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):
When “Exact Recovery” is Exact Recovery in Compressed Sensing Simulation

Bob L. Sturm

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

August 27, 2012

---

1B. L. Sturm is supported in part by Independent Postdoc Grant 11-105218 from Det Frie Forskningsråd.
Measurements $\mathbf{u}$ come from sensing $\mathbf{x}$ by sensing matrix $\Phi$: $\mathbf{u} = \Phi \mathbf{x} + \mathbf{n}$. We use a recovery algorithm to build $\hat{\mathbf{x}}$ given $\mathbf{u}$ and $\Phi$, e.g., OMP, BP.

Exact Recovery

- In theory, we have no trouble asking $\hat{\mathbf{x}} = \mathbf{x}$.
- 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 $\Omega$ index the columns of $\Phi$, and define the support of $x$ as

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

$x$ is exactly recovered with respect to support if

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

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


One exact recovery criterion in CS simulation: Support


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


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


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 $\ell_2$-norm Error

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

$x$ is exactly recovered with respect to normalized squared error if

$$\frac{\|x - \hat{x}\|_2^2}{\|x\|_2^2} \leq \epsilon^2$$

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


Another exact recovery criterion: Normalized $\ell_2$-norm Error

Another exact recovery criterion: Normalized $\ell_2$-norm Error


Fig. 1. Empirical noiseless PTCs for Bernoulli-Gaussian signals and theoretical PTC for Lasso.
Another exact recovery criterion: Normalized $\ell_2$-norm Error

Two Criteria for Exact Recovery

1. \( x \) is exactly recovered with respect to support if
   \[ S(\hat{x}) = S(x) \]  \hspace{1cm} (SC)

2. \( x \) is exactly recovered with respect to normalized squared error if
   \[ \frac{\|x - \hat{x}\|^2}{\|x\|^2} \leq \epsilon^2 \]  \hspace{1cm} (\epsilon^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 \((\hat{x}, x)\), when does “exact recovery” occur with respect to only one or both criteria?
- What is the role of \(\epsilon^2\), and how should we define it?
1 Noiseless Case
   - $x \sim \text{Bernoulli-Rademacher sparse signals}$
   - $x \sim \text{Bernoulli-Gaussian sparse signals}$
   - Simulations

2 Noisy Case
   - $x \sim \text{Bernoulli-Rademacher sparse signals}$
   - Simulations

3 Conclusions
Measurements \( u \) come from sensing \( x \) by the sensing matrix \( \Phi \), \( \|n\| = 0 \):

\[
u = \Phi x + n.
\]

- Given \( \hat{x} \), the weights minimizing the measurement modeling error are

\[
y_{ls} := \arg \min_{y'} \| u - \Phi_{S(\hat{x})} y' \|_2^2 = \Phi_{S(\hat{x})}^\dagger u.
\]

With \( \hat{x} \) composed of \( y_{ls} \), if (SC) then for any \( \epsilon^2 \in [0, 1] \) \((\epsilon^2 C)\).

- If, however, \((\epsilon^2 C)\) 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.
Consider the **best case scenario** for sparsity $s$

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

This means that

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

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

\[
\frac{1}{\|x\|^2} \sum_{n=1}^{k} x_n^2 \leq \epsilon^2. \quad (1)
\]
Bernoulli-Rademacher Signals

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

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

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

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

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

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

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

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

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

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 $\epsilon^2$ limits the number of missed detections $k$ before $((\epsilon^2 C) \land \neg (SC))$ is false in a majority sense.
Experiments for several $\epsilon^2$ (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 $\Phi$, $\|\mathbf{n}\| > 0$: 

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

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

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

Then, $(\epsilon^2 \mathcal{C})$ becomes

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

Hence, for any $\epsilon^2 \in (0, 1]$ we can find an $\mathbf{n}$ such that ($(\text{SC}) \wedge \neg (\epsilon^2 \mathcal{C})$).

This is different from noiseless case.
Define $v := \Phi_{S(x)}^\dagger n$, and assume its $|S(x)|$ elements are iid $\mathcal{N}(0, \sigma_v^2)$ and independent of $y$. Define the chi-squared-distributed rv

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

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

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

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

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

Given (SC), if $\epsilon^2 \geq \sigma_v^2$ then $(\epsilon^2C)$ in a majority sense.
Assume $s$ non-zero elements of $x \sim \mathcal{N}(0, \sigma_y^2)$, independent of $v$. Define

$$X_s := \sum_{n=1}^{s} \left[ \frac{x_n}{\sigma_y} \right]^2 \sim \chi^2(s).$$

(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 $(\epsilon^2C)$ given (SC) is

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

(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 $(\epsilon^2C)$ in a majority sense.
Experiments for several SNR (legend) given (SC)

(a) Rademacher (theoretical)

(b) Zero-mean Gaussian (theoretical)

(c) Zero-mean Laplacian (empirical)

(d) Zero-mean Uniform (empirical)
Noisy Case (assuming not (SC))

Consider \((\epsilon^2 C)\) is true but not (SC), and best case scenario for sparsity \(s\):
- \(S(x) = \{1, 2, \ldots, s\}\);
- \(\hat{x}\) lacks the first \(0 < k < s\) elements, i.e., for \(n \in \{1, \ldots, k\}\)(\(\hat{x}_n = 0\));
- \(\hat{x}\) has the others perturbed by \(v\): \(n \in \Omega \setminus \{1, \ldots, k\}\)(\(\hat{x}_n = x_n + v_n\)).

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

Assume \(x\) and \(v\) are independent, \((\epsilon^2 C)\) given not (SC) becomes

\[
\frac{1}{\|x\|^2} \left[ \sum_{n=1}^{k} x_n^2 + \sum_{n=1}^{s-k} v_n^2 \right] \leq \epsilon^2. \tag{9}
\]
Bernoulli-Rademacher Signals (assuming not (SC))

Define the rv

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

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

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

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

For Bernoulli-Rademacher signals in the best case scenario:

If \( \frac{\epsilon^2 s - k}{\sigma_v^2} < s - k \), then \((\epsilon^2 C)\) is false in a majority sense.
Experiments for several $\epsilon^2$ (labeled) & SNR (legend)

(a) Rademacher (theoretical)

(b) Zero-mean Gaussian (empirical)
Summary and Conclusion

• In theory, we can test for exact recovery with $\hat{x} = x$.
• 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. $x$ is exactly recovered with respect to support if
     \[ S(\hat{x}) = S(x) \]  
     (SC)
  2. $x$ is exactly recovered with respect to normalized squared error if
     \[ \frac{\|x - \hat{x}\|^2_2}{\|x\|^2_2} \leq \epsilon^2. \]  
     ($\epsilon^2$C)
• We have shown that each does not necessarily imply the other.
• $\epsilon^2$ limits the acceptable number of missed detections.

See the paper for more useful tips!