Aalborg Universitet

When 'exact recovery' is exact recovery in compressed sensing simulation

Sturm, Bob L.

Published in:
Proceedings of the European Signal Processing Conference

Publication date:
2012

Document Version
Accepted author manuscript, peer reviewed version

WHEN "EXACT RECOVERY" IS EXACT RECOVERY IN COMPRESSED SENSING SIMULATION

Bob L. Sturm

Dept. Architecture, Design and Media Technology, Aalborg University Copenhagen
A.C. Meyers Vænge 15, DK-2450 Copenahgen SV, Denmark, E-mail: bst@create.aau.dk

ABSTRACT

In a simulation of compressed sensing (CS), one must test whether the recovered solution $\hat{\mathbf{x}}$ is the true solution $\mathbf{x}$, i.e., "exact recovery." Most CS simulations employ one of two criteria: 1) the recovered support is the true support; or 2) the normalized squared error is less than ϵ^2. We analyze these exact recovery criteria independent of any recovery algorithm, but with respect to signal distributions that are often used in CS simulations. That is, given a pair $(\hat{\mathbf{x}}, \mathbf{x})$, when does "exact recovery" occur with respect to only one or both of these criteria for a given distribution of $\mathbf{x}$? We show that, in a best case scenario, ϵ^2 sets a maximum allowed missed detection rate in a majority sense.

***Index Terms*—** compressed sensing, exact recovery

1. INTRODUCTION

CS is, by definition, a low-cost signal acquisition system, and being such, it is concerned first and foremost with the guarantees of recovering sensed signals [1, 2]. Essentially, CS takes m measurements of an $N \gg m$ length signal $\mathbf{x}$ by non-adaptively projecting it onto a set of sensing functions, e.g., the rows of a matrix $\mathbf{\Phi} \in \mathbb{R}^{m\times N}$. Recovery then finds $\hat{\mathbf{x}}$ from $\mathbf{\Phi}$ and the measurements $\mathbf{u} = \mathbf{\Phi}\mathbf{x}+\mathbf{n}$, where $\mathbf{n}$ is noise. The problem is underdetermined, but when $\mathbf{x}$ is known to be sparse, or compressible, recovery becomes possible given certain conditions are met [1, 2].

Dozens of algorithms have been repurposed or proposed for CS recovery (see [3–5] for overviews). Their practical implementation, testing and comparison requires a useful criterion to determine for a given pair $(\hat{\mathbf{x}}, \mathbf{x})$ if $\hat{\mathbf{x}} \stackrel{?}{=} \mathbf{x}$, i.e., "exact recovery." Two particular criteria have been widely used when evaluating the performance of CS recovery algorithms. The first criterion declares $\mathbf{x}$ "exactly recovered" when $\hat{\mathbf{x}}$ has no false alarms or missed detections. Let Ω index the columns of $\mathbf{\Phi}$. Define the support of $\mathbf{x}$ by $S(\mathbf{x}) := \{i \in \Omega : x_i \neq 0\}$, where x_i is the ith element of $\mathbf{x}$. If

$$S(\hat{\mathbf{x}}) = S(\mathbf{x}) \quad \text{(SC)}$$

BLS is supported in part by: Independent Postdoc Grant 11-105218 from Det Frie Forskningsråd; and EPSRC Platform Grant EP/E045235/1 at the Centre for Digital Music of Queen Mary University of London.

then $\mathbf{x}$ is "exactly recovered" with respect to the uniform support recovery criterion. In this case, each component of the original signal has the same cost if missed, no matter its value. This criterion is used in, e.g., [1, 6–10]. The second criterion involves the ℓ_2-norm of the error. If for a $0 \le \epsilon^2 < 1$,

$$\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2}{\|\mathbf{x}\|_2^2} \le \epsilon^2 \quad (\epsilon^2\text{C})$$

then $\mathbf{x}$ is "exactly recovered" with respect to the normalized ℓ_2-norm criterion with parameter ϵ^2. This criterion has been used in, e.g., [11–14].

Unsurprisingly, a particular $(\hat{\mathbf{x}}, \mathbf{x})$ can satisfy one criterion but not the other, thus calling into question whether the results of two CS simulations using these different criteria are comparable. We now analyze these two criteria, and the role of ϵ^2, independent of the recovery algorithm one employs.

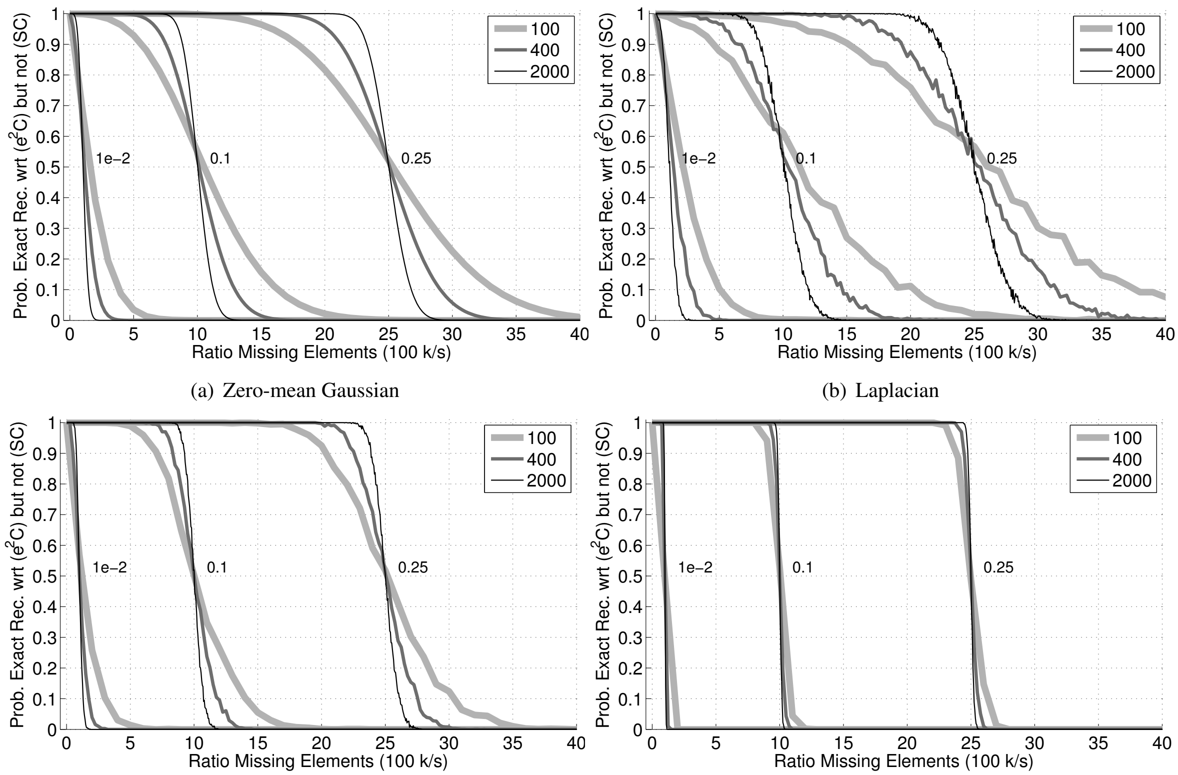
2. EXACT RECOVERY CRITERIA

In general, neither criterion is necessarily true or false given the other is true or false. For any $\mathbf{x}$ and any $\epsilon^2 \in [0, 1]$, we can produce a $\hat{\mathbf{x}}$ such that (SC), but not (ϵ^2C). For instance, given $\mathbf{x}$, let $\hat{\mathbf{x}} = \delta(1+\sqrt{2\epsilon^2 - \epsilon^4})\mathbf{x}$ where $\delta > 1$. Then, (SC) is true but (ϵ^2C) is not. For any $\epsilon^2 \in (0, 1]$, given an $\mathbf{x}$ with an element $0 < |x_i|^2 \le \epsilon^2\|\mathbf{x}\|_2^2$, then we can discard it from $\mathbf{x}$ to create a $\hat{\mathbf{x}}$ satisfying (ϵ^2C) but not (SC). There are instances, however, when one must be true if the other is true. First for the noiseless condition, and then for noise, we explore the equivalence between the two criteria. This provides a way to interpret ϵ^2, and to see when the conditions are equivalent in a majority sense for a best case scenario.

2.1. Noiseless case $\|\mathbf{n}\| = 0$

Given $\hat{\mathbf{x}}$, the weights minimizing the measurement modeling error are $\hat{\mathbf{y}} := \arg\min_{\mathbf{y}'} \|\mathbf{u} - \mathbf{\Phi}_{S(\hat{\mathbf{x}})}\mathbf{y}'\|_2^2 = \mathbf{\Phi}^\dagger_{S(\hat{\mathbf{x}})}\mathbf{u}$. In this case, if $\hat{\mathbf{x}}$ is composed of the least-squares weights, if (SC) then for any $\epsilon^2 \in [0, 1]$ (ϵ^2C). If, however, (ϵ^2C) is true for $\epsilon^2 = 0$, (SC) necessarily follows. Now we look at the behavior of these criteria for signals distributed various ways.

Without loss of generality, consider the true support of $\mathbf{x}$ to be its first s elements, i.e, $S(\mathbf{x}) = \{1, 2, \dots, s\}$. Consider that our solution $\hat{\mathbf{x}}$ is missing the first $0 < k < s$ of these



(a) Zero-mean Gaussian

(b) Laplacian

(c) Uniform

(d) Bimodal Gaussian

Fig. 1. Best case scenario probability of exact recovery with respect to (ϵ^2C) but not (SC) for sparse vectors distributed in four ways, for various ϵ^2 (labeled) and sparsities s (legend). Empirical results in (b-d) are averaged over 1000 trials.

elements, i.e., for $n \in \{1,\ldots,k\}(\hat{x}_n = 0)$, but that for $n \in \Omega\backslash\{1,\ldots,k\}(\hat{x}_n = x_n)$. This means $S(\hat{\mathbf{x}}) \subset S(\mathbf{x})$, i.e., $\hat{\mathbf{x}}$ has no false detections, and that the missed detections do not influence our estimation of the values of the recovered support. We call this the "best case scenario." In this case, (ϵ^2C) but not (SC) becomes

$$\frac{1}{\|\mathbf{x}\|_2^2}\sum_{n=1}^{k} x_n^2 \le \epsilon^2. \quad (1)$$

We now find the probability of (ϵ^2C) but not (SC) for a few distributions. Consider the s non-zero elements of $\mathbf{x}$ to be distributed Rademacher, i.e., iid equiprobable in $\{-1, 1\}$. This type of sparse signal is often used in simulations of CS recovery [6, 11, 12]. In this case, $\|\mathbf{x}\|_2^2 = s$, so

$$P\left\{\frac{1}{\|\mathbf{x}\|_2^2}\sum_{n=1}^{k} x_n^2 \le \epsilon^2\right\} = \begin{cases} 1, & k/s \le \epsilon^2 \\ 0, & \text{else.}\end{cases} \quad (2)$$

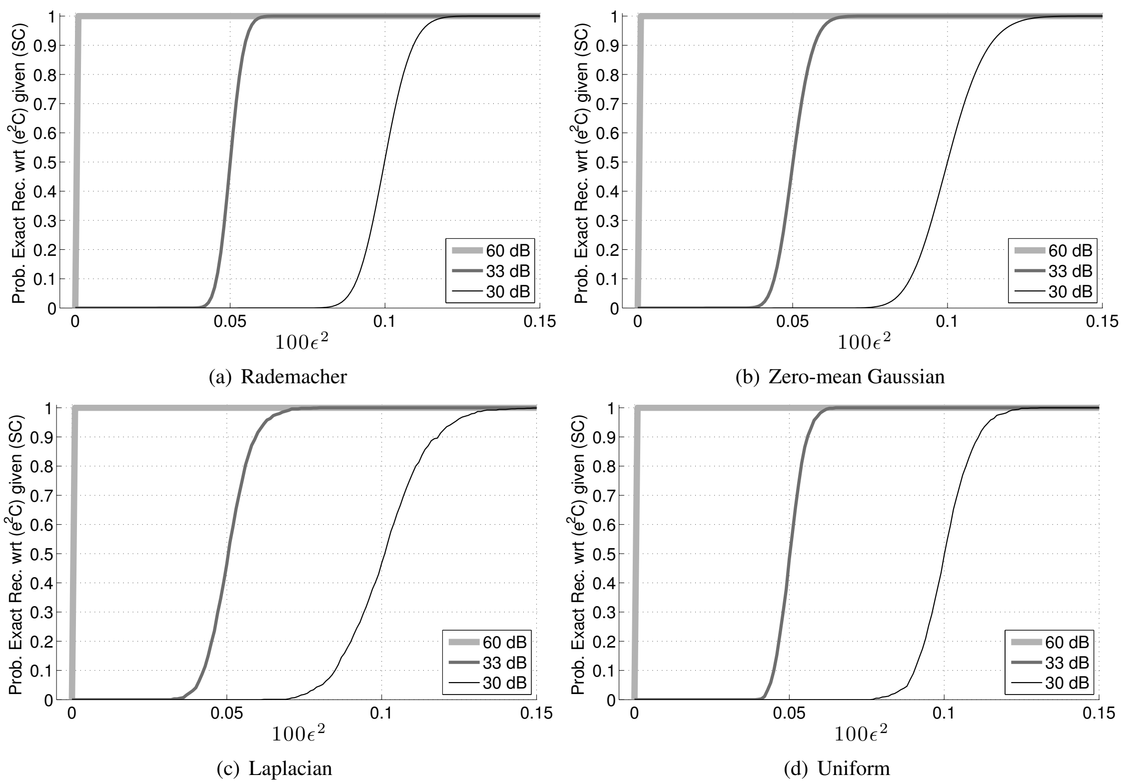
This shows for Rademacher sparse signals how to interpret ϵ^2 in the best case scenario: it is the ratio of the maximum number of missed detections to the signal sparsity. Thus, in the recovery of signals distributed Rademacher, as long as $s < \epsilon^{-2}$, if (ϵ^2C) then (SC). Otherwise, if $k = 1$, (ϵ^2C) can be true while (SC) is not. For the test problems in [12], $s < 800$ and $\epsilon^2 = 10^{-4}$, thus it is clear for their simulations that a solution satisfying (ϵ^2C) must also satisfy (SC). However, if the sparsity $s > 10000$ for this ϵ^2, then the two conditions are no longer equivalent.

Now consider the s non-zero elements of $\mathbf{x} \sim \mathcal{N}(0, \sigma_y^2)$, i.e., iid zero-mean Gaussian with variance $\sigma_y^2 > 0$. (Note the distinction between the variance of all the elements in $\mathbf{x}$, and just its non-zero elements in $\mathbf{y}$.) This is another signal distribution used extensively in testing CS recovery algorithms, e.g., [3,12,14]. Define the independent random variables (rvs)

$$Y_k := \sum_{n=1}^{k} [x_n/\sigma_y]^2 \sim \chi^2(k) \quad (3)$$

$$Z_{s-k} := \sum_{n=k+1}^{s} [x_n/\sigma_y]^2 \sim \chi^2(s-k) \quad (4)$$

both of which are distributed chi-squared with the argument as a parameter. Notice that $\|\mathbf{x}\|_2^2 = (Y_k + Z_{s-k})\sigma_y^2$. Since Y_k and Z_{s-k} are independent, $F_{k,s-k} := [Y_k/k]/[Z_{s-k}/(s-k)] \sim \mathcal{F}(k, s-k)$, or F-distributed with parameters k and $s-k$. Thus, the probability of (ϵ^2C) but not (SC) in the best



(a) Rademacher (b) Zero-mean Gaussian (c) Laplacian (d) Uniform

Fig. 2. Best case scenario probability of exact recovery by (ϵ^2C) given (SC), as a function of ϵ^2 ($s = 400$) for measurements in zero-mean Gaussian white noise at three signal to noise ratios (σ_y^2/σ_v^2, legend) distributed in four ways. Empirical results in (c, d) are averaged over 1000 trials.

case scenario for an s-sparse signal distributed $\mathcal{N}(0, \sigma_y^2)$ is

$$P\left\{\frac{1}{\|\mathbf{x}\|_2^2}\sum_{n=1}^{k} x_n^2 \le \epsilon^2\right\} = P\left\{Y_k < \frac{\epsilon^2}{1-\epsilon^2} Z_{s-k}\right\}$$
$$= P\left\{F_{k,s-k} < \frac{\epsilon^2}{1-\epsilon^2}\frac{1-k/s}{k/s}\right\}. \quad (5)$$

Figure 1(a) plots (5) for three sparsities and ϵ^2. It is clear that the two exact recovery conditions become identical as ϵ^2 becomes small. If $k/s > \epsilon^2$ then for $s \ge 2k$ $P\{F_{k,s-k} < 1 + \delta\} > 0.5$ for $\delta > 0$. Thus, here the parameter ϵ^2 prescribes the maximum number of missed support elements in the best case scenario before (ϵ^2C) given not (SC) is false in a majority sense for sparse signals distributed $\mathcal{N}(0, \sigma_y^2)$. Figure 1(b-d) show empirically-derived (also best case scenario) probabilities for other distributions, all of which have similar behavior. We see for all distributions tested and $s \ge 2k$, that when $k/s < \epsilon^2$, (ϵ^2C) but not (SC) is true in a majority sense.

2.2. Noisy case $\|\mathbf{n}\| > 0$

Unlike for the noiseless case, with a least-squares solution in the noisy case, if (SC) then not necessarily (ϵ^2C) for any $\epsilon^2 \in [0, 1]$. To see this, assume (SC), and that $\hat{\mathbf{x}}$ is built from $\hat{\mathbf{y}} = \arg\min_{\mathbf{y}'} \|\mathbf{u} - \mathbf{n} - \mathbf{\Phi}_{S(\hat{\mathbf{x}})}\mathbf{y}'\|_2^2 = \mathbf{\Phi}^\dagger_{S(\mathbf{x})}(\mathbf{u} - \mathbf{n})$. Define $\mathbf{y}$ as the non-zero elements of $\mathbf{x}$. In this case, (ϵ^2C) becomes

$$\frac{\|\mathbf{y} - \mathbf{\Phi}^\dagger_{S(\mathbf{x})}\mathbf{u}\|_2^2}{\|\mathbf{y}\|_2^2} = \frac{\|\mathbf{\Phi}^\dagger_{S(\mathbf{x})}\mathbf{n}\|_2^2}{\|\mathbf{y}\|_2^2} \le \epsilon^2. \quad (6)$$

Hence, for any $\epsilon^2 \in (0, 1]$ we can find an $\mathbf{n}$ such that (SC) is true but not (ϵ^2C). As in the noiseless case, though, only for $\epsilon^2 = 0$ if (ϵ^2C) then (SC).

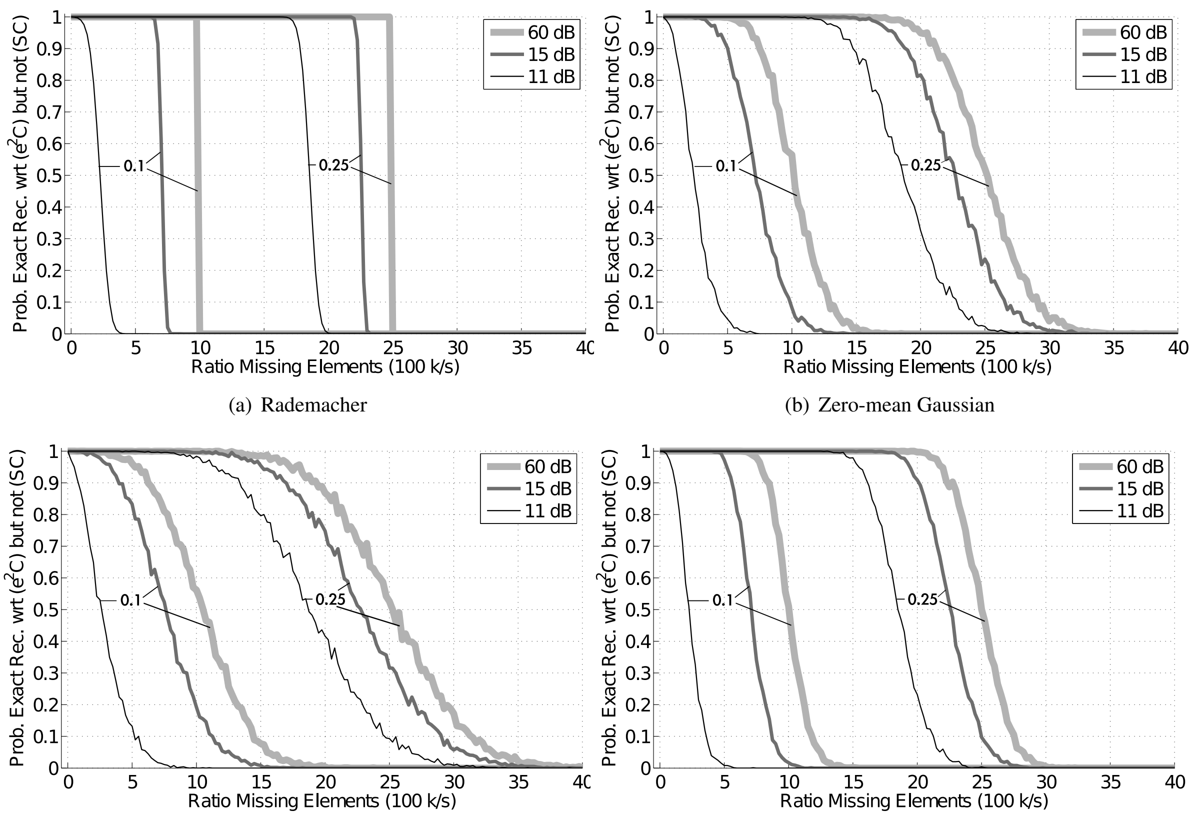
We now find the probability of (ϵ^2C) whether or not (SC) for a few signal distributions. Define $\mathbf{v} := \mathbf{\Phi}^\dagger_{S(\mathbf{x})}\mathbf{n}$, and assume its $|S(\mathbf{x})|$ elements are iid $\mathcal{N}(0, \sigma_v^2)$ and independent of $\mathbf{y}$. Define the chi-squared-distributed rv

$$V_s := \sum_{n=1}^{s} [v_n/\sigma_v]^2 \sim \chi^2(s). \quad (7)$$

If the s non-zero elements of $\mathbf{x}$ are distributed Rademacher, then the probability of (ϵ^2C) given (SC)

$$P\left\{\frac{1}{\|\mathbf{y}\|_2^2}\sum_{n=1}^{s} v_n^2 \le \epsilon^2\right\} = P\left\{V_s < \frac{\epsilon^2 s}{\sigma_v^2}\right\}. \quad (8)$$

Since $P\{V_s < s + \delta\} > 0.5$ for $\delta > 0$, if $\epsilon^2 \ge \sigma_v^2$ then, given (SC) is true, (ϵ^2C) is true in a majority sense.



(a) Rademacher

(b) Zero-mean Gaussian

(c) Laplacian

(d) Uniform

Fig. 3. Best case scenario probability of exact recovery with respect to (ϵ^2C) but not (SC) for sparse vectors ($s = 400$) distributed in four ways, and measurements in zero-mean white Gaussian noise, at three sparse signal to noise ratios (σ_y^2/σ_v^2, legend) and two ϵ^2 (labeled). Empirical results in (b-d) averaged over 1000 trials.

Now, assume the s non-zero elements of $\mathbf{x} \sim \mathcal{N}(0, \sigma_y^2)$, and still independent of $\mathbf{v}$. Define

$$X_s := \sum_{n=1}^{s} [x_n/\sigma_y]^2 \sim \chi^2(s). \tag{9}$$

The ratio V_s/X_s is an F-distributed rv

$$W_{s,s} := V_s/X_s \sim \mathcal{F}(s,s). \tag{10}$$

Thus, the probability of (ϵ^2C) given (SC) is

$$P\left\{\frac{1}{\|\mathbf{y}\|_2^2}\sum_{n=1}^{s} v_n^2 \le \epsilon^2\right\} = P\left\{W_{s,s} < \frac{\sigma_y^2}{\sigma_v^2}\epsilon^2\right\}. \tag{11}$$

Here we see that since $P\{W_{s,s} < 1+\delta\} > 0.5$ for $\delta > 0$, if $\epsilon^2 \ge \sigma_v^2/\sigma_y^2$, given (SC) is true, then (ϵ^2C) is true in a majority sense.

For three signal to noise power ratios, σ_y^2/σ_v^2, Figure 2 shows the probability of exact recovery with respect to (ϵ^2C) given (SC) as a function of ϵ^2: (a) shows (8); (b) shows (11); and (c,d) show empirical results for $\mathbf{y}$ distributed Laplacian and Uniform. As ϵ^2 increases, it becomes easier to meet (ϵ^2C) given (SC); but as noise increases, ϵ^2 must likewise increase.

Now, consider (ϵ^2C) is true but not (SC). As before, consider that $\hat{\mathbf{x}}$ is missing the first $0 < k < s$ of the true elements, $n \in \{1,\ldots,k\}(\hat{x}_n = 0)$, but $n \in \Omega\backslash\{1,\ldots,k\}(\hat{x}_n = x_n + v_n)$. This "best case scenario" means we have no false detections, and that for the support recovered, we find their exact values perturbed only by the noise. Thus, assuming $\mathbf{x}$ and $\mathbf{v}$ are independent, (ϵ^2C) given not (SC) becomes

$$\frac{1}{\|\mathbf{x}\|_2^2}\left[\sum_{n=1}^{k} x_n^2 + \sum_{n=1}^{s-k} v_n^2\right] \le \epsilon^2. \tag{12}$$

Define the rv

$$G_{s-k} := \sum_{n=1}^{s-k} [v_n/\sigma_v]^2 \sim \chi^2(s-k). \tag{13}$$

When the non-zero elements of $\mathbf{x}$ are distributed Rademacher, and $v_n \sim \mathcal{N}(0, \sigma_v^2)$, ($\epsilon^2$C) given not (SC) becomes

$$P\left\{\frac{1}{\|\mathbf{x}\|_2^2}\left[\sum_{n=1}^{k} x_n^2 + \sum_{n=1}^{s-k} v_n^2\right] \le \epsilon^2\right\} = P\left\{G_{s-k} < \frac{\epsilon^2 s - k}{\sigma_v^2}\right\}. \tag{14}$$

Since $P\{G_{s-k} < s-k+\delta\} > 0.5$ for $\delta > 0$, we see (ϵ^2C) is false in a majority sense when $(\epsilon^2 s - k)/\sigma_v^2 < s-k$. Figure 3(a) shows (14) as a function of k/s for two different ϵ^2 and three signal to noise power ratios; and Fig. 3(b-d) show empirical results for other distributions. We see here that as the noise power increases for a given ϵ^2, in the best case scenario, the probability that (ϵ^2C) is true in a majority sense given (SC) is false decreases.

3. CONCLUSION

Regardless of whether CS is applied in practice to exactly sparse signals, much significant work in simulating, testing and comparing CS recovery algorithms has used exactly sparse signals, and either (ϵ^2C) or (SC) as a criterion for "exact recovery" [1, 6–14]. The significance of ϵ^2, and the relationship between (ϵ^2C) and (SC) had yet to be analyzed. In this paper, we show that ϵ^2 can be interpreted within the context of CS recovery in a best case scenario as a maximum acceptable missed detection rate in a majority sense, in both noiseless and noisy measurements.

For noiseless measurements, if the weights of the recovered solution are least-squares optimal, then $(\text{SC}) \to \forall\epsilon \in [0,1](\epsilon^2\text{C})$. Only if $\epsilon^2 = 0$, $(\epsilon^2\text{C}) \to (\text{SC})$. For s-sparse signals with non-zero elements distributed Rademacher, and $\|\mathbf{n}\| = 0$, if $s < \epsilon^{-2}$, $(\epsilon^2\text{C}) \to (\text{SC})$. When the measurements have zero-mean WG noise with variance σ_v^2, if $\epsilon^2 > \sigma_v^2$, then in the best case scenario $P\{(\epsilon^2\text{C})|(\text{SC})\} > 0.5$. And if $\epsilon^2 < \sigma_v^2 + (k/s)(1-\sigma_v^2)$, then $P\{(\epsilon^2\text{C})|\neg(\text{SC})\} < 0.5$. For sparse signals with non-zero elements distributed zero-mean Gaussian and variance σ_y^2, when there is no measurement noise, if $k/s < \epsilon^2$ and $s \geq 2k$, then in the best case scenario $P\{(\epsilon^2\text{C})|\neg(\text{SC})\} < 0.5$. When noise is in the measurements, as above, if $\epsilon^2 > \sigma_v^2/\sigma_y^2$, then in the best case scenario $P\{(\epsilon^2\text{C})|(\text{SC})\} > 0.5$. This suggests for sparse signals distributed zero-mean Gaussian or Rademacher, a useful range over which one should define the parameter ϵ^2 such that exact recovery by (SC) and (ϵ^2C) are equivalent in a majority sense in a best case scenario when $s \geq 2k$, and k is the number of elements a solution can miss yet still be considered an "exact recovery":

$$\epsilon^2 \in \left(\frac{\sigma_v^2}{\sigma_y^2}, \frac{k}{s} + \sigma_v^2\left(1 - k/s\right)\right). \tag{15}$$

Our future work explores how particular CS recovery algorithms preclude certain pairs $(\mathbf{x}, \hat{\mathbf{x}})$ for which one condition is true and the other is false with high probability. Furthermore, we are investigating the equivalence of other "exact recovery" criteria used in CS simulations for signals that are compressible rather than exactly sparse.

4. REFERENCES

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

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

[3] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1, pp. 34–81, Feb. 2009.

[4] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, June 2010.

[5] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, 2010.

[6] J. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Info. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[7] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied Comp. Harmonic Anal.*, vol. 27, no. 3, pp. 265–274, Jan. 2009.

[8] M. J. Wainwright, "Sharp thresholds for noisy and high-dimensional recovery of sparsity using ℓ_1-constrained quadratic programming (lasso)," *IEEE Trans. Info. Theory*, vol. 55, no. 5, pp. 2183–2202, May 2009.

[9] ——, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," *IEEE Trans. Info. Theory*, vol. 55, no. 12, pp. 5728–5741, Dec. 2009.

[10] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Necessary and sufficient conditions for sparsity pattern recovery," *IEEE Trans. Info. Theory*, vol. 55, no. 12, pp. 5758–5772, Dec. 2009.

[11] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. National Academy of the Sciences*, vol. 106, no. 45, pp. 18 914–18 919, Nov. 2009.

[12] A. Maleki and D. L. Donoho, "Optimally tuned iterative reconstruction algorithms for compressed sensing," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 330–341, Apr. 2010.

[13] Y. Wang and W. Yin, "Sparse signal reconstruction via iterative support detection," *SIAM J. Imaging Sciences*, vol. 3, no. 3, pp. 462–491, 2010.

[14] J. Vila and P. Schniter, "Expectation-maximization Bernoulli-Gaussian approximate message passing," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 2011.