

# Shannon Theoretic Limits on Noisy Compressive Sampling

Mehmet Akçakaya and Vahid Tarokh

## Abstract

In this paper, we study the number of measurements required to recover a sparse signal in  $\mathbb{C}^M$  with  $L$  non-zero coefficients from compressed samples in the presence of noise. For a number of different recovery criteria, we prove that  $O(L)$  (an asymptotically linear multiple of  $L$ ) measurements are necessary and sufficient if  $L$  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. We also show that  $O(L \log(M - L))$  measurements are required in the sublinear regime ( $L = o(M)$ ).

## Index Terms

Shannon theory, compressive sampling, linear regime

## I. INTRODUCTION

Let  $\mathbb{C}$  denote the complex field and  $\mathbb{C}^M$  the  $M$ -dimensional complex space. For any  $\mathbf{x} \in \mathbb{C}^M$ , let  $\|\mathbf{x}\|_0$  denote the number of non-zero coefficients of  $\mathbf{x}$ . Whenever  $\|\mathbf{x}\|_0 = L \ll M$ , it is advantageous to measure a linear combination of the components of  $\mathbf{x}$  as

$$\mathbf{y} = \mathbf{A}\mathbf{x},$$

where  $\mathbf{A}$  is an  $N \times M$  measurement matrix.

A decoder can then recover  $\mathbf{x}$  from the observed vector by solving the  $\mathcal{L}_0$  minimization problem

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

M. Akçakaya and V. Tarokh are with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, 02138. (e-mails: {akcakaya, vahid}@deas.harvard.edu)

This data acquisition technique for sparse signals is called compressive sampling [4], [5]. However, the optimization problem for recovery is NP-hard to solve [8]. 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

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

and then establishes criteria under which the solution to this problem is also that of the  $\mathcal{L}_0$  minimization problem. By considering certain classes of Gaussian and partial Fourier ensembles, Candès and Tao showed in [4] that this recovery problem could be solved for  $L = O(M)$  with  $N = O(L)$  as long as the observations are noiseless. 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.

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

for some additive noise  $\mathbf{n} \in \mathbb{C}^N$ . 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 the linear sparsity regime  $M = \beta L$  for  $\beta > 2$ . It was shown in [1] that  $\beta > 2$  is required even in the noiseless setting for the unique recovery of the signal.

Wainwright considered this problem with  $\mathbf{n}$  being Gaussian noise in [10], and derived information theoretic limits on the noisy problem for a specific performance metric and a decoder that decodes to the closest subspace, showing that for the linear sparsity regime, the number of measurements required is also  $O(L)$ . In [11], Wainwright studied the  $\mathcal{L}_1$  constrained quadratic programming algorithm (LASSO) in the noisy setting and showed that in this case the number of measurements required is  $N = O(L \log(M-L))$ . Therefore there is a gap between what is achievable theoretically with an information theoretic decoder and what is achievable with a practical decoder based on  $\mathcal{L}_1$  regularization. The total power of the signal,

$$\|\mathbf{x}\|_2^2 = P$$

grows unboundedly as a function of  $N$  according to the analysis in [10]. The reason for this requirement is that at high dimensions, the performance metric in consideration is too stringent for an average case analysis.

In this note, we consider various performance metrics, some of which are of more Shannon theoretic spirit. We use a decoder based on joint typicality. Although such a decoder may not be computationally feasible in practice, it enables us to characterize the performance limits on the sparse representation

problem. Using this decoder, we first derive a result similar to that of [10] for the same performance metric. For the other performance metrics that are more statistical in nature, we derive results stating that the number of required measurements is  $O(L)$  and that  $P$  does not have to grow with  $N$ .

The outline of this paper is given next. In Section II, we define the problem to be considered in this paper, establish the notation and performance metrics, and state our main results and their implications. Section III and Section IV provide the proofs for the theorems stated in Section II. In Section V, we state analogous theorems for the sublinear sparsity regime,  $L = o(M)$ .

## II. MAIN RESULTS

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

$$\text{supp}(\mathbf{x}) = \{i \mid x_i \neq 0\}$$

with  $\|\mathbf{x}\|_0 = |\mathcal{I}| = L = \lfloor \frac{1}{\beta}M \rfloor$ , where  $\beta > 2$ . We also define

$$\mu(\mathbf{x}) = \min_{i \in \mathcal{I}} |x_i|. \quad (2)$$

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

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

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

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

where  $\mathbb{I}(\cdot)$  is the indicator function and  $\alpha, \gamma \in (0, 1)$ .

Error Metric 1 is referred to as the 0-1 loss metric, and it is the one considered by Wainwright [10]. Error Metric 2 is a statistical extension of Error Metric 1, and considers the recovery of most of the subspace information of  $\mathbf{x}$ . Error Metric 3 is directly from Shannon Theory and characterizes the recovery of most of the energy of  $\mathbf{x}$ .

Consider a sequence of vectors,  $\{\mathbf{x}^{(M)}\}_M$  such that  $\mathbf{x}^{(M)} \in \mathbb{C}^M$  with  $\mathcal{I}^{(M)} = \text{supp}(\mathbf{x}^{(M)})$ , where  $|\mathcal{I}^{(M)}| = L^{(M)} = \lfloor \frac{1}{\beta}M \rfloor$ . For  $\mathbf{x}^{(M)}$ , we will consider an ensemble of  $N \times M$  Gaussian measurement

matrices,  $\mathbf{A}^{(M)}$ , where  $N$  is a function of  $M$ . Since the dependence of  $L^{(M)}, \mathcal{I}^{(M)}$  and  $\mathbf{A}^{(M)}$  on  $M$  is implied by the vector  $\mathbf{x}^{(M)}$ , we will omit the superscript for brevity, and denote the support of  $\mathbf{x}^{(M)}$  by  $\mathcal{I}$ , its size by  $L$  and any measurement matrix from the ensemble by  $\mathbf{A}$ , whenever there is no ambiguity.

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

$$p_{\text{err}}(\mathcal{D}|\mathbf{x}^{(M)}) = \mathbb{E}_{\mathbf{A}}(p_{\text{err}}(\mathbf{A}|\mathbf{x}^{(M)})), \quad (6)$$

where  $p_{\text{err}}(\mathbf{A}|\mathbf{x}^{(M)}) = \mathbb{P}(\mathcal{D}(\mathbf{y}) \neq \mathcal{I})$  for  $\mathbf{y} = \mathbf{A}\mathbf{x}^{(M)} + \mathbf{n}$  and  $\mathbb{P}(\cdot)$  is the probability measure.

We say a decoder achieves *asymptotic reliable sparse recovery* if  $p_{\text{err}}(\mathcal{D}|\mathbf{x}^{(M)}) \rightarrow 0$  as  $M \rightarrow \infty$ . Similarly we say asymptotic reliable sparse recovery is not possible if  $p_{\text{err}}(\mathcal{D}|\mathbf{x}^{(M)})$  stays bounded away from 0 as  $M \rightarrow \infty$ .

We also use the notation

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

for either  $f(x) = g(x) = 0$  or for non-decreasing non-negative 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  $g(x) \succ f(x)$ .

*Theorem 2.1:* (Achievability for Error Metric 1) Let a sequence of sparse vectors,  $\{\mathbf{x}^{(M)} \in \mathbb{C}^M\}_M$  with  $\|\mathbf{x}^{(M)}\|_0 = L = \lfloor \frac{1}{\beta} M \rfloor$ , where  $\beta > 2$  be given. Then asymptotic reliable recovery is possible for  $\{\mathbf{x}^{(M)}\}$  with respect to Error Metric 1 if  $\frac{L\mu^4(\mathbf{x}^{(M)})}{\log L} \rightarrow \infty$  as  $L \rightarrow \infty$  and

$$N \succ C_1 L \quad (7)$$

for some constant  $C_1 > 1$  that depends only on  $\beta, \mu(\mathbf{x}^{(M)})$  and  $\nu$ .

*Proof:* The proof is given in Section III-C.1. □

*Corollary 2.2:* Let the conditions of Theorem 2.1 be satisfied. Then for any Gaussian measurement matrix,  $\mathbf{A}$ , and for Error Metric 1,  $-\log \mathbb{P}(p_{\text{err}}(\mathbf{A}|\mathbf{x}^{(M)}) \geq \xi) / \log L \rightarrow \infty$  as  $L \rightarrow \infty$  for any  $\xi \in (0, 1]$ .

*Proof:* Markov's Inequality implies

$$\mathbb{P}(p_{\text{err}}(\mathbf{A}|\mathbf{x}^{(M)}) \geq \xi) \leq \frac{\mathbb{E}_{\mathbf{A}}(p_{\text{err}}(\mathbf{A}|\mathbf{x}^{(M)}))}{\xi} = \frac{p_{\text{err}}(\mathcal{D}|\mathbf{x}^{(M)})}{\xi}.$$

As shown in the proof of Theorem 2.1,  $-\log p_{\text{err}}(\mathcal{D}|\mathbf{x}^{(M)}) / \log L \rightarrow \infty$  as  $L \rightarrow \infty$ , yielding the desired result. □

*Theorem 2.3:* (Converse for Error Metric 1) Let a sequence of sparse vectors,  $\{\mathbf{x}^{(M)} \in \mathbb{C}^M\}_M$  with  $\|\mathbf{x}^{(M)}\|_0 = L = \lfloor \frac{1}{\beta}M \rfloor$ , where  $\beta > 2$  be given. Then asymptotic reliable recovery is not possible for  $\{\mathbf{x}^{(M)}\}$  with respect to Error Metric 1 if

$$N \prec C_2 \frac{L}{\log P} \quad (8)$$

for some constant  $C_2 > 0$  that depends only on  $\beta, P$  and  $\nu$ .

*Proof:* The proof is given in Section IV-A.1.  $\square$

*Corollary 2.4:* Let a sequence of sparse vectors,  $\{\mathbf{x}^{(M)} \in \mathbb{C}^M\}_M$  with  $\|\mathbf{x}^{(M)}\|_0 = L = \lfloor \frac{1}{\beta}M \rfloor$ , where  $\beta > 2$  be given. Then for  $\xi > 0$ , for any Gaussian measurement matrix,  $\mathbf{A}$ , and for Error Metric 1,  $\mathbb{P}(p_{\text{err}}(\mathbf{A}|\mathbf{x}^{(M)}) \rightarrow 1)$  goes to 1 exponentially fast as a function of  $M$  if  $N \prec \hat{C}_2 \frac{L}{\log P}$ , where  $\hat{C}_2 < C_2$  is a positive constant that depends only on  $\beta, P, \nu$  and  $\xi$ .

*Proof:* The proof is given in Section IV-A.1.  $\square$

*Theorem 2.5:* (Achievability for Error Metric 2) Let a sequence of sparse vectors,  $\{\mathbf{x}^{(M)} \in \mathbb{C}^M\}_M$  with  $\|\mathbf{x}^{(M)}\|_0 = L = \lfloor \frac{1}{\beta}M \rfloor$ , where  $\beta > 2$  be given such that  $L\mu^2(\mathbf{x}^{(M)})$  and  $P$  are constant. Then asymptotic reliable recovery is possible for  $\{\mathbf{x}^{(M)}\}$  with respect to Error Metric 2 if

$$N \succ C_3 L \quad (9)$$

for some constant  $C_3 > 1$  that depends only on  $\alpha, \beta, \mu(\mathbf{x}^{(M)})$  and  $\nu$ .

*Proof:* The proof is given in Section III-C.2.  $\square$

*Corollary 2.6:* Let the conditions of Theorem 2.5 be satisfied. Then for any Gaussian measurement matrix,  $\mathbf{A}$ , and for Error Metric 2,  $\mathbb{P}(p_{\text{err}}(\mathbf{A}|\mathbf{x}^{(M)}) > \xi)$  is exponentially decaying to zero as a function of  $M$  for any  $\xi \in (0, 1]$ .

*Proof:* As shown in the proof of Theorem 2.5,  $p_{\text{err}}(\mathcal{D}|\mathbf{x}^{(M)})$  decays exponentially fast in  $M$ . Applying Markov's Inequality, yields the desired result.  $\square$

*Theorem 2.7:* (Converse for Error Metric 2) Let a sequence of sparse vectors,  $\{\mathbf{x}^{(M)} \in \mathbb{C}^M\}_M$  with  $\|\mathbf{x}^{(M)}\|_0 = L = \lfloor \frac{1}{\beta}M \rfloor$ , where  $\beta > 2$  be given such that  $P$  is constant. Then asymptotic reliable recovery is not possible for  $\{\mathbf{x}^{(M)}\}$  with respect to Error Metric 2 if

$$N \prec C_4 L \quad (10)$$

for some constant  $C_4 \geq 0$  that depends only on  $\alpha, \beta, P$  and  $\nu$ .

*Proof:* The proof is given in Section IV-A.2.  $\square$

*Corollary 2.8:* Let a sequence of sparse vectors,  $\{\mathbf{x}^{(M)} \in \mathbb{C}^M\}_M$  with  $\|\mathbf{x}^{(M)}\|_0 = L = \lfloor \frac{1}{\beta}M \rfloor$ , where  $\beta > 2$  be given such that  $P$  is constant. Then for  $\xi > 0$ , for any Gaussian measurement matrix,  $\mathbf{A}$ , and

for Error Metric 2,  $\mathbb{P}(p_{err}(\mathbf{A}|\mathbf{x}^{(M)}) \rightarrow 1)$  goes to 1 exponentially fast as a function of  $M$  if  $N \succ \hat{C}_4 L$ , where  $\hat{C}_4 \leq C_4$  is a non-negative constant that depends only on  $\alpha, \beta, P, \nu$  and  $\xi$ .

*Proof:* The proof is analogous to the proof of Corollary 2.4.  $\square$

*Theorem 2.9:* (Achievability for Error Metric 3) Let a sequence of sparse vectors,  $\{\mathbf{x}^{(M)} \in \mathbb{C}^M\}_M$  with  $\|\mathbf{x}^{(M)}\|_0 = L = \lfloor \frac{1}{\beta} M \rfloor$ , where  $\beta > 2$  be given such that  $P$  is constant. Then asymptotic reliable recovery is possible for  $\{\mathbf{x}^{(M)}\}$  with respect to Error Metric 3 if

$$N \succ C_5 L \quad (11)$$

for some constant  $C_5 > 1$  that depends only on  $\beta, \gamma, P$  and  $\nu$ .

*Proof:* The proof is given in Section III-C.3.  $\square$

*Corollary 2.10:* Let the conditions of Theorem 2.9 be satisfied. Then for any Gaussian measurement matrix,  $\mathbf{A}$ , and for Error Metric 3,  $\mathbb{P}(p_{err}(\mathbf{A}|\mathbf{x}^{(M)}) > \xi)$  is exponentially decaying to zero as a function of  $M$  for any  $\xi \in (0, 1]$ .

*Proof:* The proof is analogous to the proof of Corollary 2.6.  $\square$

*Theorem 2.11:* (Converse for Error Metric 3) Let a sequence of sparse vectors,  $\{\mathbf{x}^{(M)} \in \mathbb{C}^M\}_M$  with  $\|\mathbf{x}^{(M)}\|_0 = L = \lfloor \frac{1}{\beta} M \rfloor$ , where  $\beta > 2$  be given such that  $P$  is constant and the non-zero terms decay to zero at the same rate. Then asymptotic reliable recovery is not possible for  $\{\mathbf{x}^{(M)}\}$  with respect to Error Metric 3 if

$$N \prec C_6 L \quad (12)$$

for some constant  $C_6 \geq 0$  that depends only on  $\beta, \gamma, P, \mu(\mathbf{x}^{(M)})$  and  $\nu$ .

*Proof:* The proof is given in Section IV-A.3.  $\square$

*Corollary 2.12:* Let a sequence of sparse vectors,  $\{\mathbf{x}^{(M)} \in \mathbb{C}^M\}_M$  with  $\|\mathbf{x}^{(M)}\|_0 = L = \lfloor \frac{1}{\beta} M \rfloor$ , where  $\beta > 2$  be given such that  $P$  is constant and the non-zero terms decay to zero at the same rate. Then for  $\xi > 0$ , for any Gaussian measurement matrix,  $\mathbf{A}$ , and for Error Metric 3,  $\mathbb{P}(p_{err}(\mathbf{A}|\mathbf{x}^{(M)}) \rightarrow 1)$  goes to 1 exponentially fast as a function of  $M$  if  $N \succ \hat{C}_6 L$ , where  $\hat{C}_6 \leq C_6$  is a non-negative constant that depends only on  $\beta, \gamma, P, \mu(\mathbf{x}^{(M)}), \nu$  and  $\xi$ .

*Proof:* The proof is analogous to the proof of Corollary 2.4.  $\square$

#### A. Discussion of The Results

Theorem 2.1 implies that for Error Metric 1,  $O(L)$  measurements are sufficient for asymptotic reliable sparse recovery. There is a clear gap between this number of measurements and  $O(L \log(M - L))$

measurements required by  $\mathcal{L}_1$  constrained quadratic programming [11]. In this proof, it is required that  $\frac{L\mu^4(\mathbf{x}^{(M)})}{\log L} \rightarrow \infty$  as  $L \rightarrow \infty$ , which implies that  $P$  grows without bound as a function of  $N$ .

Theorems 2.5 and 2.9 show that for Error Metrics 2 and 3, the number of required measurements to achieve asymptotic reliable sparse recovery is  $N = O(L)$ . In this case  $P$  remains constant, which is a much less stringent requirement than that of Theorem 2.1. Converses to these theorems are established in Theorems 2.3, 2.7 and 2.11, which demonstrate that  $O(L)$  measurements are asymptotically necessary.

Finally we note that Corollaries 2.6 and 2.10 imply that with overwhelming probability (i.e. the probability goes to 1 exponentially fast as a function of  $M$ ) a given  $N \times M$  Gaussian measurement matrix  $\mathbf{A}$  can be used for asymptotic reliable sparse recovery (respectively for Error Metrics 2 and 3) as long as  $N = O(L)$ . Similarly Corollaries 2.8 and 2.12 prove that a given Gaussian matrix  $\mathbf{A}$  will have  $p_{err}(\mathbf{A}|\mathbf{x}^{(M)}) \rightarrow 1$  (respectively for Error Metrics 2 and 3) with overwhelming probability as long as the number of measurements is less than specified constant multiples of  $L$ . Corollaries 2.2 and 2.4 are similar in nature.

### III. ACHIEVABILITY PROOFS

#### A. Notation

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

#### B. Joint Typicality

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

*Definition 3.1:* (Joint Typicality) We say an  $N \times 1$  noisy observation vector,  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$  and a set of indices  $\mathcal{J} \subset \{1, 2, \dots, M\}$ , with  $|\mathcal{J}| = L$ , are  $\delta$ -jointly typical if  $\text{rank}(\mathbf{A}_{\mathcal{J}}) = L$  and

$$\left| \frac{1}{N} \|\Pi_{\mathbf{A}_{\mathcal{J}}}^{\perp} \mathbf{y}\|^2 - \frac{N-L}{N} \nu^2 \right| < \delta, \quad (13)$$

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

*Lemma 3.2:* For an index set  $\mathcal{I} \subset \{1, 2, \dots, M\}$  with  $|\mathcal{I}| = L$ ,

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

*Lemma 3.3:*

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

$$\mathbb{P}\left(\left|\frac{1}{N}\|\Pi_{\mathbf{A}_{\mathcal{I}}}^{\perp}\mathbf{y}\|^2 - \frac{N-L}{N}\nu^2\right| > \delta\right) \leq 2 \exp\left(-\frac{\delta^2}{4\nu^4} \frac{N^2}{N-L + \frac{2\delta}{\nu^2}N}\right). \quad (14)$$

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

$$\mathbb{P}\left(\left|\frac{1}{N}\|\Pi_{\mathbf{A}_{\mathcal{J}}}^{\perp}\mathbf{y}\|^2 - \frac{N-L}{N}\nu^2\right| < \delta\right) \leq \exp\left(-\frac{N-L}{4}\left(\frac{\sum_{k \in \mathcal{I} \setminus \mathcal{J}} |x_k|^2 - \delta'}{\sum_{k \in \mathcal{I} \setminus \mathcal{J}} |x_k|^2 + \nu^2}\right)^2\right), \quad (15)$$

where

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

*Proof:* We first note that for

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} = \sum_{i \in \mathcal{I}} x_i \mathbf{a}_i + \mathbf{n},$$

we have

$$\Pi_{\mathbf{A}_{\mathcal{I}}}^{\perp}\mathbf{y} = \Pi_{\mathbf{A}_{\mathcal{I}}}^{\perp}\mathbf{n},$$

and

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

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

$$\|\Pi_{\mathbf{A}_{\mathcal{I}}}^{\perp}\mathbf{y}\|^2 = \|\mathbf{D}\mathbf{n}'\|^2,$$

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

$$\|\mathbf{D}\mathbf{n}'\|^2 = |n'_1|^2 + \cdots + |n'_{N-L}|^2.$$

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

$$\|\Pi_{\mathbf{A}_{\mathcal{J}}}^{\perp}\mathbf{y}\|^2 = \|\mathbf{D}\mathbf{w}\|^2 = |w_1|^2 + \cdots + |w_{N-L}|^2,$$

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

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

Let  $\Omega_1 = \frac{\|\mathbf{Dn}'\|^2}{\nu^2}$  and  $\Omega_2 = \frac{\|\mathbf{Dw}\|^2}{\sigma_{\mathcal{J}}^2}$ . We note that both  $\Omega_1$  and  $\Omega_2$  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,

$$\begin{aligned} \mathbb{P}\left(\left|\frac{1}{N}\|\mathbf{\Pi}_{\mathbf{A}_x}^\perp \mathbf{y}\|^2 - \frac{N-L}{N}\nu^2\right| > \delta\right) &= \mathbb{P}\left(\left|\Omega_1 - (N-L)\right| > \frac{\delta}{\nu^2}N\right) \\ &= \mathbb{P}\left(\Omega_1 - (N-L) < -\frac{\delta}{\nu^2}N\right) + \mathbb{P}\left(\Omega_1 - (N-L) > \frac{\delta}{\nu^2}N\right), \end{aligned} \quad (16)$$

and

$$\begin{aligned} \mathbb{P}\left(\left|\frac{1}{N}\|\mathbf{\Pi}_{\mathbf{A}_x}^\perp \mathbf{y}\|^2 - \frac{N-L}{N}\nu^2\right| < \delta\right) &= \mathbb{P}\left(\left|\Omega_2 - (N-L)\frac{\nu^2}{\sigma_{\mathcal{J}}^2}\right| < \frac{\delta}{\sigma_{\mathcal{J}}^2}N\right) \\ &\leq \mathbb{P}\left(\Omega_2 - (N-L) < -(N-L)\left(1 - \frac{\nu^2}{\sigma_{\mathcal{J}}^2}\right) + \frac{\delta}{\sigma_{\mathcal{J}}^2}N\right) \end{aligned} \quad (17)$$

For a chi-square random variable,  $\Omega$  with  $(N - L)$  degrees of freedom [3], [7],

$$\mathbb{P}\left(\Omega - (N-L) \leq -2\sqrt{(N-L)\lambda}\right) \leq e^{-\lambda}, \quad (18)$$

and

$$\mathbb{P}\left(\Omega - (N-L) \geq 2\sqrt{(N-L)\lambda} + 2\lambda\right) \leq e^{-\lambda}. \quad (19)$$

By replacing  $\Omega = \Omega_1$  and

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

in Equation (18) 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 Equation (19), we obtain using Equation (16)

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

Similarly by replacing  $\Omega = \Omega_2$  and

$$\begin{aligned}\lambda &= \left( \frac{\sqrt{N-L}}{2} \left( 1 - \frac{\nu^2}{\sigma_{\mathcal{J}}^2} \right) - \frac{\delta}{\sigma_{\mathcal{J}}^2} \frac{N}{2\sqrt{(N-L)}} \right)^2 \\ &= \left( \frac{\sqrt{N-L}}{2} \left( 1 - \frac{\nu^2}{\sigma_{\mathcal{J}}^2} - \frac{\delta}{\sigma_{\mathcal{J}}^2} \frac{N}{N-L} \right) \right)^2\end{aligned}$$

in Equation (18), we obtain using Equation (17)

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

□

### C. Proofs of Theorems For Different Error Metrics

We define the event

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

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

We also define the error event

$$E_0 = \{ \text{rank}(\mathbf{A}_{\mathcal{I}}) < L \},$$

which results in an order reduction in the model, and implies that the decoder is looking through subspaces of incorrect dimension. By Lemma 3.2, we have  $\mathbb{P}(E_0) = 0$ .

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

1) *Proof of Theorem 2.1 (Error Metric 1)*: Clearly the decoder fails if  $E_0$  or  $E_{\mathcal{I}}^C$  occur or when one of  $E_{\mathcal{J}}$  occurs for  $\mathcal{J} \neq \mathcal{I}$ . Thus

$$\begin{aligned}p_{\text{err}}(\mathcal{D}|\mathbf{x}) &= \mathbb{P}(E_0 \cup E_{\mathcal{I}}^C \cup \bigcup_{\mathcal{J}, \mathcal{J} \neq \mathcal{I}, |\mathcal{J}|=L} E_{\mathcal{J}}) \\ &\leq \mathbb{P}(E_{\mathcal{I}}^C) + \sum_{\mathcal{J}, \mathcal{J} \neq \mathcal{I}, |\mathcal{J}|=L} \mathbb{P}(E_{\mathcal{J}})\end{aligned}$$

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 2.1, we have  $L\mu^4(\mathbf{x})$  grows faster than  $\log L$ . We note that this requirement is milder than that of [10], where the growth requirement is on  $\mu^2(\mathbf{x})$  rather than  $\mu^4(\mathbf{x})$ . Since the decoder needs to distinguish between even the smallest non-overlapping coordinates, we let  $\delta' = \zeta \mu^2(\mathbf{x})$  for  $0 < \zeta < 1$ . For computational convenience, we will only consider  $2/3 < \zeta < 1$ .

By Lemma 3.3,

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

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

Again by Lemma 3.3, 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  $\sum_{k \in \mathcal{I} \setminus \mathcal{J}} |x_k|^2 \geq (L-K)\mu^2(\mathbf{x})$ , we have

$$\mathbb{P}(E_{\mathcal{J}}) \leq \exp \left( - \frac{N-L}{4} \left( \frac{(L-K)\mu^2(\mathbf{x}) - \delta'}{(L-K)\mu^2(\mathbf{x}) + \nu^2} \right)^2 \right), \quad (20)$$

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

The condition of Theorem 2.1 on  $\mu(\mathbf{x})$  implies that  $\mathbb{P}(E_{\mathcal{J}}) \rightarrow 0$  for all  $K$ . We note that this condition also implies  $P \rightarrow \infty$  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  $\text{rank}(\mathbf{A}_{\mathcal{J}}) = L$ ) is upper-bounded by

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

Thus

$$\begin{aligned} p_{\text{err}}(\mathcal{D}|\mathbf{x}) &\leq 2 \exp \left( - \frac{\zeta^2 C_0}{\nu^2} \frac{L \mu^4(\mathbf{x})}{\nu^2 + 2\zeta \mu^2(\mathbf{x})} \right) \\ &\quad + \sum_{K=0}^{L-1} \binom{L}{L-K} \binom{M-L}{L-K} \exp \left( - \frac{N-L}{4} \left( \frac{(L-K)\mu^2(\mathbf{x}) - \delta'}{(L-K)\mu^2(\mathbf{x}) + \nu^2} \right)^2 \right) \\ &= 2 \exp \left( - \frac{\zeta^2 C_0}{\nu^2} \frac{L \mu^4(\mathbf{x})}{\nu^2 + 2\zeta \mu^2(\mathbf{x})} \right) + \sum_{K'=1}^L \binom{L}{K'} \binom{M-L}{K'} \exp \left( - \frac{N-L}{4} \left( \frac{(K')\mu^2(\mathbf{x}) - \delta'}{(K')\mu^2(\mathbf{x}) + \nu^2} \right)^2 \right) \end{aligned}$$

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

$$\exp \left( K' \log \left( \frac{L}{K'} \right) \right) \leq \binom{L}{K'} \leq \exp \left( K' \log \left( \frac{Le}{K'} \right) \right) \quad (21)$$

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

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

We upper bound the whole summation by maximizing the function

$$\begin{aligned} f(z) &= Lz \log \frac{e}{z} + Lz \log \frac{(\beta-1)e}{z} - C_0 L \left( \frac{Lz\mu^2(\mathbf{x}) - \delta'}{Lz\mu^2(\mathbf{x}) + \nu^2} \right)^2 \\ &= -2Lz \log z + Lz(2 + \log(\beta-1)) - C_0 L \left( \frac{Lz\mu^2(\mathbf{x}) - \zeta\mu^2(\mathbf{x})}{Lz\mu^2(\mathbf{x}) + \nu^2} \right)^2 \end{aligned} \quad (22)$$

for  $z \in [\frac{1}{L}, 1]$ . 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 3.4:* Let  $g(z)$  be a twice differentiable function on  $[a, b]$  that has a continuous second derivative. If  $g(a) < 0$ ,  $g(b) < 0$ , and  $g'(a) < 0$ ,  $g'(b) > 0$ , 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  $g''(x_2) = 0$ .  $\square$

*Lemma 3.5:* 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.  $\square$

*Lemma 3.6:* For  $L$  sufficiently large,  $f(z)$  (see Equation (22)) is negative for all  $z \in [\frac{1}{L}, 1]$ . 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  $\overset{\approx}{\asymp}$  for denoting the behavior of  $f(z)$  for large  $L$ , and  $\prec$  and  $\succ$  for inequalities that hold asymptotically.

$$f\left(\frac{1}{L}\right) = 2 \log L + 2 + \log(\beta-1) - C_0 L \left( \frac{\mu^2(\mathbf{x})(1-\zeta)}{\mu^2(\mathbf{x}) + \nu^2} \right)^2 \prec 0 \quad (23)$$

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

$$\begin{aligned} f(1) &= L(2 + \log(\beta-1)) - C_0 L \left( \frac{\mu^2(\mathbf{x})(L-\zeta)}{L\mu^2(\mathbf{x}) + \nu^2} \right)^2 \\ &\overset{\approx}{\asymp} L(2 + \log(\beta-1) - C_0) \prec 0. \end{aligned} \quad (24)$$

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

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

Also,

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

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

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

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

Additionally,

$$\begin{aligned} f''(z) &= -\frac{2L}{z} - 2C_0 L^2 \mu^4(\mathbf{x})(\nu^2 + \zeta \mu^2(\mathbf{x})) \left( \frac{-2Lz\mu^2(\mathbf{x}) + \nu^2 + 3\zeta\mu^2(\mathbf{x})}{(Lz\mu^2(\mathbf{x}) + \nu^2)^4} \right) L \\ &= \frac{-2L}{z(Lz\mu^2(\mathbf{x}) + \nu^2)^4} \left( (Lz\mu^2(\mathbf{x}) + \nu^2)^4 + C_0 L^2 \mu^4(\mathbf{x})(\nu^2 + \zeta \mu^2(\mathbf{x}))(-2Lz\mu^2(\mathbf{x}) + \nu^2 + 3\zeta\mu^2(\mathbf{x}))z \right) \end{aligned} \quad (25)$$

Thus,

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

and

$$\begin{aligned} f''(1) &= -2L \left( 1 + C_0 L^2 \mu^4(\mathbf{x})(\nu^2 + \zeta \mu^2(\mathbf{x})) \frac{-2L\mu^2(\mathbf{x}) + \nu^2 + 3\zeta\mu^2(\mathbf{x})}{(L\mu^2(\mathbf{x}) + \nu^2)^4} \right) \\ &\approx -2L \left( 1 - 2C_0 \frac{\nu^2 + \zeta\mu^2(\mathbf{x})}{L\mu^2(\mathbf{x})} \right) < 0. \end{aligned}$$

Since  $f(z)$  is twice differentiable function on  $[\frac{1}{L}, 1]$  with a continuous second derivative, Lemma 3.4 implies that  $f''(z)$  crosses 0 at least twice in this interval. Next we examine the polynomial (see Equation (25)),

$$p(z) = (Lz\mu^2(\mathbf{x}) + \nu^2)^4 + 2C_0 L^2 \mu^4(\mathbf{x})(\nu^2 + \zeta \mu^2(\mathbf{x}))(-2Lz\mu^2(\mathbf{x}) + \nu^2 + 3\zeta\mu^2(\mathbf{x}))z.$$

Since  $p(z)$  satisfies the conditions of Lemma 3.5, we conclude that it has at most two positive roots, and thus at most two roots of  $p(z)$  can lie in  $[\frac{1}{L}, 1]$ . In other words  $f''(z)$  can cross 0 for  $z \in [\frac{1}{L}, 1]$  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  $z_0^{(2)} = 1$ .  $\square$

Thus we have,

$$\begin{aligned} p_{\text{err}}(\mathcal{D}|\mathbf{x}) &\leq 2 \exp\left(-\frac{\zeta^2 C_0}{\nu^2} \frac{L\mu^4(\mathbf{x})}{\nu^2 + 2\zeta\mu^2(\mathbf{x})}\right) + \sum_{K=0}^{L-1} \exp(\max\{f(z_0^{(1)}), f(z_0^{(2)})\}) \\ &= 2 \exp\left(-\frac{\zeta^2 C_0}{\nu^2} \frac{L\mu^4(\mathbf{x})}{\nu^2 + 2\zeta\mu^2(\mathbf{x})}\right) + \exp\left(\log L + \max\left\{f\left(\frac{1}{L}\right), f(1)\right\}\right) \end{aligned}$$

From Equations (23) and (24), it is clear that  $\log(L) + \max\left\{f\left(\frac{1}{L}\right), f(1)\right\} \rightarrow -\infty$  as  $L \rightarrow \infty$ . Hence with the conditions of Theorem 2.1,  $p_{\text{err}}(\mathcal{D}|\mathbf{x}) \rightarrow 0$  as  $L \rightarrow \infty$ .

2) *Proof of Theorem 2.5 (Error Metric 2)*: For asymptotic reliable recovery with Error Metric 2, we require that  $\mathbb{P}(E_{\mathcal{J}})$  goes to 0 for only  $K \leq (1 - \alpha)L$  with  $\alpha \in (0, 1)$ . By a re-examination of Equation (20), we observe that the right hand side of

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

converges to 0 asymptotically, even when  $L\mu^2(\mathbf{x})$  converges to a constant. In this case  $P$  does not have to grow with  $N$ . We let  $\delta > 0$  (and hence  $\delta'$ ) be a constant, and let  $N = (4\hat{C}_3 + 1)L$  for

$$\hat{C}_3 > \beta \left(\frac{\alpha L\mu^2(\mathbf{x}) + \nu^2}{\alpha L\mu^2(\mathbf{x}) - \delta'}\right)^2. \quad (26)$$

Given the decay rate of  $\mu^2(\mathbf{x})$  and that  $\delta' > 0$  is arbitrary, we note that this constant only depends on  $\alpha, \beta, \mu(\mathbf{x})$  and  $\nu$ . Hence

$$\begin{aligned} p_{\text{err}}(\mathcal{D}|\mathbf{x}) &\leq \mathbb{P}(E_{\mathcal{I}}^C) + \sum_{K=0}^{(1-\alpha)L} \binom{L}{L-K} \binom{M-L}{L-K} \exp\left(-\frac{N-L}{4} \left(\frac{(L-K)\mu^2(\mathbf{x}) - \delta'}{(L-K)\mu^2(\mathbf{x}) + \nu^2}\right)^2\right) \\ &\leq 2 \exp\left(-\frac{\delta^2}{4\nu^4} \frac{4\hat{C}_3 + 1}{4\hat{C}_3 + \frac{2\delta}{\nu^2}(4\hat{C}_3 + 1)} N\right) \\ &\quad + \sum_{K'=\alpha L}^L \exp\left(LH\left(\frac{K'}{L}\right) + (M-L)H\left(\frac{K'}{M-L}\right) - \hat{C}_3 L \left(\frac{K'\mu^2(\mathbf{x}) - \delta'}{K'\mu^2(\mathbf{x}) + \nu^2}\right)^2\right), \end{aligned}$$

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 Equation (26), we see  $p_{\text{err}}(\mathcal{D}|\mathbf{x}) \rightarrow 0$  exponentially fast as  $L \rightarrow \infty$ .

3) *Proof of Theorem 2.9 (Error Metric 3):* An error occurs for Error Metric 3 if

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

Thus we can bound the error event for  $\mathcal{J}$  from Lemma 3.3 as

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

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

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

For  $N > C_5 L$ , a similar argument to that of Section III-C.2 proves that  $p_{\text{err}}(\mathcal{D}|\mathbf{x}) \rightarrow 0$  exponentially fast as  $L \rightarrow \infty$ , where  $C_5$  depends only on  $\beta, \gamma, P$  and  $\nu$ .

#### IV. PROOFS OF CONVERSES

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

##### A. Genie-Aided Decoding and Connection with Noisy Communication Systems

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

Clearly we have

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

1) *Proof of Theorem 2.3 (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  $\mathbf{H} = [x_{i_1} x_{i_2} \dots x_{i_L}] = \mathbf{x}_{\mathcal{I}}^T$ . The encoder,  $\mathfrak{E}_1 : \{0, 1\}^M \rightarrow \mathbb{C}^{L \times N}$ , 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  $\mathfrak{D}_1 : \mathbb{C}^N \rightarrow \{0, 1\}^M$  such that its output  $\hat{\mathbf{c}}$  has weight  $L$ .

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

$$\mathfrak{E}_1 = \left\{ \left( \begin{array}{c} \mathbf{z}_1^{\mathcal{J}} \quad \mathbf{z}_2^{\mathcal{J}} \quad \dots \quad \mathbf{z}_N^{\mathcal{J}} \end{array} \right) \middle| \mathcal{J} \subset \{1, 2, \dots, M\}, |\mathcal{J}| = L \right\},$$

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

$$y_k = \mathbf{H}\mathbf{z}_k^{\mathcal{J}} + n_k \quad \text{for } k = 1, 2, \dots, N,$$

where  $y_k$  and  $n_k$  are the  $k^{\text{th}}$  coordinates of  $\mathbf{y}$  and  $\mathbf{n}$  respectively. The average signal power is  $\mathbb{E}(\|\mathbf{z}_k^{\mathcal{J}}\|^2) = L$ , 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 [9]

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

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

$$\frac{1}{M+1} \exp \left( MH \left( \frac{L}{M} \right) \right) \leq \binom{M}{L} \leq \exp \left( MH \left( \frac{L}{M} \right) \right), \quad (27)$$

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 2.4, 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 [6]. If  $\mathbf{A}$  is chosen from a Gaussian distribution, then by Inequality (19),

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

for any  $\mathcal{J} \subset \{1, 2, \dots, M\}$ ,  $|\mathcal{J}| = L$  and for  $k = 1, 2, \dots, N$ . Let  $\Lambda = 2 \sqrt{\beta H \left( \frac{1}{\beta} \right) + \xi} + 2 \left( \beta H \left( \frac{1}{\beta} \right) + \xi \right)$  for  $\xi > 0$ . By the union bound over all  $\binom{M}{L}$  possible index sets  $\mathcal{J}$  and  $k = 1, 2, \dots, N$ ,

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

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

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

2) *Proof of Theorem 2.7 (Error Metric 2)*: For any given  $\mathbf{x}$  with  $\|\mathbf{x}\|_0 = L$ , we will prove the contrapositive. Let  $P_{e_2}^{(M)}$  denote the probability of error with respect to Error Metric 2 for  $\mathbf{x} \in \mathbb{C}^M$ . We show that  $N \succ C_4 L$  if  $P_{e_2}^{(M)} \rightarrow 0$ .

Consider a single input single output system,  $\mathcal{S}$ , whose input is  $\mathbf{c} \in \{0, 1\}^M$ , and whose output is  $\hat{\mathbf{c}} \in \{0, 1\}^M$ , such that  $\|\mathbf{c}\|_0 = \|\hat{\mathbf{c}}\|_0 = L$ , and  $\|\mathbf{c} - \hat{\mathbf{c}}\|_0 \leq 2\alpha L$ . The last condition states that the support of  $\mathbf{c}$  and that of  $\hat{\mathbf{c}}$  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  $d(\mathbf{c}, \hat{\mathbf{c}}) = \frac{1}{M} \sum_{k=1}^M d_H(c_i, \hat{c}_i)$ , where  $\mathbf{c}$  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 IV-A.1. Since the source is transmitted within distortion  $\frac{2\alpha}{\beta}$  over the MISO channel, we have [2]

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

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

*Lemma 4.1*: Let  $\alpha \in (0, 1]$  and  $\beta > 2$ , and let

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

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

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

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

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.  $\square$

Thus we have

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

where the first inequality follows since given  $\hat{\mathbf{c}}$ ,  $\mathbf{c}$  is among  $\sum_{K=0}^{\alpha L} \binom{L}{K} \binom{M-L}{K}$  possible binary vectors within Hamming distance  $2\alpha L$  from  $\hat{\mathbf{c}}$ . The second inequality follows from Inequality (27), and the third inequality follows by Lemma 4.1.

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

$$C_{\alpha,\beta} = \begin{cases} \beta H\left(\frac{1}{\beta}\right) - H(\alpha) - (\beta-1)H\left(\frac{\alpha}{\beta-1}\right) & \text{if } \alpha \leq \frac{\beta-1}{\beta} \\ 0 & \text{if } \alpha > \frac{\beta-1}{\beta} \end{cases} \quad (28)$$

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 Theorem 2.7.

3) *Proof of Theorem 2.11 (Error Metric 3):* For Error Metric 3, we assume that  $\rho(\mathbf{x}) = \max_{i \in \mathcal{I}} |x_i|$  and  $\mu(\mathbf{x}) = \min_{i \in \mathcal{I}} |x_i|$  both decay at rate  $O\left(\sqrt{\frac{1}{L}}\right)$ . Thus  $P$  is constant. In the absence of this assumption, some terms of  $\mathbf{x}$  can be asymptotically dominated by noise. Such terms are unimportant for recovery purposes, and therefore could be replaced by zeros (in the definition of  $\mathbf{x}$ ) with no significant harm.

Let  $\alpha(\gamma, \mathbf{x}) = \min\left(\frac{\gamma P}{L\mu^2(\mathbf{x})}, 1\right)$ . Let  $P_{e_3}^{(M)}$  denote the probability of error with respect to Error Metric 3 for  $\mathbf{x} \in \mathbb{C}^M$ . 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  $\mathcal{I} = \text{supp}(\mathbf{x})$ . This implies that  $|\mathcal{I} \setminus \mathcal{J}| \leq \alpha(\gamma, \mathbf{x})L$ . Thus  $P_{e_3}^{(M)} = 0$  implies that  $P_{e_2}^{(M)} = 0$  when recovering  $\alpha(\gamma, \mathbf{x})$  fraction of the support of  $\mathbf{x}$ . As shown in Section IV-A.2, reliable recovery of  $\mathbf{x}$  is

not possible if

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

where  $C_{\alpha(\gamma, \mathbf{x}), \beta}$  is a constant (as defined in Equation (28)) that only depends on  $\gamma, \beta, \mu(\mathbf{x})$  and  $P$  for a given  $\mathbf{x}$ .

## V. SUBLINEAR REGIME

For completeness, we also state the equivalent theorems, when  $L = o(M)$ . The proofs follow the same steps as those in the linear regime. For the proofs of converse results, we use the bounds from Equation (21) instead of those of Equation (27).

*Theorem 5.1:* (Achievability for Error Metric 1) Let a sequence of sparse vectors,  $\{\mathbf{x}^{(M)} \in \mathbb{C}^M\}_M$  with  $\|\mathbf{x}^{(M)}\|_0 = L = o(M)$  be given. Then asymptotic reliable recovery is possible for  $\{\mathbf{x}^{(M)}\}$  with respect to Error Metric 1 if  $L\mu^4(\mathbf{x}^{(M)}) \rightarrow \infty$  as  $L \rightarrow \infty$  and

$$N \succ C'_1 L \log(M - L) \quad (29)$$

for some constant  $C'_1 > 0$  that depends only on  $\mu(\mathbf{x}^{(M)})$  and  $\nu$ .

*Proof:* The proof is similar to that of Theorem 2.1, with  $f(z)$  replaced by

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

The behavior of  $k(z), k'(z)$  and  $k''(z)$  at the endpoints  $\{\frac{1}{L}, 1\}$ , is the same as that in the proof of Theorem 2.1 whenever  $N = C'_1 L \log(M - L)$ . The result follows.  $\square$

*Theorem 5.2:* (Converse for Error Metric 1) Let a sequence of sparse vectors,  $\{\mathbf{x}^{(M)} \in \mathbb{C}^M\}_M$  with  $\|\mathbf{x}^{(M)}\|_0 = L = o(M)$  be given. Then asymptotic reliable recovery is not possible for  $\{\mathbf{x}^{(M)}\}$  with respect to Error Metric 1 if

$$N \prec C'_2 \frac{L \log(M - L)}{\log P} \quad (30)$$

for some constant  $C'_2 > 0$  that depends only on  $P$  and  $\nu$ .

*Proof:* The proof is similar to that of Theorem 2.3.  $\square$

*Theorem 5.3:* (Achievability for Error Metric 2) Let a sequence of sparse vectors,  $\{\mathbf{x}^{(M)} \in \mathbb{C}^M\}_M$  with  $\|\mathbf{x}^{(M)}\|_0 = L = o(M)$  be given such that  $L\mu^2(\mathbf{x}^{(M)})$  and  $P$  are constant. Then asymptotic reliable recovery is possible for  $\{\mathbf{x}^{(M)}\}$  with respect to Error Metric 2 if

$$N \succ C'_3 L \log(M - L) \quad (31)$$

for some constant  $C'_3 > 0$  that depends only on  $\alpha$ ,  $\mu(\mathbf{x}^{(M)})$  and  $\nu$ .

*Proof:* The proof is similar to that of Theorem 2.5.  $\square$

*Theorem 5.4:* (Converse for Error Metric 2) Let a sequence of sparse vectors,  $\{\mathbf{x}^{(M)} \in \mathbb{C}^M\}_M$  with  $\|\mathbf{x}^{(M)}\|_0 = L = o(M)$  be given such that  $P$  is constant. Then asymptotic reliable recovery is not possible for  $\{\mathbf{x}^{(M)}\}$  with respect to Error Metric 2 if

$$N \prec C'_4 L \log(M - L) \quad (32)$$

for some constant  $C'_4 > 0$  that depends only on  $\alpha$ ,  $P$  and  $\nu$ .

*Proof:* We have the following technical lemma,

*Lemma 5.5:* 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  $M$ .  $\square$

*Continuation of the proof of the theorem:* Thus we have,

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

where the first inequality follows from Inequality (21), and the second inequality follows by Lemma 5.5 for sufficiently large  $M$ . The rest of the proof is analogous to that of Theorem 2.7.  $\square$

*Theorem 5.6:* (Achievability for Error Metric 3) Let a sequence of sparse vectors,  $\{\mathbf{x}^{(M)} \in \mathbb{C}^M\}_M$  with  $\|\mathbf{x}^{(M)}\|_0 = L = o(M)$  be given such that  $P$  is constant. Then asymptotic reliable recovery is possible for  $\{\mathbf{x}^{(M)}\}$  with respect to Error Metric 3 if

$$N \succ C'_5 L \log(M - L) \quad (33)$$

for some constant  $C'_5 > 0$  that depends only on  $\gamma$ ,  $P$  and  $\nu$ .

*Proof:* The proof is similar to that of Theorem 2.9.  $\square$

*Theorem 5.7:* (Converse for Error Metric 3) Let a sequence of sparse vectors,  $\{\mathbf{x}^{(M)} \in \mathbb{C}^M\}_M$  with  $\|\mathbf{x}^{(M)}\|_0 = L = o(M)$  be given such that  $P$  is constant and the non-zero terms decay to zero at the same rate. Then asymptotic reliable recovery is not possible for  $\{\mathbf{x}^{(M)}\}$  with respect to Error Metric 3 if

$$N \prec C'_6 L \log(M - L) \quad (34)$$

for some constant  $C'_6 \geq 0$  that depends only on  $\gamma, P, \mu(\mathbf{x}^{(M)})$  and  $\nu$ .

*Proof:* As in the proof of Theorem 2.11, we let  $\alpha(\gamma, \mathbf{x}) = \min\left(\frac{\gamma P}{L\mu^2(\mathbf{x})}, 1\right)$ , and conclude that  $P_{e_3}^{(M)} = 0$  implies that  $P_{e_2}^{(M)} = 0$  when recovering  $\alpha(\gamma, \mathbf{x})$  fraction of the support of  $\mathbf{x}$ . The rest of the proof is analogous to that of Theorem 5.4.  $\square$

## REFERENCES

- [1] M. Akçakaya and V. Tarokh, "A Frame Construction and A Universal Distortion Bound for Sparse Representations", Accepted for publication in IEEE Trans. on Sig. Proc.
- [2] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Prentice-Hall, 1971.
- [3] L. Birgé and P. Massart, "Minimum Contrast Estimators on Sieves: Exponential Bounds and Rates of Convergence", *Bernoulli*, vol. 4, no. 3, pp. 329-375, Sept. 1998.
- [4] E. J. Candès and T. Tao, "Decoding by Linear Programming", *IEEE Trans. on Inf. Theory*, vol. 51, no. 12, pp. 4203-4215, Dec. 2005.
- [5] D. L. Donoho, "Compressed Sensing", *IEEE Trans. on Inf. Theory*, vol. 52, no. 4, pp. 1289-1306, April 2006.
- [6] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, 1968.
- [7] B. Laurent and P. Massart, "Adaptive Estimation of a Quadratic Functional by Model Selection", *Annals of Statistics*, vol. 28, no. 5, pp. 1303-1338, Oct. 2000.
- [8] J. A. Tropp, "Topics in Sparse Approximation", Ph.D. dissertation, Computational and Applied Mathematics, UT-Austin, August 2004.
- [9] B. Vucetic and J. Yuan, *Space-Time Coding*, John Wiley & Sons, 2003.
- [10] M. J. Wainwright, "Information-Theoretic Limits on Sparsity Recovery in the High-Dimensional and Noisy Setting", Technical Report, UC Berkeley, Department of Statistics, January 2007.
- [11] M. J. Wainwright, "Sharp thresholds for noisy and high-dimensional recovery of sparsity using  $\mathcal{L}_1$ -constrained quadratic programming". Technical report, UC Berkeley, Department of Statistics, May 2006.