Enhanced Compressed Sensing based on Iterative Support Detection

Wotao Yin

Department of Computational and Applied Mathematics
Rice University

Joint work with Yilun Wang

Supported by ONR and NSF

Notation

- x: sparse signal, has $\leq k$ nonzero entries
- $b = Ax$: CS measurements

Notation

- x: sparse signal, has $\leq k$ nonzero entries
- $b = Ax$: CS measurements

- ℓ_0-problem: $\min \|x\|_0$, s.t. $Ax = b$. Exact recovery needs $m \geq 2k$ for Gaussian A
- ℓ_1-problem: $\min \|x\|_1$, s.t. $Ax = b$. Needs a much bigger m Also called *Basis Pursuit*

Outline

1. Overview
   - The Approach
   - Simple Examples
2. Theoretical Results
   - Summary
   - The Null Space Property
   - Recoverability Improvement
3. Numerical Results
   - Noiseless measurements
   - Noisy measurements
   - A failed case
4. Conclusions

Approach

Goal: to beat the ℓ_1-minimization, i.e., basis pursuit

- Recover x from less measurements
- Remain computationally tractable

Approach

Goal: to beat the ℓ_1-minimization, i.e., basis pursuit

- Recover x from less measurements
- Remain computationally tractable

Iterative approach:

- If ℓ_1-minimization fails, detect *good* in the (wrong) solution

Approach

Goal: to beat the ℓ_1-minimization, i.e., basis pursuit

- Recover x from less measurements
- Remain computationally tractable

Iterative approach:

- If ℓ_1-minimization fails, detect *good* in the (wrong) solution
- Remove the discoveries from the ℓ_1-norm.

Approach

Goal: to beat the ℓ_1-minimization, i.e., basis pursuit

- Recover x from less measurements
- Remain computationally tractable

Iterative approach:

- If ℓ_1-minimization fails, detect *good* in the (wrong) solution
- Remove the discoveries from the ℓ_1-norm.
  T: remaining entries, $\|x_T\|_1 = \sum_{i \in T} |x_i|$,
  T^C: discoveries = correct $\cup$ wrong, out of ℓ_1-norm.
  $t = |T|$

Approach

Goal: to beat the ℓ_1-minimization, i.e., basis pursuit

- Recover x from less measurements
- Remain computationally tractable

Iterative approach:

- If ℓ_1-minimization fails, detect *good* in the (wrong) solution
- Remove the discoveries from the ℓ_1-norm.
  T: remaining entries, $\|x_T\|_1 = \sum_{i \in T} |x_i|$,
  T^C: discoveries = correct $\cup$ wrong, out of ℓ_1-norm.
  $t = |T|$
- Solve

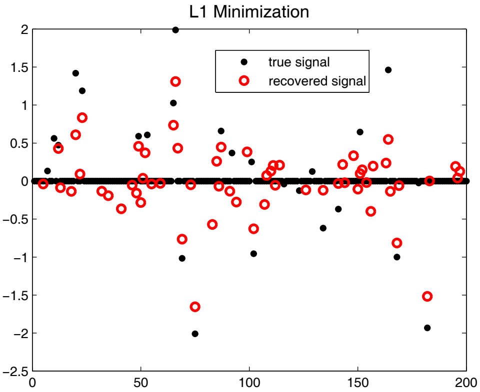
Truncate ℓ_1-problem: $\min_x \|x_T\|_1$, s.t. $Ax = b$.

A Simple Example

Setup:

- $n = 200$, $k = 25$, $m = 2k = 50$, A is Gaussian random

Basis pursuit result: $x^{(1)}$

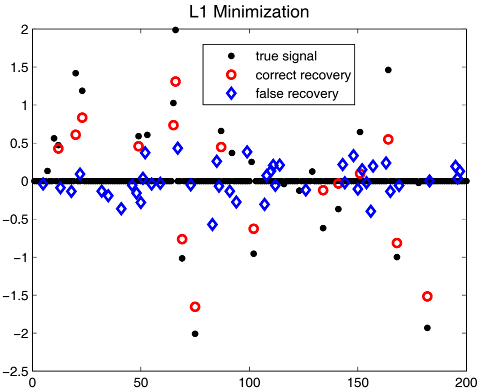


A Simple Example

Setup:

- $n = 200$, $k = 25$, $m = 2k = 50$, A is Gaussian random

Basis pursuit result: $x^{(1)}$

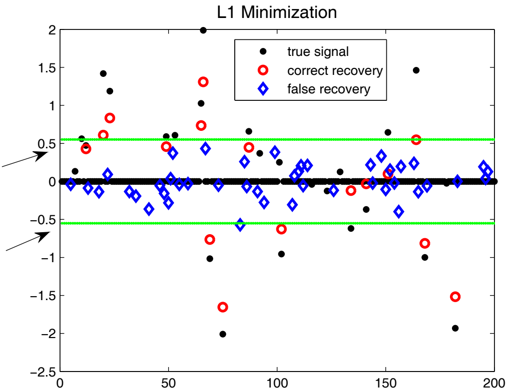


A Simple Example

Setup:

- $n = 200$, $k = 25$, $m = 2k = 50$, A is Gaussian random

Basis pursuit result: $x^{(1)}$, threshold $\epsilon = \|x^{(1)}\|_\infty/3$

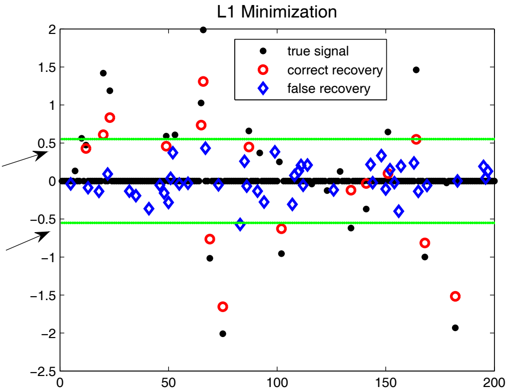


A Thresholding Framework

- Initialize: $j \leftarrow 1$ and $T = \{1, 2, \ldots, n\}$.
- While *not converged* do
  1. Truncated ℓ_1-minimization:
  $$x^{(j)} \leftarrow \min \|x_T\|_1 \quad \text{s.t. } Ax = b.$$
  2. Support detection *by thresholding*:
  $$\epsilon \leftarrow \|x^{(j)}\|_\infty / 3^j,$$
  $$T \leftarrow \{i : |x_i^{(j)}| < \epsilon\}.$$

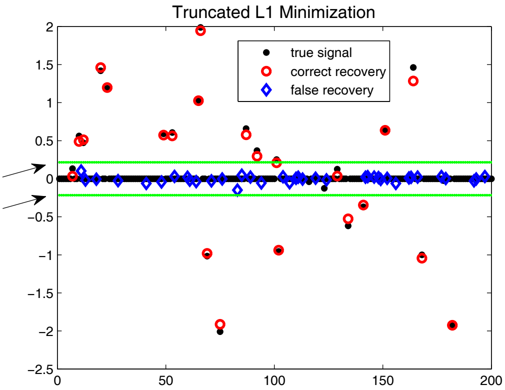
Results of Iterative Thresholding

Basis pursuit result: $x^{(1)}$, threshold $\epsilon = \|x^{(1)}\|_\infty/3$



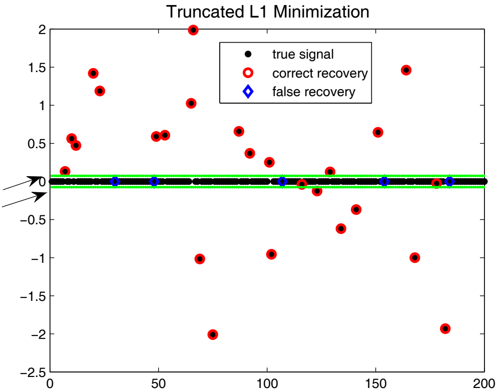
Results of Iterative Thresholding

Truncated ℓ_1-result: $x^{(2)}$, reduced threshold $\epsilon = \|x^{(2)}\|_\infty / 3^2$



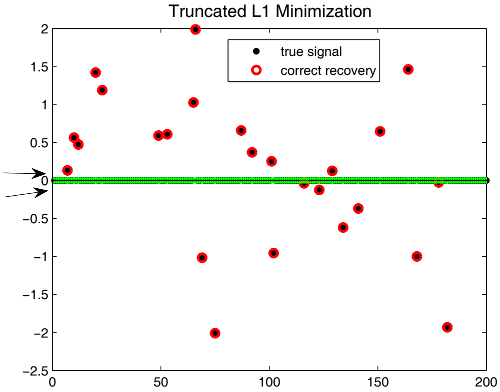
Results of Iterative Thresholding

Truncated ℓ_1-result: $x^{(3)}$, reduced threshold $\epsilon = \|x^{(3)}\|_\infty / 3^3$



Results of Iterative Thresholding

Truncated ℓ_1-result: $x^{(4)}$, exact recovery!

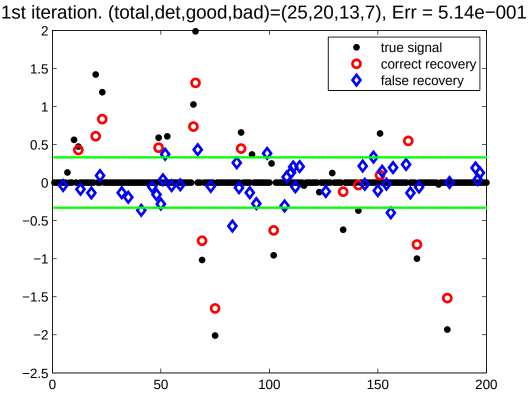


Robustness Test

Try tighter thresholds:

$$\epsilon = \|x^{(j)}\|_\infty / 5^j.$$

$j = 1$, basis pursuit result:

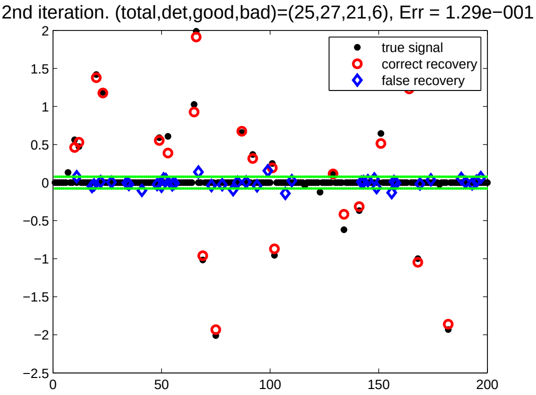


Robustness Test

Try tighter thresholds:

$$\epsilon = \|x^{(j)}\|_\infty / 5^j.$$

$j = 2$, truncated ℓ_1-minimization result:

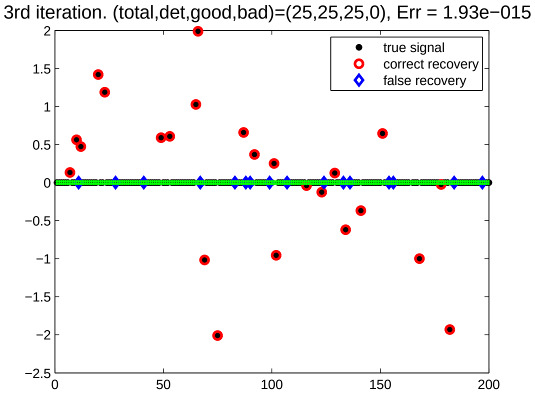


Robustness Test

Try tighter thresholds:

$$\epsilon = \|x^{(j)}\|_\infty / 5^j.$$

$j = 3$, truncated ℓ_1-minimization result:

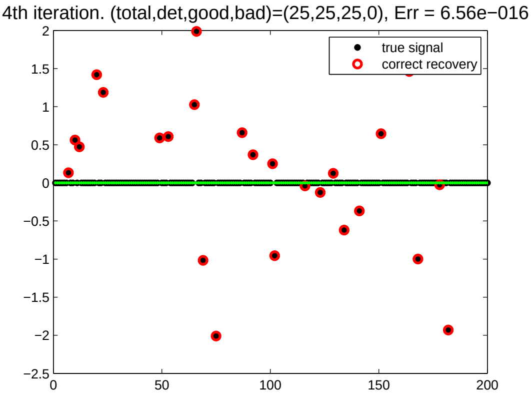


Robustness Test

Try tighter thresholds:

$$\epsilon = \|x^{(j)}\|_\infty / 5^j.$$

$j = 4$, truncated ℓ_1-minimization result: exact recovery!



Outline

1. Overview
   - The Approach
   - Simple Examples
2. Theoretical Results
   - Summary
   - The Null Space Property
   - Recoverability Improvement
3. Numerical Results
   - Noiseless measurements
   - Noisy measurements
   - A failed case
4. Conclusions

Summary of Results

- Q: Condition for an exact recovery?

  Truncated Null Space Property (T-NSP) holds with $\gamma < 1$.

Summary of Results

- Q: Condition for an exact recovery?

  Truncated Null Space Property (T-NSP) holds with $\gamma < 1$.
- Q: How good is a support discovery?

  $= \gamma^{(j)} - \gamma^{(j+1)}$. To have $\gamma^{(j)} > \gamma^{(j+1)}$:

  (inc. correct discoveries) / (inc. wrong discoveries) $> \gamma^{(j)}$

Summary of Results

- Q: Condition for an exact recovery?

  Truncated Null Space Property (T-NSP) holds with $\gamma < 1$.
- Q: How good is a support discovery?

  $= \gamma^{(j)} - \gamma^{(j+1)}$. To have $\gamma^{(j)} > \gamma^{(j+1)}$:

  (inc. correct discoveries) / (inc. wrong discoveries) $> \gamma^{(j)}$
- Q: Not knowing the exact solution
  - how to have enough correct discoveries?
    thresholding for fast decaying signals. Ranking is not as robust.

Summary of Results

- Q: Condition for an exact recovery?

  Truncated Null Space Property (T-NSP) holds with $\gamma < 1$.
- Q: How good is a support discovery?

  $= \gamma^{(j)} - \gamma^{(j+1)}$. To have $\gamma^{(j)} > \gamma^{(j+1)}$:

  (inc. correct discoveries) / (inc. wrong discoveries) $> \gamma^{(j)}$
- Q: Not knowing the exact solution
  - how to have enough correct discoveries?
    thresholding for fast decaying signals. Ranking is not as robust.
  - how to measure improvement?
    compute the size of tail.

Summary of Results

- Q: Condition for an exact recovery?

  Truncated Null Space Property (T-NSP) holds with $\gamma < 1$.
- Q: How good is a support discovery?

  $= \gamma^{(j)} - \gamma^{(j+1)}$. To have $\gamma^{(j)} > \gamma^{(j+1)}$:

  (inc. correct discoveries) / (inc. wrong discoveries) $> \gamma^{(j)}$
- Q: Not knowing the exact solution
  - how to have enough correct discoveries?
    thresholding for fast decaying signals. Ranking is not as robust.
  - how to measure improvement?
    compute the size of tail.
  - when to stop?
    tail is zero or small enough.

Null Space Property

- A sufficient condition for $\min\{\|x\|_1 : Ax = b\}$ to yield the right x.

Null Space Property

- A sufficient condition for $\min\{\|x\|_1 : Ax = b\}$ to yield the right x.
- Observe $\{x : Ax = b\} = x + \mathcal{N}(A)$. We need $\|x\|_1 < \|x + v\|_1$ for all $v \in \mathcal{N}(A)$.

Null Space Property

- A sufficient condition for $\min\{\|x\|_1 : Ax = b\}$ to yield the right x.
- Observe $\{x : Ax = b\} = x + \mathcal{N}(A)$. We need $\|x\|_1 < \|x + v\|_1$ for all $v \in \mathcal{N}(A)$.
- Let $S = \{i : x_i \neq 0\}$.

$$
\begin{aligned}
\|x + v\|_1 &= \|x_S + v_S\|_1 + \|0 + v_{S^C}\|_1 \\
&= (\|x_S + v_S\|_1 - \|x_S\|_1 + \|v_S\|_1) + \|x_S\|_1 + \|v_{S^C}\|_1 - \|v_S\|_1 \\
&= \underbrace{(\|x_S + v_S\|_1 - \|x_S\|_1 + \|v_S\|_1)}_{\geq 0} + \|x\|_1 + \underbrace{\|v_{S^C}\|_1 - \|v_S\|_1}_{\geq 0??}.
\end{aligned}
$$

We need $\|v_S\|_1 < \|v_{S^C}\|_1$.

Null Space Property

- A sufficient condition for $\min\{\|x\|_1 : Ax = b\}$ to yield the right x.
- Observe $\{x : Ax = b\} = x + \mathcal{N}(A)$. We need $\|x\|_1 < \|x + v\|_1$ for all $v \in \mathcal{N}(A)$.
- Let $S = \{i : x_i \neq 0\}$.

$$\begin{aligned} \|x+v\|_1 &= \|x_S + v_S\|_1 + \|0 + v_{S^C}\|_1 \\ &= (\|x_S + v_S\|_1 - \|x_S\|_1 + \|v_S\|_1) + \|x_S\|_1 + \|v_{S^C}\|_1 - \|v_S\|_1 \\ &= \underbrace{(\|x_S + v_S\|_1 - \|x_S\|_1 + \|v_S\|_1)}_{\geq 0} + \|x\|_1 + \underbrace{\|v_{S^C}\|_1 - \|v_S\|_1}_{\geq 0??}. \end{aligned}$$

We need $\|v_S\|_1 < \|v_{S^C}\|_1$.

- A necessary condition for uniform exact recovery for all $|S|$-sparse signals.

Null Space Property

Definition (Cohen-Dahmen-DeVore and others)

$A \in \mathbb{R}^{m \times n}$ has the *Null Space Property* (*NSP*) with order L and $\gamma > 0$ if

$$\|v_S\|_1 \leq \gamma \|v_{S^c}\|_1, \quad \forall |S| \leq L, v \in \mathcal{N}(A).$$

Null Space Property

Definition (Cohen-Dahmen-DeVore and others)

$A \in \mathbb{R}^{m\times n}$ has the *Null Space Property* (*NSP*) with order L and $\gamma > 0$ if

$$\|v_S\|_1 \le \gamma \|v_{S^c}\|_1, \quad \forall |S| \le L, v \in \mathcal{N}(A).$$

Usage:

- A has NSP with L and $\gamma < 1 \Rightarrow$
  - Uniform exact recovery of all L-sparse signals
  - For $k > L$, a recovery error bound for k-sparse signals

Null Space Property

Definition (Cohen-Dahmen-DeVore and others)

$A \in \mathbb{R}^{m \times n}$ has the *Null Space Property* (*NSP*) with order L and $\gamma > 0$ if

$$\|v_S\|_1 \leq \gamma \|v_{S^c}\|_1, \quad \forall |S| \leq L, v \in \mathcal{N}(A).$$

Usage:

- A has NSP with L and $\gamma < 1 \Rightarrow$
  - Uniform exact recovery of all L-sparse signals
  - For $k > L$, a recovery error bound for k-sparse signals
- Minimal γ is monotonic in L

Null Space Property

Definition (Cohen-Dahmen-DeVore and others)

$A \in \mathbb{R}^{m \times n}$ has the *Null Space Property* (*NSP*) with order L and $\gamma > 0$ if

$$\|v_S\|_1 \leq \gamma \|v_{S^c}\|_1, \quad \forall |S| \leq L, v \in \mathcal{N}(A).$$

Usage:

- A has NSP with L and $\gamma < 1 \Rightarrow$
  - Uniform exact recovery of all L-sparse signals
  - For $k > L$, a recovery error bound for k-sparse signals
- Minimal γ is monotonic in L
- NSP is weaker than RIP and can be obtained from RIP

Null Space Property

Definition (Cohen-Dahmen-DeVore and others)

$A \in \mathbb{R}^{m \times n}$ has the *Null Space Property* (*NSP*) with order L and $\gamma > 0$ if

$$\|v_S\|_1 \leq \gamma \|v_{S^c}\|_1, \quad \forall |S| \leq L, v \in \mathcal{N}(A).$$

Usage:

- A has NSP with L and $\gamma < 1 \Rightarrow$
  - Uniform exact recovery of all L-sparse signals
  - For $k > L$, a recovery error bound for k-sparse signals
- Minimal γ is monotonic in L
- NSP is weaker than RIP and can be obtained from RIP
- NSP is more essential than RIP for basis pursuit (left multiplying A by a nonsingular matrix changes RIP but not NSP)

Truncated Null Space Property

Definition (Y.-Wang)

$A \in \mathbb{R}^{m \times n}$ has the *Truncated Null Space Property* (T-NSP) with t, L, and γ, written as T-NSP(t, L, γ), if

$$\|v_S\|_1 \leq \gamma \|v_{T \setminus S}\|_1, \quad \forall S \subset T, |S| \leq L, |T| = t, v \in \mathcal{N}(A).$$

Intuitively, T-NSP(t, L, γ) $\Leftrightarrow$ all length-t subvectors of $v \in \mathcal{N}(A)$ satisfy the inequality of NSP(L, γ)

Truncated Null Space Property

Definition (Y.-Wang)

$A \in \mathbb{R}^{m \times n}$ has the *Truncated Null Space Property* (T-NSP) with t, L, and γ, written as T-NSP(t, L, γ), if

$$\|v_S\|_1 \leq \gamma \|v_{T \setminus S}\|_1, \quad \forall S \subset T, |S| \leq L, |T| = t, v \in \mathcal{N}(A).$$

Intuitively, T-NSP(t, L, γ) $\Leftrightarrow$ all length-t subvectors of $v \in \mathcal{N}(A)$ satisfy the inequality of NSP(L, γ)

Theorem (Y.-Wang)

For T given , if A satisfies T-NSP$(|T|, L, \gamma)$ where $\gamma < 1$, then truncated ℓ_1-minimization over the support of T yields an exact recovery.

Recoverability Improvement

Theorem (Y.-Wang)

Suppose A satisfies both T-NSP(t^1, L^1, γ^1) and T-NSP(t^2, L^2, γ^2) where $t^2 < t^1$ and γ^1 and γ^2 are minimal. Then,

$$\frac{L^1 - L^2}{(t^1 - t^2) - (L^1 - L^2)} > \gamma^1 \Longrightarrow \gamma^2 < \gamma^1.$$

Recoverability Improvement

Theorem (Y.-Wang)

Suppose A satisfies both T-NSP(t^1, L^1, γ^1) and T-NSP(t^2, L^2, γ^2) where $t^2 < t^1$ and γ^1 and γ^2 are minimal. Then,

$$\frac{L^1 - L^2}{(t^1 - t^2) - (L^1 - L^2)} > \gamma^1 \Longrightarrow \gamma^2 < \gamma^1.$$

Interpretation:

- $t^1 = |T^1|$: numbers of entries in T before detection
  $t^2 = |T^2|$: numbers of entries in T after detection

Recoverability Improvement

Theorem (Y.-Wang)

Suppose A satisfies both T-NSP(t^1, L^1, γ^1) and T-NSP(t^2, L^2, γ^2) where $t^2 < t^1$ and γ^1 and γ^2 are minimal. Then,

$$\frac{L^1 - L^2}{(t^1 - t^2) - (L^1 - L^2)} > \gamma^1 \Longrightarrow \gamma^2 < \gamma^1.$$

Interpretation:

- $t^1 = |T^1|$: numbers of entries in T before detection
  $t^2 = |T^2|$: numbers of entries in T after detection
- $t^1 - t^2 =$ decrease in $|T|$ = increase in the total discoveries

Recoverability Improvement

Theorem (Y.-Wang)

Suppose A satisfies both T-NSP(t^1, L^1, γ^1) *and T-NSP*(t^2, L^2, γ^2) *where* $t^2 < t^1$ *and* γ^1 *and* γ^2 *are minimal. Then,*

$$\frac{L^1 - L^2}{(t^1 - t^2) - (L^1 - L^2)} > \gamma^1 \Longrightarrow \gamma^2 < \gamma^1.$$

Interpretation:

- $t^1 = |T^1|$: numbers of entries in T before detection
  $t^2 = |T^2|$: numbers of entries in T after detection
- $t^1 - t^2 =$ decrease in $|T|$ = increase in the total discoveries
- $L^1 - L^2 =$ increase in the correct discoveries

Recoverability Improvement

Theorem (Y.-Wang)

Suppose A satisfies both T-NSP(t^1, L^1, γ^1) *and T-NSP*(t^2, L^2, γ^2) *where* $t^2 < t^1$ *and* γ^1 *and* γ^2 *are minimal. Then,*

$$\frac{L^1 - L^2}{(t^1 - t^2) - (L^1 - L^2)} > \gamma^1 \Longrightarrow \gamma^2 < \gamma^1.$$

Interpretation:

- $t^1 = |T^1|$: numbers of entries in T before detection
  $t^2 = |T^2|$: numbers of entries in T after detection
- $t^1 - t^2 =$ decrease in $|T|$ = increase in the total discoveries
- $L^1 - L^2 =$ increase in the correct discoveries
- $\gamma^2 < \gamma^1$: recoverability improved (recall $\gamma < 1 \Rightarrow$ exact recovery)
- To improve, it is sufficient to have
  (inc. corr. discoveries) / (inc. false discoveries) $> \gamma^1$

RICE

Recoverability Improvement

Theorem (Y.-Wang)

Suppose A satisfies both T-NSP(t^1, L^1, γ^1) and T-NSP(t^2, L^2, γ^2) where $t^2 < t^1$ and γ^1 and γ^2 are minimal. Then,

$$\frac{L^1 - L^2}{(t^1 - t^2) - (L^1 - L^2)} > \gamma^1 \Longrightarrow \gamma^2 < \gamma^1.$$

Interpretation:

- $t^1 = |T^1|$: numbers of entries in T before detection
  $t^2 = |T^2|$: numbers of entries in T after detection
- $t^1 - t^2 =$ decrease in $|T|$ = increase in the total discoveries
- $L^1 - L^2 =$ increase in the correct discoveries
- $\gamma^2 < \gamma^1$: recoverability improved (recall $\gamma < 1 \Rightarrow$ exact recovery)
- To improve, it is sufficient to have
  (inc. corr. discoveries) / (inc. false discoveries) $> \gamma^1$
- Result is in dependent of support detectors.

Results for Random Sampling

Theorem (Y.-Wang, an extension to Candés-Tao and Zhang)

For Gaussian random A (or any rank-m matrix A such that $BA^\top = 0$ where $B \in \mathbb{R}^{(n-m)\times m}$ is Gaussian random), a sufficient condition for exact recovery ***with high probability*** *is*

$$\|x_T\|_0 < \frac{C^2}{4} \frac{m-d}{1+\log\frac{n-d}{m-d}},$$

where $d = n - |T|$ and C is an independent constant.

Results for Random Sampling

Theorem (Y.-Wang, an extension to Candés-Tao and Zhang)

For Gaussian random A (or any rank-m matrix A such that $BA^\top = 0$ where $B \in \mathbb{R}^{(n-m)\times m}$ is Gaussian random), a sufficient condition for exact recovery ***with high probability*** *is*

$$\|x_T\|_0 < \frac{C^2}{4} \frac{m-d}{1+\log\frac{n-d}{m-d}},$$

where $d = n - |T|$ and C is an independent constant.

Application: Bound C and show that

$$-1 < \frac{\partial RHS}{\partial d} < 0,$$

leaving room for incorrect discoveries.

Outline

Numerical Results

Experiment 1: noiseless measurements

- $n = 100$, $m = 50$
- $k = 9, \ldots, 21$. Each k had 200 trials.
- x: sparse Gaussian signals
- A: Gaussian random
- Successful recovery declared if $\|x^{(j)} - x\|_\infty \leq 10^{-3}$
- Thresholds: $\epsilon = \|x^{(j)}\|_\infty / 2^j$

Numerical Results

Experiment 1: noiseless measurements

- $n = 100$, $m = 50$
- $k = 9, \ldots, 21$. Each k had 200 trials.
- x: sparse Gaussian signals
- A: Gaussian random
- Successful recovery declared if $\|x^{(j)} - x\|_\infty \le 10^{-3}$
- Thresholds: $\epsilon = \|x^{(j)}\|_\infty / 2^j$

Empirical exact recovery conditions:

- Basis pursuit:

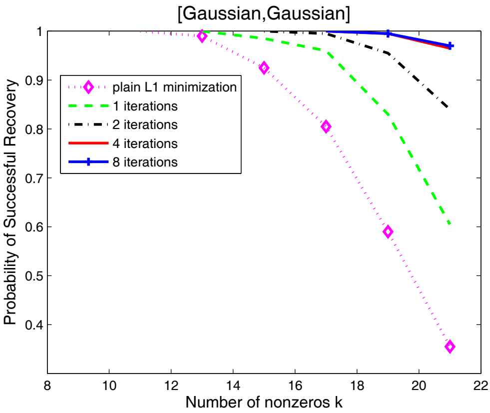
$$k \le \frac{m}{5}$$

- With iterative support detection:

$$k \le \frac{m}{3}$$

Numerical Results

Percentage of Successful Recoveries

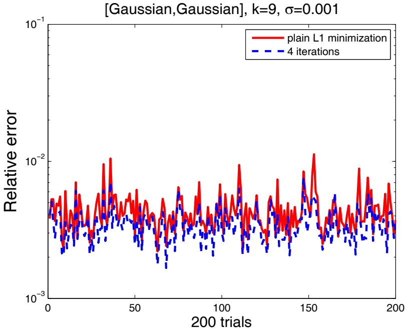


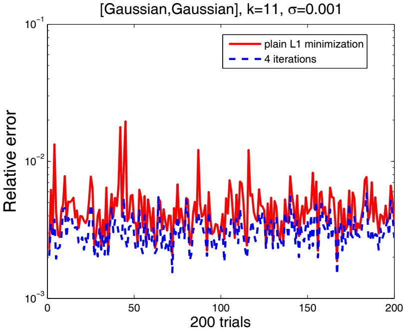
Numerical Results

Experiment 2: noisy measurements

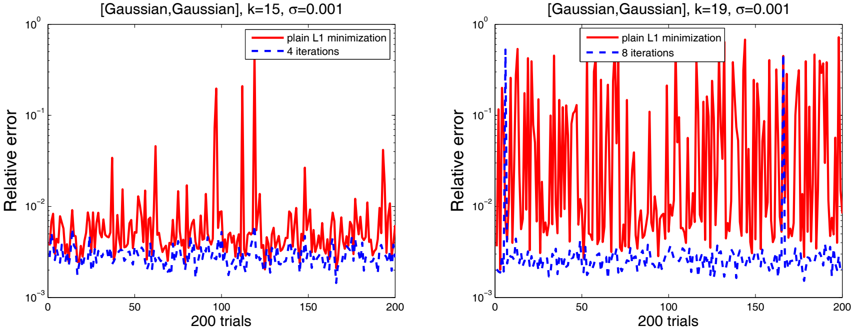
- $n = 100$, $m = 50$
- $k = 9, 11, 15, 19$. Each k had 200 trials.
- x: sparse Gaussian signals
- A: Gaussian random
- $b = Ax + z$, where $z \sim N(0, 0.001)$
- Logarithms of relative errors of $x^{(j)}$ to x are plotted
- Thresholds: $\epsilon = \|x^{(j)}\|_\infty / 2^j$

Numerical Results



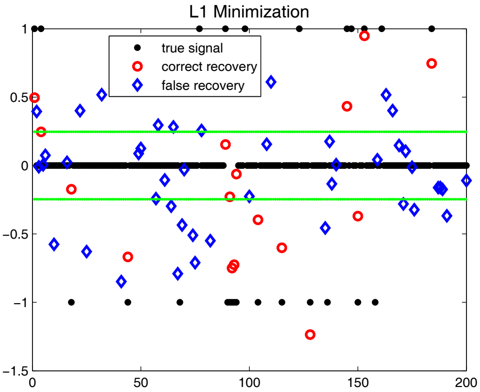


Numerical Results



Numerical Results

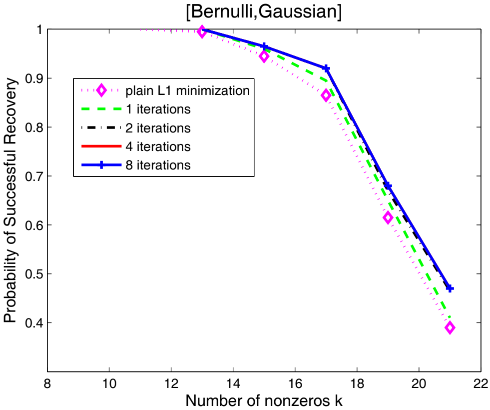
Experiment 3: sparse signals with nonzero $= \pm 1$, noiseless measurements



Excessive false detections!

Numerical Results

Experiment 3: signals with Bernoulli nonzeros, noiseless measurements



Little improvement over basis pursuit.

Outline

Conclusions

Conclusions

- Effective support detection improves CS recovery

Conclusions

Conclusions

- Effective support detection improves CS recovery
- In particular, iterative thresholding is effective on sparse signals with fast decaying distribution of nonzero values

Conclusions

Conclusions

- Effective support detection improves CS recovery
- In particular, iterative thresholding is effective on sparse signals with fast decaying distribution of nonzero values
- Computationally tractable
  - one ℓ_1-minimization per iteration, can be warm-started
  - only a small number of iterations are needed

On-going Work

On-going work

- other types of signals: images, video, etc.

On-going Work

On-going work

- other types of signals: images, video, etc.
- other priors: $\|\Phi x\|_p$, $p \leq 1$, and $TV(x)$

On-going Work

On-going work

- other types of signals: images, video, etc.
- other priors: $\|\Phi x\|_p$, $p \leq 1$, and $TV(x)$
- further theoretical analysis is underway

On-going Work

On-going work

- other types of signals: images, video, etc.
- other priors: $\|\Phi x\|_p$, $p \leq 1$, and $TV(x)$
- further theoretical analysis is underway
- apply to greedy algorithms (OMP, ROMP, CoSaMP, ...).

Acknowledgements

- Colleagues
  Rice: Rich Baraniuk, Volkan Cevher, Kevin Kelly, Yin Zhang
  Columbia: Donald Goldfarb, Shiqian Ma, Zaiwen Wen
  UCLA: Stan Osher, Jerome Darbon, Bin Dong, Yu Mao
  Los Alamos: Rick Chartrand, Simon Morgan
  Alabama: Weihong Guo
- Students: Junfeng Yang, Yilun Wang
- Funding Agencies: ONR, NSF

CS Resources: `www.dsp.ece.rice.edu/cs`

Our algorithms: `www.caam.rice.edu/~optimization/L1`

Thank You!