a_3, $a_0 > 0$. Then $p(z)$ can have at most two positive roots.

Proof: Let $r_p^{(1)}$, $r_p^{(2)}$, $r_p^{(3)}$, $r_p^{(4)}$ be the roots of $p(z)$, counting multiplicities. Since

$$r_p^{(1)} r_p^{(2)} r_p^{(3)} r_p^{(4)} = \frac{a_0}{a_4} > 0$$

the number of positive roots must be even, and since

$$r_p^{(1)} + r_p^{(2)} + r_p^{(3)} + r_p^{(4)} = -\frac{a_3}{a_4} < 0$$

not all the roots could be positive. The result follows. □

Lemma 2.6: For L sufficiently large, $f(z)$ (see (20)) is negative for all $z \in \left[\frac{1}{L}, 1\right]$. Moreover, the endpoints of the interval $z_0^{(1)} = \frac{1}{L}$ and $z_0^{(2)} = 1$ are its local maxima.

Proof: We first confirm that $f(z)$ is negative at the endpoints of the interval. We use the notation $\vec{\approx}$ for denoting the behavior of $f(z)$ for large L, and $\prec$ and $\succ$ for inequalities that hold asymptotically.

$$f\left(\frac{1}{L}\right) = 2\log L + 2 + \log(\beta-1) - C_0 L\left(\frac{\mu^2(\boldsymbol{x})(1-\zeta)}{\mu^2(\boldsymbol{x})+\nu^2}\right)^2 \prec 0 \tag{21}$$

for sufficiently large L, since $L\mu^4(\boldsymbol{x})$ grows faster than $\log L$. Also, for large L, we have

$$\begin{aligned} f(1) &= L(2+\log(\beta-1)) - C_0 L\left(\frac{\mu^2(\boldsymbol{x})(L-\zeta)}{L\mu^2(\boldsymbol{x})+\nu^2}\right)^2 \\ &\vec{\approx} L(2+\log(\beta-1)-C_0) \prec 0. \end{aligned} \tag{22}$$

We now examine the derivative of $f(z)$, given by

$$f'(z) = -2L\log z + L\log(\beta-1) - 2C_0 L^2\mu^4(\boldsymbol{x})(\nu^2+\zeta\mu^2(\boldsymbol{x}))\frac{Lz-\zeta}{(Lz\mu^2(\boldsymbol{x})+\nu^2)^3}.$$

Also

$$\begin{aligned} f'\left(\frac{1}{L}\right) &= 2L\log L + L\log(\beta-1) - 2C_0 L^2\mu^4(\boldsymbol{x})(\nu^2+\zeta\mu^2(\boldsymbol{x}))\frac{1-\zeta}{(\mu^2(\boldsymbol{x})+\nu^2)^3} \\ &\vec{\approx} L\left(2\log L + \log(\beta-1) - 2\hat{C}_0\frac{L\mu^4(\boldsymbol{x})}{(\mu^2(\boldsymbol{x})+\nu^2)^2}\right) \\ &\prec 0 \end{aligned}$$

for sufficiently large L, since $L\mu^4(\boldsymbol{x})$ grows faster than $\log L$. Similarly

$$\begin{aligned} f'(1) &= L\log(\beta-1) - 2C_0 L^2\mu^4(\boldsymbol{x})(\nu^2+\zeta\mu^2(\boldsymbol{x}))\frac{L-\zeta}{(L\mu^2(\boldsymbol{x})+\nu^2)^3} \\ &\vec{\approx} L\log(\beta-1) - 2C_0\frac{1}{\mu^2(\boldsymbol{x})}(\nu^2+\zeta\mu^2(\boldsymbol{x})) \succ 0 \end{aligned}$$

since $\frac{1}{\mu^2(\boldsymbol{x})}$ grows slower than $\sqrt{\frac{L}{\log L}}$.

Additionally, we get $f''(z)$ defined in (23) at the top of the following page. Thus

$$f''\left(\frac{1}{L}\right) = -2L\Bigg(L + C_0 L^2\mu^4(\boldsymbol{x})(\nu^2+\zeta\mu^2(\boldsymbol{x})) \times \frac{-2\mu^2(\boldsymbol{x})+\nu^2+3\zeta\mu^2(\boldsymbol{x})}{(\mu^2(\boldsymbol{x})+\nu^2)^4}\Bigg) \prec 0$$

$$f''(z) = -\frac{2L}{z} - 2C_0L^2\mu^4(\boldsymbol{x})(\nu^2+\zeta\mu^2(\boldsymbol{x}))\left(\frac{-2Lz\mu^2(\boldsymbol{x})+\nu^2+3\zeta\mu^2(\boldsymbol{x})}{(Lz\mu^2(\boldsymbol{x})+\nu^2)^4}\right)L$$
$$= \frac{-2L}{z(Lz\mu^2(\boldsymbol{x})+\nu^2)^4}\left((Lz\mu^2(\boldsymbol{x})+\nu^2)^4 + C_0L^2\mu^4(\boldsymbol{x})(\nu^2+\zeta\mu^2(\boldsymbol{x}))(-2Lz\mu^2(\boldsymbol{x})+\nu^2+3\zeta\mu^2(\boldsymbol{x}))z\right). \quad (23)$$

and

$$f''(1) = -2L\Bigg(1 + C_0L^2\mu^4(\boldsymbol{x})(\nu^2+\zeta\mu^2(\boldsymbol{x})) \times \frac{-2L\mu^2(\boldsymbol{x})+\nu^2+3\zeta\mu^2(\boldsymbol{x})}{(L\mu^2(\boldsymbol{x})+\nu^2)^4}\Bigg)$$
$$\overrightarrow{\approx} -2L\left(1-2C_0\frac{\nu^2+\zeta\mu^2(\boldsymbol{x})}{L\mu^2(\boldsymbol{x})}\right) \prec 0.$$

Since $f(z)$ is a twice differentiable function on $\left[\frac{1}{L},1\right]$ with a continuous second derivative, Lemma 2.4 implies that $f''(z)$ crosses 0 at least twice in this interval. Next we examine the polynomial (see (23))

$$p(z) = (Lz\mu^2(\boldsymbol{x})+\nu^2)^4$$
$$+2C_0L^2\mu^4(\boldsymbol{x})(\nu^2+\zeta\mu^2(\boldsymbol{x}))(-2Lz\mu^2(\boldsymbol{x})+\nu^2+3\zeta\mu^2(\boldsymbol{x}))z.$$

Since $p(z)$ satisfies the conditions of Lemma 2.5, we conclude that it has at most two positive roots, and thus at most two roots of $p(z)$ can lie in $\left[\frac{1}{L},1\right]$. In other words, $f''(z)$ can cross 0 for $z \in \left[\frac{1}{L},1\right]$ at most twice. Combining this with the previous information, we conclude that $f''(z)$ crosses 0 exactly twice in this interval, and that $f'(z)$ crosses 0 only once, and this point is a local minima of $f(z)$. Thus, the local maxima of $f(z)$ are the endpoints $z_0^{(1)} = \frac{1}{L}$ and $z_0^{(2)} = 1$. □

Thus, we have

$$p_{\mathrm{err}}(\mathcal{D}) \le 2\exp\left(-\frac{\zeta^2C_0}{\nu^2}\frac{L\mu^4(\boldsymbol{x})}{\nu^2+2\zeta\mu^2(\boldsymbol{x})}\right) + \sum_{K=0}^{L-1}\exp\left(\max\left\{f\left(z_0^{(1)}\right), f\left(z_0^{(2)}\right)\right\}\right)$$
$$= 2\exp\left(-\frac{\zeta^2C_0}{\nu^2}\frac{L\mu^4(\boldsymbol{x})}{\nu^2+2\zeta\mu^2(\boldsymbol{x})}\right) + \exp\left(\log L + \max\left\{f\left(\frac{1}{L}\right), f(1)\right\}\right).$$

From (21) and (22), it is clear that

$$\log(L) + \max\left\{f\left(\frac{1}{L}\right), f(1)\right\} \to -\infty$$

as $L \to \infty$. Hence, with the conditions of Theorem 1.1, $p_{\mathrm{err}}(\mathcal{D}) \to 0$ as $L \to \infty$.

2) Proof of Achievability for Error Metric 2: For asymptotically reliable recovery with Error Metric 2, we require that $\mathbb{P}(E_{\mathcal{J}})$ goes to 0 for only $K \le (1-\alpha)L$ with $\alpha \in (0,1)$. By a re-examination of (18), we observe that the right-hand side of

$$\mathbb{P}(E_{\mathcal{J}}) \le \exp\left(-\frac{N-L}{4}\left(\frac{\alpha L\mu^2(\boldsymbol{x})-\delta'}{\alpha L\mu^2(\boldsymbol{x})+\nu^2}\right)^2\right)$$

converges to 0 asymptotically, even when $L\mu^2(\boldsymbol{x})$ converges to a constant. In this case, P does not have to grow with M. Since by the assumption of the theorem, $L\mu^2(\boldsymbol{x})$ and P are both constants, we can write $\hat{\alpha} = \alpha L\mu^2(\boldsymbol{x})/P$. We let $\delta > 0$ (and hence δ') be a constant, and let $N = (4\hat{C}_3+1)L$ for

$$\hat{C}_3 > \beta\left(\frac{\hat{\alpha}P+\nu^2}{\hat{\alpha}P-\delta'}\right)^2. \quad (24)$$

Given that $\delta' > 0$ is arbitrary, we note that this constant only depends on α, β, P, and ν. Hence, we define $p_{\mathrm{err}}(\mathcal{D})$ in the equation at the bottom of the page,where $H(a) = -a\log(a) - (1-a)\log(1-a)$ is the entropy function for $a \in [0,1]$. Since K' is greater than a linear factor of L and since P is a constant, and using (24), we see $p_{\mathrm{err}}(\mathcal{D}) \to 0$ exponentially fast as $L \to \infty$.

3) Proof of Achievability for Error Metric 3: An error occurs for Error Metric 3 if

$$\sum_{k\in\mathcal{I}\setminus\mathcal{J}} |x_k|^2 \ge \gamma P.$$

Thus, we can bound the error event for $\mathcal{J}$ from Lemma 2.3 as

$$\mathbb{P}(E_{\mathcal{J}}) \le \exp\left(-\frac{N-L}{4}\left(\frac{\gamma P-\delta'}{\gamma P+\nu^2}\right)^2\right).$$

$$p_{\mathrm{err}}(\mathcal{D}) \le \mathbb{P}\left(E_{\mathcal{I}}^C\right) + \sum_{K=0}^{(1-\alpha)L}\binom{L}{L-K}\binom{M-L}{L-K}\exp\left(-\frac{N-L}{4}\left(\frac{(L-K)\mu^2(\boldsymbol{x})-\delta'}{(L-K)\mu^2(\boldsymbol{x})+\nu^2}\right)^2\right)$$
$$\le 2\exp\left(-\frac{\delta^2}{4\nu^4}\frac{4\hat{C}_3+1}{4\hat{C}_3+\frac{2\delta}{\nu^2}(4\hat{C}_3+1)}N\right) + \sum_{K'=\alpha L}^{L}\exp\left(LH\left(\frac{K'}{L}\right) + (M-L)H\left(\frac{K'}{M-L}\right) - \hat{C}_3L\left(\frac{K'\mu^2(\boldsymbol{x})-\delta'}{K'\mu^2(\boldsymbol{x})+\nu^2}\right)^2\right)$$

Let $\delta' > 0$ be a fraction of γP. We denote the number of index sets $\mathcal{J} \subset \{1, 2, \ldots, M\}$ with $|\mathcal{J}| = L$ as T_* and note that $T_* \leq \binom{M}{L}$. Thus

$$p_{\text{err}}(\mathcal{D}) \leq 2\exp\left(-\frac{\delta^2}{4\nu^4}\frac{N}{N-L+\frac{2\delta}{\nu^2}N}N\right) + \binom{M}{L}\exp\left(-\frac{N-L}{4}\left(\frac{\gamma P - \delta'}{\gamma P + \nu^2}\right)^2\right).$$

For $N > C_5 L$, a similar argument to that of Section II-C.2 proves that $p_{\text{err}}(\mathcal{D}) \to 0$ exponentially fast as $L \to \infty$, where C_5 depends only on β, γ, P, and ν.

III. Proofs of Converses for the Linear Regime

Throughout this section, we will write $\boldsymbol{x}$ for $\boldsymbol{x}^{(M)}$ whenever there is no ambiguity.

Let the support of $\boldsymbol{x}$ be $\mathcal{I} = \{i_1, i_2, \ldots, i_L\}$ with $i_1 < i_2 < \cdots < i_L$. We assume a genie provides $\boldsymbol{x}_{\mathcal{I}} = (x_{i_1}, x_{i_2}, \ldots, x_{i_L})^T$ to the decoder defined in Section I-A.

Clearly, we have

$$p_{\text{err}} \geq p_{\text{err}}^{\text{genie}}.$$

A. Proof of Converse for Error Metric 1

We derive a lower bound on the probability of genie-aided decoding error for any decoder. Consider a multiple-input single-output (MISO) transmission model given by an encoder, a decoder, and a channel. The channel is specified by $\boldsymbol{H} = [x_{i_1} x_{i_2} \ldots x_{i_L}] = \boldsymbol{x}_{\mathcal{I}}^T$. The encoder, $\mathfrak{E}_1 : \{0,1\}^M \to \mathbb{C}^{L\times N}$, maps one of the $\binom{M}{L}$ possible binary vectors of (Hamming) weight L to a codeword in $\mathbb{C}^{L\times N}$. This codeword is then transmitted over the MISO channel in N channel uses. The decoder is a mapping $\mathfrak{D}_1 : \mathbb{C}^N \to \{0,1\}^M$ such that its output $\hat{\boldsymbol{c}}$ has weight L.

Let $\boldsymbol{c} \in \{0,1\}^M$ and $\text{supp}(\boldsymbol{c}) = \mathcal{J} = \{j_1, j_2, \ldots, j_L\}$ with $j_1 < j_2 < \cdots < j_L$. Let $\boldsymbol{z}_k^{\mathcal{J}} = (a_{k,j_1}, a_{k,j_2}, \ldots, a_{k,j_L})^T$, where $a_{m,n}$ is the (m,n)th term of $\boldsymbol{A}$. The codebook is specified by

$$\mathfrak{C}_1 = \{(\,\boldsymbol{z}_1^{\mathcal{J}} \quad \boldsymbol{z}_2^{\mathcal{J}} \quad \ldots \quad \boldsymbol{z}_N^{\mathcal{J}}\,) \,|\, \mathcal{J} \subset \{1,2,\ldots,M\}, |\mathcal{J}| = L\}$$

and has size $\binom{M}{L}$. The output of the channel, $\boldsymbol{y}$ is

$$y_k = \boldsymbol{H}\boldsymbol{z}_k^{\mathcal{I}} + n_k, \quad \text{for } k = 1, 2, \ldots, N$$

where y_k and n_k are the kth coordinates of $\boldsymbol{y}$ and $\boldsymbol{n}$, respectively. The average signal power is $\mathbb{E}\left(\|\boldsymbol{z}_k^{\mathcal{J}}\|^2\right) = L$, and the noise variance is $\mathbb{E}n_k^2 = \nu^2$. The capacity of this channel in N channel uses (without channel knowledge at the transmitter) is given by [23]

$$C_{\text{MISO}} = N\log\left(1 + \frac{1}{L}\frac{\mathbb{E}\left(\|\boldsymbol{z}_k^{\mathcal{J}}\|^2\right)}{\mathbb{E}n_k^2}\boldsymbol{H}\boldsymbol{H}^\dagger\right) = N\log\left(1 + \frac{P}{\nu^2}\right).$$

After N channel uses, $p_{\text{err}}^{\text{MISO}} > 0$ if $\log\binom{M}{L} > C_{\text{MISO}}$. Using

$$\frac{1}{M+1}\exp\left(MH\left(\frac{L}{M}\right)\right) \leq \binom{M}{L} \leq \exp\left(MH\left(\frac{L}{M}\right)\right) \tag{25}$$

we obtain the equivalent condition

$$N < \frac{1}{\log\left(1 + \frac{P}{\nu^2}\right)}MH\left(\frac{1}{\beta}\right) - o(M)$$

where $L = \beta M$, and $H(\cdot)$ is the entropy function.

To prove Corollary 1.4, we first show that with high probability, all codewords of a Gaussian codebook satisfy a power constraint. Combining this with the strong converse of the channel coding theorem will complete the proof [15]. If $\boldsymbol{A}$ is chosen from a Gaussian distribution, then by inequality (17)

$$\mathbb{P}\left(\frac{1}{L}\|\boldsymbol{z}_k^{\mathcal{J}}\|^2 > \left(1 + 2\left(\sqrt{\beta H\left(\frac{1}{\beta}\right) + \xi}\right) + 2\left(\beta H\left(\frac{1}{\beta}\right) + \xi\right)\right)\right) \leq \exp\left(-\left(\beta H\left(\frac{1}{\beta}\right) + \xi\right)L\right)$$

for any $\mathcal{J} \subset \{1, 2, \ldots, M\}$, $|\mathcal{J}| = L$, and for $k = 1, 2, \ldots, N$. Let

$$\Lambda = 2\sqrt{\beta H\left(\frac{1}{\beta}\right) + \xi} + 2\left(\beta H\left(\frac{1}{\beta}\right) + \xi\right), \quad \text{for } \xi > 0.$$

By the union bound over all $\binom{M}{L}$ possible index sets $\mathcal{J}$ and $k = 1, 2, \ldots, N$

$$\mathbb{P}\left(\frac{1}{L}\|\boldsymbol{z}_k^{\mathcal{J}}\|^2 < (1 + \Lambda),\, \forall\, \mathcal{J},\, k = 1, \ldots, N\right) \geq 1 - N\exp(-\xi L).$$

If the power constraint is satisfied, then the strong converse of the channel coding theorem implies that $p_{\text{err}}(\boldsymbol{A}|\boldsymbol{x})$ goes to 1 exponentially fast in M if

$$N \prec \frac{1}{\log\left(1 + \frac{P(1+\Lambda)}{\nu^2}\right)}MH\left(\frac{1}{\beta}\right).$$

B. Proof of Converse for Error Metric 2

For any given $\boldsymbol{x}$ with $\|\boldsymbol{x}\|_0 = L$, we will prove the contrapositive. Let $P_{e_2}^{(M)}$ denote the probability of error with respect to Error Metric 2 for $\boldsymbol{x} \in \mathbb{C}^M$. We show that $N \succ C_4 L$ if $P_{e_2}^{(M)} \to 0$.

Consider a single-input single-output system, $\mathcal{S}$, whose input is $\boldsymbol{c} \in \{0,1\}^M$, and whose output is $\hat{\boldsymbol{c}} \in \{0,1\}^M$, such that $\|\boldsymbol{c}\|_0 = \|\hat{\boldsymbol{c}}\|_0 = L$, and $\|\boldsymbol{c} - \hat{\boldsymbol{c}}\|_0 \leq 2\alpha L$. The last condition states that the support of $\boldsymbol{c}$ and that of $\hat{\boldsymbol{c}}$ overlap in more than $(1-\alpha)L$ locations, i.e., $P_{e_2}^{(M)} = 0$. We are interested in the rates at which one can communicate reliably over $\mathcal{S}$.

In our case, $d(\boldsymbol{c}, \hat{\boldsymbol{c}}) = \frac{1}{M}\sum_{k=1}^{M} d_H(c_i, \hat{c}_i)$, where $\boldsymbol{c}$ is i.i.d. distributed among $\binom{M}{L}$ binary vectors of length M and weight L, and $d_H(\cdot,\cdot)$ is the Hamming distance. Thus, $D \leq \frac{2\alpha L}{M} = \frac{2\alpha}{\beta}$. We also note that $\mathcal{S}$ can be viewed as consisting of an encoder

$\mathfrak{E}_1$, a MISO channel and a decoder, $\mathfrak{D}_1$ as described in Section III-A. Since the source is transmitted within distortion $\frac{2\alpha}{\beta}$ over the MISO channel, we have [3]

$$R\left(\frac{2\alpha}{\beta}\right) < C_{\text{MISO}}.$$

In order to bound $R\left(\frac{2\alpha}{\beta}\right)$, we first state a technical lemma.

Lemma 3.1: Let $\alpha \in (0,1]$ and $\beta > 2$, and let

$$\begin{aligned} c(z) &= H(z) + (\beta-1)H\left(\frac{z}{\beta-1}\right) \\ &= -2z\log(z) - (1-z)\log(1-z) \\ &\quad + (\beta-1)\log(\beta-1) - (\beta-1-z)\log(\beta-1-z) \end{aligned}$$

where $H(\cdot)$ is the entropy function. Then for $z \in [0,\alpha]$, $c(z) \geq 0$, and $c(z)$ attains its maximum at $z = \min\left(a, \frac{\beta-1}{\beta}\right)$.

Proof: By definition of $H(\cdot)$, $c(z) \geq 0$ for $z \in [0,\alpha]$. By examining

$$\begin{aligned} c'(z) &= -2\log(z) + \log(1-z) + \log(\beta-1-z) \\ &= \log\left(\frac{(1-z)(\beta-1-z)}{z^2}\right) \end{aligned}$$

it is easy to see that $c'(z) \geq 0$ for $z \in \left(0, \min\left(\alpha, \frac{\beta-1}{\beta}\right)\right]$ and $c'(z) < 0$ otherwise. □

Thus we have get the expression shown at the bottom of the page, where the first inequality follows since given $\hat{\boldsymbol{c}}$, $\boldsymbol{c}$ is among $\sum_{K=0}^{\alpha L}\binom{L}{K}\binom{M-L}{K}$ possible binary vectors within Hamming distance $2\alpha L$ from $\hat{\boldsymbol{c}}$. The second inequality follows from inequality (25), and the third inequality follows by Lemma 3.1.

Thus, $R\left(\frac{2\alpha}{\beta}\right) \geq LC_{\alpha,\beta} - o(L)$, where

$$C_{\alpha,\beta} = \begin{cases} \beta H\left(\frac{1}{\beta}\right) - H(\alpha) - (\beta-1)H\left(\frac{\alpha}{\beta-1}\right), & \text{if } \alpha \leq \frac{\beta-1}{\beta} \\ 0, & \text{if } \alpha > \frac{\beta-1}{\beta}. \end{cases} \tag{26}$$

Therefore, if $P_{e_2}^{(M)} = 0$, then

$$LC_{\alpha,\beta} - o(L) < N\log\left(1 + \frac{P}{\nu^2}\right)$$

or, equivalently, for large M

$$N \succ \frac{C_{\alpha,\beta}}{\log\left(1+\frac{P}{\nu^2}\right)}L.$$

The contrapositive statement proves the result.

C. Proof of Converse for Error Metric 3

For Error Metric 3, we assume that $\rho(\boldsymbol{x}) = \max_{i\in\mathcal{I}}|x_i|$ and $\mu(\boldsymbol{x}) = \min_{i\in\mathcal{I}}|x_i|$ both decay at rate $O\left(\sqrt{\frac{1}{L}}\right)$. Thus, P is constant.[4] In the absence of this assumption, some terms of $\boldsymbol{x}$ can be asymptotically dominated by noise. Such terms are unimportant for recovery purposes, and therefore could be replaced by zeros (in the definition of $\boldsymbol{x}$) with no significant harm.

Let $\alpha(\gamma,\boldsymbol{x}) = \min\left(\frac{\gamma P}{L\mu^2(\boldsymbol{x})}, 1\right)$. We note that by the assumption of the theorem, $P/L\mu^2(\boldsymbol{x})$ is a constant. Let $P_{e_3}^{(M)}$ denote the probability of error with respect to Error Metric 3 for $\boldsymbol{x} \in \mathbb{C}^M$. If $P_{e_3}^{(M)} = 0$ and if an index set $\mathcal{J}$ is recovered, then $\sum_{k\in\mathcal{I}\setminus\mathcal{J}}|x_k|^2 \leq \gamma P$, where $\mathcal{I} = \text{supp}(\boldsymbol{x})$. This implies that $|\mathcal{I}\setminus\mathcal{J}| \leq \alpha(\gamma,\boldsymbol{x})L$. Thus, $P_{e_3}^{(M)} = 0$ implies that $P_{e_2}^{(M)} = 0$ when recovering $\alpha(\gamma,\boldsymbol{x})$ fraction of the support of $\boldsymbol{x}$. As shown in Section III-B, reliable recovery of $\boldsymbol{x}$ is not possible if

$$N \prec \frac{C_{\alpha(\gamma,\boldsymbol{x}),\beta}}{\log\left(1+\frac{P}{\nu^2}\right)}L$$

where $C_{\alpha(\gamma,\boldsymbol{x}),\beta}$ is a constant (as defined in (26)) that only depends on γ, β, and P for a given $\boldsymbol{x}$.

IV. Proofs for the Sublinear Regime

The proofs for the sublinear regime follow the same steps as those in the linear regime. For the proofs of converse results, we use the bounds from (19) instead of those of (25). We provide outlines, highlighting any differences.

A. Outline of the Proof of Theorem 1.5

The proof is similar to that of Theorem 1.1, with $f(z)$ replaced by

$$k(z) = -2Lz\log z + 2Lz + Lz\log\left(\frac{M-L}{L}\right) - \frac{N-L}{4}\left(\frac{Lz\mu^2(\boldsymbol{x}) - \zeta\mu^2(\boldsymbol{x})}{Lz\mu^2(\boldsymbol{x}) + \nu^2}\right)^2.$$

The behavior of $k(z)$, $k'(z)$, and $k''(z)$ at the endpoints $\left\{\frac{1}{L}, 1\right\}$ is the same as that in the proof of Theorem 1.1 whenever $N = C_1'L\log(M-L)$. The result follows in an analogous way for Error Metrics 1, 2 and 3.

[4] Technically, P is bounded above and below by constants, and we note that this does not affect the proof. Without loss of generality and to ease the notation, we will take P to be a constant.

$$\begin{aligned} I(\boldsymbol{c},\hat{\boldsymbol{c}})\big|_{||\boldsymbol{c}||_0=||\hat{\boldsymbol{c}}||_0=L,||\boldsymbol{c}-\hat{\boldsymbol{c}}||_0\leq 2\alpha L} &= H(\boldsymbol{c}) - H(\boldsymbol{c}|\hat{\boldsymbol{c}})\big|_{||\boldsymbol{c}||_0=||\hat{\boldsymbol{c}}||_0=L,||\boldsymbol{c}-\hat{\boldsymbol{c}}||_0\leq 2\alpha L} \\ &\geq \log\binom{M}{L} - \log\left(\sum_{K=0}^{\alpha L}\binom{L}{K}\binom{M-L}{K}\right) \\ &\geq MH\left(\frac{1}{\beta}\right) - \log(M+1) - \log\left(\sum_{K=0}^{\alpha L}\exp\left(LH\left(\frac{K}{L}\right) + (M-L)H\left(\frac{K}{M-L}\right)\right)\right) \\ &\geq \begin{cases} MH\left(\frac{1}{\beta}\right) - \log(M+1) - \log(\alpha L+1) - L\left(H(\alpha) + (\beta-1)H\left(\frac{\alpha}{\beta-1}\right)\right), & \text{if } \alpha \leq \frac{\beta-1}{\beta} \\ 0, & \text{if } \alpha > \frac{\beta-1}{\beta} \end{cases} \end{aligned}$$

B. Outline of the Proof of Theorem 1.6

For Error Metric 1, the proof is essentially the same as that of Theorem 1.3. For Error Metric 2, we have the following technical lemma.

Lemma 4.1: Let $\alpha \in (0,1]$ and $L = o(M)$, and let

$$d(z) = 2z - 2z\log(z) + z\log\left(\frac{M-L}{L}\right).$$

Then for $z \in [0,\alpha]$, and for sufficiently large M, $d(z)$ attains its maximum at $z = \alpha$.

Proof: By examining

$$d'(z) = -2\log(z) + \log\left(\frac{M-L}{L}\right) = \log\left(\frac{M-L}{Lz^2}\right)$$

it is easy to see that $d'(z) \succ 0$ for sufficiently large M. □

Thus, we get the expression at the bottom of the page, where the first inequality follows from inequality (19), and the second inequality follows by Lemma 4.1 for sufficiently large M. The rest of the proof is analogous to that of Theorem 1.3.

For Error Metric 3, we let $\alpha(\gamma, \boldsymbol{x}) = \min\left(\frac{\gamma P}{L\mu^2(\boldsymbol{x})}, 1\right)$, and conclude that $P_{e_3}^{(M)} = 0$ implies that $P_{e_2}^{(M)} = 0$ when recovering $\alpha(\gamma, \boldsymbol{x})$ fraction of the support of $\boldsymbol{x}$. The rest of the proof is analogous to that of Theorem 1.3.

V. Proof of Distortion Characterization for Error Metric 3

To prove this result, we use eigenvalue concentration properties of Gaussian matrices, which is referred to as the restricted isometry principle (RIP) in the context of compressive sampling [5]. With our scaling, RIP states that for a given constant $\epsilon_L > 0$, with overwhelming probability, a Gaussian matrix $\boldsymbol{A}$ satisfies

$$(1-\epsilon_L)\|\boldsymbol{x}\|_2^2 \le \frac{1}{N}\|\boldsymbol{A}\boldsymbol{x}\|_2^2 \le (1+\epsilon_L)\|\boldsymbol{x}\|_2^2 \qquad (27)$$

for any $\boldsymbol{x} \in \mathbb{C}^M$ with $\|\boldsymbol{x}\|_0 \le L$, as long as $N > \hat{C}_3 L\log(M/L)$ for some constant $\hat{C}_3$ that depends on ϵ_L. ϵ_L is referred to as the restricted isometry constant of $\boldsymbol{A}$.

We first state some implications of RIP. For any index set $\mathcal{J} \subset \{1,2,\ldots,M\}$ with $|\mathcal{J}| = L$, we have

$$(1-\epsilon_L) \le \lambda_{\min}\left(\frac{1}{N}\boldsymbol{A}_{\mathcal{J}}^*\boldsymbol{A}_{\mathcal{J}}\right) \le \lambda_{\max}\left(\frac{1}{N}\boldsymbol{A}_{\mathcal{J}}^*\boldsymbol{A}_{\mathcal{J}}\right) \le (1+\epsilon_L)$$

where $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ denote the minimum and maximum eigenvalues of $(\cdot)$, respectively. It can also be shown [11], that for any two integers $L \le L'$, we have $\epsilon_L \le \epsilon_{L'}$. Also, for two disjoint index sets $\mathcal{I}, \mathcal{J} \subset \{1,2,\ldots,M\}$, we have

$$\|\boldsymbol{A}_{\mathcal{J}}^*\boldsymbol{A}_{\mathcal{I}}\boldsymbol{b}\|_2 \le \epsilon_{|\mathcal{I}|+|\mathcal{J}|}\|\boldsymbol{b}\|_2$$

for $\boldsymbol{b} \in \mathbb{C}^{|\mathcal{I}|}$ as long as $\epsilon_{|\mathcal{I}|+|\mathcal{J}|} < 1$.

Next, we introduce some more notation for this proof. We denote the pseudoinverse of $\boldsymbol{A}_{\mathcal{I}}$ by $\boldsymbol{A}_{\mathcal{I}}^\dagger = (\boldsymbol{A}_{\mathcal{I}}^*\boldsymbol{A}_{\mathcal{I}})^{-1}\boldsymbol{A}_{\mathcal{I}}^*$. For an index set $\mathcal{I} = \{i_1, i_2, \ldots, i_L\}$ with $i_1 < i_2 < \cdots < i_L$, we let $\boldsymbol{x}_{\mathcal{I}} = (x_{i_1}, x_{i_2}, \ldots, x_{i_L})^T \in \mathbb{C}^{|\mathcal{I}|}$ (as defined in Section III). We also let the restriction of $\boldsymbol{x}$ to $\mathcal{I}$ be the vector $\boldsymbol{x}|_{\mathcal{I}} \in \mathbb{C}^M$ such that

$$\boldsymbol{x}|_{\mathcal{I}} = \begin{cases} x_i & \text{if } i \in \mathcal{I} \\ 0 & \text{otherwise.} \end{cases}$$

Finally, for $\boldsymbol{v} \in \mathbb{C}^{|\mathcal{I}|}$, we define the following assignment operation:

$$\boldsymbol{x}|_{\mathcal{I}} \xleftarrow{\mathcal{I}} \boldsymbol{v}$$

which results in $x_{i_k} = v_k$ and $x_i = 0$ for $i \notin \mathcal{I}$.

A. Proof of Theorem 1.7

Given ϵ_L any matrix from the Gaussian ensemble satisfies (27) with probability $1-\exp(-\kappa_1 N)$ for some constant $\kappa_1 > 0$ if $N > \hat{C}_3 L\log(M/L)$. Similarly, given γ, from the proof of Corollary 1.2, we have that $p_{\text{err}}(\boldsymbol{A}) \le \exp(-\kappa_0' N)$ with probability $1-\exp(-\kappa_0' N)$ for some constant $\kappa_0' > 0$ if $N > C_3 L$ in the linear regime. Let $\hat{C} = \max(C_3, \hat{C}_3\log\beta)$. Then if the condition in the statement of the theorem is satisfied, with probability $1-2\exp(-\kappa N)$, any matrix $\boldsymbol{A}$ from the Gaussian ensemble will both satisfy (27) and have $p_{\text{err}}(\boldsymbol{A}) \le \exp(-\kappa_0' N)$, where $\kappa = \min(\kappa_1, \kappa_0')$ with respect to Error Metric 3 for $\boldsymbol{x}$. We next fix $\boldsymbol{A}$ to be a measurement matrix from the Gaussian ensemble satisfying RIP and $p_{\text{err}}(\boldsymbol{A}) \le \exp(-\kappa_0' N)$.

Next, we characterize $\mathbb{E}_{\boldsymbol{n}}(\Delta_{\text{genie}}^2)$. Once $\mathcal{I}$ is known at the decoder, the optimal estimate is

$$\hat{\boldsymbol{x}}_{\text{genie}}|_{\mathcal{I}} \xleftarrow{\mathcal{I}} \boldsymbol{A}_{\mathcal{I}}^\dagger\boldsymbol{y} = \boldsymbol{x}_{\mathcal{I}} + \boldsymbol{A}_{\mathcal{I}}^\dagger\boldsymbol{n}.$$

Thus

$$\Delta_{\text{genie}} = \|\boldsymbol{x} - \hat{\boldsymbol{x}}_{\text{genie}}\|_2 = \|\boldsymbol{A}_{\mathcal{I}}^\dagger\boldsymbol{n}\|_2.$$

It is easy to show that

$$\mathbb{E}_{\boldsymbol{n}}(\Delta_{\text{genie}}^2) = \nu^2\text{trace}\left((\boldsymbol{A}_{\mathcal{I}}^*\boldsymbol{A}_{\mathcal{I}})^{-1}\right).$$

It follows from RIP that

$$\frac{L}{N}\frac{\nu^2}{1+\epsilon_L} \le \mathbb{E}_{\boldsymbol{n}}(\Delta_{\text{genie}}^2) \le \frac{L}{N}\frac{\nu^2}{1-\epsilon_L}.$$

$$\begin{aligned} I(\boldsymbol{c},\hat{\boldsymbol{c}})\big|_{\|\boldsymbol{c}\|_0=\|\hat{\boldsymbol{c}}\|_0=L,\|\boldsymbol{c}-\hat{\boldsymbol{c}}\|_0\le 2\alpha L} &= H(\boldsymbol{c}) - H(\boldsymbol{c}|\hat{\boldsymbol{c}})\big|_{\|\boldsymbol{c}\|_0=\|\hat{\boldsymbol{c}}\|_0=L,\|\boldsymbol{c}-\hat{\boldsymbol{c}}\|_0\le 2\alpha L} \\ &\ge L\log\left(\frac{M}{L}\right) - \log\left(\sum_{K=0}^{\alpha L}\exp\left(K\log\left(\frac{Le}{K}\right) + K\log\left(\frac{(M-L)e}{K}\right)\right)\right) \\ &\ge L\log(M) - \alpha L\log(M-L) - o(L\log M) \ge (1-\alpha)L\log(M-L) - o(L\log M) \end{aligned}$$

With probability greater than or equal to $1-\exp(-\kappa_0' N)$, the joint typicality decoder outputs a set of indices $\mathcal{J}$ such that

$$\sum_{k\in\mathcal{I}\setminus\mathcal{J}} |x_k|^2 = \|\boldsymbol{x}_{\mathcal{I}\setminus\mathcal{J}}\|_2^2 < \gamma P.$$

If it fails, it outputs $\hat{\boldsymbol{x}}=0$. If recovery is successful, the estimate $\hat{\boldsymbol{x}}$ is given by $\hat{\boldsymbol{x}}|_{\mathcal{J}} \xleftarrow{\mathcal{J}} \boldsymbol{A}_{\mathcal{J}}^{\dagger}\boldsymbol{y}$. Let $\hat{\boldsymbol{u}}\in\mathbb{C}^M$ be the vector such that

$$\hat{\boldsymbol{u}}|_{\mathcal{J}} \xleftarrow{\mathcal{J}} (\boldsymbol{A}_{\mathcal{J}}^{*}\boldsymbol{A}_{\mathcal{J}})^{-1}\boldsymbol{A}_{\mathcal{J}}^{*}\boldsymbol{A}_{\mathcal{I}}\boldsymbol{x}_{\mathcal{I}}.$$

We have that

$$\mathbb{E}_{\boldsymbol{n}}\left(\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|_2^2\right) = \mathbb{E}_{\boldsymbol{n}}\left(\|\boldsymbol{x}-\hat{\boldsymbol{u}}\|_2^2\right) + \mathbb{E}_{\boldsymbol{n}}\left(\|\boldsymbol{A}_{\mathcal{J}}^{\dagger}\boldsymbol{n}\|_2^2\right)$$

since $\mathbb{E}_{\boldsymbol{n}}\boldsymbol{n}=\boldsymbol{0}$.

Since $\boldsymbol{x}|_{\mathcal{I}} = \boldsymbol{x}|_{\mathcal{I}\cap\mathcal{J}} + \boldsymbol{x}|_{\mathcal{I}\setminus\mathcal{J}}$, we can write

$$\begin{aligned}\hat{\boldsymbol{u}}|_{\mathcal{J}} &\xleftarrow{\mathcal{J}} \boldsymbol{A}_{\mathcal{J}}^{\dagger}\left(\boldsymbol{A}_{\mathcal{I}\cap\mathcal{J}}\boldsymbol{x}_{\mathcal{I}\cap\mathcal{J}} + \boldsymbol{A}_{\mathcal{I}\setminus\mathcal{J}}\boldsymbol{x}_{\mathcal{I}\setminus\mathcal{J}}\right)\\ &= \boldsymbol{x}|_{\mathcal{I}\cap\mathcal{J}} + \boldsymbol{A}_{\mathcal{J}}^{\dagger}\boldsymbol{A}_{\mathcal{I}\setminus\mathcal{J}}\boldsymbol{x}_{\mathcal{I}\setminus\mathcal{J}}.\end{aligned}$$

Thus, since $\boldsymbol{x}-\boldsymbol{x}|_{\mathcal{I}\cap\mathcal{J}} = \boldsymbol{x}|_{\mathcal{I}\setminus\mathcal{J}}$

$$\begin{aligned}\|\boldsymbol{x}-\hat{\boldsymbol{u}}\|_2^2 &= \|\boldsymbol{x}_{\mathcal{I}\setminus\mathcal{J}}\|_2^2 + \|(\boldsymbol{A}_{\mathcal{J}}^{*}\boldsymbol{A}_{\mathcal{J}})^{-1}\boldsymbol{A}_{\mathcal{J}}^{*}\boldsymbol{A}_{\mathcal{I}\setminus\mathcal{J}}\boldsymbol{x}_{\mathcal{I}\setminus\mathcal{J}}\|_2^2\\ &\le \gamma P + \gamma P\left(\frac{\epsilon_{|\mathcal{I}\setminus\mathcal{J}|+|\mathcal{J}|}}{1-\epsilon_L}\right)^2\\ &\le \gamma P\left(1+\left(\frac{\epsilon_{2L}}{1-\epsilon_L}\right)^2\right). \qquad (28)\end{aligned}$$

It follows that

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{n}}\left(\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|_2^2\right)\\ &\le e^{-\kappa_0' N}P + (1-e^{-\kappa_0' N})\left(\mathbb{E}_{\boldsymbol{n}}(\|\boldsymbol{x}-\hat{\boldsymbol{u}}\|_2^2) + \mathbb{E}_{\boldsymbol{n}}\|\boldsymbol{A}_{\mathcal{J}}^{\dagger}\boldsymbol{n}\|_2^2\right)\\ &= e^{-\kappa_0' N}P + (1-e^{-\kappa_0' N})\left(\mathbb{E}_{\boldsymbol{n}}(\|\boldsymbol{x}-\hat{\boldsymbol{u}}\|_2^2)\right.\\ &\qquad\left. + \nu^2 \mathrm{trace}\left((\boldsymbol{A}_{\mathcal{J}}^{*}\boldsymbol{A}_{\mathcal{J}})^{-1}\right)\right)\\ &\le e^{-\kappa_0' N}P + \gamma P\left(1+\left(\frac{\epsilon_{2L}}{1-\epsilon_L}\right)^2\right) + \frac{L}{N}\frac{\nu^2}{1-\epsilon_L}. \qquad (29)\end{aligned}$$

Finally

$$\begin{aligned}&\frac{\mathbb{E}_{\boldsymbol{n}}\left(\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|_2^2\right)}{\mathbb{E}_{\boldsymbol{n}}\left(\Delta_{\mathrm{genie}}^2\right)}\\ &\quad\le e^{-\kappa_0' N}P\frac{N}{L}\frac{1+\epsilon_L}{\nu^2}\\ &\qquad + \gamma P\left(1+\left(\frac{\epsilon_{2L}}{1-\epsilon_L}\right)^2\right)\frac{N}{L}\frac{1+\epsilon_L}{\nu^2} + \frac{1+\epsilon_L}{1-\epsilon_L}\\ &\quad\le e^{-\kappa_0' N}P\frac{1}{\hat{C}}\frac{1+\epsilon_L}{\nu^2}\\ &\qquad + \gamma P\left(1+\left(\frac{\epsilon_{2L}}{1-\epsilon_L}\right)^2\right)\frac{1}{\hat{C}}\frac{1+\epsilon_L}{\nu^2} + \frac{1+\epsilon_L}{1-\epsilon_L}. \qquad (30)\end{aligned}$$

For sufficiently large N, the first term can be made arbitrarily small, i.e., smaller than any real number $\xi > 0$. The sum of the second and the third terms is a constant that depends only on β, γ, P, ν, and ϵ_L.

In the sublinear sparsity regime, we let $\hat{C}' = \max\left(\hat{C}_3, C_3'\right)$, where $\hat{C}_3$, C_3' are constant that have been defined previously. If $N > \hat{C}' L \log(M-L)$, then with probability $1-2\exp(-\kappa N)$, any matrix $\boldsymbol{A}$ from the Gaussian ensemble will both satisfy (27) and have $p_{\mathrm{err}}(\boldsymbol{A}) \le \exp\left(-\frac{\kappa}{2}N\right)$, for some constant κ with respect to Error Metric 3. We fix $\boldsymbol{A}$ to be a measurement matrix from the Gaussian ensemble, satisfying RIP and $p_{\mathrm{err}}(\boldsymbol{A}) \le \exp\left(-\frac{\kappa}{2}N\right)$. The condition that $N > \hat{C}' L\log(M-L)$ implies that $\Delta_{\mathrm{genie}}^2 = O(1/\log M)$. In this case, the second term on the right-hand side of inequality (30) is going to contribute $O\left(\Delta_{\mathrm{genie}}^2 \log M\right)$. Thus, for the sublinear regime, recovery with respect to Error Metric 3 implies that $\mathbb{E}_{\boldsymbol{n}}\left(\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|_2^2\right) \le \tilde{C}'\mathbb{E}_{\boldsymbol{n}}\left(\Delta_{\mathrm{genie}}^2\right)\log M$ for some constant $\tilde{C}' > 0$.

REFERENCES

[1] M. Akçakaya and V. Tarokh, "A frame construction and a universal distortion bound for sparse representations," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2443–2450, Jun. 2008.

[2] B. Babadi, N. Kalouptsidis, and V. Tarokh, "Asymptotic achievability of the Cramer-Rao bound for noisy compressive sampling," *IEEE Trans. signal Process.*, vol. 57, no. 3, pp. 1233–1236, Mar. 2009.

[3] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

[4] L. Birgé and P. Massart, "Minimum contrast estimators on sieves: Exponential bounds and rates of convergence," *Bernoulli*, vol. 4, no. 3, pp. 329–375, Sep. 1998.

[5] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[6] E. J. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Ann. Statist.*, vol. 35, pp. 2313–2351, Dec. 2007.

[7] E. J. Candès and J. Romberg, "Practical signal recovery from random projections," in *Proc. XI Conf. Wavelet Applications and Signal Image Processing*, San Diego, CA, 2005.

[8] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery for incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, pp. 1207–1223, Aug. 2006.

[9] S. S. Chen, D. L. Donoho, and M. A. Saunder, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, Jul. 2003.

[10] T. M. Cover and J. M. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.

[11] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[12] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[13] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Necessary and sufficient conditions on sparsity pattern recovery," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5758–5772, Dec. 2009, , Apr. 2008.

[14] A. K. Fletcher, S. Rangan, V. K. Goyal, and K. Ramchadran, "Denoising by sparse approximation: Error bounds based on rate-distortion theory," *EURASIP J. Appl. Signal Process.*, vol. 2006, pp. 1–19, Article ID 26318.

[15] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[16] B. Laurent and P. Massart, "Adaptive estimation of a quadratic functional by model selection," *Ann. Statist.*, vol. 28, no. 5, pp. 1303–1338, Oct. 2000.

[17] D. Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comp. Harmonic Anal.*, vol. 26, pp. 301–321.

[18] D. Omidiran and M. J. Wainwright, "High-Dimensional Subset Recovery in Noise: Sparsified Measurements Without Loss of Statistical Efficiency," Dept. Statist., UC Berkeley, Berkeley, CA, Tech. Rep., 2008.

[19] G. Reeves and M. Gastpar, "Sampling bounds for sparse support recovery in the presence of noise," in *Proc. IEEE Int. Symp. Information Theory*, Toronto, ON, Canada, Jul. 2008, pp. 2187–2191.

[20] S. Sarvotham, D. Baron, and R. Baraniuk, "Measurements vs. bits: Compressed sensing meets information theory," in *Proc. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Sep. 2006.

[21] J. A. Tropp, "Topics in Sparse Approximation," Ph.D. dissertation, Computational and Applied Mathematics, UT-Austin, TX, 2004.

[22] J. A. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.

[23] B. Vucetic and J. Yuan*, Space-Time Coding*. New York: Wiley, 2003.

[24] M. J. Wainwright, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5728–5741, Dec. 2009.

[25] M. J. Wainwright, "Sharp thresholds for noisy and high-dimensional recovery of sparsity using $\mathcal{L}_1$-constrained quadratic programming," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2183–2202, May 2009, "," Dept. Statistics, UC Berkeley, Berkeley, CA, Tech, Rep., 2006.

[26] W. Wang, M. J. Wainwright, and K. Ramchandran, "Information-Theoretic Limits on Sparse Signal Recovery: Dense Versus Sparse Measurement Matrices," Dept. Statistics, UC Berkeley, Berkeley, CA, Tech. Rep., 2008.

Mehmet Akçakaya (S'04) received the Bachelor's degree in electrical engineering in 2005 with great distinction from McGill University, Montreal, QC, Canada.

He is currently working toward the Ph.D. degree in the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA. His research interests include the theory of sparse representations, compressive sampling and its applications to cardiovascular MRI.

Vahid Tarokh (M'97–SM'02–F'09) received the Ph.D. degree in electrical engineering from the University of Waterloo, Waterloo, ON, Canada, in 1995.

He is a Perkins Professor of Applied Mathematics and Hammond Vinton Hayes Senior Fellow of Electrical Engineering at Harvard University, Cambridge, MA. At Harvard, he teaches courses and supervises research in communications, networking and signal processing.

Dr. Tarokh has received a number of major awards and holds two honorary degrees.