Sampling Bounds for Sparse Support Recovery in the Presence of Noise

Galen Reeves and Michael Gastpar
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley
Berkeley, CA, 94720, USA
Email: {greeves, gastpar}@eecs.berkeley.edu

***Abstract*— It is well known that the support of a sparse signal can be recovered from a small number of random projections. However, in the presence of noise all known sufficient conditions require that the per-sample signal-to-noise ratio (SNR) grows without bound with the dimension of the signal. If the noise is due to quantization of the samples, this means that an unbounded rate per sample is needed. In this paper, it is shown that an unbounded SNR is also a necessary condition for perfect recovery, but any fraction (less than one) of the support can be recovered with bounded SNR. This means that a finite rate per sample is sufficient for partial support recovery. Necessary and sufficient conditions are given for both stochastic and non-stochastic signal models. This problem arises in settings such as compressive sensing, model selection, and signal denoising.**

I. Introduction

The task of support recovery (also known as recovery of sparsity [1], [2] or model selection [3]) is to determine which elements of some unknown sparse signal $x \in \mathbb{R}^n$ are non-zero based on a set of noisy linear observations. This problem arises in areas, such as compressive sensing, graphical model selection in statistics, and signal denoising in regression. Typically, the number of observations m is far less than the signal dimension n.

The observation model can be generally formulated as a sampling problem where each "sample" consists of a noisy inner product of x and some predetermined measurement vector $\phi_i \in \mathbb{R}^n$

$$y_i = \langle \phi_i, x \rangle + w_i \quad \text{for} \quad i = 1, \cdots, m, \tag{1}$$

where w_i is noise. When m is less than n, general inference problems are challenging. Typically, optimal estimation is computationally hard but for certain tasks, such as approximating x in the ℓ_2 sense, efficient relaxations have been shown to produce near optimal solutions [4], [5].

The task considered in this paper is estimation of the support

$$K = \{i \in \{1, \cdots n\} \ : \ x_i \neq 0\} \tag{2}$$

where $k = |K|$ is the number of non-zero elements of x. Our goal is to give fundamental (information-theoretic) bounds on the degree to which the support can be recovered in the under-sampled ($m < n$) large system setting ($n, k, m \to \infty$) with linear sparsity ($k = \Omega\, n$ for some $\Omega \in (0, 1)$). Under reasonable assumptions, one may consider a natural definition of the per-sample signal-to-noise ratio SNR. In previous results on perfect support recovery [1], [2], there exists a gap: the sufficient conditions require that the SNR grows without bound with n whereas the necessary conditions are satisfied with non-increasing SNR. This paper makes the following contributions:

- **Perfect support recovery is hard:** In Theorem 1 we show that if the SNR does not increase with the dimension of the signal, then exact recovery of the support is impossible.
- **Fractional support recovery is not as hard:** We introduce a notion of partial support recovery and show that even if the SNR does not increase with the dimension of the signal, it is still possible to recover some *fraction* (less than one) of the support. Under standard signal assumptions the fraction of errors is inversely proportional to the SNR, and Theorems 2 and 3 give necessary and sufficient conditions. If the noise is due to quantization of the samples, this means that a finite rate per sample is sufficient.
- **Stochastic versus worst-case analysis:** Previous results require a (lower) bound on the smallest non-zero signal component. This paper considers both stochastic and non-stochastic sparse signal models. Thus, we can give performance guarantees even when the smallest non-zero signal component is arbitrarily small.

Section II describes our observation model, discusses relevant research, and specifics our recovery task. Section III gives our main results and Section IV outlines the proofs.

II. Problem Formulation and Related Work

Let $\mathcal{X}$ denote a class of sparse real signals and let $\mathcal{X}_n$ denote the sub-class of signals with length n. In general, the signal class $\mathcal{X}$ may be stochastic or non-stocastic (specific examples are discussed in Section III-B). For $x \in \mathcal{X}_n$ we consider the linear observation model in which samples $y \in \mathbb{R}^m$ are taken as

$$y = \Phi x + w, \tag{3}$$

where $\Phi \in \mathbb{R}^{m \times n}$ is a sampling matrix and $w \sim \mathcal{N}(0, \sigma_w^2 I_m)$. We assume that the x has exactly k non-zero elements which are indexed by the support K, and that K is distributed uniformly over the $\binom{n}{k}$ possibilities. We further assume that the sampling matrix Φ is randomly constructed with i.i.d. rows $\phi_i \sim \mathcal{N}(0, \frac{1}{n} I_n)$.

We define $\Omega = k/n$ to be the sparsity and $\rho = m/n$ to be the sampling rate. We exclusively consider the setting of linear sparsity where $\Omega \in (0,1)$ does not depend on n, and we are interested in which sampling tasks can (and cannot) be solved in the under-sampled setting ($\rho < 1$) as $n \to \infty$.

We find it convenient to consider a sampling matrix that preserves the magnitude of x. Our choice of Φ means that $\mathbb{E}[\langle \phi_i, x\rangle^2] = ||x||^2/n$, and we consider signals whose average energy $||x||^2/n$ does not depend on n. We caution the reader that these choices are in contrast to some of the related work [1], [2] where Φ is chosen such that $\mathbb{E}[\langle \phi_i, x\rangle^2] = ||x||^2$ and hence Φ amplifies the signal x.

Section II-A gives a brief summary of relevant research (a more more extensive summary is given in [6]), Section II-B describes our error metric, and Section II-C describes an optimal estimation algorithm.

A. Related Work

In the noiseless setting, perfect support recovery requires $m = k+1$ samples using optimal, but computationally expensive, recovery algorithms [7], and requires $m = \mathcal{O}(k \log(n/k))$ samples using linear programming [8]–[10].

In the presence of noise, Compressive Sensing [4], [5] shows that for $m = \mathcal{O}(k \log(n/k))$ samples, quadratic programming can provide a signal estimate $\hat{x}$ that is stable; that is, $||\hat{x} - x||$ is bounded with respect to $||w||$. The papers [4], [11]–[13] give sufficient conditions for the support of $\hat{x}$ to be contained inside the support of x.

The work in [1], [3], [14] addresses the asymptotic performance of a particular quadratic program, the *Lasso*. Results are formulated in terms of scaling conditions for (n,k,m) and the magnitude of the smallest non-zero component of x denoted x_{min}. For the observation model considered in this paper, Wainwright [1] shows that perfect recovery (using the Lasso) is possible if and only if $m/n \to \infty$ or $x_{\text{min}} \to \infty$.

Another line of research has considered information-theoretic bounds on the asymptotic performance of optimal support recovery algorithms. For perfect support recovery, Gastpar et al. [15] lower bound the probability of success, and Wainwright [2] gives necessary and sufficient conditions for an exhaustive search algorithm. Since the submission of this paper, Fletcher et al. [16] have generalized the necessary conditions given below in Theorem 1 for all scalings of (n,k,m). More generally, support recovery with respect to some distortion measure has also been considered [17]–[20]. Aeron et al. [19] derive bounds similar to Theorems 2 and 3 in this paper for the special setting in which each element of x has only a finite number of values.

B. Partial Support Recovery

Given the true support K and any estimate $\hat{K}$ there are several natural measures for the distortion $d(K,\hat{K})$. One may consider recovery of K as a target recognition problem where for each index $i \in \{1, \cdots, n\}$ we want to determine whether or not i is in the support K.

At one extreme, minimization of

$$d_{\text{sub}}(K,\hat{K}) = \begin{cases} |K| - |\hat{K}| & \hat{K} \subseteq K \\ \infty & \hat{K} \supset K \end{cases}$$

attempts to find the largest subset $\hat{K}$ that is contained in K. The results of [4], [11]–[13] can be interpreted in terms of this metric. Roughly speaking, their results guarantee that $d_{\text{sub}}(\hat{K},K) \leq |K|$ but cannot say much more because no guarantees are given on the size $\hat{K}$.

At the other extreme, one may want to find the smallest estimate $\hat{K}$ such that $\hat{K} \supseteq K$, and in general one may formulate a Neyman-Pearson style tradeoff between the two types of errors. The focus of this paper is reconstruction at the point where the number of false positives is equal to the number of false negatives. Since we assume that $|K|$ is known a priori, we can impose this condition by requiring that $|\hat{K}| = |K|$. We use the following metric which is proportional (by a factor of two) to the total number of errors

$$d(K,\hat{K}) = |K| - |K \cap \hat{K}|.$$

Partial recovery corresponds to $d(K,\hat{K}) \leq a^*$ for some $a^* \geq 0$. There are several interesting choices for the scaling of a^* and n. For instance, if recovery is possible with $a^* = \mathcal{O}(\log n)$ then as $n \to \infty$ the average distortion $\frac{1}{k} d(K,\hat{K}) \to 0$ although the allowable number of errors $d(K,\hat{K}) \to \infty$. Our results pertain to a linear scaling where $a^* = \alpha k$ for some fixed $\alpha \in [0,1]$ that does not depend on n. The parameter α is the fractional distortion, and the requirement $\alpha = 0$ corresponds to perfect recovery.

To characterize the performance of an estimator $\hat{K}(y)$ we recall that y is a function of x and thus the performance depends on the signal class. If $\mathcal{X}_n$ is non-stochastic, the probability of error is defined as

$$P_e(\alpha, \mathcal{X}_n) = \max_{x \in \mathcal{X}_n} \mathbb{P}\left\{ d(K,\hat{K}(y)) > \alpha\, k \right\}, \qquad (4)$$

where the probability is over K, w, and Φ. If $\mathcal{X}_n$ is stochastic, then

$$P_e(\alpha, \mathcal{X}_n) = \mathbb{P}\left\{ d(K,\hat{K}(y)) > \alpha\, k \right\}, \qquad (5)$$

where the probability is over x, K, w, and Φ. An estimator $\hat{K}(y)$ is said to be *asymptotically reliable* for a class $\mathcal{X}$ if there exists some constant $c > 0$ such that $P_e(\alpha, \mathcal{X}_n) < e^{-n\,c}$. Conversely, an estimator $\hat{K}(y)$ is said to be *asymptotically unreliable* for a class $\mathcal{X}$ if there exists some constant $c > 0$ and integer N such that $P_e(\alpha, \mathcal{X}_n) > c$ for all $n \geq N$.

We remark that a weaker notion of reliable recovery is to constrain the expected distortion, that is to require that $\mathbb{E}_K[d(K,\hat{K}(y))] \leq \alpha\, k$. Although such a statement means that on average the fractional distortion is less than α, it is still possible that a linear fraction of all possible supports have resulting distortion greater than α. Our notion of asymptotically reliable recovery implies more. It says that although there may be a set of "bad" supports with resulting distortion greater than α, the size of this set is very small relative to the total number of possible supports.

C. ML Estimation

Our sufficient (achievable) conditions correspond to the performance of a maximum likelihood (ML) decoder which uses no information about the assumed signal class $\mathcal{X}$. This is the same estimator studied (for the special case of $\alpha = 0$) in Wainwright [2] and is given by

$$\hat{K}_{\mathrm{ML}}(y) = \arg\min_{|U|=k} ||[I_m - \Phi_U(\Phi_U^T\Phi_U)^{-1}\Phi_U^T]\, y||^2,$$

where Φ_U corresponds to the columns of Φ indexed by U.

We remark that ML decoding is computationally expensive for any problem of non-trivial size. However, the resulting achievable bound is interesting because it shows where there is potential for improvement in current sub-optimal recovery algorithms. Furthermore, if one is able to lower bound the performance of some efficient estimator with respect to the optimal decoder, then an achievable result is automatically attained.

III. Results

This section gives our main results. Section III-A provides necessary definitions, Section III-B states the theorems, and Section III-C gives discussion.

A. Definitions

For a given signal x we define two quantities: the per-sample signal-to-noise ratio (SNR) is given by

$$\mathsf{SNR}(x) = \frac{\mathbb{E}\left[||\Phi x||^2\right]}{\mathbb{E}\left[||w||^2\right]} = \frac{1}{n\sigma_w^2}||x||^2,$$

and the normalized magnitude of smallest αk non-zero elements is given by

$$g(\alpha, x) = \min_{U\subset K\,:\,|U|=\alpha k} \quad \frac{1}{\alpha}\frac{||x_U||^2}{||x||^2},$$

where x_U is the vector of elements indexed by the set U.

Performance guarantees for a given class $\mathcal{X}$ require good bounds on the above quantities. In our analysis, we may use any bounds $\mathsf{SNR}(\mathcal{X})$ and $g(\alpha,\mathcal{X})$ which satisfy the following requirements: if $\mathcal{X}$ is non-stochastic then

$$\mathsf{SNR}(\mathcal{X}) \leq \mathsf{SNR}(x) \quad \text{and} \quad g(\alpha,\mathcal{X}) \leq g(\alpha, x)$$

for all $x \in \mathcal{X}$, and if $\mathcal{X}$ is stochastic then there exists some $c > 0$ such that

$$\mathbb{P}\left\{\mathsf{SNR}(\mathcal{X}_n) \leq \mathsf{SNR}(x)\right\} > 1 - e^{-n\,c} \quad \text{and}$$
$$\mathbb{P}\left\{g(\alpha,\mathcal{X}_n) \leq g(\alpha, x)\right\} > 1 - e^{-n\,c}.$$

We also need an upper bound, $\beta_L(\mathcal{X})$, on the relative magnitude of the smallest element of x. If $\mathcal{X}$ is non-stochastic then

$$\beta_L(\mathcal{X}) = \lim_{n\to\infty} \inf_{x\in\mathcal{X}} \inf_{i\in K} x_i^2/\sigma_w^2,$$

and if $\mathcal{X}$ is stochastic then there must exist some constant $c > 0$ such that

$$\mathbb{P}\big\{\beta_L(\mathcal{X}_n) \geq \min_{i\in K} x_i^2/\sigma_w^2\big\} > 1 - e^{-n\,c}.$$

Finally, we define an extended version of the binary entropy function $h(p) = -p\log(p) - (1-p)\log(1-p)$. For $0 \leq \Omega \leq 1$ and $0 \leq \alpha \leq 1-\Omega$ we have

$$h(\Omega,\alpha) = \Omega h(\alpha) + (1-\Omega)\, h\left(\frac{\alpha}{1/\Omega - 1}\right). \qquad (6)$$

B. Sampling Rate vs. Fractional Distortion

We now present our main results which characterize the tradeoff between the sampling rate ρ and the fractional distortion α. All theorems pertain to linear sparsity ($\Omega \in (0,1)$) in the under-sampled ($\rho < 1$), and asymptotic ($n, k, m \to \infty$) setting.

As a baseline, it is straightforward to show (see Lemma 2.1 in [6]) that randomly choosing any k indices gives an estimate that is asymptotically reliable for any $\alpha > 1-\Omega$ and asymptotically unreliable for any $\alpha < 1-\Omega$. Thus we consider the range $0 \leq \alpha < 1-\Omega$.

We first address the task of perfect support recovery ($\alpha = 0$). In the paper [2], Wainwright gives a necessary condition for perfect support recovery. With respect to our sampling model, this condition is satisfied with $\beta_L(\mathcal{X}) < \infty$. In the following Theorem, we show that $\beta_L(\mathcal{X})$ must be infinite.

Theorem 1: For perfect recovery ($\alpha = 0$) any estimator $\hat{K}(y)$ is asymptotically unreliable if $\beta_L(\mathcal{X}) < \infty$.

Theorem 1 is very general in that it depends only on the behavior of the smallest non-zero element of x. This means that perfect recovery is not possible unless the per-sample SNR grows without bound with n.

We next address the task of fractional support recovery ($\alpha \in (0, 1-\Omega)$). For a given SNR, the following theorems provide both upper and lower bounds on the fraction of the support that can be reliably recovered.

Theorem 2 (Necessary Conditions): Asymptotically reliable recovery is impossible if

$$\rho < \frac{h(\Omega) - h(\Omega,\alpha) + \frac{1}{n}I(x;y|K)}{\frac{1}{2}\log\big(1 + \mathsf{SNR}(\mathcal{X})\big)}, \qquad (7)$$

where $I(x;y|K)$ is mutual information between x and y conditioned on K.

The conditional mutual information $I(x;y|K)$ is zero for non-stochastic signal classes, but is positive for stochastic signal classes. As demonstrated in Section III-B.2 this term can be given explicitly for some classes.

Theorem 3 (Sufficient Conditions): The estimator $\hat{K}_{\mathrm{ML}}(y)$ is asymptotically reliable if $\gamma(\alpha,\mathcal{X}) > 1$ and

$$\rho > \Omega + \max_{u\in[\alpha,1-\Omega]} \frac{2\,h(\Omega,u)}{\log\big(\gamma(u,\mathcal{X})\big) + 1/\gamma(u,\mathcal{X}) - 1}, \qquad (8)$$

where $\gamma(u,\mathcal{X}) = \mathsf{SNR}(\mathcal{X})\, u\, g(u,\mathcal{X})$.

Note that for a given signal class $\mathcal{X}$, the above bound depends only on the functions $\mathsf{SNR}(\mathcal{X})$ and $g(u,\mathcal{X})$.

In the following two sections we refine Theorems 2 and 3 for two particular signal classes: one stochastic and one non-stochastic.

1) Bounded Signals: Often it is appealing to have signal models that do not assume a distribution. Such models may arise naturally when we need a worst-case analysis. Also, resulting claims are robust in that they do not depend on the choice or parameters of an assumed distribution. We define $\mathcal{B}_n$ to be the set of all $x \in \mathbb{R}^n$ whose non-zero elements are bounded from below in magnitude. Specifically, $|x_i| \geq x_{\text{min}}$ for all $i \in K$, where x_{min} is a known constant that does not depend n. Previous work on support recovery [1], [3], [14] has focused on this class, and by definition, we have $\beta_L(\mathcal{B}) = x_{\text{min}}^2/\sigma_w^2$, $\mathsf{SNR}(\mathcal{B}) = \Omega\beta_L(\mathcal{B})$, and $g(\alpha, \mathcal{B}) = 1$. Since the bounded signal class does not necessary have a distribution the conditional mutual information $I(x; y|K)$ is zero.

A simplified (an necessarily weakened) sufficient condition is $\mathsf{SNR}(\mathcal{B})\, \alpha > e$ and

$$\rho > \Omega + \frac{2h(\Omega)}{\log\left(\mathsf{SNR}(\mathcal{B})\,\alpha/e\right)}. \tag{9}$$

This means that if we set $\rho = \Omega + 2h(\Omega)$, then with exponentially high probability in n, the fractional distortion of the estimator $\hat{K}_{\text{ML}}(y)$ obeys

$$\alpha < e^2/\mathsf{SNR}(\mathcal{B}). \tag{10}$$

2) Gaussian Signals: Support recovery becomes more challenging when the non-zero elements of x can be arbitrarily close to zero. In this setting, stochastic signal classes still allow performance guarantees. We consider the special case of a zero mean Gaussian distribution and define $\mathcal{G}_n$ to be the set of all $x \in \mathbb{R}^n$ whose non-zero elements are i.i.d. $\mathcal{N}(0, \sigma_x^2)$. We define $\beta = \sigma_x^2/\sigma_w^2$.

Since $\mathsf{SNR}(x)$ is a random variable that obeys concentration of measure, standard large deviation bounds for χ^2 variables (Lemma A.1 in [6]) show that we may choose $\mathsf{SNR}(\mathcal{G}_n)$ arbitrarily close to $\Omega\beta$. Also, we may trivially choose $\beta_L(\mathcal{G}) = \beta$ although much tighter bounds are possible. A suitable choice for $g(\alpha, \mathcal{G})$ is given by the following lemma which is proved in [6] in Section 4.4.3.

Lemma 1: For the Gaussian signal class $\mathcal{G}$ we may choose

$$g(\alpha, \mathcal{G}) = -W\left(-e^{-(2/\alpha)h(\alpha)-1}\right) > \alpha^2/e^3 \tag{11}$$

where the Lambert-W function $W(z)$ is the inverse function of $f(z) = z\,e^z$.

Furthermore, the asymptotic spectrum of the random matrix $\Phi_U^T\Phi_U$ is given by the Marcenko-Pastur law [21], and the following lemma follows from results in [22].

Lemma 2: For the Gaussian signal class $\mathcal{G}$

$$\tfrac{1}{n}I(x; y|K) \to \mathcal{V}(\mathsf{SNR}(\mathcal{G}); \rho/\Omega) \quad \text{as} \quad n \to \infty, \tag{12}$$

where

$$\begin{aligned} \mathcal{V}(\gamma; r) &= \log\left(1 + \gamma - F(\gamma, r)\right) + \tfrac{1}{r\gamma}F(\gamma, r) \\ &\quad + \tfrac{1}{r}\log\left(1 + r\gamma - F(\gamma, r)\right) \qquad \text{and} \\ F(\gamma, r) &= \tfrac{1}{4}\left(\sqrt{\gamma(1+\sqrt{r})^2 + 1} - \sqrt{\gamma(1-\sqrt{r})^2 + 1}\right)^2. \end{aligned}$$

A simplified (an necessarily weakened) sufficient condition is $\mathsf{SNR}(\mathcal{G})\,\alpha^3 > e^4$ and

$$\rho > \Omega + \frac{2\,h(\Omega)}{\log\left(\mathsf{SNR}(\mathcal{G})\,\alpha^3/e^4\right)}. \tag{13}$$

This means that if we set $\rho = \Omega + 2h(\Omega)$, then with exponentially high probability in n, the fractional distortion of the estimator $\hat{K}_{\text{ML}}(y)$ obeys

$$\alpha < (e^4/\mathsf{SNR}(\mathcal{G}))^{1/3}. \tag{14}$$

C. Illustration of Results

The bounds in Theorems 2 and 3 are shown in Figure 1 for the bounded and Gaussian signal classes. In light of Theorem 1 we see the necessary bounds are overly conservative as $\alpha \to 0$. For both signal classes, we see that recovery in the under-sampled setting with fixed SNR is possible over a range of α. However, as α becomes small, the sampling rate increases without bound. Also, we see that the upper and lower bond are reasonably tight for values of α that are not near 0 or $1 - \Omega$. These results show that if we accept a small fraction of errors, only a small number of samples is needed.

IV. Analysis

The section provides proof outlines. The full proofs can be found in [6].

A. Proof of Theorem 1

We consider a modified problem in which the estimator has access to additional information about x and show that optimal recovery is asymptotically unreliable. For a given signal x, let $i_0 = \arg\min_{i \in K}|x_i|$, and assume that the decoder knows the signal x_K and the set $K_1 = K\backslash i_0$, that is every element of the support except for i_0. All that remains is to determine which of the remaining $n - k + 1$ indices belongs in K. Note that $y - \Phi_{K_0}x_{K_0} \sim \mathcal{N}(x_{i_0}a_i, \sigma_w^2 I)$ and thus the MAP estimate of i_0 is given by

$$\begin{aligned} \hat{i}_0 &= \arg\min_{j \in K_1^\perp} ||y - \Phi_{K_1}x_{K_1} - x_{i_0}a_j||^2 \\ &= \arg\min_{j \in K_1^\perp} ||w + x_{i_0}a_{i_0} - x_{i_0}a_j||^2. \end{aligned}$$

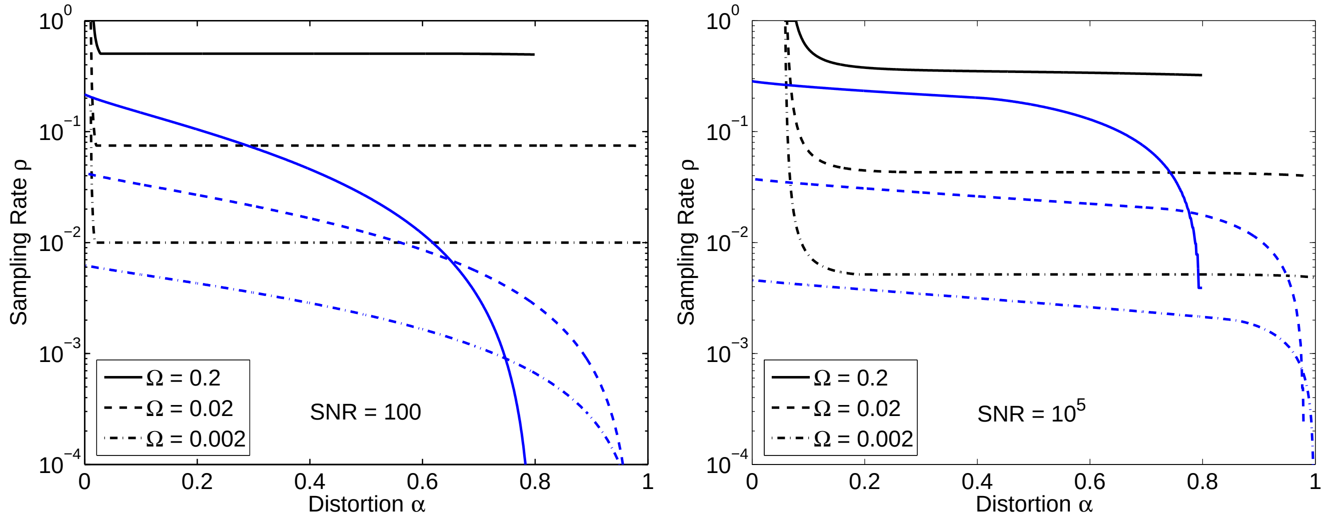
For this decoder, an error occurs if there exists $j \in K^\perp$ such that

$$||w + x_{i_0}a_{i_0} - x_{i_0}a_j||^2 < ||w||^2.$$

Using properties of chi squared random variables, it is possible to lower bound the probability of the above event with some positive constant when $x_{i_0}^2/\sigma_w^2 < \infty$.

B. Proof of Theorem 2

The key step is to note that $h(\Omega) - h(\Omega, \alpha)$ is the bit rate required to describe the support to within distortion α. The rest of the bound follows from the bound given by Gastpar and Bresler [15].



(a) (b)

Fig. 1. Sufficient (bold) and necessary (light) sampling densities ρ (log scale) as a function of the fractional distortion α for various Ω for bounded signal class (a) and the Gaussian signal class (b).

C. Proof of Theorem 3

The main technical result underlying the proof is the following lemma which relates the desired error probability to the large deviations behavior of multiple independent chi-squared variables.

Lemma 3: For given parameters (n, k, m, α), signal class $\mathcal{X}_n$, and any scalar $t > 0$ we have

$$P_e(\alpha, \mathcal{X}_n) \leq \mathbb{P}\{\chi^2(m-k) > t\} + \sum_{a=\lfloor \alpha k \rfloor}^{\lceil k^2/n \rceil} \Big[e^{-n\,c} + \binom{k}{a}\binom{n-k}{a} \mathbb{P}\left\{\chi^2(m-k) < \tau(a)\, t\right\} \Big],$$

where $\tau(a) = \left[\mathsf{SNR}(\mathcal{X})(a/k) g(a/k, \mathcal{X})\right]^{-1}$ and $\chi^2(d)$ denotes a chi-squared variable with d degrees of freedom.

ACKNOWLEDGMENT

We would like to thank Martin Wainwright for helpful discussions and pointers. This work was supported in part by ARO MURI No. W911NF-06-1-0076.

REFERENCES

[1] M. Wainwright, "Sharp thresholds for high-dimensional and noisy recovery of sparsity," in *Proc. Allerton Conf. on Comm., Control, and Computing*, Monticello, IL, Sep 2006.

[2] ——, "Information-theoretic bounds on sparsity recovery in the high-dimensional and noisy setting," in *Proc. IEEE Int. Symposium on Information Theory*, Nice, France, Jun 2007.

[3] P. Zhao and B. Yu, "On model selection consistency of Lasso," *J. of Machine Learning Research*, vol. 51(10), pp. 2541–2563, Nov 2006.

[4] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52(1), pp. 6–18, Jan 2006.

[5] E. Candes, J. Romberg, and T. Tao, "Stable Signal Recovery from Incomplete and Inaccurate Measurements," *Comm. on Pure and Applied Math.*, vol. 59(8), pp. 1207–1223, 2006.

[6] G. Reeves, "Sparse signal sampling using noisy linear projections," Master's thesis, EECS Department, University of California, Berkeley, Jan 2008. [Online]. Available: http://www.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS-2008-3.html

[7] P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," in *Proc. IEEE Int. Conf. Acoust. Speech Sig. Proc.*, vol. 3, Atlanta, GA, May 1996, pp. 1689–1692.

[8] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. of Scientific Computing*, vol. 20(1), pp. 33–61, 1998.

[9] D. Donoho, "Compressed Sensing," *IEEE Trans. Inf. Theory*, vol. 52(4), pp. 1289–1306, Apr. 2006.

[10] E. Candes, J. Romberg, and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52(12), pp. 5406–5425, Dec. 2006.

[11] J. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 52(3), pp. 1030–1051, 2006.

[12] J. Fuchs, "Recovery of exact sparse representations in the presence of noise," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Montreal, QC, Canada, May 2004, pp. 533–536.

[13] ——, "Recovery of exact sparse representations in the presence of bounded noise," *IEEE Trans. Inf. Theory*, vol. 51(10), pp. 3601–3608, Oct 2005.

[14] N. Meinshausen and P. Buhlmann, "Consistent neighborhood selection for high-dimensional graphs with Lasso," *Annals of Stat.*, vol. 34(3), 2006.

[15] M. Gastpar and Y. Bresler, "On the necessary density for spectrum-blind nonuniform sampling subject to quantization," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Istanbul, Turkey, Jun 2000, pp. 248–351.

[16] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Necessary and sufficient conditions on sparsity pattern recovery," May 2008, arXiv:0804.1839v1 [cs.IT].

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

[18] A. Fletcher, S. Rangan, and V. Goyal, "Rate-Distortion bounds for sparse approximation," in *Proc. IEEE Stat. Signal Processing Workshop*, Madison, WI, Aug 2007, pp. 254–258.

[19] S. Aeron, M. Zhao, and V. Saligrama, "On sensing capacity of sensor networks for the class of linear observation, fixed snr models," Jun 2007, arXiv:0704.3434v3 [cs.IT].

[20] M. Akcakaya and V. Tarokh, "Shannon theoretic limits on noisy compressive sampling," Nov 2007, arXiv:0711.0366v1 [cs.IT].

[21] V. Marcenko and L. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Math. USSR-Sbornik*, vol. 1, pp. 457–483, 1967.

[22] A. Tulino and S. Verdu, *Random Matrix Theory and Wireless Communications*. Hanover, MA: now Publisher Inc., 2004.