

Analysis of Denoising by Sparse Approximation with Random Frame Asymptotics

Alyson K. Fletcher
Univ. of California, Berkeley
alyson@eecs.berkeley.edu

Sundeep Rangan
Flarion Technologies
s.rangan@flarion.com

Vivek K Goyal
Mass. Inst. of Tech.
vgoyal@mit.edu

Kannan Ramchandran
Univ. of California, Berkeley
kannanr@eecs.berkeley.edu

Abstract—If a signal x is known to have a sparse representation with respect to a frame, the signal can be estimated from a noise-corrupted observation y by finding the best sparse approximation to y . This paper analyzes the mean squared error (MSE) of this denoising scheme and the probability that the estimate has the same sparsity pattern as the original signal. The first main result is an MSE bound that depends on a new bound on approximating a Gaussian signal as a linear combination of elements of an overcomplete dictionary. This bound may be of independent interest for source coding. Further analyses are for dictionaries generated randomly according to a spherically-symmetric distribution and signals expressible with single dictionary elements. Easily-computed approximations for the probability of selecting the correct dictionary element and the MSE are given. In the limit of large dimension, these approximations have simple forms. The asymptotic expressions reveal a critical input signal-to-noise ratio (SNR) for signal recovery.

I. INTRODUCTION

Estimating a signal from a noise-corrupted observation of the signal is a recurring task in science and engineering. This paper explores the limits of estimation performance in the case where the only *a priori* structure on the signal $x \in \mathbb{R}^N$ is that it has known sparsity K with respect to a given set of vectors $\Phi = \{\varphi_i\}_{i=1}^M \subset \mathbb{R}^N$. The set Φ is called a *dictionary* and is generally a *frame* (i.e., $M \geq N$) [1]. The sparsity of K with respect to Φ means that the signal x lies in the set

$$\Phi_K = \left\{ v \in \mathbb{R}^N \mid v = \sum_{i=1}^M \alpha_i \varphi_i, \text{ at most } K \text{ nonzero } \alpha_i\text{'s} \right\}.$$

Consider the problem of estimating a signal $x \in \Phi_K$ from the noisy observation $y = x + d$ where $d \in \mathbb{R}^N$ has the $\mathcal{N}(0, \sigma^2 I_N)$ distribution. The set Φ_K is the union of at most $J = \binom{M}{K}$ subspaces of dimension K . We henceforth assume $M > K$ (thus $J > 1$); if not, the model reduces to the classical case of knowing a single subspace that contains x . In this paper, the denoising operation is based only on the geometry of the signal model Φ_K and the distribution of d .

With the addition of the noise d , the observed vector y will (almost surely) not be represented sparsely, i.e., not be in Φ_K . Intuitively, a good estimate for x is the point from Φ_K that is closest to y in Euclidean distance. Formally, because the probability density function of d is a strictly decreasing function of $\|d\|_2$, this is the maximum likelihood estimate of

x given y . We will write

$$\hat{x}_{\text{SA}} = \operatorname{argmin}_{x \in \Phi_K} \|y - x\|_2 \quad (1)$$

for this estimate and call it the optimal K -term approximation of y . Henceforth we omit the subscript 2 indicating the norm.

Finding \hat{x}_{SA} can be viewed as a two-step procedure: first, find the subspace spanned by K elements of Φ that contains \hat{x}_{SA} ; then, project y to that subspace. The identification of a subspace and the orthogonality of $y - \hat{x}_{\text{SA}}$ to that subspace is used in our analyses. Let $\mathcal{P}_K = \{P_i\}_i$ be the set of the projections onto subspaces spanned by K of the M vectors in Φ . Then \mathcal{P}_K has at most $J = \binom{M}{K}$ elements, and $\hat{x}_{\text{SA}} = P_T y$, where $T = \operatorname{argmax}_i \|P_i y\|$.

The main results of this paper are bounds on the per-component mean-squared estimation error $\frac{1}{N} \mathbf{E} [\|x - \hat{x}_{\text{SA}}\|_2^2]$ for denoising via sparse approximation. These bounds only depend on (N, M, K) . Some results hold for all Φ and others are for randomly generated Φ . To the best of our knowledge, the results differ from any in the literature in several ways:

- (a) We study mean-squared estimation error for additive Gaussian noise. In contrast, analyses such as [2] impose a deterministic bound on the norm of the noise.
- (b) We concentrate on having dependence solely on dictionary size rather than more fine-grained properties of the dictionary. In particular, most signal recovery results in the literature are based on noise being bounded above by a function of the *coherence* of the dictionary.
- (c) We use source coding thought experiments in bounding estimation performance. This technique may be useful in answering other related questions, especially in sparse approximation source coding.

Our preliminary results were first presented in [3], with further details in [4]. Proofs of theorems of this paper appear in [5].

A. Related Work

Computing optimal K -term approximations for general Φ is an NP-complete problem. A greedy heuristic has been known as *matching pursuit* in the signal processing literature since the work of Mallat and Zhang [6]. Also, Chen, Donoho and Saunders [7] proposed a convex relaxation of the approximation problem (1) called *basis pursuit*.

Two related discoveries have touched off a flurry of recent research:

- (a) *Stability of sparsity*—Under certain conditions, the positions of the nonzero entries in a sparse representation of a signal are stable: applying optimal sparse approximation to a noisy observation of the signal will give a coefficient vector with the original support. Typical results are upper bounds (functions of the norm of the signal and the coherence of the dictionary) on the norm of the noise that allows a guarantee of stability (see, *e.g.*, [2]).
- (b) *Effectiveness of heuristics*—Both basis pursuit and matching pursuit are able to find optimal sparse approximations, under certain conditions on the dictionary and the sparsity of signal (see, *e.g.* [2], [8], [9]).

To contrast: in this paper we consider noise with unbounded support and thus a positive probability of failing to satisfy a sufficient condition for stability as in (a) above; and we do not address algorithmic issues in finding sparse approximations.

This paper uses quantization and rate–distortion theory only as a proof technique; there are no encoding rates because the problem is purely one of estimation. However, the “negative” results on representing white Gaussian signals with frames presented here should be contrasted with the “positive” encoding results of Goyal *et al.* [10]. The positive results of [10] are limited to low rates (and hence signal-to-noise ratios that are usually uninteresting). Note also that probability of error results for basis pursuit appear in the manuscript [11].

B. Preview of Results and Outline

As motivation, consider a set of numerical results that qualitatively reflect our main results. In these experiments, N , M , and K are small because of the high complexity of computing optimal approximations and because a large number of independent trials is needed to get adequate precision. Each data point shown is the average of 100 000 trials.

Consider a true signal $x \in \mathbb{R}^4$ ($N = 4$) that has an exact 1-term representation ($K = 1$) with respect to M -element dictionary Φ . We observe $y = x + d$ with $d \sim \mathcal{N}(0, \sigma^2 I_4)$ and compute estimate \hat{x}_{SA} from (1). The signal is generated with unit norm so that the signal-to-noise ratio (SNR) is $1/\sigma^2$ or $-10 \log_{10} \sigma^2$ dB.

We used dictionaries that are M maximally separated unit vectors in \mathbb{R}^N , where separation is measured by the minimum pairwise angle among the vectors and their negations. These are Grassmannian packings [12] in the simplest case of packing one-dimensional subspaces (lines).

Figure 1 shows the MSE as a function of σ for several values of M . (For visual clarity, MSE/σ^2 is plotted.) For small values of σ , the MSE is $(1/4)\sigma^2$. This is an example of the general statement that the MSE is $(K/N)\sigma^2$ for small σ . For large values of σ , the scaled MSE approaches a constant value:

$$\lim_{\sigma \rightarrow \infty} \frac{\text{MSE}}{\sigma^2} = g_{K,M},$$

where $g_{K,M}$ is a slowly increasing function of M and $\lim_{M \rightarrow \infty} g_{K,M} = 1$. This limiting value makes sense because in the limit $\hat{x}_{\text{SA}} \approx y = x + d$ and each component of d has variance σ^2 ; the denoising does not do anything. The

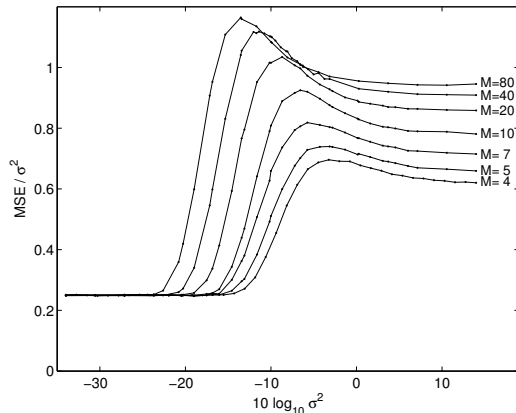


Fig. 1. Performance of denoising by sparse approximation when the true signal $x \in \mathbb{R}^4$ has an exact 1-term representation with respect to a dictionary that is an optimal M -element Grassmannian packing.

characterization of the dependence of $g_{K,M}$ on K and M is the main contribution of Section II.

Another apparent pattern in Figure 1 that we would like to explain is the transition between low and high SNR behavior. The transition occurs at smaller values of σ for larger values of M . Also, MSE/σ^2 can exceed 1, so in fact the sparse approximation procedure can *increase* the noise. We are not able to characterize the transition well for general frames. However, in Section III we obtain results for large frames that are generated by choosing vectors uniformly at random from the unit sphere in \mathbb{R}^N . There we get a sharp transition between low and high SNR behavior.

II. RATE-DISTORTION ANALYSIS AND LOW SNR BOUND

We now establish a bound on the performance of sparse approximation denoising that applies for any dictionary Φ . This bound qualitatively explains the low-SNR performance (right-hand side asymptotes) shown in Figure 1.

A. Sparse Approximation of a Gaussian Source

Before addressing the denoising performance of sparse approximation, we give an *approximation* result for Gaussian signals. This result is a *lower bound* on the MSE when sparsely approximating a Gaussian signal; it is the basis for an *upper bound* on the MSE for denoising when the SNR is low.

Theorem 1: Let Φ be an M -element dictionary, let $J = \binom{M}{K}$, and let $v \in \mathbb{R}^N$ have the distribution $\mathcal{N}(\bar{v}, \sigma^2 I_N)$. If \hat{v} is the optimal K -sparse approximation of v with respect to Φ , then

$$\frac{1}{N} \mathbf{E} [\|v - \hat{v}\|^2] \geq \sigma^2 \frac{c_1}{1 - c_1} \left(1 - \frac{K}{N}\right) \quad (2)$$

where

$$c_1 = J^{-2/(N-K)} \left(\frac{K}{N}\right)^{K/(N-K)}.$$

Remarks:

- (i) Theorem 1 shows that for any Φ , there is an approximation error lower bound that depends only on the frame

size M , the dimension of the signal N , and the dimension of the signal model K .

- (ii) As $M \rightarrow \infty$ with K and N fixed, $c_1 \rightarrow 0$. This is consistent with the fact that it is possible to drive the approximation error to zero by letting the dictionary grow.
- (iii) The decay of c_1 as M increases is slow. To see this, define a sparsity measure $\alpha = K/N$ and a redundancy factor $\rho = M/N$. Now using the approximation

$$\binom{\rho N}{\alpha N} \approx \left(\frac{\rho}{\alpha}\right)^{\alpha N} \left(\frac{\rho}{\rho - \alpha}\right)^{(\rho - \alpha)N},$$

we can compute the limit

$$\lim_{N \rightarrow \infty} c_1 = \left[\left(\frac{\alpha}{\rho}\right)^{2\alpha} \left(1 - \frac{\alpha}{\rho}\right)^{2(\rho - \alpha)} \alpha^\alpha \right]^{1/(1 - \alpha)}.$$

Thus the decay of the lower bound in (2) as ρ is increased behaves as $\rho^{-2\alpha/(1 - \alpha)}$. This is slow when α is small.

B. Bounds on Denoising MSE

To create an analogy between the approximation problem considered in Section II-A and the denoising problem, let $\bar{v} = x$, $v - \bar{v} = d$, and $v = y$. These correspondences fit perfectly, since $d \sim \mathcal{N}(0, \sigma^2 I_N)$ and we apply sparse approximation to y to get \hat{x}_{SA} . Theorem 1 gives the bound

$$\frac{1}{N} \mathbf{E} [\|y - \hat{x}_{\text{SA}}\|^2] \geq \sigma^2 \cdot \frac{c_1}{1 - c_1} \cdot \left(1 - \frac{K}{N}\right)$$

where c_1 is defined as before.

In the case that x and \hat{x}_{SA} are in the same subspace, $d - \hat{d}$ is orthogonal to \hat{d} so $\|d\|^2 = \|\hat{d}\|^2 + \|d - \hat{d}\|^2$. Thus knowing $\mathbf{E} [\|d\|^2] = N\sigma^2$ and having a lower bound on $\mathbf{E} [\|\hat{d}\|^2]$ immediately gives an upper bound on the MSE.

The interesting case is when x and \hat{x}_{SA} are not necessarily in the same subspace. Recalling that T is the index of the subspace selected in sparse approximation, orthogonally decompose d as $d = d_T \oplus d_{T^\perp}$ with d_T in the selected subspace and similarly decompose \hat{d} . Then $\hat{d}_T = d_T$ and the expected squared norm of this component can be bounded above as in the previous paragraph. Unfortunately, $\|\hat{d}_{T^\perp}\|$ can be larger than $\|d_{T^\perp}\|$ in proportion to $\|x\|$. The worst case is for $\|\hat{d}_{T^\perp}\| = 2\|d_{T^\perp}\|$, when y lies equidistant from the subspace of x and the subspace of \hat{x}_{SA} .

From this analysis we obtain the weak bound

$$\frac{1}{N} \mathbf{E} [\|x - \hat{x}_{\text{SA}}\|^2] \leq 4\sigma^2 \quad (3)$$

and the limiting low SNR bound

$$\frac{1}{N} \mathbf{E} [\|x - \hat{x}_{\text{SA}}\|^2]_{|x=0} \leq \sigma^2 \cdot \left(1 - \frac{c_1}{1 - c_1} \cdot \left(1 - \frac{K}{N}\right)\right). \quad (4)$$

III. ANALYSIS FOR ISOTROPIC RANDOM FRAMES

In general, the performance of sparse approximation denoising is given by

$$\frac{1}{N} \int_{R^N} \left\| x - \left(\underset{\hat{x} \in \Phi_K}{\operatorname{argmin}} \|x + \eta - \hat{x}\|_2 \right) \right\|^2 f(\eta) d\eta$$

where $f(\cdot)$ is the density of the noise d . While this expression does not give any fresh insight, it does remind us that the performance depends on every element of Φ . In this section, we improve greatly upon (3) with an analysis that depends on each dictionary element being an independent random vector and on the dictionary being large. The results are expectations over both the noise d and the dictionary itself. In addition to analyzing the MSE, we also analyze the probability of error in the subspace selection, *i.e.*, the probability that x and \hat{x}_{SA} lie in different subspaces. In light of the simulations reported in [4], [5], we expect these analyses to qualitatively match the performance of a variety of dictionaries.

A. Modeling Assumptions

This section specifies the precise modeling assumptions in analyzing denoising performance with large, isotropic random frames. Though the results are limited to the case of $K = 1$, the model is described for general K . Difficulties in extending the results to general K are described in the concluding comments of the paper. While many practical problems involve $K > 1$, the analysis of the $K = 1$ case presented here illustrates a number of unexpected qualitative phenomena, some of which have been observed for higher values of K .

The model (including the definition of \hat{x}_{SA}) is unchanged from earlier in the paper except that the dictionary Φ and signal x are random:

- (a) *Dictionary generation:* The dictionary Φ consists of M i.i.d. random vectors uniformly distributed on the unit sphere in \mathbb{R}^N .
- (b) *Signal generation:* The true signal x is a linear combination of the first K dictionary elements so that

$$x = \sum_{i=1}^K \alpha_i \varphi_i,$$

for some random coefficients $\{\alpha_i\}$. The coefficients $\{\alpha_i\}$ are independent of the dictionary except that x is normalized to have $\|x\|^2 = N$ for all realizations of the dictionary and coefficients.

- (c) *Noise:* The noisy signal y is given by $y = x + d$ where, as before, $d \sim \mathcal{N}(0, \sigma^2 I_N)$. The noise d is independent of Φ and x . We will let $\gamma = 1/\sigma^2$, which is the input SNR because of the scaling of x .

For the special case when M and N are large and $K = 1$, we will estimate two quantities:

Definition 1: The *subspace selection error probability* p_{err} is defined as

$$p_{\text{err}} = \Pr(T \neq j_{\text{true}}), \quad (5)$$

where T is the subspace selection index and j_{true} is the index of the subspace containing the true signal x , *i.e.*, j_{true} is the index of the subset $\{1, 2, \dots, K\}$.

Definition 2: The *normalized expected MSE* is defined as

$$E_{\text{MSE}} = \frac{1}{N\sigma^2} \mathbf{E} [\|x - \hat{x}_{\text{SA}}\|^2] = \frac{\gamma}{N} \mathbf{E} [\|x - \hat{x}_{\text{SA}}\|^2]. \quad (6)$$

The term ‘‘expected MSE’’ emphasizes that the expectation in (6) is over not just d , but also Φ and x .

B. Analyses of Subspace Selection Error and MSE

The first result is an easy-to-compute approximation for the subspace selection error probability. While the result is only proven for the case of $K = 1$, K is left in the expressions to indicate the precise role of this parameter.

Theorem 2: Consider the model described in Section III-A. When $K = 1$ and M and N are large, the subspace selection error probability defined in (5) is approximated well by

$$\hat{p}_{\text{err}} = 1 - \int_0^\infty f_r(u) \exp\left(-\left(\frac{Cau}{1+au}\right)^r\right) du, \quad (7)$$

where

$$C = \left(\frac{J-1}{r\beta(r,s)}\right)^{1/r}, \quad r = \frac{N-K}{2}, \quad s = \frac{K}{2}, \quad (8)$$

$$a = \frac{(N-K)\sigma^2}{N} = \frac{N-K}{N\gamma}, \quad (9)$$

$f_r(u)$ is the probability distribution

$$f_r(u) = r^r \Gamma(r) u^{r-1} e^{-ru}, \quad u \in [0, \infty), \quad (10)$$

$\beta(r, s)$ is the beta function, and $\Gamma(r)$ is the Gamma function.

It is interesting to evaluate \hat{p}_{err} in two limiting cases. First, suppose that $J = 1$. This corresponds to the situation where there is only one subspace. In this case, $C = 0$ and (7) gives $\hat{p}_{\text{err}} = 0$. This is expected, since with one subspace there is no chance of a subspace selection error.

At the other extreme, suppose that N , K , and γ are fixed and $M \rightarrow \infty$. Then $C \rightarrow \infty$ and $\hat{p}_{\text{err}} \rightarrow 1$. This is expected since as the size of the frame increases, the number of possible subspaces increases and the probability of error increases.

The next result approximates the normalized expected MSE with a double integral. The integrand is relatively simple to evaluate and decays quickly as $\rho \rightarrow \infty$ and $u \rightarrow \infty$ so numerically approximating the double integral is not difficult.

Theorem 3: Consider the model described in Section III-A. When $K = 1$ and M and N are large, the normalized expected MSE defined in (6) is given approximately by

$$\hat{E}_{\text{MSE}} = \frac{K}{N} + \int_0^\infty \int_0^\infty f_r(u) g_r(\rho) F(\rho, u) d\rho du, \quad (11)$$

where $g_r(\rho)$ is the probability distribution

$$g_r(\rho) = rC^r r^{r-1} \exp(-C\rho)^r, \quad (12)$$

$$F(\rho, u) = \begin{cases} \gamma(au(1-\rho) + \rho), & \text{if } \rho(1+au) < au; \\ 0, & \text{otherwise,} \end{cases}$$

and $f_r(u)$, C , r , and a are defined in (8)–(10).

C. Numerical Examples

We now present simulation results to examine the accuracy of the approximations in Theorems 2 and 3. Three pairs of (N, M) values were used: (5,1000), (10,100), and (10,1000). For each integer SNR from -10 dB to 35 dB, the subspace selection and normalized MSE were measured for 5×10^5 independent experiments. The resulting empirical probabilities of subspace selection error and normalized expected MSEs are

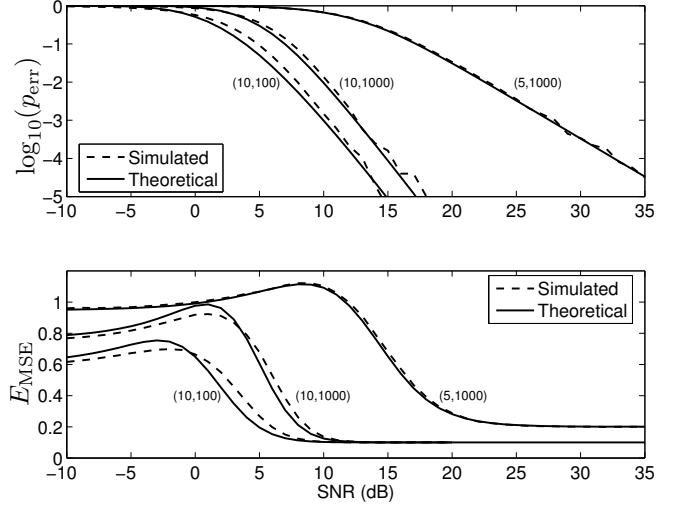


Fig. 2. Simulation of subspace selection error probability and normalized expected MSE for isotropic random dictionaries. Calculations were made for integer SNRs (in dB), with 5×10^5 independent simulations per data point. In all cases $K = 1$. The curve pairs are labeled by (N, M) . Simulation results are compared to the estimates from Theorems 2 and 3.

shown in Figure 2. Plotted alongside the empirical results are the estimates \hat{p}_{err} and \hat{E}_{MSE} from (7) and (11).

Comparing the theoretical and measured values in Figure 2, we see that the theoretical values match the simulation closely over the entire SNR range. Also note that the bottom panel of Figure 2 shows qualitatively the same behavior as Figure 1 (the direction of the horizontal axis is reversed). In particular, $E_{\text{MSE}} \approx \frac{K}{N}$ for high SNR and the low SNR behavior depends on M and N as described by (4).

D. Asymptotic Analysis

The estimates \hat{p}_{err} and \hat{E}_{MSE} are not difficult to compute numerically, but the expressions (7) and (11) provide little direct insight. It is thus interesting to examine the asymptotic behavior of \hat{p}_{err} and \hat{E}_{MSE} as N and M grow. The following theorem gives an asymptotic expression for the limiting value of the error probability function.

Theorem 4: Consider the function $\hat{p}_{\text{err}}(N, M, K, \gamma)$ defined in (7). For a fixed $K = 1$, γ , and C ,

$$\lim_{\substack{N, M \rightarrow \infty \\ C \text{ constant}}} \hat{p}_{\text{err}} = \begin{cases} 1, & \text{if } \gamma < \gamma_{\text{crit}}; \\ 0, & \text{if } \gamma > \gamma_{\text{crit}}, \end{cases}$$

where γ_{crit} is the *critical SNR*,

$$\gamma_{\text{crit}} = C - 1 = \left(\frac{J-1}{r\beta(r,s)}\right)^{1/r} - 1. \quad (13)$$

The theorem shows that, asymptotically, there is a critical SNR γ_{crit} above which the error probability goes to one and below which the probability is zero. Thus, even though the frame is random, the error event asymptotically becomes deterministic. A similar result holds for the asymptotic MSE.

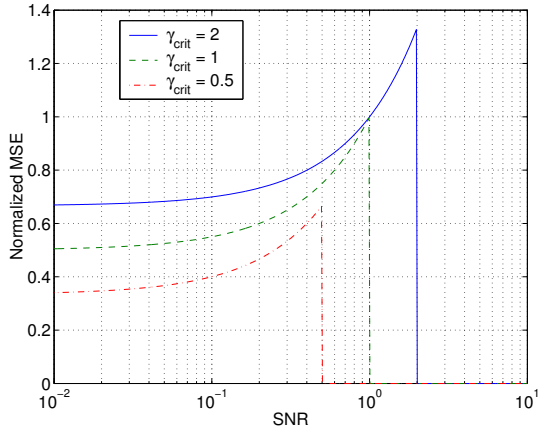


Fig. 3. Asymptotic normalized MSE as $N \rightarrow \infty$ (from Theorem 5) for various critical SNRs γ_{crit} .

Theorem 5: Consider the function $\hat{E}_{\text{MSE}}(M, N, K, \gamma)$ defined in (11). For fixed K , γ , and C ,

$$\lim_{\substack{N, M \rightarrow \infty \\ C \text{ constant}}} \hat{E}_{\text{MSE}} = \begin{cases} \hat{E}_{\text{lim}}(\gamma), & \text{if } \gamma < \gamma_{\text{crit}}; \\ 0, & \text{if } \gamma > \gamma_{\text{crit}}, \end{cases} \quad (14)$$

where γ_{crit} is given in (13) and $\hat{E}_{\text{lim}}(\gamma) = \frac{\gamma + \gamma_{\text{crit}}}{1 + \gamma_{\text{crit}}}$.

The asymptotic normalized MSE is plotted in Figure 3 for various values of the critical SNR γ_{crit} . When $\gamma > \gamma_{\text{crit}}$, the normalized MSE is zero. This is expected: from Theorem 4, when $\gamma > \gamma_{\text{crit}}$, the estimator will always pick the correct subspace. We know that for a fixed subspace estimator, the normalized MSE is K/N . Thus, as $N \rightarrow \infty$, the normalized MSE approaches zero.

What is perhaps surprising is the behavior for $\gamma < \gamma_{\text{crit}}$. In this regime, the normalized MSE actually *increases* with increasing SNR. At the critical level, $\gamma = \gamma_{\text{crit}}$, the normalized MSE approaches its maximum value

$$\max \hat{E}_{\text{lim}} = \frac{2\gamma_{\text{crit}}}{1 + \gamma_{\text{crit}}}.$$

When $\gamma_{\text{crit}} > 1$, the limit of the normalized MSE $\hat{E}_{\text{lim}}(\gamma)$ satisfies $\hat{E}_{\text{lim}}(\gamma) > 1$. Consequently, the sparse approximation results in noise *amplification* instead of noise reduction. In the worst case, as $\gamma_{\text{crit}} \rightarrow \infty$, $\hat{E}_{\text{lim}}(\gamma) \rightarrow 2$. Thus, sparse approximation can result in a noise amplification by a factor as large as 2. Contrast this with the factor of 4 in (3), which seems to be a very weak bound.

IV. COMMENTS AND CONCLUSIONS

This paper has addressed properties of denoising by sparse approximation that are geometric in that the signal model is membership in a specified union of subspaces, without a probability density on that set. The denoised estimate is the feasible signal closest to the noisy observed signal.

The first main result (Theorem 1) is a bound on the performance of sparse approximation applied to a Gaussian

signal. This lower bound on mean-squared approximation error is used to determine an upper bound on denoising MSE in the limit of low input SNR.

The remaining results apply to the expected performance when the dictionary itself is isotropically random. Easy-to-compute estimates for the probability that the subspace containing the true signal is selected and for the MSE are given (Theorems 2 and 3). Unfortunately, these results are proven only for the case of $K = 1$. The main technical difficulty in extending these results to general K is that the distances to the various subspaces are not mutually independent [5].

Asymptotic analysis ($N \rightarrow \infty$) of the situation with a random dictionary reveals a critical value of the SNR (Theorems 4 and 5). Below the critical SNR, the probability of selecting the subspace containing the true signal approaches zero and the expected MSE approaches a constant with a simple, closed form; above the critical SNR, the probability of selecting the subspace containing the true signal approaches one and the expected MSE approaches zero.

Sparsity with respect to a randomly generated dictionary is a strange model for naturally-occurring signals. However, most indications are that a variety of dictionaries lead to performance that is qualitatively similar to that of random dictionaries. Also, sparsity with respect to randomly generated dictionaries occurs when the dictionary elements are produced as the random instantiation of a communication channel. Both of these observations require further investigation.

REFERENCES

- [1] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1992.
- [2] D. L. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inform. Th.*, Feb. 2004, submitted.
- [3] A. K. Fletcher and K. Ramchandran, "Estimation error bounds for frame denoising," in *Proc. Wavelets: Appl. in Sig. & Image Proc. X, part of SPIE Int. Symp. on Optical Sci. & Tech.*, vol. 5207, San Diego, CA, Aug. 2003, pp. 40–46.
- [4] A. K. Fletcher, S. Rangan, V. K. Goyal, and K. Ramchandran, "Denoising by sparse approximation: Error bounds based on rate-distortion theory," Univ. California, Berkeley, Electron. Res. Lab. Memo., Sept. 2004.
- [5] A. K. Fletcher, "Estimation via sparse approximation: Error bounds and random frame analysis," Master's thesis, Univ. California, Berkeley, May 2005.
- [6] 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.
- [7] 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.
- [8] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inform. Th.*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
- [9] —, "Just relax: Convex programming methods for subset selection and sparse approximation," Univ. of Texas at Austin, ICES Report 0404, Feb. 2004.
- [10] V. K. Goyal, M. Vetterli, and N. T. Thao, "Quantized overcomplete expansions in \mathbb{R}^N : Analysis, synthesis, and algorithms," *IEEE Trans. Inform. Th.*, vol. 44, no. 1, pp. 16–31, Jan. 1998.
- [11] M. Elad and M. Zibulevsky, "A probabilistic study of the average performance of basis pursuit," *IEEE Trans. Inform. Th.*, Dec. 2004, submitted.
- [12] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc.: Packings in Grassmannian spaces," *Experimental Mathematics*, vol. 5, no. 2, pp. 139–159, 1996.