RATE-DISTORTION BOUNDS FOR SPARSE APPROXIMATION

Alyson K. Fletcher[1], *Sundeep Rangan*[2], *and Vivek K Goyal*[3]

[1]University of California, Berkeley
[2]QUALCOMM Flarion Technologies
[3]Massachusetts Institute of Technology

ABSTRACT

Sparse signal models arise commonly in audio and image processing. Recent work in the area of compressed sensing has provided estimates of the performance of certain widely-used sparse signal processing techniques such as basis pursuit and matching pursuit. However, the optimal achievable performance with sparse signal approximation remains unknown. This paper provides bounds on the ability to estimate a sparse signal in noise. Specifically, we show that there is a critical minimum signal-to-noise ratio (SNR) that is required for reliable detection of the sparsity pattern of the signal. We furthermore relate this critical SNR to the asymptotic mean squared error of the maximum likelihood estimate of a sparse signal in additive Gaussian noise. The critical SNR is a simple function of the problem dimensions.

Index Terms— basis pursuit, compressed sensing, estimation, matching pursuit, maximum likelihood, unions of subspaces

1. INTRODUCTION

Sparse signal models arise in a variety of applications. A simple mathematical model for sparse signals is as follows: Suppose T is an $N \times M$ matrix and $\mathbf{X}$ is the set of vectors

$$\mathbf{X} = \{x = Tu \mid u \text{ has at most } K \text{ non-zero components}\}. \quad (1)$$

When $M \geq N$ and T has rank N, the columns of T are an overcomplete set and comprise a *frame* in $\mathbb{R}^N$. The vector $x \in \mathbf{X}$ is said to be K-*sparse* with respect to T.

A fundamental problem in sparse signal processing is the estimation of sparse signals in noise. Specifically, suppose x is an unknown signal that is K-sparse with respect to some frame T. Suppose that y is a noisy version of x given by

$$y = x + d, \quad (2)$$

where d is additive noise. The problem is to estimate x from y.

One natural estimate is the K-sparse vector closest to the observed signal y. (Under a probabilistic model for d in which the probability density is maximum at $d = 0$, this is the maximum likelihood (ML) estimate of x.) Unfortunately, finding this sparse approximation estimator is well-known to be NP-hard [1]. Most research has thus focused on approximate methods, including iterative techniques such as matching pursuit [2] and convex programming-based methods such as basis pursuit [3].

Support provided by a University of California President's Postdoctoral Fellowship, the Texas Instruments Leadership Universities Program, and the Centre Bernoulli, École Polytechnique Fédérale de Lausanne.

While both matching pursuit and basis pursuit are computationally simple and appear to often work well in practice, their performance is difficult to quantify precisely. Important developments in recent years include a number of explicit conditions that guarantee that basis pursuit or matching pursuit recovers the sparsity pattern of the unknown signal u in (1) [4–11]. The conditions to guarantee the recovery of the sparsity pattern u are generally stated in terms of the magnitude of the noise d and the so-called *coherence* of T. These performance guarantees and further insights from [12] have motivated an application to distributed and universal compression known as *compressed sensing* [13, 14].

If the above results can be considered as lower bounds on the achievable performance of sparse approximation with practical algorithms, the purpose of this paper is to establish *upper bounds* that apply to all algorithms. Specifically, for the sparse approximation problem, we consider an optimal, but computationally intractable, ML estimator that exhaustively searches all possible K-dimensional subspaces for the sparse signal. For this estimator, we provide a simple lower bound on the minimum SNR required to detect the correct subspace with vanishing probability of error. The bound is given in terms of the problem dimensions.

Some of the results follow from an information-theoretic analysis along the lines of our earlier paper [15] and the more recent paper [16]. Proofs are outlined in Section 6; details have been omitted due to space constraints.

2. INDEPENDENT SUBSPACE SIGNAL MODELS

Our analysis is based on the following generalization of frame-based signals in (1). We define as a (J, N, K)-*subspace set* any set

$$\mathbf{S} = \{S_1, \ldots, S_J\} \quad (3)$$

where each element S_j is a K-dimensional subspace of $\mathbb{R}^N$. The *sparsity ratio* of the set is $\alpha = K/N$. We call the union of the subspaces

$$\mathbf{X} = \left\{x \in \mathbb{R}^N \mid x \in S_j, \text{ for some } j = 1, \ldots, J\right\} \quad (4)$$

the (J, N, K)-*subspace signal set* generated by $\mathbf{S}$.

With these definitions, the sparse signal model (1) is a particular example of a (J, N, K)-signal set. Specifically, any K-sparse vector x in (1) is the span of at most K of the M columns of T, and therefore belongs to one of $\binom{M}{K}$ subspaces of $\mathbb{R}^N$ of dimension K. Therefore, if we let $J = \binom{M}{K}$ and denote the possible subspaces containing x by S_j with $j = 1, \ldots, J$, then $\mathbf{X}$ in (1) is precisely the (J, N, K)-signal set generated by the set of S_js. In particular, any bound that applies to all (J, N, K)-signal sets will apply to the frame model (1).

1-4244-1198-X/07/$25.00 ©2007 IEEE

Definition 1 *A (J, N, K)-subspace set (3) is called* independent and uniformly generated *if*

(a) the subspaces S_j are random, independently and identically distributed; and

(b) each S_j is the range of a random $N \times K$ matrix A_j, where the NK components of A_j are i.i.d. Gaussian scalars with zero mean and unit variance.

Part (b) of the definition is equivalent to the subspaces S_js being *rotationally invariant* in that each K-dimensional subspace, S_j, is identically distributed to US_j for any $N \times N$ orthogonal matrix U. It is important to recognize that an independent and uniformly generated signal model in Definition 1 is *not* equivalent to the frame model (1) for some random matrix T. If the components of the matrix T in the frame model are i.i.d. Gaussian variables with zero mean, then each of the $J = \binom{M}{K}$ subspaces of spanned by K columns of T will satisfy the rotationally invariant property in part (b) of Definition 1. However, the random subspaces will not be independent since, in general, they will share common column vectors. Nevertheless, we will see that independent and uniformly generated subspace signal sets provide certain lower bounds that apply to all subspace sets including those generated by frames.

3. APPROXIMATION ERROR BOUNDS

Our first result provides a lower bound on the ability of any (J, N, K)-signal model to approximate a Gaussian random vector. Let $y \sim \mathcal{N}(0, I_N)$ be an N-dimensional Gaussian vector with zero mean and unit variance. Given any (J, N, K)-subspace signal set, $\mathbf{X}$, we can define the *relative approximation error* of y with respect to $\mathbf{X}$ as

$$\rho_{\min} = \frac{1}{\|y\|^2} \min_{x \in \mathbf{X}} \|y - x\|^2. \tag{5}$$

The approximation error $\rho_{\min}$ is between 0 and 1, with a value closer to 0 implying a lower approximation error.

The first result provides a lower bound on $\rho_{\min}$. To state the result, given N and J, define the *subspace rate* as

$$R_J = \frac{1}{N} \log_2 J, \tag{6}$$

which represents the number of bits per dimension to index the J subspaces. Also, for scalars p and $q \in (0, 1)$, let $D(p, q)$ denote the binary *Kullback-Leibler distance* [17] given by

$$D(p, q) = p \log_2 \left(\frac{p}{q}\right) + (1 - p) \log_2 \left(\frac{1 - p}{1 - q}\right). \tag{7}$$

With these two definitions, we have the following bound.

Theorem 1 *Let $y \sim \mathcal{N}(0, I_N)$.*

(a) For any (J, N, K) subspace signal set, $\mathbf{X}$, the mean relative approximation is bounded below as:

$$\mathbf{E}\rho_{\min} \geq \rho^*_{\min}, \tag{8}$$

where $\rho^*_{\min}$ is the unique solution to

$$2R_J = D(\alpha, 1 - \rho^*_{\min}), \quad \rho^*_{\min} \in (0, 1 - \alpha). \tag{9}$$

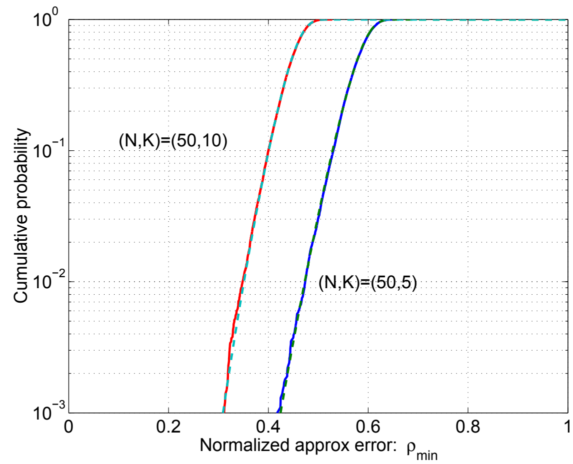

Fig. 1. Approximation of a Gaussian random vector with an independent and uniformly generated subspace set. Plotted are theoretical (solid lines) and simulated (dashed lines) cumulative distribution functions of $\rho_{\min}$ with $N = 50$, $R_J = 0.25$ and $K \in \{5, 10\}$. Simulated values are based on 10 000 Monte Carlo trials.

(b) Suppose $\mathbf{X}$ is an independent and uniformly generated (J, N, K)-subspace set as in Definition 1. Then, the limit of the approximation error $\rho_{\min}$ with $N \to \infty$ and $\alpha = K/N$ and R_J in (6) held constant is given by

$$\lim_{N \to \infty} \rho_{\min} = \rho^*_{\min}. \tag{10}$$

Part (a) of the theorem provides a simple lower bound on the relative approximation error of a Gaussian random vector with respect to an arbitrary subspace signal model. In particular, the bound applies when $\mathbf{X}$ is a frame. Part (b) shows the result is tight in that the bound can be achieved for certain large random subspace signal sets. However, the result does not imply that the bound can be achieved for large random *frames*. Indeed for frames, it can be shown that the bound is not tight, and the exact characterization of the approximation error using frames remains open.

The proof of Theorem 1 relies on a derivation of the distribution of $\rho_{\min}$ for the case of an independent and uniformly generated subspace set. Specifically, $\rho_{\min}$ has the cumulative distribution

$$F_J(\rho) = \Pr(\rho_{\min} \leq \rho) = 1 - (1 - F(\rho))^J \tag{11}$$

where $F(\cdot)$ is the cumulative distribution function of a beta(r,s) random variable with $r = (N - K)/2$ and $s = K/2$. The corresponding p.d.f. of $\rho_{\min}$ is given by

$$f_J(\rho) = \frac{\partial F_J(\rho)}{\partial \rho} = J f(\rho)(1 - F(\rho))^{J-1}. \tag{12}$$

Two examples of the distribution of $\rho_{\min}$ are shown in Figure 1. The convergence of $\rho_{\min}$ to $\rho^*_{\min}$ is illustrated in Figure 2. The theorem correctly implies convergence to $\rho^*_{\min} \approx 0.513$.

4. SUBSPACE DETECTION BOUNDS

We next consider the estimation of subspace-based signals in the presence of noise. Specifically, suppose $\mathbf{X}$ is a (J, N, K)-subspace

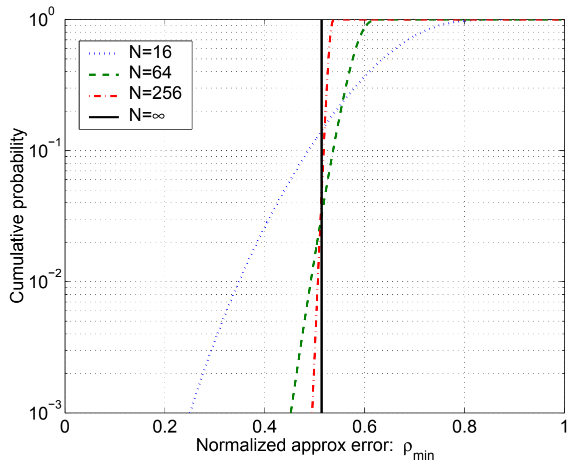

Fig. 2. Convergence of $\rho_{\min}$ for a uniformly and independently generated sparse signal model. Plotted is the c.d.f. of the approximation error $\rho_{\min}$ for various values of N with $\alpha = 0.1$ and $R_J = 0.5$.

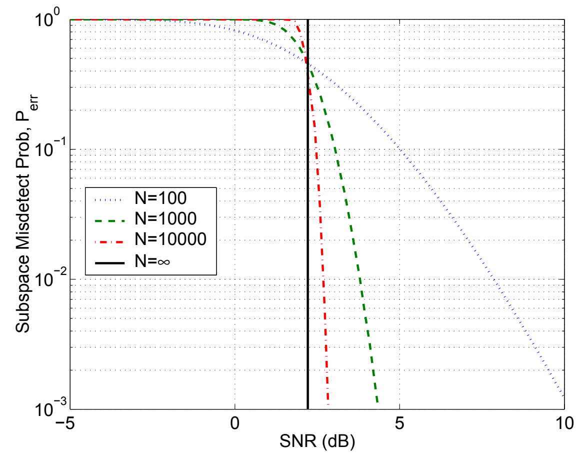

Fig. 3. Convergence of the probability of sparsity pattern detection. P_{err} is plotted as a function of the SNR γ for various values of N. In all cases, $\alpha = 0.1$ and $R_J = 0.5$.

signal set. Given an unknown $x \in \mathbf{X}$, let y be a noisy observation of the form (2), where $d \sim \mathcal{N}(0, I_N)$ is additive Gaussian noise. Since $x \in \mathbf{X}$, x belongs to one of J subspaces of dimensional K, S_j. We consider the problem of detecting the subspace to which x belongs. For the frame model (1), the problem is equivalent to detecting which K components of the vector u are non-zero.

We analyze the following ML estimator. Let θ be the index of the subspace containing x. Under the assumption that d is additive Gaussian noise, the ML estimate for θ is given by

$$\hat{\theta} = \operatorname*{argmax}_{j \in \{1,\ldots,J\}} \|P_j y\|, \tag{13}$$

where P_j is the orthogonal projection operator onto S_j. Similar to the approximation problem, this ML estimator is computationally infeasible since it involves an exhaustive search over all J subspaces. However, we will ignore this consideration since we are interested here in information-theoretic limits.

Our analysis of the ML estimator considers the average error under the following model for the unknown signal x.

Assumption 1 *The signal x in (2) is of the form*

$$x = V_\theta u, \tag{14}$$

where θ is an unknown subspace index, uniformly distributed on the set $\{1, \ldots, J\}$; each V_j is an orthogonal $N \times K$ matrix, with independent and rotationally invariant distributions; and u is a Gaussian random vector with $u \sim \mathcal{N}(0, I_K \gamma/\alpha)$ for some constant $\gamma > 0$.

Here, the constant $\gamma > 0$ represents the SNR, since it can be easily verified that $\gamma = \mathbf{E}\|x\|^2/\mathbf{E}\|d\|^2$. The following result provides a bound on the average probability of error in terms of the signal dimensions J, N and K and the SNR γ.

Theorem 2 *Consider the subspace detection problem above.*

(a) Let γ_{crit} be the solution to

$$R_J = \frac{1}{2}\log_2(1 + \gamma_{\text{crit}}) - \frac{\alpha}{2}\log_2\left(1 + \frac{\gamma_{\text{crit}}}{\alpha}\right), \tag{15}$$

where R_J is defined in (6). Then, if $\mathbf{X}$ is independent and uniformly generated, the limit of the error probability P_{err} as $N \to \infty$ with γ, $\alpha = K/N$ and R_J held constant is given by

$$\lim_{N\to\infty} P_{\text{err}} = \begin{cases} 0, & \gamma > \gamma_{\text{crit}}; \\ 1, & \gamma < \gamma_{\text{crit}}. \end{cases} \tag{16}$$

(b) For any (J, N, K)-subspace signal set $\mathbf{X}$, the probability of error is lower bounded by

$$\left(1 - H\left(\tfrac{P_{\text{err}}}{2}\right)\right)\left[R_J + \frac{\alpha}{2}\log_2\left(1 + \frac{\gamma}{\alpha}\right)\right] \le \frac{\log_2(1+\gamma)}{2},$$

where $H(p)$ is the binary entropy.

Part (b) of the theorem provides a simple lower bound on the probability of error, P_{err}, in terms of the signal dimensions and SNR. One way of interpreting the inequality in part (b) is as follows. The binary entropy $H(p)$ satisfies $H(p) \to 0$ when $p \to 0$. Therefore, if we require the probability of error to vanish, we must have

$$R_J \le \frac{1}{2}\log_2(1+\gamma) - \frac{\alpha}{2}\log_2\left(1 + \frac{\gamma}{\alpha}\right).$$

Since the right hand side of this inequality increases with γ, we must have that $\gamma \ge \gamma_{\text{crit}}$, where γ_{crit} is defined in (15). Therefore, γ_{crit} represents a minimum SNR required for reliable detection of the correct subspace.

The result in part (a) thus shows that the minimum SNR bound is tight in that, for large random subspace signal models, the subspace can be reliably detected at this SNR.

The convergence of P_{err} to a step function is illustrated in Figure 3. This reinforces the interpretation of γ_{crit} as a critical SNR.

5. ESTIMATION ERROR BOUNDS

Our final results are bounds on the mean squared error (MSE) of the ML estimator described in the previous section. We measure the performance of estimate $\hat{x}$ via the *average normalized MSE*:

$$\text{MSE} = \frac{\mathbf{E}\|x - \hat{x}\|^2}{\mathbf{E}\|d\|^2} = \frac{1}{N}\mathbf{E}\|x - \hat{x}\|^2. \tag{17}$$

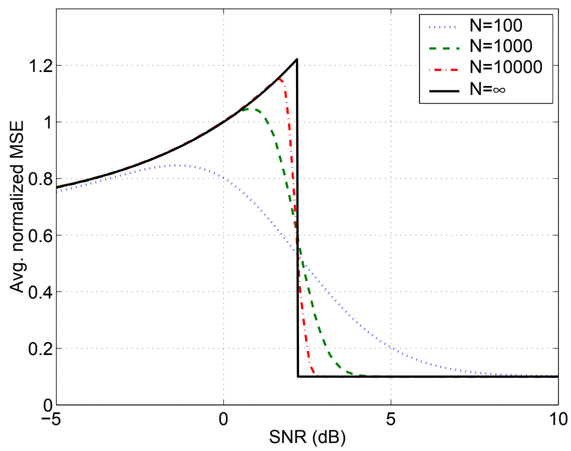

Fig. 4. Convergence of the average normalized MSE. The average normalized MSE in Theorem 3 is plotted as a function of the SNR γ for various values of N. In all cases, $\alpha = 0.1$ and $R_J = 0.5$.

Theorem 3 *Consider the signal model in Assumption 1. Then the average normalized MSE of the ML estimate $\hat{x}$ is given by*

$$MSE = \alpha + \int_0^1 \int_0^1 f(z) f_{J-1}(\rho) H(\rho, z) \mathbf{1}_{\{\rho < g(z)\}} \, d\rho \, dz + \epsilon, \tag{18}$$

where $f_{J-1}(\rho)$ is the p.d.f. in (12),

$$g(z) = \frac{z}{1 + (1-z)\gamma/\alpha}, \tag{19}$$

$$H(\rho, z) = (1-\rho)z + (1+\gamma/\alpha)\rho(1-z) - 2\alpha\rho, \tag{20}$$

and ϵ is bounded by

$$|\epsilon|^2 \le 8K/N^2 = 8\alpha/N. \tag{21}$$

Note that the MSE expression in Theorem 3 is not exact but rather has an $O(1/N)$ error term. Therefore, it is asymptotically exact for large signal dimensions N. For any fixed N, (18) is relatively easy to compute numerically. This expression yields an asymptotic MSE given in our final theorem.

Theorem 4 *Consider the MSE in Theorem 3, and let $\gamma_{\rm crit}$ be the solution to (15). Then the limit of MSE in (18) as $N \to \infty$ with γ, $\alpha = K/N$ and R_J in (6) held constant is given by*

$$\lim_{N \to \infty} MSE = \alpha + (1-\alpha)\frac{\gamma + \gamma_{\rm crit}}{1 + \gamma_{\rm crit}} \mathbf{1}_{\{\gamma < \gamma_{\rm crit}\}}. \tag{22}$$

The convergence of the MSE is illustrated in Figure 4. For the values of α and R_J in the figure, the solution to (15) results in a critical SNR $\gamma_{\rm crit}$ of 2.21 dB. It can be seen that as $N \to \infty$, for $\gamma > \gamma_{\rm crit}$, the MSE approaches $\alpha = 0.1$. For $\gamma < \gamma_{\rm crit}$ the asymptotic MSE grows with γ and achieves values with MSE > 1.

6. PROOFS

6.1. Proof of Theorem 1

A full proof would require some tedious details that we omit. For any $u \in (0, 1-\alpha)$, define the function

$$\phi(u) = 2R_J - D(\alpha, 1-u). \tag{23}$$

With this definition, (9) is equivalent to $\phi(\rho^*_{\min}) = 0$. We use the following two lemmas.

Lemma 1 *Given any $\alpha \in (0,1)$ and $R_J > 0$, there exists a unique solution $\rho^*_{\min} \in (0, 1-\alpha)$ to $\phi(\rho^*_{\min}) = 0$. Moreover,*

$$\begin{array}{ll} \phi(\rho) < 0 & \textit{for all } \rho \in (0, \rho^*_{\min})\textit{; and} \\ \phi(\rho) > 0 & \textit{for all } \rho \in (\rho^*_{\min}, 1-\alpha). \end{array} \tag{24}$$

Lemma 2 *Let $f(u)$ be the beta(r,s) p.d.f. with $r = (N-K)/2$ and $s = K/2$. Then*

$$\lim_{N \to \infty} \frac{2}{N} \log_2(J f(u)) = \phi(u).$$

Lemma 1 proves the existence and uniqueness of $\rho^*_{\min}$ satisfying (9). To prove the limit in distribution, $\rho \to \rho^*_{\min}$, we must show that

$$\lim_{N \to \infty} F_J(\rho) = \begin{cases} 0, & \text{if } \rho < \rho^*_{\min}; \\ 1, & \text{if } \rho > \rho^*_{\min}, \end{cases} \tag{25}$$

where $F_J(\rho)$ is the c.d.f. of $\rho_{\min}$ in (11).

Combining Lemmas 1 and 2 shows that for any $u \in (0, 1-\alpha)$,

$$\lim_{N \to \infty} J f(u) = \begin{cases} 0, & \text{if } u < \rho^*_{\min}; \\ \infty, & \text{if } u > \rho^*_{\min}. \end{cases}$$

Integrating this limit,

$$\lim_{N \to \infty} J F(\rho) = \lim_{N \to \infty} \int_0^\rho J f(u) \, du = \begin{cases} 0, & \text{if } \rho < \rho^*_{\min}; \\ \infty, & \text{if } \rho > \rho^*_{\min}. \end{cases} \tag{26}$$

Now suppose that $\rho > \rho^*_{\min}$. From (26), $JF(\rho) \to \infty$ as $N \to \infty$, and consequently,

$$(1 - F(\rho))^J \le \exp(-JF(\rho)) \to 0.$$

Therefore,

$$\lim_{N \to \infty} F_J(\rho) = 1 - \lim_{N \to \infty} (1 - F(\rho))^J = 1 \tag{27}$$

for all $\rho > \rho^*_{\min}$.

Next, suppose $\rho < \rho^*_{\min}$. Let $\epsilon > 0$. From (26), there exists an $N_0 > 0$ such that for all $N > N_0$, $JF(\rho) < \epsilon$. Therefore,

$$\lim_{N \to \infty} (1 - F(\rho))^J \ge \lim_{N \to \infty} (1 - \epsilon/J)^J = \exp(-\epsilon).$$

Since this is true for all $\epsilon > 0$, we must have

$$\lim_{N \to \infty} (1 - F(\rho))^J = 1,$$

and consequently

$$\lim_{N \to \infty} F_J(\rho) = 1 - \lim_{N \to \infty} (1 - F(\rho))^J = 0 \tag{28}$$

for all $\rho < \rho^*_{\min}$. The limits (27) and (28) together prove (25). Therefore, $\rho_{\min} \to \rho^*_{\min}$ in distribution. □

6.2. Proof of Theorem 2

One can show

$$P_{\mathrm{err}} = \Pr\left(\rho_{\min} < g(Z)\right), \tag{29}$$

where Z is a beta($(N-K)/2$,$K/2$) random variable and g is given in (19). From Theorem 1, $\rho_{\min} \to \rho^*_{\min}$ where $\rho^*_{\min}$ is the unique solution to (9). Also, $\lim_{N\to\infty} Z = \mathbf{E}Z = 1-\alpha$, where the convergence is in mean square. Substituting these limits into (29),

$$\lim_{N\to\infty} P_{\mathrm{err}} = \begin{cases} 0, & \text{if } \rho^*_{\min} > g(1-\alpha); \\ 1, & \text{if } \rho^*_{\min} < g(1-\alpha). \end{cases}$$

So, the theorem will be proven if we can show that the condition $\rho^*_{\min} > g(1-\alpha)$ is equivalent to $\gamma > \gamma_{\mathrm{crit}}$, where γ_{crit} is defined in (15).

To this end, let $\rho^* = (1-\alpha)/(1+\gamma_{\mathrm{crit}})$. Then,

$$1 - \rho^* = (\gamma_{\mathrm{crit}} + \alpha)/(1+\gamma_{\mathrm{crit}}).$$

Using the definition of the Kullback-Leibler distance function in (7),

$$\begin{aligned} D(\alpha, 1-\rho^*) &= \alpha \log_2\left(\frac{\alpha}{1-\rho^*}\right) + (1-\alpha)\log_2\left(\frac{1-\alpha}{\rho^*}\right) \\ &= \log_2(1+\gamma_{\mathrm{crit}}) - \alpha\log_2\left(1+\frac{\gamma_{\mathrm{crit}}}{\alpha}\right) = 2R_J. \end{aligned}$$

Thus, $2R_J = D(\alpha, 1-\rho^*)$. But, by Theorem 1, $\rho^*_{\min}$ is the unique solution to $2R_J = D(\alpha, 1-\rho^*_{\min})$, Therefore,

$$\rho^*_{\min} = \rho^* = (1-\alpha)/(1+\gamma_{\mathrm{crit}}). \tag{30}$$

Now, using the definition of $g(z)$ in (19),

$$g(1-\alpha) = \frac{1-\alpha}{1+\gamma}.$$

Therefore,

$$\rho^*_{\min} > g(1-\alpha) \Longleftrightarrow \frac{1-\alpha}{1+\gamma_{\mathrm{crit}}} > \frac{1-\alpha}{1+\gamma} \Longleftrightarrow \gamma > \gamma_{\mathrm{crit}}.$$

So $\gamma > \gamma_{\mathrm{crit}}$ if and only if $\rho^*_{\min} > g(1-\alpha)$. □

6.3. Proof of Theorem 3

In the integral in (18), $f_{J-1}(\rho)$ and $f(z)$ are the p.d.f.'s of $\rho_{\min}$ and the beta random variable Z as before. Therefore, the MSE in (18) can be rewritten as

$$\mathrm{MSE} = \alpha + \mathbf{E}\left[H(\rho_{\min}, Z)\mathbf{1}_{\{\rho_{\min} < g(Z)\}}\right] + \epsilon. \tag{31}$$

Similar to the proof of Theorem 2, we know that $Z \to 1-\alpha$ and $\rho_{\min} \to \rho^*_{\min}$, where $\rho^*_{\min}$ is the solution to (9). Also, because of (21), the error term ϵ in (18) goes to zero as $N \to \infty$. Therefore,

$$\lim_{N\to\infty} \mathrm{MSE} = \alpha + H(\rho^*_{\min}, 1-\alpha)\mathbf{1}_{\{\rho^*_{\min} < g(1-\alpha)\}}. \tag{32}$$

Substituting (30) into the definition of $H(\rho^*_{\min}, 1-\alpha)$ in (20),

$$\begin{aligned} H(\rho^*_{\min}, 1-\alpha) &= (1-\rho^*_{\min})(1-\alpha) + (\alpha+\gamma)\rho^*_{\min} - 2\alpha\rho^*_{\min} \\ &= 1-\alpha + \left[\alpha - 1 + \alpha + \gamma - 2\alpha\right]\rho^*_{\min} \\ &= 1-\alpha + (\gamma-1)\frac{(1-\alpha)}{1+\gamma_{\mathrm{crit}}} \\ &= (1-\alpha)\frac{\gamma_{\mathrm{crit}}+\gamma}{1+\gamma_{\mathrm{crit}}}. \end{aligned}$$

Also, from the proof of Theorem 2, we know that $\rho^*_{\min} < g(1-\alpha)$ is equivalent to $\gamma < \gamma_{\mathrm{crit}}$. Substituting these limits into (32) gives

$$\lim_{N\to\infty} \mathrm{MSE} = \alpha + (1-\alpha)\frac{\gamma_{\mathrm{crit}}+\gamma}{1+\gamma_{\mathrm{crit}}}\mathbf{1}_{\{\gamma<\gamma_{\mathrm{crit}}\}},$$

which proves (22). □

7. REFERENCES

[1] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Computing*, vol. 24, no. 2, pp. 227–234, Apr. 1995.

[2] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Proc.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[3] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.

[4] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2558–2567, Sept. 2002.

[5] D. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. Nat. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.

[6] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inform. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

[7] J.-J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 1341–1344, June 2004.

[8] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[9] D. L. Donoho and M. Elad, "On the stability of basis pursuit in the presence of noise," *Signal Proc.*, vol. 86, no. 3, pp. 511–532, Mar. 2006.

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

[11] M. J. Wainwright, "Sharp thresholds for high-dimensional and noisy recovery of sparsity," Tech. Rep., Univ. of California, Berkeley, Dept. of Statistics, May 2006, arXiv:math.ST/0605740 v1 30 May 2006.

[12] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

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

[14] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[15] A. K. Fletcher, S. Rangan, V. K Goyal, and K. Ramchandran, "Denoising by sparse approximation: Error bounds based on rate–distortion theory," *EURASIP J. Appl. Sig. Proc.*, vol. 2006, pp. 1–19, Mar. 2006.

[16] S. Sarvotham, D. Baron, and R. G. Baraniuk, "Measurements vs. bits: Compressed sensing meets information theory," in *Proc. 44th Ann. Allerton Conf. on Commun., Control and Comp.*, Monticello, IL, Sept. 2006.

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, 1991.