DIFFERENCES BETWEEN OBSERVATION AND SAMPLING ERROR IN SPARSE SIGNAL RECONSTRUCTION

Galen Reeves and Michael Gastpar

Department of Electrical Engineering and Computer Sciences, UC Berkeley

ABSTRACT

The field of Compressed Sensing has shown that a relatively small number of random projections provide sufficient information to accurately reconstruct sparse signals. Inspired by applications in sensor networks in which each sensor is likely to observe a noisy version of a sparse signal and subsequently add sampling error through computation and communication, we investigate how the distortion differs depending on whether noise is introduced before sampling (observation error) or after sampling (sampling error). We analyze the optimal linear estimator (for known support) and an ℓ_1 constrained linear inverse (for unknown support). In both cases, observation noise is shown to be less detrimental than sampling noise and low sampling rates. We also provide sampling bounds for a non-stochastic ℓ_∞ bounded noise model.

Index Terms— compressed sensing, random matrices, sparsity, sensor networks, ℓ_1-minimization,

1. INTRODUCTION

Recently, researchers have shown that accurate reconstruction of high dimensional signals is possible from a relatively small number of random projections. Initial work in this area [1, 2], termed *compressed sensing*, showed that if a signal $x \in \mathbb{R}^n$ is k-sparse (i.e. is non-zero only on a set K with $|K| = k$) then it can be recovered perfectly using $m \ll n$ noiseless linear samples of the form $y_i = \langle \phi_i, x\rangle$ for $i = 1, \cdots, m$, where ϕ_i are random vectors uniformly distributed on the $(n-1)$ dimensional unit sphere. These results are possible with $m = \mathcal{O}(n \log(n/k))$ and no prior information on the location of the support K.

Ensuing work has addressed how things change when each sample is corrupted by some small noise. For the task of support recovery, [3] shows that m must be super-linear in either n or k. For signal recovery, [4, 5] show that near optimal signal reconstruction is possible with $m = \mathcal{O}(k \log(n))$, and [6, 7] show that more general scalings admit reconstructions that are stable in the ℓ_2 norm, that is their reconstruction error is within a factor of the noise magnitude.

This paper is inspired by the application of compressed sensing for use in sensor networks. In [8, 9] the authors describe setups in which n sensors sample some compressible phenomena x. Then, m random projections of x are computed over the network and communicated to a receiver such that each sample picks up some small sampling noise $w_i, i = 1, \cdots m$.

We investigate how things differ in the presence of what we call observation noise, that is small noise components added to the phenomena x_j prior to being observed at each of the n sensors. It is tempting to argue that noise added to the signal prior to sampling can be "pushed through" the sampling process and equivalently be considered as noise added after the sampling process. In this paper, we show that such a consideration is inappropriate.

In Section 2, we determine the distortion of the optimal linear estimator when the support K is known and all non-zero signals have i.i.d. elements. In this setting, the effects of observation noise are *correlated* and exist in the *same subspace* as x. For $m < n$ samples, we verify that the correlation renders observation noise less detrimental than sampling noise. However, for $m > n$, the distortion due to sampling error continues to decrease whereas the distortion due to observation error is necessarily constant.

In many interesting situations, such as sensor networks, the support K is not known a priori. In Section 3, we analyze the reconstruction performance of an ℓ_1 constrained linear inverse, termed Basis Pursuit, under this setting. We present upper bounds on the distortion for both the observation and the sampling noise model. These bounds suggest that observation noise is less detrimental than sampling noise, which is confirmed to some degree by numerical simulations.

We address the problem of support recovery in Section 4 for the situation where the each element of the observation noise is bounded in magnitude and give a set of necessary scalings which permit perfect recovery.

1.1. Sampling Model

In the most general sampling model considered in this paper, a k-sparse signal $x \in \mathbb{C}^n$ is observed via

$$y = \Phi(x + w_o) + w_s \tag{1}$$

where w_o is some observation error applied directly to the signal x, and w_s is some sampling error. The entries of the $m \times n$ sampling matrix Φ are i.i.d. circularly symmetric complex Gaussians with zero mean and variance[1] $1/n$. The support K is uniformly distributed over the $\binom{n}{k}$ possibilities. For a vector or matrix A the notation A_K refers to the part of A supported on K, and and the notation K^c refers to the complement of K. We analyze the effects of the two kinds of noise separately by comparing two simplified versions of (1), one with only sampling noise, denoted y_s, and one with only observation noise, denoted y_o.

2. OPTIMAL LINEAR ESTIMATOR OF X_K GIVEN K

In this section, we study the problem of recovering x_K when K is known. This is interesting when K can be accurately determined

The work of G. Reeves and M. Gastpar was supported in part by ARO MURI No. W911NF-06-1-0076

[1] This is in contrast to other work in which the matrix elements typically have variance $1/m$. Our choice is made so that $E||\Phi w_0|| = E||w_s||$.

or when K is known a priori but there is no control over how the samples are taken. Additionally, the analysis of this case shows the fundamental differences between the types of noise regardless of uncertainty in the support.

We consider signals x_K, w_s, and w_o to have i.i.d. elements with variances σ_x^2 and σ_w^2 respectively. Because we chose our sampling matrix to have unit normed rows (in expectation) each sampling model has the same per-sample signal to noise ratio (SNR) denoted by $\beta = \sigma_x^2/\sigma_w^2$.

Conditioned on K, the normalized distortion for the linear least squares estimator is given by

$$D = \min_{M \in \mathbb{C}^{k \times m}} \frac{1}{k\sigma_x^2} E||x_K - My||^2 \tag{2}$$

$$= \frac{1}{k\sigma_x^2} E \text{ tr}\left\{K_x - K_{xy}K_y^{-1}K_{yx}\right\} \tag{3}$$

where K_x, K_{xy}, and K_y are the covariance matrices of x_K and y. The expectation in (3) is taken over the sampling matrix Φ.

The first step toward calculating the distortion is to recast (3) in terms of a single random matrix. It is straightforward to show that the distortion due to sampling noise can be written as

$$D_s = \frac{1}{k} E \text{ tr}\left\{\left(I_k + \beta\Phi_K^H\Phi_K\right)^{-1}\right\} \tag{4}$$

where $(\frac{n}{m})\Phi_K^*\Phi_K$ is a central Wishart matrix.

For observation noise, on the other hand, we must consider whether or not the covariance matrix $T = \Phi_{K^c}\Phi_{K^c}^*$ corresponding to the effect of w_{o,K^c} is invertible. For $m < n - k$, the matrix T is full rank and the resulting distortion is

$$D_{o,1} = \frac{\beta}{1+\beta}\left[\frac{1}{\beta} + \frac{1}{k}E \text{ tr}\left\{\left(I_k + (1+\beta)\Phi_K^*T^{-1}\Phi_K\right)^{-1}\right\}\right] \tag{5}$$

where $(\frac{n-k}{k})\Phi_K^*T^{-1}\Phi_K$ is a central F matrix.

For $n - k < m < n$, the matrix T has rank $n - k$ and is not invertible. The distortion can be determined using $\tilde{T} = \Phi_{K^c}^*\Phi_{K^c}$ which has the same non-zero eigenvalues as T, and an $n-k \times n-m$ random matrix G whose i.i.d. elements have the same distribution as the elements of Φ. The resulting distortion is

$$D_{o,2} = \frac{\beta}{1+\beta}\left[\frac{1}{\beta} + \frac{1}{k}E \text{ tr}\left\{\left(I_{m-n} + (1+\beta)G^*\tilde{T}^{-1}G\right)^{-1}\right\}\right] \tag{6}$$

where $(\frac{m}{m-n})G^*\tilde{T}^{-1}G$ is a central F matrix. Finally, for $m \geq n$, the matrix Φ is invertible and the distortion is simply $1/(1+\beta)$.

The next step involves evaluating the expectations in (4), (5), and (6). We begin with the following transform from [10].

Definition 1. *Let W be a nonnegative definite random matrix. Its η-transform for $\gamma \geq 0$ is*

$$\eta_W(\gamma) = \int_0^\infty \frac{1}{1+\gamma x} f_W(x)dx \tag{7}$$

where $f_W(x)$ is the marginal density distribution of an unordered eigenvalue of W.

For an $n \times n$ Hermitian random matrix W, the η-transform is exactly the function we need, that is

$$\eta_W(\beta) = \frac{1}{n}E \text{ tr}\{(I + \beta W)^{-1}\}. \tag{8}$$

Accordingly, all that remains is to determine the marginal distribution of the unordered eigenvalues of the random matrices in (4), (5), and (6). This can be achieved by considering the asymptotic limits as $n, k, m \to \infty$ with $k/n \to \Omega$ and $m/n \to \rho$. In this setting, it has been shown that for both the Wishart [11] and F matrices [12], the probability distributions converge to non-random continuous functions with closed form expressions. Lemma 1 gives the η-transform for a Wishart matrix [10]. Lemma 2 gives the η-transform for an F matrix which we have calculated using the pdf derived in [12].

Lemma 1. *Let $H \in \mathbb{C}^{m \times k}$ have zero mean i.i.d. entries with variance one. If $k/m \to r$ as $k, m \to \infty$, then the central Wishart matrix $(\frac{1}{m})H^*H$ has*

$$\eta_{WS}(\beta, r) = 1 - \frac{F_1(\beta, r)}{4\beta r} \tag{9}$$

with

$$F_1(x, r) = \left(\sqrt{x(1+\sqrt{r})^2 + 1} - \sqrt{x(1-\sqrt{r})^2 + 1}\right)^2.$$

Lemma 2. *Let $H \in \mathbb{C}^{m \times k}$ and $M \in \mathbb{C}^{m \times p}$ ($m < p$) have zero mean i.i.d. Gaussian entries with unit variance. If $m/k \to r_1$ and $m/p \to r_2 \in (0, 1)$ as $k, m, p \to \infty$, then the central F matrix $(\frac{1}{k})H^*((\frac{1}{p})MM^*)^{-1}H$ has*

$$\eta_{FM}(\beta, r_1, r_2) = \frac{1}{1-r_2} + \left(\frac{\beta}{\beta - r_2/r_1}\right)F_2(\beta) + \left(\frac{r_2/r_1}{r_2/r_1 - \beta}\right)F_2(r_2/r_1) \tag{10}$$

with

$$a = \left(\frac{1-\sqrt{1-(1-r_1)(1-r_2)}}{1-r_2}\right)^2 \quad b = \left(\frac{1+\sqrt{1-(1-r_1)(1-r_2)}}{1-r_2}\right)^2$$

$$F_2(x) = \frac{1-r1}{2} - \frac{1}{1-r_2} + \frac{1-r_2}{2x}(\sqrt{(1+ax)(1+bx)} - 1)$$

Using Lemmas 1 and 2 we can specify the asymptotic distortion for each type of noise.

Theorem 1. *The asymptotic distortion of the optimal linear estimator of a signal x with sparsity $\Omega = k/n$ and sampling rate $\rho = m/n$ is given by*

$$D_s = \eta_{WS}\left(\rho\beta, \frac{\rho}{\Omega}\right) \tag{11}$$

$$D_o = \begin{cases} D_{o,1} & \rho \leq 1 - \Omega \\ D_{o,2} & 1 - \Omega < \rho \leq 1 \\ 1/(1+\beta) & 1 < \rho \end{cases} \tag{12}$$

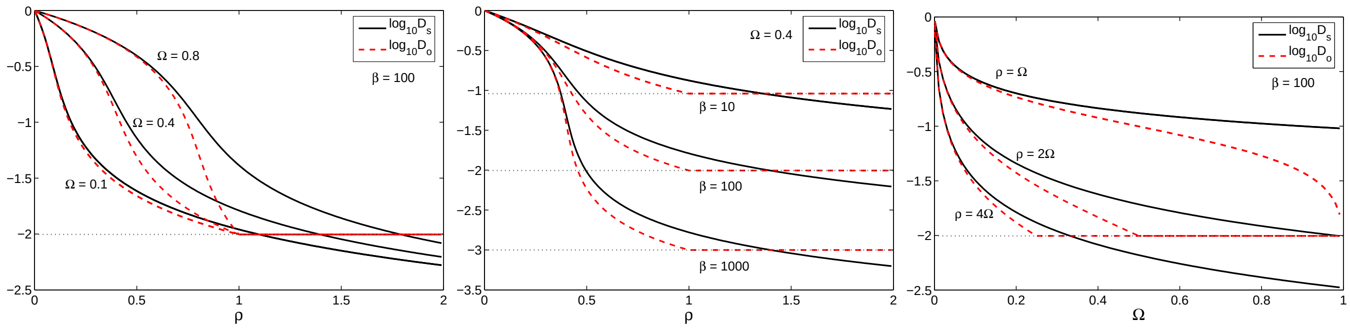
$$D_{o,1} = \frac{\beta}{1+\beta}\left[\frac{1}{\beta} + \eta_{FM}\left(\frac{\Omega}{1-\Omega}(\beta+1), \frac{\rho}{\Omega}, \frac{\rho}{1-\Omega}\right)\right]$$

$$D_{o,2} = \frac{\beta}{1+\beta}\left[\frac{1}{\beta} + \frac{1-\rho}{\Omega}\eta_{FM}\left(\frac{1-\rho}{\rho}(\beta+1), \frac{1-\Omega}{1-\rho}, \frac{1-\Omega}{\rho}\right)\right]$$

depending on whether the error, with SNR = β, is due to sampling or observation noise.

The distortions in Theorem 1 are shown in Figure 1. Figure 1(a) shows $\log_{10} D$ as a function of ρ for various Ω, Figure 1(b) shows $\log_{10} D$ as a function of ρ for various β, and Figure 1(c) shows $\log_{10} D$ as a function of Ω for various scalings of $\rho(\Omega)$.

The results shown in Figure 1 are in line with our intuition. For $\rho < 1$ the noise due to observation noise is correlated and is less



(a) (b) (c)

Fig. 1. The distortion, $\log_{10} D$, achieved by the optimal linear estimator under sampling noise (solid) and observation noise (dashed).

detrimental to recovery. On the other hand, for $\rho > 1$, D_o remains constant whereas $D_s \to 0$ as $\rho \to \infty$. Figure 1(b) shows that these results are consistent for a range of SNR. We remark that in all cases, the differences between the types of noise becomes very small as Ω becomes small. This occurs because $\eta_{\text{FM}}(\beta) \to \eta_{\text{WM}}(\beta)$, as the ratio r_2 of the F-Matrix goes to zero.

3. RECOVERY OF X_K WHEN K IS UNKNOWN

While Section 2 provides insight into what information can be learned from noisy random projections, applications in sensor networks require a practical strategy for recovering x without any knowledge of K. A great deal of research in this setting has demonstrated practical algorithms whose performance is comparable (by various constants and logarithmic factors) to what can be achieved when K is known [4, 5, 6, 7].

Remarkably, most of the results hold for arbitrary x_K. To allow our exposition to be consistent with other sections in this paper we consider x to be bounded in magnitude, that is x_K is any signal such that $||x||^2 \leq k\sigma_x^2$. Ultimately, removing this assumption does not change the nature of our results. Furthermore, for the sake of simplicity, we will consider x, w_o, w_s, and Φ to be real valued although this too is not a necessary restriction.

Many interesting results hold for scalings in which k is sublinear in n, that is $k/n \to 0$ as $k, n \to \infty$. In this setting, the authors of [4] and [5] specifically address the reconstruction i.i.d. Gaussian sampling error. In [5], Corollary 2 states that the error bounds for the given algorithm differ under the presence of observation noise.

However, in this paper we focus on linear scalings in which $k/n \to \Omega \in (0, 1)$ as $k, n \to \infty$, and we analyze the ℓ_1 constrained linear inverse reconstruction method known as Basis Pursuit [6, 7]. We will use the results from Theorem 1 of [6], which show the reconstruction error may be bounded even when $m < n$. The theorem is formulated in terms of the *restricted isometry constant* (defined below) and presented in a slightly altered version in Lemma 3 of this paper.

Definition 2. *Given an $n \times m$ matrix A, the k restricted isometry constant δ_k is the smallest quantity such that the all the eigenvalues of the $k \times k$ matrix $A_K^* A_K$ are in the interval $(1 - \delta_k, 1 + \delta_k)$ for all possible subsets K with $|K| \leq k$.*

Lemma 3. *Let $y = Ax + w$ where $x \in \mathbb{R}^n$ is k-sparse and $A \in \mathbb{R}^{m \times n}$ with unit normed rows. If there exists $s > k$ such that $s - k > s\delta_{k+s} + k\delta_s$, then the solution $\hat{x}$ to the ℓ_1 constrained linear inverse*

$$\min ||x||_1 \quad \textit{s.t.} \quad ||\Phi x - y||_2^2 \leq ||w||^2 \tag{13}$$

obeys

$$\frac{1}{n}||\hat{x} - x||^2 \leq C(k, n, m)\frac{1}{m}||w||^2 \tag{14}$$

with

$$C(k, n, m) = \inf_{s>k} \frac{2\sqrt{1 + k/s}}{\sqrt{1 - \delta_{k+s}} - \sqrt{k/s}\sqrt{1 + \delta_s}} \tag{15}$$

Although Lemma 3 actually holds for error w that is bounded in magnitude, we will instead assume that both w_o and w_s are zero mean i.i.d. Gaussians with variances σ_w^2. Because of concentration inequalities for χ^2 variables, the probability that either $||w_o||^2$ or $||w_s||^2$ is greater than its mean plus a standard deviation is very small as $m, n \to \infty$. Also, we remark that the function $C(n, k, m)$ in Lemma 3 has not been completely optimized.

To use the bound in Lemma 3, we must consider how the restricted isometry constant behaves as a function of the sampling matrix. For random matrices with i.i.d. elements, very strong results have been attained as the dimensions become large. The asymptotic restricted isometry constant of a random Gaussian matrix is described in Lemma 3.1 of [13] and reiterated below in Lemma 4 of this paper. Note that in the asymptotic setting, C becomes a function of just Ω and ρ.

Lemma 4. *Let $A \in \mathbb{R}^{m \times n}$ have zero mean i.i.d. Gaussian elements with variance $1/m$. As $n, k, m \to \infty$ with $k/n \to \Omega$ and $m/n \to \rho$, then with very high probability,*

$$\delta_\Omega < [1 + f(\Omega, \rho)]^2 - 1 \tag{16}$$

with

$$f(\Omega, \rho) = \sqrt{1/\rho}\left(\sqrt{\Omega} + \sqrt{2H(\Omega)}\right) \tag{17}$$

where $H(x) = -x\ln(x) - (1 - x)\ln(1 - x)$.

With these results in hand, we can begin to analyze an upper bound on the distortion. We use the fact that $||x||^2/n$ is bounded by $(k/n)\sigma_x^2$ to normalize the expected reconstruction error as

$$D = \frac{1}{k\sigma_x^2} E\left[||\hat{x} - x||^2\right] \tag{18}$$

where the expectation is taken over w and Φ. As the dimensions become large, $||w_s||^2, ||\Phi w_o||^2 \to m\sigma^2$, and Lemmas 3 and 4 say that with very high probability, under the influence of sampling noise, the distortion is upper bounded by B_s which is given as

$$D \leq B_s = \min\{1, C(\Omega, \rho)\frac{1}{\beta\Omega}\}. \tag{19}$$

The function $C(\Omega, \rho) \to 2$ as $\rho \to \infty$ and is finite for all $\rho > C_0(\Omega)H(\Omega)$, where the factor $C_0(\Omega)$ depends only on Ω. Figure 2 shows $C(\Omega, \rho)$ as a function of ρ for $\Omega^* = 1 \times 10^{-4}$ on a logarithmic scale. The bound descends rapidly for ρ just a little larger than $C_0(\Omega^*)H(\Omega^*) \approx 0.23$ and reaches approximately 5.2 for $\rho = 1$.

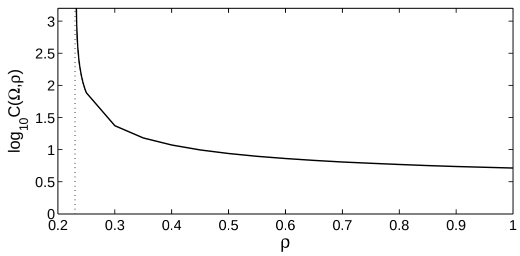


Fig. 2. $\log_{10} C(\Omega, \rho)$ as a function of ρ for $\Omega = 1 \times 10^{-4}$.

We now address how the bound changes in the presence of observation noise. Because $E||\Phi w_o|| = E||w_s||$, and because $C(\Omega, \rho)$ and $C_0(\Omega)$ do not depend on the noise, we know that the bound will not increase. To see if it can be lowered we appeal to a geometric argument. Let the signal z equal $x_K + w_{oK}$ on the the support and equal x everywhere else. We note that $||w_{o,K}|| \to \sqrt{k}\,\sigma_w$ and $||\Phi_{K^c} w_{o,K^c}|| \to \sqrt{n-k}\,\sigma_w$. Using (14) we can bound the distance between $\hat{x}$ and z as

$$d_z = ||\hat{x}_o - z|| \le \sqrt{C(\Omega, \rho)}\sqrt{n-k}\,\sigma_w \tag{20}$$

where the subscript on $\hat{x}$ indicates that it was estimated based on y_o. Using the triangle inequality, we conclude that

$$||\hat{x}_o - x|| \le \left(\sqrt{k} + \sqrt{C(\Omega, \rho)}\sqrt{n-k}\right)\sigma_w. \tag{21}$$

Comparing (14) and (21) leads to the following inequality for the distortion bounds.

Theorem 2. *The asymptotic upper bound on the distortion due to observation noise, B_o, can be related to the asymptotic upper bound due to sampling noise, B_s, by the following expression*

$$B_o \le \min\{1, g(\Omega, C(\Omega, \rho))\}B_s \tag{22}$$

where the function

$$g(\Omega, C(\Omega, \rho)) = \left(\sqrt{\frac{\Omega}{C^*(\Omega, \rho)}} + \sqrt{1-\Omega}\right)^2 \tag{23}$$

is less than one for

$$C(\Omega, \rho) > \frac{\Omega}{(1 - \sqrt{1-\Omega})^2}. \tag{24}$$

3.1. Heuristic Interpretation of the Results

The functions $C(\Omega, \rho)$ and $C_0(\Omega)$ from Lemmas 3 and 4 are loose upper bounds which correspond to a broader class of noise signals than we address in this paper. As a consequence, the given bounds may indicate the behavior of the true limits on the distortion, but are likely overly pessimistic. For instance, $C_0(\Omega)H(\Omega)$ is less than one only for very small values of Ω (less than 5×10^{-4}) even though simulations under non-adversarial stochastic noise have shown stable recovery to occur at much high sparsity levels [6, 13].

Motivated by the results of Section 2, we expect the differences between sampling and observation error to diminish for very small Ω. To proceed with our analysis over a broader range of parameters, we introduce the function $C^*(\Omega, \rho)$ which corresponds to the true average case distortion of the recovery algorithm used in Lemma 3. Based on the behavior of $C(\beta, \Omega)$, we make the assumptions that $C^*(\Omega, \rho)$ is large (approximately $\beta\Omega$) for all ρ less than some critical sampling rate $C_0^*(\Omega)H(\Omega)$ and that it descends rapidly to some reasonable value as ρ is increased beyond the sampling threshold.

Figure 3 shows the function g from Theorem 2 as a function of Ω for various fixed values of C^*. The figure shows that for small Ω and C^*, Theorem 2 does not establish a difference between the distortions. However, if C^* is large relative to Ω, then distortion under observation noise will be significantly lower than distortion under sampling noise.

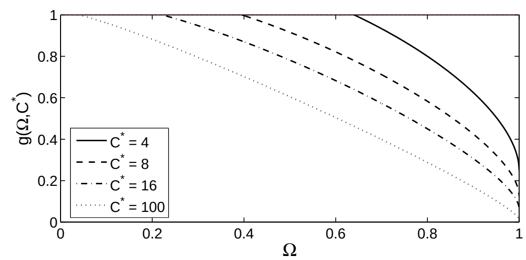


Fig. 3. The factor $g(\Omega, C^*)$ as a function of Ω for fixed C^*.

Finally, we remark that given the assumptions, the results shown in Figure 3 are still conservative because the bound corresponds to the worst case scenario for the observation noise. In the ideal case where the reconstruction error is mostly orthogonal to $w_{o,K}$ we have

$$\frac{D_o}{D_s} \approx \frac{\Omega}{C^*(\Omega, \rho)} + 1 - \Omega \tag{25}$$

which is less than one for all $C^*(\Omega, \rho) > 1$. The performance of this idealized situation mirrors what was shown in Section 2 for the optimal linear estimator.

3.2. Numerical Simulations

In this section we present some numerical simulations to show the difference in reconstruction between sampling and observation noise. Using ℓ_1-*Magic* [14], we implemented the ℓ_1 constrained linear inverses shown in (13). Appealing to concentration inequalities for χ^2 variables, an upper bound on the magnitude of $||w||^2$ was chosen to be $\sigma_w^2(m+2\sqrt{2m})$ and $(1-\Omega)\sigma_w^2(m+2\sqrt{2m})$ for sampling noise and observation noise respectively.

We measured the distortion over the range of sparsity levels $\Omega \in [0.05, 0.25]$ where the sampling rate varied as $\rho(\Omega) = 1.77H(\Omega)$ and the SNR was fixed at $\beta = 1000$. In the simulations $n = 500$ and 100 runs were performed for each data point.

Figure 4(a) shows the unnormalized (with respect to Ω) distortion $\Omega D = \frac{1}{n\beta}||\hat{x} - x||^2$. Also shown, is the baseline $1/\beta$ which corresponds to the error of the estimate $\hat{x} = x + w_o$. Figure 4(b) shows the ratio D_o/D_s.The simulations show that for the chosen scaling of ρ, the distortion due to observation is less than the distortion due to sampling error of the same magnitude, and that the relative difference increases with Ω. Furthermore, for the range of Ω shown, it appears that the ratio between the distortions scales roughly like $1 - \Omega$ plus some constant, which is consistent with the heuristic ratio (25) in the event that $C^*(\Omega, \text{const}\, H(\Omega))$ scales roughly like Ω.

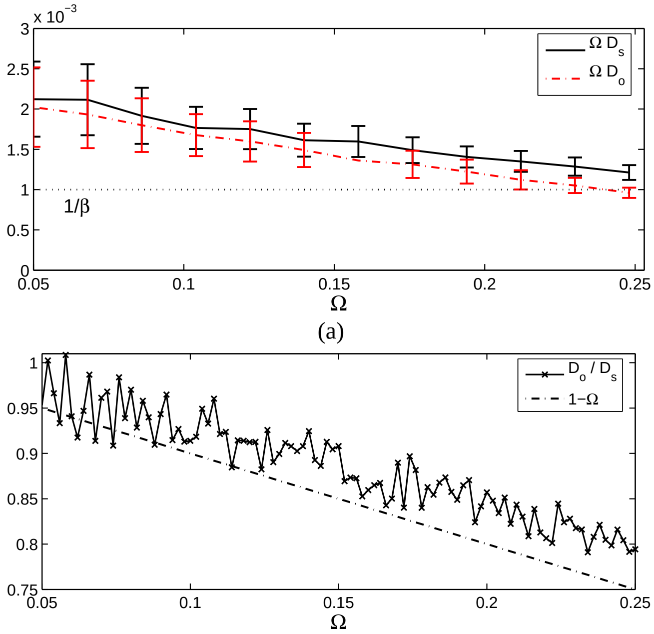


Fig. 4. Reconstruction errors for sampling and observation noise. (a) Comparison of $\frac{1}{n\beta}||\hat{x} - x||^2$ as a function of Ω where $\rho = 1.77H(\Omega)$; bars show the standard deviation. (b) The ratio of the distortions shown in (a) compared with $1 - \Omega$.

4. RECOVERY OF K WITH ℓ_∞ BOUNDED OBSERVATION NOISE

In tasks such as subset selection for regression and structure estimation in graphical models, the primary quantity of interest is the support [3]. In this section, we consider recovery of K under under the influence of non-stochastic ℓ_∞ bounded observation noise. We assume that each element of w_o is bounded in magnitude, that is $||w_o||_\infty < b$, and that $\min_{i \in K} |x_i| \geq 1$. This models the general situation where x contains a few very large components whose locations we would like to recover. The distortion metric is the probability of error $P_e(\hat{K} \neq K)$ where the probability is taken over Φ and K. We analyze recovery in the absence of sampling noise ($w_s = 0$).

This noise model is interesting because for any $b < 1/2$, perfect support recovery is possible for $m \geq n$ samples regardless of K. In this case, K can be determined by thresholding each element of $\Phi^{-1}y = x + w_o$. At the same time, as $b \to 0$, results from noiseless compressed sensing [7] show that it is possible to determine x, and accordingly K, with very high precision for $m \ll n$ provided that $m > 2k$. For $m < n$ however, the question remains, how must (b, n, k, m) scale to guarantee a given P_e? A sufficient condition is given by the following theorem.

Theorem 3. *Let $C_0'(\cdot)$ and $C'(\cdot,\cdot)$ be the well behaved functions defined in Theorem 2 of [6], and let*

$$m(\epsilon, s) = (1 + \epsilon) C_0'(s/n) H(s/n). \qquad (26)$$

For all $s \geq k$, reconstruction with $P_e < \mathcal{O}(n^{-\epsilon})$ can be achieved with $m = m(\epsilon, s)$ samples provided that $b = ||w_o||_\infty$ obeys

$$b < \frac{1}{2}\left(1 + C'(k/n, m/n)\frac{(n-s)}{\sqrt{s}}\right)^{-1}. \qquad (27)$$

Basically, this theorem asserts that b must decrease like $1/\sqrt{n}$ to allow perfect recovery of K, a much more stringent condition than in the stochastic case. Although the noise is bounded in the ℓ_∞ norm, its magnitude can grow like $\sqrt{n}b$, and $\max_{||v||_\infty \leq b} ||\Phi v||^2$ grows roughly like $n\, b^2 \lambda_{\max}$ where $\lambda_{\max}$ is the largest eigenvalue of Φ.

Proof of Theorem 3 . We define z_S to be the best $|S|$-term approximation of $x + w_o$, i.e. $z_S = (x + w_o)_S$, where $|S| = s$. By the restrictions on w_o and x, $K \subset S$ for any $s \geq k$. The proof follows from Theorem 2 in [6] which guarantees that

$$||\hat{x} - (x + w_o)|| \leq C'(k/n, m/n)||x - z_S||_1/\sqrt{s} \qquad (28)$$

for a well behaved function $C'(\cdot,\cdot)$. The ℓ_1 norm the on the right hand side can be bounded as $||x - z_S|| \leq (n - s)b$.

Thresholding $\hat{x}$ will yield the correct support provided that the error on any given element is small enough, that is

$$||\hat{x} - (x + w_o)|| < 1/2 - b. \qquad (29)$$

Applying this requirement to (28) provides the necessary scaling on b. □

5. REFERENCES

[1] D. Donoho, “Compressed Sensing,” *Info. Theory, IEEE Trans.*, vol. 52(4), pp. 1289–1306, Apr. 2006.

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

[3] 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.

[4] E. Candes and T. Tao, “The Dantzig selector: statistical estimation when p is much larger than n,” *ArXiv Mathematics e-prints*, June 2005.

[5] J. Haupt and R. Nowak, “Signal Reconstruction From Noisy Random Projections,” *Info. Theory, IEEE Trans.*, vol. 52(9), pp. 4036–4048, sep 2006.

[6] 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.

[7] D. Donoho, M. Elad, and V. Temlyakov, “Stable Recovery of Sparse Overcomplete Representations in the Presence of Noise,” *Info. Theory, IEEE Trans.*, vol. 52(1), pp. 6–18, jan 2006.

[8] M. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, “Compressive wireless sensing,” in *Proc. 5th Int. Conf. on Info. Processing in Sensor Networks*, Nashville, TN, Apr. 2006, pp. 134–142.

[9] M. Rabbat, J. Haupt, A. Singh, and R. Nowak, “Decentralized Compression and Predistribution via Randomized Gossiping,” in *Proc. 5th Int. Conf. on Info. Processing in Sensor Networks*, Nashville, TN, Apr. 2006, pp. 51–59.

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

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

[12] J. Silverstein, “The limiting eigenvalue distribution of a multivariate F-matrix,” *SIAM J. of Math. Analysis*, vol. 30, pp. 641–646, 1985.

[13] E. Candes and T. Tao, “Decoding by Linear Programming,” *Info. Theory, IEEE Trans.*, vol. 51(12), pp. 4203–4215, Dec. 2005.

[14] E. Candes, “L1-magic: Recovery of sparse signals,” *http://www.acm.caltech.edu/l1magic/*.