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
Early version, also known as pre-print

Link to publication from Aalborg University

Citation for published version (APA):
Sturm, B. L. (2012). When 'exact recovery' is exact recovery in compressed sensing simulation. *Proceedings of the European Signal Processing Conference*, *2012*, 979-983. http://www.scopus.com/inward/record.url?eid=2-s2.0-84869754566&partnerID=40&md5=7acd2d5ff6239a41f89f0083ebcdc6a4

When "Exact Recovery" is Exact Recovery in Compressed Sensing *Simulation*

Bob L. Sturm[1]

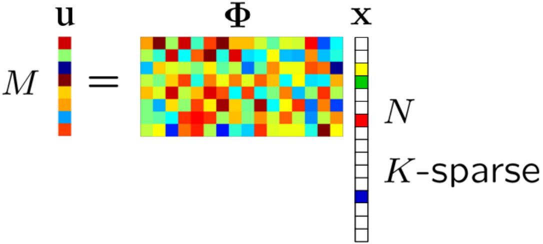
Department of Architecture, Design and Media Technology
Aalborg University Copenhagen
A.C. Meyers Vænge 15, DK-2450, Denmark

August 27, 2012

[1]B. L. Sturm is supported in part by Independent Postdoc Grant 11-105218 from Det Frie Forskningsråd.

Setup

Measurements $\mathbf{u}$ come from sensing $\mathbf{x}$ by sensing matrix $\mathbf{\Phi}$: $\mathbf{u} = \mathbf{\Phi}\mathbf{x} + \mathbf{n}$.
We use a recovery algorithm to build $\hat{\mathbf{x}}$ given $\mathbf{u}$ and $\mathbf{\Phi}$, e.g., OMP, BP.



Exact Recovery

- In theory, we have no trouble asking $\hat{\mathbf{x}} \stackrel{?}{=} \mathbf{x}$.
- In practice, we must use a different criterion.
- **At least two different criteria have been used in the simulation of compressed sensing recovery algorithms.**

One exact recovery criterion in CS simulation: Support

Let Ω index the columns of $\mathbf{\Phi}$, and define the support of $\mathbf{x}$ as

$$S(\mathbf{x}) := \{i \in \Omega : x_i \neq 0\}.$$

$\mathbf{x}$ is exactly recovered with respect to support if

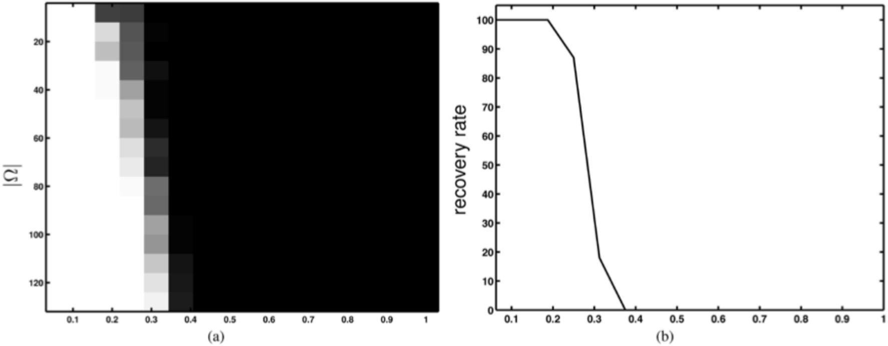
$$S(\hat{\mathbf{x}}) = S(\mathbf{x}). \tag{SC}$$

This has been used in simulations of CS recovery in, e.g.,

- 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.
- 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.
- 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.

One exact recovery criterion in CS simulation: Support

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.



For $N = 512$. (a) Empirical prob. exact recovery as fun. of M (ord.), K/M (abs.). White is 1.0. (b) Empirical prob. of exact recovery for $M = 64$ as function of K/M.

One exact recovery criterion in CS simulation: Support

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

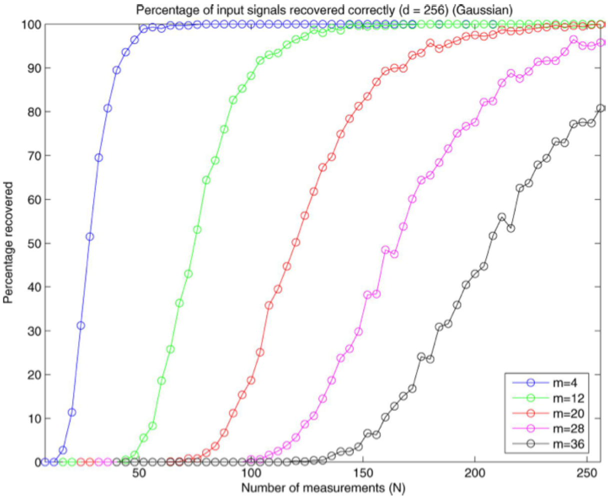


Fig. 1. The percentage of 1000 input signals correctly recovered as a function of the number N of measurements for different sparsity levels m in dimension $d = 256$.

One exact recovery criterion in CS simulation: Support

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.

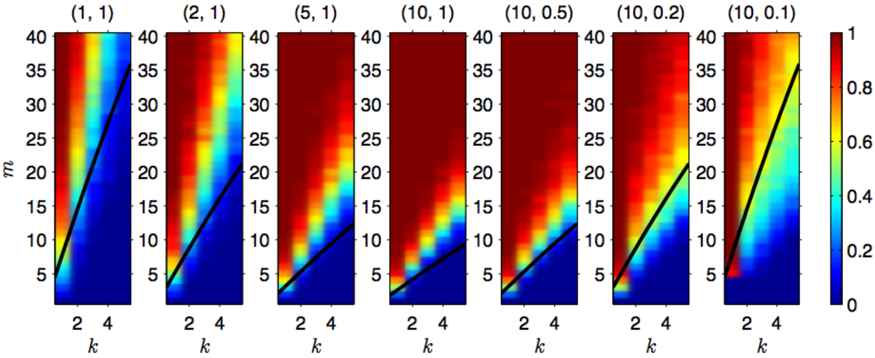


Fig. 1. Simulated success probability of ML detection for $n = 20$ and many values of k, m, SNR, and MAR. Each subfigure gives simulation results for $k \in \{1, 2, \ldots, 5\}$ and $m \in \{1, 2, \ldots, 40\}$ for one (SNR, MAR) pair. Each subfigure heading gives (SNR, MAR). Each point represents at least 500 independent trials. Overlaid on the color-intensity plots is a black curve representing (6).

Another exact recovery criterion: Normalized ℓ_2-norm Error

Define a $0 \leq \epsilon^2 < 1$.

$\mathbf{x}$ is exactly recovered with respect to normalized squared error if

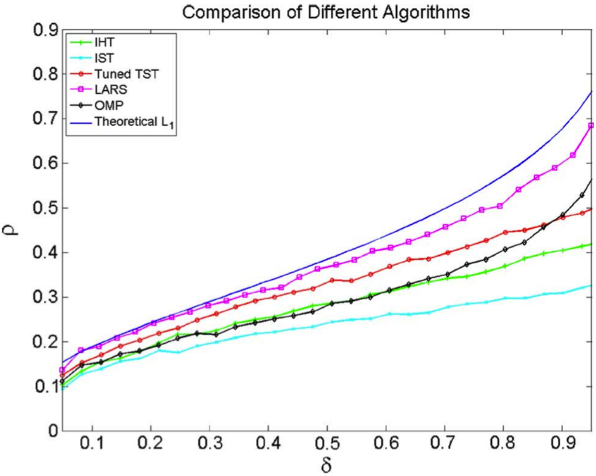
$$\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2}{\|\mathbf{x}\|_2^2} \leq \epsilon^2 \qquad (\epsilon^2\text{C})$$

This has been used in simulations of CS recovery in, e.g.,

- 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.
- J. Vila and P. Schniter, "Expectation-maximization Bernoulli-Gaussian approximate message passing," in Proc. Asilomar Conf. Signals, Syst., Comput., Pacific Grove, CA, Nov. 2011.
- Y. Wang and W. Yin, "Sparse signal reconstruction via iterative support detection," SIAM J. Imaging Sciences, vol. 3, no. 3, pp. 462-491, 2010.

Another exact recovery criterion: Normalized ℓ_2-norm Error

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. $(\epsilon = 0.01.)$



Another exact recovery criterion: Normalized ℓ_2-norm Error

J. Vila and P. Schniter, "Expectation-maximization Bernoulli-Gaussian approximate message passing," in Proc. Asilomar Conf. Signals, Syst., Comput., Pacific Grove, CA, Nov. 2011. $(\epsilon = 0.01.)$

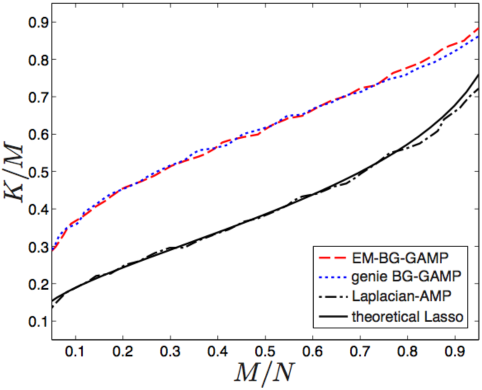
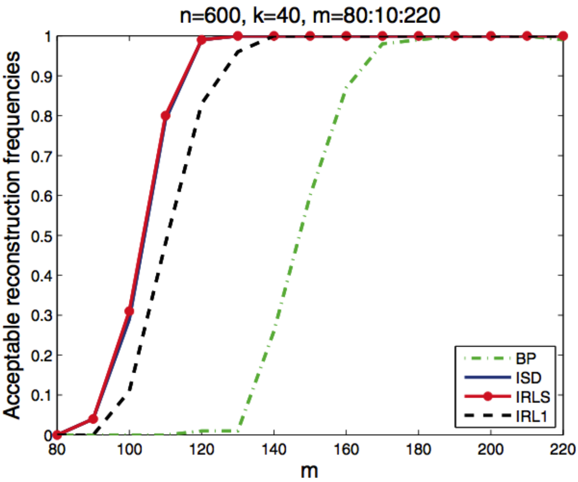


Fig. 1. Empirical noiseless PTCs for Bernoulli-Gaussian signals and theoretical PTC for Lasso.

Another exact recovery criterion: Normalized ℓ_2-norm Error

Y. Wang and W. Yin, "Sparse signal reconstruction via iterative support detection," SIAM J. Imaging Sci., vol. 3, no. 3, 2010. ($\epsilon =?.$)



Two Criteria for Exact Recovery

1. $\mathbf{x}$ is exactly recovered *with respect to support* if
$$S(\hat{\mathbf{x}}) = S(\mathbf{x}) \tag{SC}$$
2. $\mathbf{x}$ is exactly recovered *with respect to normalized squared error* if
$$\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2}{\|\mathbf{x}\|_2^2} \leq \epsilon^2 \tag{ϵ^2C}$$

One does not necessarily imply the other. There are instances, however, when one must be true if the other is true.

My Aims

With regards to running and comparing *simulations of CS recovery*:

- Given a pair $(\hat{\mathbf{x}}, \mathbf{x})$, when does "exact recovery" occur with respect to only one or both criteria?
- What is the role of ϵ^2, and how should we define it?

Presentation Outline

1. Noiseless Case
    - $\mathbf{x} \sim$ Bernoulli-Rademacher sparse signals
    - $\mathbf{x} \sim$ Bernoulli-Gaussian sparse signals
    - Simulations
2. Noisy Case
    - $\mathbf{x} \sim$ Bernoulli-Rademacher sparse signals
    - Simulations
3. Conclusions

Noiseless Case

Measurements $\mathbf{u}$ come from sensing $\mathbf{x}$ by the sensing matrix $\mathbf{\Phi}$, $\|\mathbf{n}\| = 0$:

$$\mathbf{u} = \mathbf{\Phi x} + \mathbf{n}.$$

- Given $\hat{\mathbf{x}}$, the weights minimizing the *measurement modeling error* are

$$\mathbf{y}_{\text{ls}} := \arg\min_{\mathbf{y}'} \|\mathbf{u} - \mathbf{\Phi}_{S(\hat{\mathbf{x}})}\mathbf{y}'\|_2^2 = \mathbf{\Phi}^{\dagger}_{S(\hat{\mathbf{x}})}\mathbf{u}.$$

  With $\hat{\mathbf{x}}$ composed of $\mathbf{y}_{\text{ls}}$, if (SC) then for any $\epsilon^2 \in [0, 1]$ (ϵ^2C).
- If, however, (ϵ^2C) for $\epsilon^2 = 0$ then necessarily (SC).

Now we analyze the behavior of these criteria for signals distributed Bernoulli-Rademacher, Gaussian, and empirically in other ways.

Noiseless Case

Consider the **best case scenario** for sparsity s

- $S(\mathbf{x}) = \{1, 2, \ldots, s\}$;
- $\hat{\mathbf{x}}$ lacks the first $0 < k < s$ elements, i.e., for $n \in \{1, \ldots, k\}(\hat{x}_n = 0)$;
- $\hat{\mathbf{x}}$ has all the others, i.e., $n \in \Omega \backslash \{1, \ldots, k\}(\hat{x}_n = x_n)$.

This means that

- $S(\hat{\mathbf{x}}) \subset S(\mathbf{x})$, i.e., $\hat{\mathbf{x}}$ has no false detections;
- the missed detections do not influence our estimation of the values of the recovered support.

In this case, (ϵ^2C) and not (SC) becomes for $1 \leq k \leq s$

$$\frac{1}{\|\mathbf{x}\|_2^2} \sum_{n=1}^{k} x_n^2 \leq \epsilon^2. \tag{1}$$

Bernoulli-Rademacher Signals

If $\mathbf{x} \sim$ Bernoulli-Rademacher, its non-zero elements are iid equiprobable in $\{-1, 1\}$. In this case, $\|\mathbf{x}\|_2^2 = s$, so

$$P\{(\epsilon^2\mathrm{C}) \wedge \neg(\mathrm{SC})\} = \begin{cases} 1, & k/s \leq \epsilon^2 \\ 0, & \text{else.} \end{cases} \tag{2}$$

For Bernoulli-Rademacher sparse signals *in the best case scenario*:

The parameter ϵ^2 limits the number of missed detections k for a sparsity s.

- As long as $s < \epsilon^{-2}$ for $\mathbf{x} \sim$ Bernoulli-Rademacher, (ϵ^2C) $\rightarrow$ (SC).
- In Maleki et al. 2010, where $s < 800$ and $\epsilon^2 = 10^{-4}$, (ϵ^2C) $\rightarrow$ (SC). However, if for this ϵ^2 the sparsity $s > 10000$, then the two conditions are no longer equivalent.

Bernoulli-Gaussian Signals

Let the s **non-zero elements** of $\mathbf{x} \sim \mathcal{N}(0, \sigma_y^2)$ with variance $\sigma_y^2 > 0$. Define the independent chi-squared rvs

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

Since Y_k and Z_{s-k} are independent, $F_{k,s-k} := [Y_k/k]/[Z_{s-k}/(s-k)] \sim \mathcal{F}(k, s-k)$. Thus, in the best case scenario

$$P\{(\epsilon^2\mathrm{C}) \wedge \neg(\mathrm{SC})\} = P\left\{F_{k,s-k} < \frac{\epsilon^2}{1-\epsilon^2}\frac{1-k/s}{k/s}\right\}. \tag{3}$$

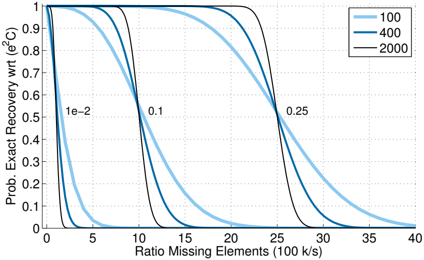
If $k/s > \epsilon^2$, then, for $s \geq 2k$, $P\{F_{k,s-k} < 1+\delta\} > 0.5$ for $\delta > 0$.

For Bernoulli-Gaussian signals *in the best case scenario*:

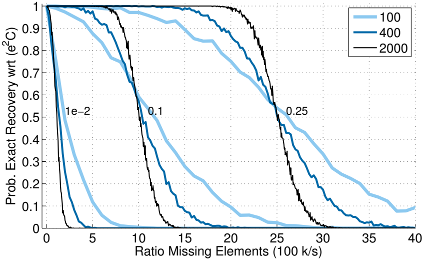
The parameter ϵ^2 limits the number of missed detections k before $((\epsilon^2\mathrm{C}) \wedge\neg\ (\mathrm{SC}))$ is false in a majority sense.

Experiments for several ϵ^2 (labeled) & sparsities (legend)

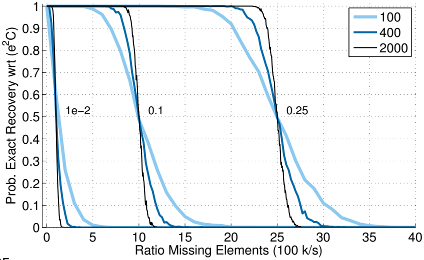
(a) Zero-mean Gaussian (theoretical)



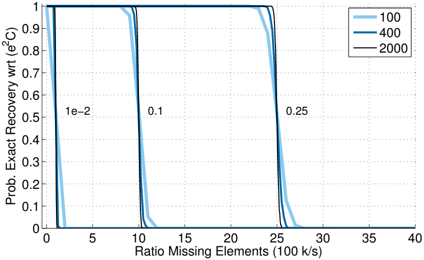
(b) Laplacian (empirical)



(c) Uniform (empirical)



(d) Bimodal Gaussian (empirical)



Noisy Case (assuming (SC))

Measurements $\mathbf{u}$ come from sensing $\mathbf{x}$ by the sensing matrix $\mathbf{\Phi}$, $\|\mathbf{n}\| > 0$:

$$\mathbf{u} = \mathbf{\Phi}\mathbf{x} + \mathbf{n}.$$

Assume (SC), and $\hat{\mathbf{x}}$ is built from $\mathbf{\Phi}^{\dagger}_{S(\mathbf{x})}\mathbf{u}$. The weights in real solution are

$$\mathbf{y} := \arg\min_{\mathbf{y}'} \|\mathbf{u} - \mathbf{n} - \mathbf{\Phi}_{S(\mathbf{x})}\mathbf{y}'\|_2^2 = \mathbf{\Phi}^{\dagger}_{S(\mathbf{x})}(\mathbf{u} - \mathbf{n}).$$

Then, (ϵ^2C) becomes

$$\frac{\|\mathbf{y} - \mathbf{\Phi}^{\dagger}_{S(\mathbf{x})}\mathbf{u}\|_2^2}{\|\mathbf{y}\|_2^2} = \frac{\|\mathbf{\Phi}^{\dagger}_{S(\mathbf{x})}(\mathbf{u} - \mathbf{n}) - \mathbf{\Phi}^{\dagger}_{S(\mathbf{x})}\mathbf{u}\|_2^2}{\|\mathbf{y}\|_2^2} = \frac{\|\mathbf{\Phi}^{\dagger}_{S(\mathbf{x})}\mathbf{n}\|_2^2}{\|\mathbf{y}\|_2^2} \leq \epsilon^2. \quad (4)$$

Hence, for any $\epsilon^2 \in (0, 1]$ we can find an $\mathbf{n}$ such that ((SC) $\wedge \neg$ (ϵ^2C)).
This is different from noiseless case.

Bernoulli-Rademacher Signals Given (SC)

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

$$V_s := \sum_{n=1}^{s} [v_n/\sigma_v]^2 \sim \chi^2(s). \tag{5}$$

If s elements of $\mathbf{x} \sim$ Rademacher, the probability of (ϵ^2C) given (SC)

$$P\{(\epsilon^2\mathrm{C})|(\mathrm{SC})\} = P\left\{V_s < \frac{\epsilon^2 s}{\sigma_v^2}\right\}. \tag{6}$$

Note $P\left\{V_s < s + \delta\right\} > 0.5$ for $\delta > 0$.

For Bernoulli-Rademacher signals, *in the best case scenario*:

Given (SC), if $\epsilon^2 \geq \sigma_v^2$ then (ϵ^2C) in a majority sense.

Bernoulli-Gaussian Signals Given (SC)

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

$$X_s := \sum_{n=1}^{s} [x_n/\sigma_y]^2 \sim \chi^2(s). \tag{7}$$

The ratio V_s/X_s is an F-distributed rv $W_{s,s} := V_s/X_s \sim \mathcal{F}(s, s)$. Thus, the probability of (ϵ^2C) given (SC) is

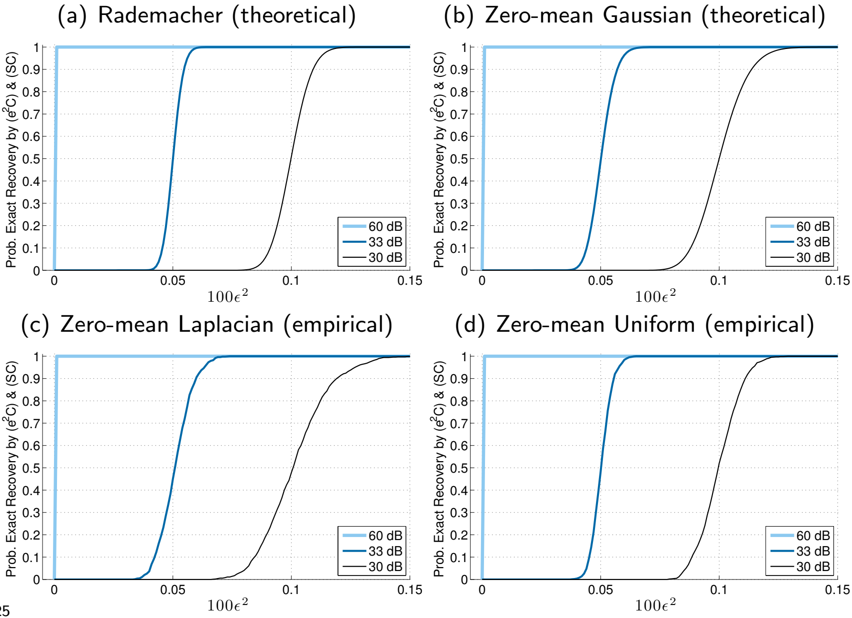
$$P\{(\epsilon^2\mathrm{C})|(\mathrm{SC})\} = P\left\{W_{s,s} < \frac{\sigma_y^2}{\sigma_v^2}\epsilon^2\right\}. \tag{8}$$

Note $P\{W_{s,s} < 1 + \delta\} > 0.5$ for $\delta > 0$.

For Bernoulli-Gaussian signals, *in the best case scenario*:

Given (SC), if $\epsilon^2 \geq \sigma_v^2/\sigma_y^2$ then (ϵ^2C) in a majority sense.

Experiments for several SNR (legend) given (SC)



Noisy Case (assuming not (SC))

Consider (ϵ^2C) is true but not (SC), and best case scenario for sparsity s:

- $S(\mathbf{x}) = \{1, 2, \ldots, s\}$;
- $\hat{\mathbf{x}}$ lacks the first $0 < k < s$ elements, i.e., for $n \in \{1, \ldots, k\}(\hat{x}_n = 0)$;
- $\hat{\mathbf{x}}$ has the others perturbed by $\mathbf{v}$: $n \in \Omega \backslash \{1, \ldots, k\}(\hat{x}_n = x_n + v_n)$.

This means that:

- $S(\hat{\mathbf{x}}) \subset S(\mathbf{x})$, i.e., $\hat{\mathbf{x}}$ has no false detections;
- missed detections do not influence estimations of support recovered;
- values of true detections perturbed only by the noise.

Assume $\mathbf{x}$ and $\mathbf{v}$ are independent, (ϵ^2C) given not (SC) becomes

$$\frac{1}{\|\mathbf{x}\|_2^2} \left[\sum_{n=1}^{k} x_n^2 + \sum_{n=1}^{s-k} v_n^2 \right] \leq \epsilon^2. \tag{9}$$

Bernoulli-Rademacher Signals (assuming not (SC))

Define the rv

$$G_{s-k} := \sum_{n=1}^{s-k} [v_n/\sigma_v]^2 \sim \chi^2(s-k). \tag{10}$$

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

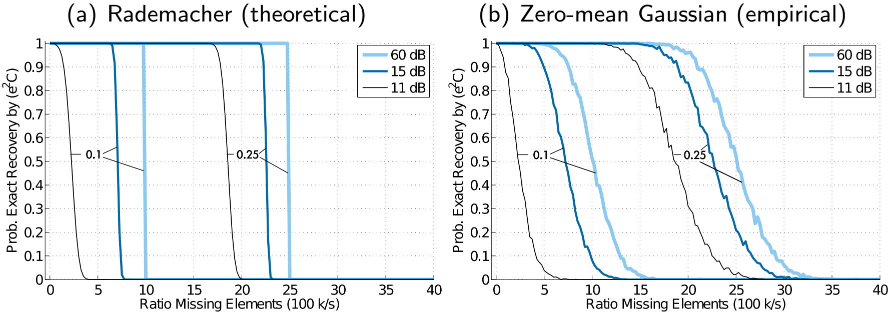
$$P\{(\epsilon^2\mathrm{C}) \wedge \neg(\mathrm{SC})\} = P\left\{G_{s-k} < \frac{\epsilon^2 s - k}{\sigma_v^2}\right\}. \tag{11}$$

Note $P\{G_{s-k} < s-k+\delta\} > 0.5$ for $\delta > 0$.

For Bernoulli-Rademacher signals *in the best case scenario*:

$$\text{If } \frac{\epsilon^2 s - k}{\sigma_v^2} < s-k, \text{then } (\epsilon^2\mathrm{C}) \text{ is false in a majority sense.}$$

Experiments for several ϵ^2 (labeled) & SNR (legend)



Summary and Conclusion

- In theory, we can test for exact recovery with $\hat{\mathbf{x}} \stackrel{?}{=} \mathbf{x}$.
- In practice (finite precision), we must use a different criterion.
- In the *simulation* of compressed sensing recovery algorithms, two different exact recovery criteria have been used:
  1. $\mathbf{x}$ is exactly recovered *with respect to support* if
  $$S(\hat{\mathbf{x}}) = S(\mathbf{x}) \tag{SC}$$
  2. $\mathbf{x}$ is exactly recovered *with respect to normalized squared error* if
  $$\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2}{\|\mathbf{x}\|_2^2} \leq \epsilon^2. \tag{ϵ^2C}$$
- We have shown that

  each does not necessarily imply the other

  ϵ^2 limits the acceptable number of missed detections.

See the paper for more useful tips!