COMPRESSIVE SAMPLING FOR SIGNAL DETECTION

Jarvis Haupt and Robert Nowak

University of Wisconsin - Madison
Department of Electrical and Computer Engineering
jdhaupt@wisc.edu, nowak@ece.wisc.edu

ABSTRACT

Compressive sampling (CS) refers to a generalized sampling paradigm in which observations are inner products between an unknown signal vector and user-specified test vectors. Among the attractive features of CS is the ability to reconstruct any sparse (or nearly sparse) signal from a relatively small number of samples, even when the observations are corrupted by additive noise. However, the potential of CS in other signal processing applications is still not fully known. This paper examines the performance of CS for the problem of signal detection. A *generalized restricted isometry property* (GRIP) is introduced, which guarantees that angles are preserved, in addition to the usual norm preservation, by CS. The GRIP is leveraged to derive error bounds for a CS matched filtering scheme, and to show that the scheme is robust to signal mismatch.

Index Terms— Signal detection

1. INTRODUCTION

Compressive sampling (CS) is a generalization of conventional point-sampling in which samples are inner products between an unknown signal vector and a set of user-defined test vectors. Recent theoretical results establish CS as a universal sampling procedure, in the sense that for certain random ensembles of test vectors, CS is an effective way to encode the salient information in *any* sparse (or nearly sparse) signal. Further, these projection samples can be used to estimate the unknown signal to a controllable mean-squared error, even in the presence of noise [1–4]. These results are remarkable since the number of samples required for reconstruction can be far fewer than the ambient dimension in which the signal vector is observed. For this reason CS has been proposed as a viable candidate in many practical applications, such as wideband communications monitoring systems, where the goal is to detect and/or intercept communication signals over a frequency range so large that conventional Nyquist sampling is technologically impossible.

In this paper, we consider a problem in which the goal is to reliably detect the presence of a signal vector from observations corrupted by additive noise. We consider a generalized sampling model where the observations are described by inner products between the unknown vector and user-specified test vectors, subject to the condition that the number of observations is much less than the ambient signal dimension. On one hand, given *a priori* knowledge of the target signal structure and accurate characterization of the noise, optimal sampling schemes can be designed based on classical matched filtering results. For example, if the noise is Gaussian and independent across observations then matching the test vectors to the signal is the best approach. Additionally, this "matched sampling" is robust to signal mismatch in the sense that if the target signal is only approximately known, the achievable error performance degrades gracefully as a function of the quality of the approximation. Despite these desirable properties, one obvious problem with this approach is that since it requires specialized observations (matched to a given signal), the observations cannot be reused effectively, for example, to detect whether a second (different) signal is present.

In contrast, we propose a detector that collects a set of universal samples, obtained without prior knowledge of the signal structure. We first establish that any signal can be detected from such samples, but with lower reliability than in the known signal case (matched filtering). However, assuming that at a later time through some auxiliary channel, information about the nature of a signal of interest (which may have been present when the samples were collected) is obtained, a stronger result is possible. Specifically, we show that universal samples, together with this "future knowledge," can be used in a matched filtering approach to reliably detect *any* sparse signals with error performance and robustness comparable to the ideal case.

The fundamental idea of CS matched filtering has arisen previously in the study of various dimensionality reduction problems. For example, in [5], an investigation of projection sampling to speed up kernel methods led to a bound on the difference between the inner product of two vectors and the inner product between their projections by what is essentially a CS ensemble. More recently, the authors of [6] arrive at the same result in bounding the performance of a scheme which estimates linear functions from CS observations. For our purposes we formulate a *generalized restricted isometry property* (GRIP) with which we will establish the stated results. In addition, we point out that GRIP may provide additional discriminating power compared to these existing results, and may provide a means to analyze the performance of CS detection in the presence of a strongly correlated interferer.

To clarify the exposition in the following sections, we make a brief summary of notation here. We use bold-face capital letters ($\boldsymbol{A}$) to denote matrices. Vectors will be written using bold-face lowercase letters ($\boldsymbol{f}$) or superscripted upper-case letters (*e.g.*, $\boldsymbol{A}^{(j)}$, which denotes the jth row vector of the matrix $\boldsymbol{A}$), and the ith component of $\boldsymbol{f} \in \mathbb{R}^n$ is given by $\boldsymbol{f}(i)$ for $i \in \{1, \ldots, n\}$. We define the support of a vector as $\text{supp}(\boldsymbol{f}) \triangleq \{i \in \{1, \ldots, n\} : \boldsymbol{f}(i) \neq 0\}$, and when $|\text{supp}(\boldsymbol{f})| \leq m \in \mathbb{N}$ we say that $\boldsymbol{f}$ is m-sparse. The inner product of two vectors $\boldsymbol{f}, \boldsymbol{g} \in \mathbb{R}^n$ is given by $\langle \boldsymbol{f}, \boldsymbol{g} \rangle \triangleq \sum_{i=1}^{n} \boldsymbol{f}(i)\boldsymbol{g}(i)$. The notation $\|\boldsymbol{f}\|^2 \triangleq \langle \boldsymbol{f}, \boldsymbol{f} \rangle$ denotes the standard Euclidean distance, also called the energy of $\boldsymbol{f}$.

The remainder of the paper is organized as follows. The sampling model is defined in Section 2, which also provides an overview of CS reconstruction theory and the notion of restricted isometry. The ideal detector is examined in Section 3.1, a universal detec-

This research was supported by the NSF under grant CCR-0350213, and by the DARPA Analog-to-Information program.

tor using CS is analyzed in Section 3.2, and our proposed future-knowledge sparse signal detection scheme is presented and analyzed in Section 3.3. Simulation results are presented and discussed in Section 4, and some concluding remarks are made in Section 5.

2. PRELIMINARIES

Suppose that an observer is allowed to make only a limited number of noisy observations of an unknown vector $\boldsymbol{f} \in \mathbb{R}^n$, where each observation is the inner product between the signal vector $\boldsymbol{f}$ and a sampling vector chosen by and known to the observer. These *projection samples* are described by

$$y_j = \left\langle \boldsymbol{A}^{(j)}, \boldsymbol{f} \right\rangle + w_j, \tag{1}$$

for $j = 1, \ldots, k$, where $k < n$, $\boldsymbol{A}^{(j)} \in \mathbb{R}^n$ are the sampling vectors, and $\{w_j\}$ is a collection of independent and identically distributed (i.i.d.) $\mathcal{N}(0, \sigma^2)$ noises. We will employ this sampling model, which is a natural generalization of conventional point sampling where each $\boldsymbol{A}^{(j)}$ would be a vector whose entries are all zero except for the entry corresponding to the desired sample location. In general, it is common to impose a unit energy restriction on the sampling vectors.

Compressive sampling (CS) refers to the case where the representation of $\boldsymbol{f}$ in some orthonormal basis (*e.g.*, wavelet) is m-sparse and the measurement vector ensemble $\{\boldsymbol{A}^{(j)}\}_{j=1}^k$ satisfies a special *restricted isometry property*, which is formally stated below. In this setting, an estimate $\widehat{\boldsymbol{f}}_k$ can be obtained from $\{y_j, \boldsymbol{A}^{(j)}\}_{j=1}^k$ that satisfies

$$\mathbb{E}\left[\frac{\|\widehat{\boldsymbol{f}}_k - \boldsymbol{f}\|^2}{n}\right] \leq C\left(\frac{k}{m \log n}\right)^{-1},$$

where C is a constant that depends on σ^2, and the expectation is over the distribution of the noise (and the test vectors if they are randomly generated) [3,4]. This result implies that CS nearly achieves the parametric estimation rate in expectation, and hence is almost as effective (within a log factor) as sampling the nonzero locations of $\boldsymbol{f}$ directly.

The success of CS reconstruction relies on the fact that certain sampling vector ensembles exhibit a degree of incoherence with *any* sparse signal. This roughly says that an ensemble of CS measurements preserves the distances between all sparse vectors. This concept is formalized in [7] using the *restricted isometry constant*, defined below.

Definition 1 (Restricted Isometry Constant) *Let $\boldsymbol{A}$ be a $k \times n$ matrix. For a subset $T \subset \{1, \ldots, n\}$, let $\boldsymbol{A}_T$ denote the $k \times |T|$ submatrix formed by retaining the columns indexed by the elements of T. The* restricted isometry constant *is the smallest number ϵ_S such that*

$$(1 - \epsilon_S)\frac{k}{n}\|\boldsymbol{f}_T\|^2 \leq \|\boldsymbol{A}_T \boldsymbol{f}_T\|^2 \leq (1 + \epsilon_S)\frac{k}{n}\|\boldsymbol{f}_T\|^2 \tag{2}$$

holds for all sets T with $|T| \leq S$ and all vectors $\boldsymbol{f}_T \in \mathbb{R}^{|T|}$.

This property implies that the matrix $\boldsymbol{A}$ approximately preserves the length of any m-sparse vector $\boldsymbol{f}$ where $m \leq S$. We say that a matrix $\boldsymbol{A}$ satisfies a *restricted isometry property* (RIP) of order S if (2) holds for some constant $\epsilon_S \in (0, 1)$. Prior work verified that matrices whose entries are i.i.d. realizations of certain random variables satisfy a RIP with constant $\epsilon_S = 1/2$ with very high probability when k is on the order of S [1,2].

The analysis in later sections uses a *generalized restricted isometry property* (GRIP), given by the following Theorem (see [8] for a complete proof).

Theorem 1 (GRIP) *If a matrix $\boldsymbol{A}$ satisfies satisfies RIP of order S with $\epsilon_S \leq 1/3$, then for any sparse vectors $\boldsymbol{f}$ and $\boldsymbol{g}$ supported on T and separated by an acute angle α (i.e., $\langle \boldsymbol{f}, \boldsymbol{g}\rangle = \|\boldsymbol{f}\|\|\boldsymbol{g}\|\cos(\alpha)$), then*

$$\begin{aligned}(1 - \epsilon_S)\frac{k}{n}&\|\boldsymbol{f}_T\|\|\boldsymbol{g}_T\|\cos\left[(1 + \theta_S)\alpha\right] \\ &\leq \langle \boldsymbol{A}_T\boldsymbol{f}_T, \boldsymbol{A}_T\boldsymbol{g}_T\rangle \leq (1 + \epsilon_S)\frac{k}{n}\|\boldsymbol{f}_T\|\|\boldsymbol{g}_T\|\cos\left[(1 - \theta_S)\alpha\right],\end{aligned}$$

where $\theta_S = \theta(\epsilon_S) = c\sqrt{\epsilon_S}$ with a small constant $c > 0$.

This result is established by showing that RIP implies relative angle preservation, which means that for two vectors $\boldsymbol{f}$ and $\boldsymbol{g}$ separated by an acute angle α, the image angle $\widehat{\alpha}$ between $\boldsymbol{A}\boldsymbol{f}$ and $\boldsymbol{A}\boldsymbol{g}$ satisfies $(1 - \theta)\alpha \leq \widehat{\alpha} \leq (1 + \theta)\alpha$, for a small value $\theta \in (0, 1)$.

The proof of Theorem 1 proceeds using a key observation from [9]. In that work, it is shown that approximate preservation of the lengths of the vectors $\{\boldsymbol{f}, \boldsymbol{g}, \boldsymbol{f} - \boldsymbol{g}\}$ (collectively, the triangle defined by the vectors $\boldsymbol{f}$ and $\boldsymbol{g}$) along with a set of properly-chosen "stabilizing vectors" is sufficient to guarantee that the heights of the triangle defined by $\boldsymbol{f}$ and $\boldsymbol{g}$ are approximately preserved. Relative angle preservation follows. The final step in establishing Theorem 1 is to show that RIP implies relative length preservation of all of the required stabilizing vectors.

It is worthwhile to comment that similar bounds to the ones we obtain in Section 3.3 can be derived using the results in [5,6]. However, for those results, relative angle preservation is only possible if ϵ in proportion to α. In contrast, GRIP guarantees relative angle preservation for a fixed ϵ. This key difference will be discussed again in Section 5, in the context of an interesting extension of the detection problem.

3. PROJECTION SAMPLING AND SIGNAL DETECTION

3.1. Signal Detection From Matched Samples

In accordance with the model defined in (1), we let $\boldsymbol{f} \in \mathbb{R}^n$ be the signal to be detected, $\boldsymbol{y} \in \mathbb{R}^k$ is the vector of observations, and $\boldsymbol{A}$ is a $k \times n$ matrix whose rows are the vectors $\boldsymbol{A}^{(j)}$. The observations when the signal is absent are noise only, independent, and distributed as

$$H_0 : P_0(y_j) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y_j^2}{2\sigma^2}\right),$$

and when the signal is present, the observations are independent and distributed as

$$H_1 : P_1(y_j) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\left(y_j - \langle \boldsymbol{A}^{\boldsymbol{j}}, \boldsymbol{f}\rangle\right)^2}{2\sigma^2}\right],$$

for $j = 1, \ldots, k$.

Using standard techniques from classical detection theory, it is straightforward to show that the optimal test statistic in this setting is of the form $\boldsymbol{y}^T\boldsymbol{A}\boldsymbol{f}$, which means that for some threshold τ, the detector decides that the signal is present if $\boldsymbol{y}^T\boldsymbol{A}\boldsymbol{f} > \tau$ and decides the signal is absent otherwise. Define the total probability of error P_{ERROR} to be the sum of the false alarm probability P_{FA} (the probability that the detector announces the signal is present when it is not) and the miss probability P_{M} (the probability that the detector announces the signal is not present when it is). We would like to optimize this total error by balancing the contributions from each error term. From [10], this optimal error obeys $P_{\text{ERROR}} \leq \prod_{j=1}^{k} \Gamma_j^*$, where $\Gamma_j^* = \min_{\lambda \in [0,1]} \Gamma_j(\lambda)$ and

$$\Gamma_j(\lambda) = \int_{-\infty}^{\infty} P_0^{\lambda}(y_j) P_1^{1-\lambda}(y_j) dx.$$

In this case, we have $P_{\text{ERROR}} \leq \exp\left(-\|\boldsymbol{A}\boldsymbol{f}\|^2/8\sigma^2\right)$. If the rows of $\boldsymbol{A}$ are given by $\widehat{\boldsymbol{f}}/\|\widehat{\boldsymbol{f}}\|$ for some approximation $\widehat{\boldsymbol{f}}$ such that $\langle \boldsymbol{f}, \widehat{\boldsymbol{f}} \rangle \geq 0$, then

$$P_{\text{ERROR}} \leq \exp\left[-knS\cos^2(\alpha)/8\right], \tag{3}$$

where α is the angle between $\boldsymbol{f}$ and $\widehat{\boldsymbol{f}}$, and $S = ||\boldsymbol{f}||^2/n\sigma^2$ is the signal to noise ratio (SNR). This result illustrates the robustness of the detector that employs matched sampling – the error performance degrades gracefully as the quality of the approximation decreases (*i.e.*, as α increases).

3.2. Arbitrary Signal Detection From Compressive Samples

We now consider the problem of detecting an arbitrary but unknown signal in the presence of additive white Gaussian noise. Again let $\boldsymbol{f} \in \mathbb{R}^n$ denote the signal vector. Since the signal is unknown we cannot match the samples to it, so instead we compressively sample. That is, we assume that the entries of the test functions $\boldsymbol{A}^{(j)}$ defined in (1) are i.i.d. $\mathcal{N}(0, 1/n)$. Under the null hypothesis, observations will be noise only, independent, and distributed as

$$H_0 : P_0(y_j) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y_j^2}{2\sigma^2}\right).$$

Because the sum of independent Gaussian random variables is itself Gaussian, the presence of the signal amounts to a variance shift in the observations, which under the alternative hypothesis are independent and distributed as

$$H_1 : P_1(y_j) = \frac{1}{\sqrt{2\pi\left(\sigma^2 + \|x\|^2/n\right)}} \exp\left(-\frac{y_j^2}{2\left(\sigma^2 + \|x\|^2/n\right)}\right),$$

for $j = 1, \ldots, k$. In this setting, it is straightforward to show that the optimal detector is an energy detector of the form $\|\boldsymbol{y}\|^2$. Again we follow the approach in [10] to obtain

$$P_{\text{ERROR}} \leq \left[\exp(1/2)\sqrt{\frac{\log(1+S)}{S(1+S)^{1/S}}}\right]^k. \tag{4}$$

This result shows that any unknown signal $\boldsymbol{f}$ can be reliably detected using CS without actual knowledge of $\boldsymbol{f}$.

3.3. Sparse Signal Detection From Compressive Samples

The previous two sections illustrated the performance of detectors on two ends of a spectrum. On one end, we see that certain ensembles of test functions are sufficient to detect arbitrary but unknown signals with a rate that is exponential in the number of observations, but only polynomial in SNR, as shown in (4). On the other hand, some useful knowledge of the signal to be detected results in a robust detector whose error probability decays exponentially in the number of observations *and* in SNR as shown in (3). Here we examine a CS matched filtering approach that exploits the sparsity of the signal to be detected and some amount of future knowledge to offer both universality and robustness, as well as error an probability that is exponential in both the number of observations and the SNR.

We proceed by utilizing the form of the ideal detector, but instead of correlating the observation vector with the projection of the true signal ($\boldsymbol{y}^T\boldsymbol{A}\boldsymbol{f}$) we instead correlate the observation vector with the projection of an approximation of the true signal. Specifically, let $\widehat{\boldsymbol{f}}$ denote our approximation of $\boldsymbol{f}$ obtained through a "future knowledge" channel. We will examine the error performance of the detector given by $\boldsymbol{y}^T\boldsymbol{A}\widehat{\boldsymbol{f}}$. Again, we assume that the approximation $\widehat{\boldsymbol{f}}$ is aligned with the true signal $\boldsymbol{f}$ so that $\langle \boldsymbol{f}, \widehat{\boldsymbol{f}} \rangle \geq 0$ (*i.e.*, α is acute).

We examine the error performance by considering the distribution of the test statistic $T = \boldsymbol{y}^T\boldsymbol{A}\widehat{\boldsymbol{f}}$ directly. Under the null hypothesis, $T \sim \mathcal{N}\left(0, ||\boldsymbol{A}\widehat{\boldsymbol{f}}||^2\sigma^2\right)$, and when the signal is present, $T \sim \mathcal{N}\left(\langle \boldsymbol{A}\boldsymbol{f}, \boldsymbol{A}\widehat{\boldsymbol{f}} \rangle, ||\boldsymbol{A}\widehat{\boldsymbol{f}}||^2\sigma^2\right)$. Thus, the error probabilities are obtained by integrating tails of the appropriate Gaussian distributions. Now assume that the vectors $\boldsymbol{f}$ and $\widehat{\boldsymbol{f}}$ are sufficiently sparse and that the matrix $\boldsymbol{A}$ satisfies GRIP with constants ϵ and θ. We use the distributions that give the worst separation and error behavior by choosing those with the lowest mean and highest variance. Thus, under the null hypothesis we have $T \sim \mathcal{N}\left(0, (1+\epsilon)(k/n)||\widehat{\boldsymbol{f}}||^2\sigma^2\right)$, and when the signal is present, $T \sim \mathcal{N}\left((1-\epsilon)(k/n)\|\boldsymbol{f}\|\|\widehat{\boldsymbol{f}}\|\cos\left[(1+\theta)\alpha\right], (1+\epsilon)(k/n)||\widehat{\boldsymbol{f}}||^2\sigma^2\right)$, where α denotes the acute angle between $\boldsymbol{f}$ and $\widehat{\boldsymbol{f}}$. For a positive detection threshold $\tau < (1-\epsilon)(k/n)\|\boldsymbol{f}\|\|\widehat{\boldsymbol{f}}\|\cos\left[(1+\theta)\alpha\right]$ we utilize a bound on the the integral of Gaussian tails [11] to obtain

$$P_{\text{FA}} \leq \frac{1}{2}\exp\left[\frac{-\tau^2}{2\sigma^2(1+\epsilon)(k/n)||\widehat{\boldsymbol{f}}||^2}\right]$$

and

$$P_{\text{M}} \leq \frac{1}{2}\exp\left[\frac{-\left((1-\epsilon)(k/n)\|\boldsymbol{f}\|\|\widehat{\boldsymbol{f}}\|\cos\left[(1+\theta)\alpha\right]-\tau\right)^2}{2\sigma^2(1+\epsilon)(k/n)||\widehat{\boldsymbol{f}}||^2}\right].$$

Setting $\tau = (1/2)(1-\epsilon)(k/n)\|\boldsymbol{f}\|\|\widehat{\boldsymbol{f}}\|\cos\left[(1+\theta)\alpha\right]$ makes the two errors equal, giving

$$\begin{aligned} P_{\text{ERROR}} &\leq \exp\left[-\frac{(1-\epsilon)^2(k/n)\|\boldsymbol{f}\|^2\cos^2\left[(1+\theta)\alpha\right]}{8\sigma^2(1+\epsilon)}\right] \\ &= \exp\left[-\frac{kS}{8}\,\frac{(1-\epsilon)^2\cos^2\left[(1+\theta)\alpha\right]}{1+\epsilon}\right], \end{aligned}$$

valid for all angles for which $\tau = (1/2)(1-\epsilon)(k/n)\|\boldsymbol{f}\|\|\widehat{\boldsymbol{f}}\|\cos\left[(1+\theta)\alpha\right]$ is positive.

When ϵ and θ are small, this bound is approximately given by $P_{\text{ERROR}} \leq \exp\left[-kS\cos^2(\alpha)/8\right]$. Comparing this approximation with (3), we see that the only difference is a factor of n in the rate. This arises because the amount of signal energy captured by each CS observation of the signal $\boldsymbol{f}$, in expectation, is given by $\|f\|^2/n$, while optimal sampling captures the full signal energy $\|f\|^2$. The loss by a factor of n illustrates the price to be paid for the universality that CS provides.

4. SIMULATION RESULTS

In this section we examine the dependence of the minimum probability of error on the number of observations and SNR for the detectors that employ CS. To illustrate the robustness of the CS matched filter detector we consider two approximations, the best one-term approximation of $\boldsymbol{f}$ and $\boldsymbol{f}$ itself. For the realizations used in the simulations, the quality of the one term approximation $\widehat{\boldsymbol{f}}$ is such that $\cos(\alpha) = 0.85$.

In the first simulation, $n = 1000$ and $\boldsymbol{f} \in \mathbb{R}^n$ is a vector with three nonzero entries chosen at random, but normalized so that $\|\boldsymbol{f}\|^2 = 1$. For each detector, and for each value of k (the number of observations), 4000 realizations of the test statistic for the signal-absent and signal-present cases were computed, each using a randomly generated sampling matrix [with i.i.d. $\mathcal{N}(0, 1/n)$ entries] and noise vector [with i.i.d. $\mathcal{N}(0, 0.025)$ entries]. For each case, we used the empirical histograms of the distributions under H_0 and H_1 to computed the minimum total error probability.

The results are shown in Fig. 1, where the curves drawn with diamond markers indicate the theoretical bounds, and the curves with x markers show the simulation results. Errors corresponding to the energy detector, CS matched filter with approximate knowledge, and CS matched filter with exact knowledge are drawn using the dash-dot, dashed, and solid lines, respectively. We see that the empirical results agree with the theory.

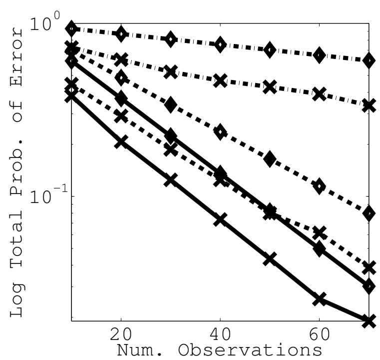

Fig. 1. Empirical probability of error vs. number of observations for a fixed SNR.

The second simulation illustrates the scaling behavior with respect to SNR. Again, $n = 1000$, and this time the number of observations was fixed at 100. For each value of σ^2, 1000 realizations of sampling matrices with i.i.d. $\mathcal{N}(0, 1/n)$ entries and vectors of i.i.d. $\mathcal{N}(0, \sigma^2)$ additive noise were generated, and test statistics for each hypothesis were computed. The minimum average probability of error was again determined empirically. The results depicted in Fig. 2 illustrate that the theoretical results again agree with the predicted error behavior. It is interesting to note that the dependence of the detector performance on the quality of approximation is noticeable in both simulations – the better approximation yields faster error decay.

5. CONCLUSIONS AND EXTENSIONS

The theory and simulations above showed that a universal CS matched filtering scheme is an effective way to detect *any* sparse signal with both the robustness and the exponential error performance of the ideal detector. A relevant extension of the problem examined here would be to determine the performance of CS matched filtering in the presence of an interferer ξ that is close to the true signal (*i.e.*, the angle between the signal and interferer is small, but nonzero). The GRIP may provide additional insight into this problem because it guarantees a relative bound on projected angles, so that for a fixed

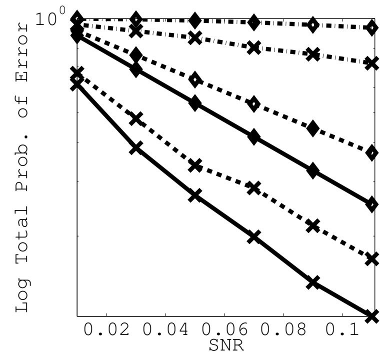

Fig. 2. Empirical probability of error vs. SNR for a fixed number of observations.

ϵ, the images of the signal and interferer will remain angularly separated. In contrast, the existing bounds [5,6] imply additive deviation bounds on the projected angle, and would require decreasing ϵ in proportion to α to guarantee angular separation. A complete analysis of this case is left for future work.

6. REFERENCES

[1] E. 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.

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

[3] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 4036–4048, Sept. 2006.

[4] E. Candès and T. Tao, "The Dantzig selector: statistical estimation when p is much larger than n," *Ann. Stat.*, 2005, submitted.

[5] D. Achlioptas, F. McSherry, and B. Schölkopf, "Sampling techniques for kernel methods," in *Annual Advances in Neural Inform. Processing Systems 14*, 2002, pp. 335–342.

[6] M. Davenport, M. Wakin, and R. Baraniuk, "The compressive matched filter," Tech. Rep. TREE 0610, Rice University, 2006.

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

[8] J. Haupt and R. Nowak, "A generalized restricted isometry property," Tech. Rep. ECE-07-1, University of Wisconsin - Madison, 2007.

[9] A. Magen, "Dimensionality reductions that preserve volumes and distance to affine spaces, and their algorithmic applications," in *Proc. of the 6th Intl. Workshop on Randomization and Approximation Techniques (RANDOM '02)*, London, UK, 2002, pp. 239–253.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.

[11] S. Verdu, *Multiuser Detection*, UK: Cambridge University Press, 1998.