Recovering Structured Signals from Noisy Measurements: Where Least-Squares Meets Compressed Sensing

Babak Hassibi

joint work with
Samet Oymak and Christos Thrampoulidis

California Institute of Technology

Matheon Workshop on Compressed Sensing and it Applications
Technical University of Berlin
December 11, 2013

Outline

- **Introduction**
  - least-squares
- **Noiseless Structured Signal Recovery**
  - compressed sensing
  - convex relaxation
- **Phase Transitions**
  - escape-through-mesh, Gaussian widths, statistical dimension
  - non-uniform sparsity, low rank matrix recovery
  - connection to denoising
- **Structured Signal Recovery in Noise**
  - generalized LASSO
  - formulae for the squared error
- **Conclusion**

Least-Squares

Suppose we are confronted with the *noisy* measurements:

$$y = Ax + z,$$

where $A \in \mathcal{R}^{m \times n}$ is the measurement matrix, $y \in \mathcal{R}^m$ is the measurement vector, $x \in \mathcal{R}^n$ is the unknown desired signal, and $z \in \mathcal{R}^n$ is the unknown noise vector.

Least-Squares

Suppose we are confronted with the *noisy* measurements:

$$y = Ax + z,$$

where $A \in \mathcal{R}^{m \times n}$ is the measurement matrix, $y \in \mathcal{R}^m$ is the measurement vector, $x \in \mathcal{R}^n$ is the unknown desired signal, and $z \in \mathcal{R}^n$ is the unknown noise vector.

In the general case, to be meaningful, we require that

$$m \geq n.$$

Least-Squares

Suppose we are confronted with the *noisy* measurements:

$$y = Ax + z,$$

where $A \in \mathcal{R}^{m \times n}$ is the measurement matrix, $y \in \mathcal{R}^m$ is the measurement vector, $x \in \mathcal{R}^n$ is the unknown desired signal, and $z \in \mathcal{R}^n$ is the unknown noise vector.

In the general case, to be meaningful, we require that

$$m \geq n.$$

A popular method for recovering x, is the least-squares criterion

$$\min_x \|y - Ax\|_2^2,$$

whose solution is famously given by

$$\hat{x} = \left(A^T A\right)^{-1} A^T y.$$

Least-Squares

The squared error is then given by

$$\|x - \hat{x}\|_2^2 = z^T A \left(A^T A\right)^{-2} A^T z.$$

Least-Squares

The squared error is then given by

$$\|x - \hat{x}\|_2^2 = z^T A \left(A^T A\right)^{-2} A^T z.$$

- The **mean square error**: Assume z has iid $N(0, \sigma^2)$ entries. Then

$$E\|x - \hat{x}\|_2^2 = \sigma^2 \text{trace} \left(A^T A\right)^{-1}.$$

Least-Squares

The squared error is then given by

$$\|x - \hat{x}\|_2^2 = z^T A \left(A^T A\right)^{-2} A^T z.$$

- The **mean square error**: Assume z has iid $N(0, \sigma^2)$ entries. Then

$$E\|x - \hat{x}\|_2^2 = \sigma^2 \text{trace} \left(A^T A\right)^{-1}.$$

In fact, when m and n grow, we do not even need the expectation: the squared error concentrates around $\sigma^2 \text{trace} \left(A^T A\right)^{-1}$.

Least-Squares

The squared error is then given by

$$\|x - \hat{x}\|_2^2 = z^T A \left(A^T A\right)^{-2} A^T z.$$

- The **mean square error**: Assume z has iid $N(0, \sigma^2)$ entries. Then

$$E\|x - \hat{x}\|_2^2 = \sigma^2 \text{trace} \left(A^T A\right)^{-1}.$$

In fact, when m and n grow, we do not even need the expectation: the squared error concentrates around $\sigma^2 \text{trace} \left(A^T A\right)^{-1}$.When A has iid $N(0, \frac{1}{m})$ entries, $A^T A$ is a *Wishart matrix* whose asymptotic eigendistribution is well known.

Least-Squares

The squared error is then given by

$$\|x - \hat{x}\|_2^2 = z^T A \left(A^T A\right)^{-2} A^T z.$$

- The **mean square error**: Assume z has iid $N(0, \sigma^2)$ entries. Then

$$E\|x - \hat{x}\|_2^2 = \sigma^2 \text{trace} \left(A^T A\right)^{-1}.$$

In fact, when m and n grow, we do not even need the expectation: the squared error concentrates around $\sigma^2 \text{trace} \left(A^T A\right)^{-1}$.When A has iid $N(0, \frac{1}{m})$ entries, $A^T A$ is a *Wishart matrix* whose asymptotic eigendistribution is well known.Using this, we obtain

$$\frac{\|x - \hat{x}\|_2^2}{m\sigma^2} \to \frac{n}{m - n}.$$

Least-Squares

- The squared error for a **fixed** z:

Least-Squares

- The squared error for a **fixed** z:

$$\|x - \hat{x}\|_2^2 = z^T A \left(A^T A\right)^{-2} A^T z \leq \frac{\|\mathrm{Proj}(z, \mathrm{Range}(A))\|_2^2}{\sigma_{\min}^2(A)}.$$

Least-Squares

- The squared error for a **fixed** z:

$$\|x - \hat{x}\|_2^2 = z^T A \left(A^T A\right)^{-2} A^T z \leq \frac{\|\mathrm{Proj}(z, \mathrm{Range}(A))\|_2^2}{\sigma_{\min}^2(A)}.$$

When A has iid $N(0, \frac{1}{m})$ entries, we have

  - $\sigma_{\min}(A) \approx 1 - \sqrt{\frac{n}{m}}$.

Least-Squares

- The squared error for a **fixed** z:

$$\|x - \hat{x}\|_2^2 = z^T A \left(A^T A\right)^{-2} A^T z \leq \frac{\|\mathrm{Proj}(z, \mathrm{Range}(A))\|_2^2}{\sigma_{\min}^2(A)}.$$

When A has iid $N(0, \frac{1}{m})$ entries, we have

  - $\sigma_{\min}(A) \approx 1 - \sqrt{\frac{n}{m}}$.
  - $\|\mathrm{Proj}(z, \mathrm{Range}(A))\|_2^2 \approx \frac{n}{m}\|z\|_2^2$.

Least-Squares

- The squared error for a **fixed** z:

$$\|x - \hat{x}\|_2^2 = z^T A \left(A^T A\right)^{-2} A^T z \leq \frac{\|\mathrm{Proj}(z, \mathrm{Range}(A))\|_2^2}{\sigma_{\min}^2(A)}.$$

When A has iid $N(0, \frac{1}{m})$ entries, we have

  - $\sigma_{\min}(A) \approx 1 - \sqrt{\frac{n}{m}}$.
  - $\|\mathrm{Proj}(z, \mathrm{Range}(A))\|_2^2 \approx \frac{n}{m}\|z\|_2^2$.

Thus,

$$\frac{\|x - \hat{x}\|_2^2}{\|z\|_2^2} \rightarrow \leq \left(\frac{\sqrt{n}}{\sqrt{m} - \sqrt{n}}\right)^2.$$

Least-Squares

- The **worst-case square error:**

Least-Squares

- The **worst-case square error:**

$$\|x - \hat{x}\|_2^2 = z^T A \left(A^T A\right)^{-2} A^T z \leq \frac{\|z\|_2^2}{\sigma_{\min}^2(A)}.$$

Least-Squares

- The **worst-case square error:**

$$\|x - \hat{x}\|_2^2 = z^T A \left(A^T A\right)^{-2} A^T z \leq \frac{\|z\|_2^2}{\sigma_{\min}^2(A)}.$$

When A has iid $N(0, \frac{1}{m})$ entries, we have

  - $\sigma_{\min}(A) \approx 1 - \sqrt{\frac{n}{m}}$.

Least-Squares

- The **worst-case square error:**

$$\|x - \hat{x}\|_2^2 = z^T A \left(A^T A\right)^{-2} A^T z \leq \frac{\|z\|_2^2}{\sigma_{\min}^2(A)}.$$

When A has iid $N(0, \frac{1}{m})$ entries, we have

  - $\sigma_{\min}(A) \approx 1 - \sqrt{\frac{n}{m}}$.

Thus,

$$\frac{\|x - \hat{x}\|_2^2}{\|z\|_2^2} \rightarrow \leq \left(\frac{\sqrt{m}}{\sqrt{m} - \sqrt{n}}\right)^2.$$

Least-Squares in Summary

Assume $y = Ax + z$ and that A has iid $N(0, \frac{1}{m})$ entries.

Least-Squares in Summary

Assume $y = Ax + z$ and that A has iid $N(0, \frac{1}{m})$ entries.

- when z has zero-mean iid Gaussian entries: $\frac{\|x-\hat{x}\|_2^2}{\|z\|_2^2} \to \frac{n}{m-n}$.
- when z is arbitrary, but independent of A: $\frac{\|x-\hat{x}\|_2^2}{\|z\|_2^2} \to \leq \left(\frac{\sqrt{n}}{\sqrt{m}-\sqrt{n}}\right)^2$
- when z can be chosen according to A: $\frac{\|x-\hat{x}\|_2^2}{\|z\|_2^2} \to \leq \left(\frac{\sqrt{m}}{\sqrt{m}-\sqrt{n}}\right)^2$

Least-Squares in Summary

Assume $y = Ax + z$ and that A has iid $N(0, \frac{1}{m})$ entries.

- when z has zero-mean iid Gaussian entries: $\frac{\|x-\hat{x}\|_2^2}{\|z\|_2^2} \to \frac{n}{m-n}$.
- when z is arbitrary, but independent of A: $\frac{\|x-\hat{x}\|_2^2}{\|z\|_2^2} \to \leq \left(\frac{\sqrt{n}}{\sqrt{m}-\sqrt{n}}\right)^2$
- when z can be chosen according to A: $\frac{\|x-\hat{x}\|_2^2}{\|z\|_2^2} \to \leq \left(\frac{\sqrt{m}}{\sqrt{m}-\sqrt{n}}\right)^2$

The above all hold with high probability in A.

Least-Squares in Summary

Assume $y = Ax + z$ and that A has iid $N(0, \frac{1}{m})$ entries.

- when z has zero-mean iid Gaussian entries: $\frac{\|x-\hat{x}\|_2^2}{\|z\|_2^2} \to \frac{n}{m-n}$.
- when z is arbitrary, but independent of A: $\frac{\|x-\hat{x}\|_2^2}{\|z\|_2^2} \to \leq \left(\frac{\sqrt{n}}{\sqrt{m}-\sqrt{n}}\right)^2$
- when z can be chosen according to A: $\frac{\|x-\hat{x}\|_2^2}{\|z\|_2^2} \to \leq \left(\frac{\sqrt{m}}{\sqrt{m}-\sqrt{n}}\right)^2$

The above all hold with high probability in A.

Note that

$$\frac{n}{m-n} \leq \left(\frac{\sqrt{n}}{\sqrt{m}-\sqrt{n}}\right)^2 \leq \left(\frac{\sqrt{m}}{\sqrt{m}-\sqrt{n}}\right)^2.$$

Structured Signals

- We are increasingly confronted with very large data sets where we need to extract some *signal-of-interest*

Structured Signals

- We are increasingly confronted with very large data sets where we need to extract some *signal-of-interest*
  - machine learning, image processing, signal processing, statistics, etc.

Structured Signals

- We are increasingly confronted with very large data sets where we need to extract some *signal-of-interest*
  - machine learning, image processing, signal processing, statistics, etc.
  - sensor networks, social networks, DNA microarrays, etc.

Structured Signals

- We are increasingly confronted with very large data sets where we need to extract some *signal-of-interest*
  - machine learning, image processing, signal processing, statistics, etc.
  - sensor networks, social networks, DNA microarrays, etc.
- On the face of it, this could lead to the *curse of dimensionality*

Structured Signals

- We are increasingly confronted with very large data sets where we need to extract some *signal-of-interest*
  - machine learning, image processing, signal processing, statistics, etc.
  - sensor networks, social networks, DNA microarrays, etc.
- On the face of it, this could lead to the *curse of dimensionality*
- Fortunately, in many applications, the signal of interest lives in a manifold of *much lower dimension* than that of the original ambient space

Structured Signals

- We are increasingly confronted with very large data sets where we need to extract some *signal-of-interest*
  - machine learning, image processing, signal processing, statistics, etc.
  - sensor networks, social networks, DNA microarrays, etc.
- On the face of it, this could lead to the *curse of dimensionality*
- Fortunately, in many applications, the signal of interest lives in a manifold of *much lower dimension* than that of the original ambient space
- In this setting, it is important to have signal recovery algorithms that are computationally efficient and that need not access the entire data directly (hence compressed recovery)

Structured Signals

- We are increasingly confronted with very large data sets where we need to extract some *signal-of-interest*
  - machine learning, image processing, signal processing, statistics, etc.
  - sensor networks, social networks, DNA microarrays, etc.
- On the face of it, this could lead to the *curse of dimensionality*
- Fortunately, in many applications, the signal of interest lives in a manifold of *much lower dimension* than that of the original ambient space
- In this setting, it is important to have signal recovery algorithms that are computationally efficient and that need not access the entire data directly (hence compressed recovery)
- The goal of compressed sampling is to perform sampling and sensing simultaneously

Noisy Compressed Sensing

Consider a "desired" signal $x \in \mathcal{R}^n$, which is k-sparse, i.e., has only $k < n$ (often $k \ll n$) non-zero entries. Suppose we make m noisy measurements of x using the $m \times n$ measurement matrix A to obtain

$$y = Ax + z.$$

How many measurements m do we need to find a good estimate of x?

Noisy Compressed Sensing

Consider a "desired" signal $x \in \mathcal{R}^n$, which is k-sparse, i.e., has only $k < n$ (often $k \ll n$) non-zero entries. Suppose we make m noisy measurements of x using the $m \times n$ measurement matrix A to obtain

$$y = Ax + z.$$

How many measurements m do we need to find a good estimate of x? .

- Suppose each set of m columns of A are linearly independent. Then, if $m > k$, we can always find the *sparsest* solution to

$$\min_x \|y - Ax\|_2^2,$$

via exhaustive search of $\begin{pmatrix} n \\ k \end{pmatrix}$ *such least-squares problems*

Noisy Compressed Sensing

Consider a "desired" signal $x \in \mathcal{R}^n$, which is k-sparse, i.e., has only $k < n$ (often $k \ll n$) non-zero entries. Suppose we make m noisy measurements of x using the $m \times n$ measurement matrix A to obtain

$$y = Ax + z.$$

How many measurements m do we need to find a good estimate of x? .

- Suppose each set of m columns of A are linearly independent. Then, if $m > k$, we can always find the *sparsest* solution to

$$\min_x \|y - Ax\|_2^2 ,$$

via exhaustive search of $\begin{pmatrix} n \\ k \end{pmatrix}$ *such least-squares problems*

This would give the normalized square errors:

$$\frac{k}{m-k} \quad , \quad \left(\frac{\sqrt{k}}{\sqrt{m} - \sqrt{k}} \right)^2 \quad , \quad \left(\frac{\sqrt{m}}{\sqrt{m} - \sqrt{k}} \right)^2 .$$

Noisy Compressed Sensing

Noisy Compressed Sensing

But can we do this more efficiently? And for what values of m?

Noisy Compressed Sensing

But can we do this more efficiently? And for what values of m?

There are also problems (such as low rank matrix recovery) where it is not possible to enumerate all structured signals.

LASSO

The LASSO algorithm was introduced by Tibshirani in 1996:

$$\hat{x} = \arg\min_{x} \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1,$$

where $\lambda \geq 0$ is a regularization parameter.

LASSO

The LASSO algorithm was introduced by Tibshirani in 1996:

$$\hat{x} = \arg\min_{x} \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1,$$

where $\lambda \geq 0$ is a regularization parameter.

- $\lambda = 0$ yields the least-squares problem

LASSO

The LASSO algorithm was introduced by Tibshirani in 1996:

$$\hat{x} = \arg\min_{x} \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1,$$

where $\lambda \geq 0$ is a regularization parameter.

- $\lambda = 0$ yields the least-squares problem
- $\lambda \to \infty$ yields ℓ_1 minimization

LASSO

The LASSO algorithm was introduced by Tibshirani in 1996:

$$\hat{x} = \arg\min_{x} \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1,$$

where $\lambda \geq 0$ is a regularization parameter.

- $\lambda = 0$ yields the least-squares problem
- $\lambda \to \infty$ yields ℓ_1 minimization

Generalized LASSO

The generalized LASSO algorithm can be used to enforce other types of structures

$$\hat{x} = \arg\min_{x} \frac{1}{2} \|y - Ax\|_2^2 + \lambda f(x),$$

where $f(\cdot)$ is a *convex* regularizer.

Generalized LASSO

The generalized LASSO algorithm can be used to enforce other types of structures

$$\hat{x} = \arg\min_{x} \frac{1}{2} \|y - Ax\|_2^2 + \lambda f(x),$$

where $f(\cdot)$ is a *convex* regularizer.

- $f(\cdot) = \|\cdot\|_1$ encourages sparsity

Generalized LASSO

The generalized LASSO algorithm can be used to enforce other types of structures

$$\hat{x} = \arg\min_{x} \frac{1}{2} \|y - Ax\|_2^2 + \lambda f(x),$$

where $f(\cdot)$ is a *convex* regularizer.

- $f(\cdot) = \|\cdot\|_1$ encourages sparsity
- $f(\cdot) = \|\cdot\|_\star$ encourages low rankness

Generalized LASSO

The generalized LASSO algorithm can be used to enforce other types of structures

$$\hat{x} = \arg\min_{x} \frac{1}{2} \|y - Ax\|_2^2 + \lambda f(x),$$

where $f(\cdot)$ is a *convex* regularizer.

- $f(\cdot) = \|\cdot\|_1$ encourages sparsity
- $f(\cdot) = \|\cdot\|_\star$ encourages low rankness
- $f(\cdot) = \|\cdot\|_{1,2}$ (the mixed ℓ_1/ℓ_2 norm) encourages block-sparsity

Generalized LASSO

The generalized LASSO algorithm can be used to enforce other types of structures

$$\hat{x} = \arg\min_x \frac{1}{2} \|y - Ax\|_2^2 + \lambda f(x),$$

where $f(\cdot)$ is a *convex* regularizer.

- $f(\cdot) = \|\cdot\|_1$ encourages sparsity
- $f(\cdot) = \|\cdot\|_\star$ encourages low rankness
- $f(\cdot) = \|\cdot\|_{1,2}$ (the mixed ℓ_1/ℓ_2 norm) encourages block-sparsity
- etc.

Generalized LASSO

The generalized LASSO algorithm can be used to enforce other types of structures

$$\hat{x} = \arg\min_x \frac{1}{2} \|y - Ax\|_2^2 + \lambda f(x),$$

where $f(\cdot)$ is a *convex* regularizer.

- $f(\cdot) = \|\cdot\|_1$ encourages sparsity
- $f(\cdot) = \|\cdot\|_\star$ encourages low rankness
- $f(\cdot) = \|\cdot\|_{1,2}$ (the mixed ℓ_1/ℓ_2 norm) encourages block-sparsity
- etc.

Often the value $f(\cdot)$ at the *true* x_0 is known a priori. In this case, we can alternatively solve

$$\begin{array}{ll} \min_x & \|y - Ax\|_2^2 \\ \text{subject to} & f(x) \leq f(x_0) \end{array}.$$

The Squared Error of Generalized LASSO

- The LASSO algorithm has been extensively studied

The Squared Error of Generalized LASSO

- The LASSO algorithm has been extensively studied
- However, most performance bounds are rather loose

The Squared Error of Generalized LASSO

- The LASSO algorithm has been extensively studied
- However, most performance bounds are rather loose
- Can we give performance bounds similar to those that we gave for least-squares?

The Squared Error of Generalized LASSO

- The LASSO algorithm has been extensively studied
- However, most performance bounds are rather loose
- Can we give performance bounds similar to those that we gave for least-squares?

Turns out *we can*.

The Squared Error of Generalized LASSO

- The LASSO algorithm has been extensively studied
- However, most performance bounds are rather loose
- Can we give performance bounds similar to those that we gave for least-squares?

Turns out *we can*. But to do so, we need to tell an earlier story....

Noiseless Compressed Sensing

Consider a "desired" signal $x \in \mathcal{R}^n$, which is k-sparse, i.e., has only $k < n$ (often $k \ll n$) non-zero entries. Suppose we make m measurements of x using the $m \times n$ measurement matrix A to obtain

$$y = Ax.$$

How many measurements m do we need to recover x?

Noiseless Compressed Sensing

Consider a "desired" signal $x \in \mathcal{R}^n$, which is k-sparse, i.e., has only $k < n$ (often $k \ll n$) non-zero entries. Suppose we make m measurements of x using the $m \times n$ measurement matrix A to obtain

$$y = Ax.$$

How many measurements m do we need to recover x? .

- Suppose each set of m columns of A are linearly independent. Then we can always find x uniquely from y—*via exhaustive search of* $\binom{n}{k}$ *systems of linear equations*—if $m > k$

But can we do this more efficiently? And for what values of m?

l_1 Optimization

The seminal work of Candes, Tao and Donoho has shown that under certain conditions the ℓ_1 optimization

$$\min \|x\|_1 \quad \text{subject to } y = Ax$$

where $\|x\|_1 = \sum_i |x_i|$, can *exactly* recover the solution, thus avoiding an exponential search.

l_1 Optimization

The seminal work of Candes, Tao and Donoho has shown that under certain conditions the ℓ_1 optimization

$$\min \|x\|_1 \quad \text{subject to } y = Ax$$

where $\|x\|_1 = \sum_i |x_i|$, can *exactly* recover the solution, thus avoiding an exponential search.

- Candes and Tao showed that if A satisfies certain *restricted isometry* conditions, then ℓ_1 optimization works for small enough k
  - gives "order optimal", but very loose bounds

l_1 Optimization

The seminal work of Candes, Tao and Donoho has shown that under certain conditions the ℓ_1 optimization

$$\min \|x\|_1 \quad \text{subject to } y = Ax$$

where $\|x\|_1 = \sum_i |x_i|$, can *exactly* recover the solution, thus avoiding an exponential search.

- Candes and Tao showed that if A satisfies certain *restricted isometry* conditions, then ℓ_1 optimization works for small enough k
  - gives "order optimal", but very loose bounds
- Necessary and sufficient conditions can be developed to obtain sharp bounds on k (Donoho-Tanner, Xu-H, Stojnic, Chandrasekaran-Parrilo-Willsky, Oymak-H)
  - we will get into this

Phase Transitions for Convex Relaxation

How any measurements m do we need to *efficiently* recover a k-sparse x from $y = Ax$?

Phase Transitions for Convex Relaxation

How any measurements m do we need to *efficiently* recover a k-sparse x from $y = Ax$?

- First answered by Donoho and Tanner (2005) in the compressed sensing context (using neighborly polytopes—very cumbersome calculations)

Phase Transitions for Convex Relaxation

How any measurements m do we need to *efficiently* recover a k-sparse x from $y = Ax$?

- First answered by Donoho and Tanner (2005) in the compressed sensing context (using neighborly polytopes—very cumbersome calculations)
- Extended to robustness and weighted ℓ_1 by Xu-H in 2007 (using Grassman angles)

Phase Transitions for Convex Relaxation

How any measurements m do we need to *efficiently* recover a k-sparse x from $y = Ax$?

- First answered by Donoho and Tanner (2005) in the compressed sensing context (using neighborly polytopes—very cumbersome calculations)
- Extended to robustness and weighted ℓ_1 by Xu-H in 2007 (using Grassman angles)
- New framework developed by Stojnic in 2009 (using escape-through-mesh and Gaussian widths)

Phase Transitions for Convex Relaxation

How any measurements m do we need to *efficiently* recover a k-sparse x from $y = Ax$?

- First answered by Donoho and Tanner (2005) in the compressed sensing context (using neighborly polytopes—very cumbersome calculations)
- Extended to robustness and weighted ℓ_1 by Xu-H in 2007 (using Grassman angles)
- New framework developed by Stojnic in 2009 (using escape-through-mesh and Gaussian widths)
  - rederived results for sparse vectors; new results for block-sparse vectors

Phase Transitions for Convex Relaxation

How any measurements m do we need to *efficiently* recover a k-sparse x from $y = Ax$?

- First answered by Donoho and Tanner (2005) in the compressed sensing context (using neighborly polytopes—very cumbersome calculations)
- Extended to robustness and weighted ℓ_1 by Xu-H in 2007 (using Grassman angles)
- New framework developed by Stojnic in 2009 (using escape-through-mesh and Gaussian widths)
  - rederived results for sparse vectors; new results for block-sparse vectors
  - much simpler derivation

Phase Transitions for Convex Relaxation

How any measurements m do we need to *efficiently* recover a k-sparse x from $y = Ax$?

- First answered by Donoho and Tanner (2005) in the compressed sensing context (using neighborly polytopes—very cumbersome calculations)
- Extended to robustness and weighted ℓ_1 by Xu-H in 2007 (using Grassman angles)
- New framework developed by Stojnic in 2009 (using escape-through-mesh and Gaussian widths)
  - rederived results for sparse vectors; new results for block-sparse vectors
  - much simpler derivation
- Allowed the development of a general framework (Chandrasekaran-Parrilo-Willsky, 2010)

Phase Transitions for Convex Relaxation

How any measurements m do we need to *efficiently* recover a k-sparse x from $y = Ax$?

- First answered by Donoho and Tanner (2005) in the compressed sensing context (using neighborly polytopes—very cumbersome calculations)
- Extended to robustness and weighted ℓ_1 by Xu-H in 2007 (using Grassman angles)
- New framework developed by Stojnic in 2009 (using escape-through-mesh and Gaussian widths)
  - rederived results for sparse vectors; new results for block-sparse vectors
  - much simpler derivation
- Allowed the development of a general framework (Chandrasekaran-Parrilo-Willsky, 2010)
  - relation to denoising (Oymak-H, 2013), nuclear norm (Oymak-H, 2010)

Phase Transitions for Convex Relaxation

How any measurements m do we need to *efficiently* recover a k-sparse x from $y = Ax$?

- First answered by Donoho and Tanner (2005) in the compressed sensing context (using neighborly polytopes—very cumbersome calculations)
- Extended to robustness and weighted ℓ_1 by Xu-H in 2007 (using Grassman angles)
- New framework developed by Stojnic in 2009 (using escape-through-mesh and Gaussian widths)
  - rederived results for sparse vectors; new results for block-sparse vectors
  - much simpler derivation
- Allowed the development of a general framework (Chandrasekaran-Parrilo-Willsky, 2010)
  - relation to denoising (Oymak-H, 2013), nuclear norm (Oymak-H, 2010)
  - tightness of Gaussian widths Stojnic, 2013 (for ℓ_1), Amelunxen-McCoy-Tropp, 2013 (more generally)

Phase Transitions for Convex Relaxation

We will consider a general framework.

Consider a structured signal x_0, with a structure-capturing atomic norm $f(\cdot) = \|\cdot\|$. We have access to *linear measurements* $y = \mathcal{A}(x_0) \in R^m$, and would like to know when we can recover the signal x_0 from the convex problem

$$\min \|x\| \quad \text{subject to } \mathcal{A}(x) = \mathcal{A}(x_0)?$$

Phase Transitions for Convex Relaxation

We will consider a general framework.

Consider a structured signal x_0, with a structure-capturing atomic norm $f(\cdot) = \|\cdot\|$. We have access to *linear measurements* $y = \mathcal{A}(x_0) \in R^m$, and would like to know when we can recover the signal x_0 from the convex problem

$$\min \|x\| \quad \text{subject to } \mathcal{A}(x) = \mathcal{A}(x_0)?$$

- For sparse signals we have the ℓ_1 norm; for nonuniform sparse signals the weighted ℓ_1 norm; for low rank matrices the nuclear norm

Phase Transitions for Convex Relaxation

We will consider a general framework.

Consider a structured signal x_0, with a structure-capturing atomic norm $f(\cdot) = \|\cdot\|$. We have access to *linear measurements* $y = \mathcal{A}(x_0) \in R^m$, and would like to know when we can recover the signal x_0 from the convex problem

$$\min \|x\| \quad \text{subject to } \mathcal{A}(x) = \mathcal{A}(x_0)?$$

- For sparse signals we have the ℓ_1 norm; for nonuniform sparse signals the weighted ℓ_1 norm; for low rank matrices the nuclear norm

Let $\mathcal{U}(x_0) = \{z, \|x_0 + z\| \leq \|x_0\|\}$. Then x_0 is the unique solution of the above convex problem iff:

Phase Transitions for Convex Relaxation

We will consider a general framework.

Consider a structured signal x_0, with a structure-capturing atomic norm $f(\cdot) = \|\cdot\|$. We have access to *linear measurements* $y = \mathcal{A}(x_0) \in R^m$, and would like to know when we can recover the signal x_0 from the convex problem

$$\min \|x\| \quad \text{subject to } \mathcal{A}(x) = \mathcal{A}(x_0)?$$

- For sparse signals we have the ℓ_1 norm; for nonuniform sparse signals the weighted ℓ_1 norm; for low rank matrices the nuclear norm

Let $\mathcal{U}(x_0) = \{z, \|x_0 + z\| \leq \|x_0\|\}$. Then x_0 is the unique solution of the above convex problem iff:

$$\mathcal{N}(\mathcal{A}) \cap \mathcal{U}(x_0) = \{0\}.$$

A Bit of Geometry: Subgradients and the Polar Cone

Note that $\mathcal{N}(\mathcal{A})$ is a linear subspace and that therefore the condition can be rewritten as

$$\mathcal{N}(\mathcal{A}) \cap \text{cone}(\mathcal{U}(x_0)) = \{0\}.$$

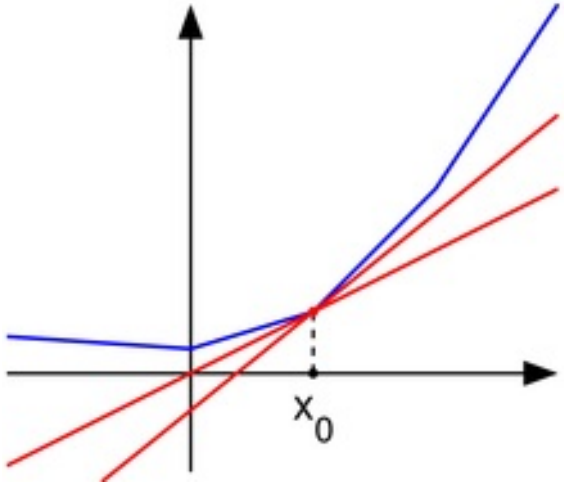
A Bit of Geometry: Subgradients and the Polar Cone

Note that $\mathcal{N}(\mathcal{A})$ is a linear subspace and that therefore the condition can be rewritten as

$$\mathcal{N}(\mathcal{A}) \cap \text{cone}(\mathcal{U}(x_0)) = \{0\}.$$

We can characterize $\text{cone}(\mathcal{U}(x_0))$ through the subgradient of the convex function $\|\cdot\|$:

$$\partial\|x_0\| = \{v | v^T(x - x_0) \leq \|x\| - \|x_0\|, \forall x\}.$$



A Bit of Geometry: Subgradients and the Polar Cone

It is now straightforward to see that

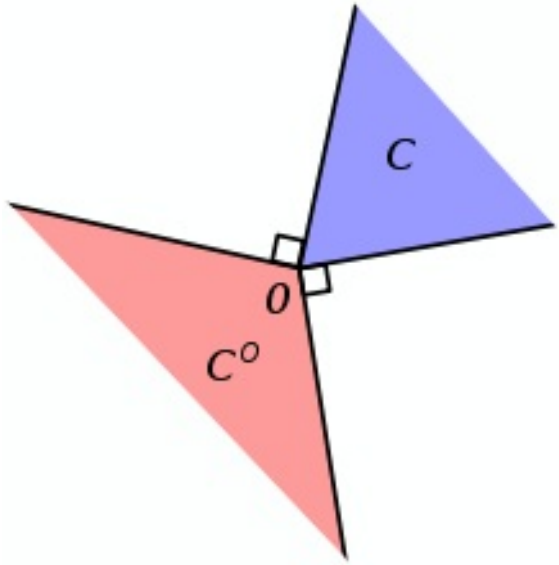
$$\text{cone}(\mathcal{U}(x_0)) = \{z | v^T z \leq 0, \forall v \in \partial \|x_0\|\}.$$

A Bit of Geometry: Subgradients and the Polar Cone

It is now straightforward to see that

$$\text{cone}(\mathcal{U}(x_0)) = \{z | v^T z \leq 0, \forall v \in \partial\|x_0\|\}.$$

But this is simply the *polar cone* of $\partial\|x_0\|$.



Thus, we can recover x_0 from the convex problem iff:

$$\mathcal{N}(\mathcal{A}) \cap (\partial\|x_0\|)^O = \{0\}.$$

A Bit of Geometry: Subgradients and the Polar Cone

- Thus, recovery depends on the null space of the measurement matrix and the polar cone of the subgradient (at the point we want to recover).

A Bit of Geometry: Subgradients and the Polar Cone

- Thus, recovery depends on the null space of the measurement matrix and the polar cone of the subgradient (at the point we want to recover).
  - computing the subgradient is often straightforward: for a k-sparse $x = \begin{bmatrix} x_S \\ 0 \end{bmatrix}$ it is

A Bit of Geometry: Subgradients and the Polar Cone

- Thus, recovery depends on the null space of the measurement matrix and the polar cone of the subgradient (at the point we want to recover).
  - computing the subgradient is often straightforward: for a k-sparse $x = \begin{bmatrix} x_S \\ 0 \end{bmatrix}$ it is

$$\partial \|x_0\|_1 = \left\{ \begin{bmatrix} \text{sign}(x_S) \\ v \end{bmatrix}, \|v\|_\infty \leq 1 \right\}.$$

A Bit of Geometry: Subgradients and the Polar Cone

- Thus, recovery depends on the null space of the measurement matrix and the polar cone of the subgradient (at the point we want to recover).
  - computing the subgradient is often straightforward: for a k-sparse $x = \begin{bmatrix} x_S \\ 0 \end{bmatrix}$ it is

$$\partial \|x_0\|_1 = \left\{ \begin{bmatrix} \text{sign}(x_S) \\ v \end{bmatrix}, \|v\|_\infty \leq 1 \right\}.$$

- However, checking the condition $\mathcal{N}(\mathcal{A}) \cap (\partial \|x_0\|)^O = \{0\}$ for a *specific* $\mathcal{A}$ is difficult.
- Therefore the focus has been on checking whether the condition holds for a *family* of random $\mathcal{A}$'s with high probability.

The Gaussian Measurement Ensemble

- It is customary to assume that the measurement matrix $\mathcal{A}$ is composed of iid zero-mean unit-variance entries

The Gaussian Measurement Ensemble

- It is customary to assume that the measurement matrix $\mathcal{A}$ is composed of iid zero-mean unit-variance entries
- This makes the nullspace $\mathcal{N}(\mathcal{A})$ *rotationally-invariant*.

The Gaussian Measurement Ensemble

- It is customary to assume that the measurement matrix $\mathcal{A}$ is composed of iid zero-mean unit-variance entries
- This makes the nullspace $\mathcal{N}(\mathcal{A})$ *rotationally-invariant*.
- The thresholds obtained from the Gaussian matrix ensemble are often *universal* and so determine the thresholds for other matrix ensembles.

Escape Through a Mesh

Theorem (Escape Through a Mesh - Gordon 1988)

Let C be a subset of unit sphere S^{n-1} in $\mathbb{R}^n$ and $\mathbf{g}$ be a vector with i.i.d. standard normal entries. Further, let H be an $n-m$ dimensional subspace distributed uniformly over Grassmannian w.r.t Haar measure. Then,

$$\mathbb{P}(H \cap C = \emptyset) \geq 1 - 3.5 \exp(-(\sqrt{m} - \frac{1}{4\sqrt{m}} - \omega(C))^2)$$

where $\omega(C)$ stands for the Gaussian width of C and is defined as:

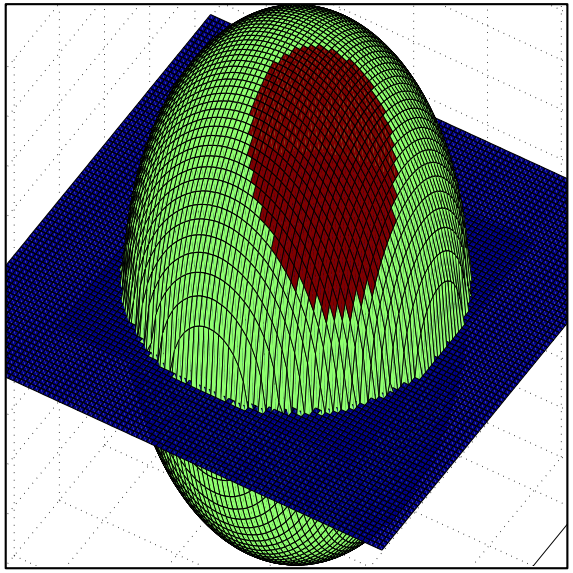
$$\omega(C) = \mathbb{E}[\sup_{\mathbf{v} \in C} \langle \mathbf{g}, \mathbf{v} \rangle]$$

Interpretation

- Question: When $\mathbf{A}$ has i.i.d. Gaussian entries what is the chance of:

$$\text{Null}(\mathbf{A}) \cap (\partial \|x_0\|)^O = \{0\}?$$

- Answer
  - Green ball: Unit sphere S^{n-1}.
  - Blue plane: Random subspace Null($\mathbf{A}$).
  - Red mesh: Undesired set $C = (\partial \|x_0\|)^O \cap S^{n-1}$.



Gaussian Width Calculation

- To ensure recovery: choose $m \geq \omega((\partial\|x_0\|)^O \cap S^{n-1})^2$
- It is critical to estimate the GW accurately

Lemma

Let $\partial\|\mathbf{x}_0\|$ be the subdifferential of $\|\cdot\|$ at $\mathbf{x}_0$ and $\mathbf{g}$ be a vector with i.i.d. standard normal entries. Then

$$\omega((\partial\|x_0\|)^O \cap S^{n-1}) = \mathbb{E}[dist(\mathbf{g}, \partial\|\mathbf{x}_0\|)] \approx \sqrt{\mathbb{E}[dist(\mathbf{g}, \partial\|\mathbf{x}_0\|)^2]}. \quad (0.1)$$

Gaussian Width Calculation

- To ensure recovery: choose $m \geq \omega((\partial\|x_0\|)^O \cap S^{n-1})^2$
- It is critical to estimate the GW accurately

Lemma

Let $\partial\|\mathbf{x}_0\|$ be the subdifferential of $\|\cdot\|$ at $\mathbf{x}_0$ and $\mathbf{g}$ be a vector with i.i.d. standard normal entries. Then

$$\omega((\partial\|x_0\|)^O \cap S^{n-1}) = \mathbb{E}[dist(\mathbf{g}, \partial\|\mathbf{x}_0\|)] \approx \sqrt{\mathbb{E}[dist(\mathbf{g}, \partial\|\mathbf{x}_0\|)^2]}. \quad (0.1)$$

Even though Gordon's lemma is a sufficient condition, Stojnic and Amelnuxen-McCoy-Tropp show that the squared Gaussian width is a *tight bound*.

Gaussian Width Calculation

- To ensure recovery: choose $m \geq \omega((\partial\|x_0\|)^O \cap S^{n-1})^2$
- It is critical to estimate the GW accurately

Lemma

Let $\partial\|\mathbf{x}_0\|$ be the subdifferential of $\|\cdot\|$ at $\mathbf{x}_0$ and $\mathbf{g}$ be a vector with i.i.d. standard normal entries. Then

$$\omega((\partial\|x_0\|)^O \cap S^{n-1}) = \mathbb{E}[dist(\mathbf{g}, \partial\|\mathbf{x}_0\|)] \approx \sqrt{\mathbb{E}[dist(\mathbf{g}, \partial\|\mathbf{x}_0\|)^2]}. \quad (0.1)$$

Even though Gordon's lemma is a sufficient condition, Stojnic and Amelnuxen-McCoy-Tropp show that the squared Gaussian width is a *tight bound*. AMT call the squared Gaussian width the *statistical dimension* of the signal.

Example: Nuclear Norm Minimization (NNM)

Nuclear norm: $\|X\|_\star = \sum_i \sigma_i(X)$

- Many applications: Netflix, minimal order controllers, graphical models, etc.
- Tightest convex relaxation to rank minimization problem

$$\min \|X\|_\star \quad \text{subject to} \quad \mathcal{A}(X) = y$$

- Closely related to ℓ_1 minimization
- Polynomial time algorithm (an SDP)
- Known to have noise robustness
- Unlike non-convex algorithms NNM can be made more complicated.
  - coupling with ℓ_1: Robust PCA
    $$\min \|L\|_\star + \lambda \|S\|_1 \quad \text{subject to} \quad X = L + S$$

Prior Work

Let $\mathbf{y} = \mathcal{A}(X) \in \mathbb{R}^m$, $X \in \mathbb{R}^{n \times n}$.
Number of samples for successful recovery is a useful measure of performance

- i.i.d. measurements
  - Recht, Fazel, Parrilo "Guaranteed minimum rank ..." (2008): $m = O(rnlog(n))$
  - Candes, Plan "Tight oracle bounds ..." (2010): $m = O(rn)$
  - First null space analysis: Recht, Xu, H "Null Space Conditions and ..." (2009)
- Observe entries at random (matrix completion)
  - Candes, Recht "Exact matrix completion via ..." (2009): $m = O(rn^{5/4}log(n))$.
  - Best known: $m = nrlog^2(n)$.
    - ★ Recht "A simpler approach ..." (2009)
    - ★ Gross "Recovering low rank matrices ..." (2011)

Strong and Weak Robustness

Let M^* denote optimal sol'n of "$\min \|X\|_\star \quad \text{subject to} \quad \mathbf{A}(X) = \mathbf{A}(M)$".

Definition (Strong robustness)

$\mathcal{A}(\cdot)$ is strongly robust with ϵ_0, r if the following holds for all M:

$$\|M - M^*\|_\star < \frac{2\|M - M^r\|_\star}{\epsilon_0}$$

$M - M^r$ is basically the tail of M.

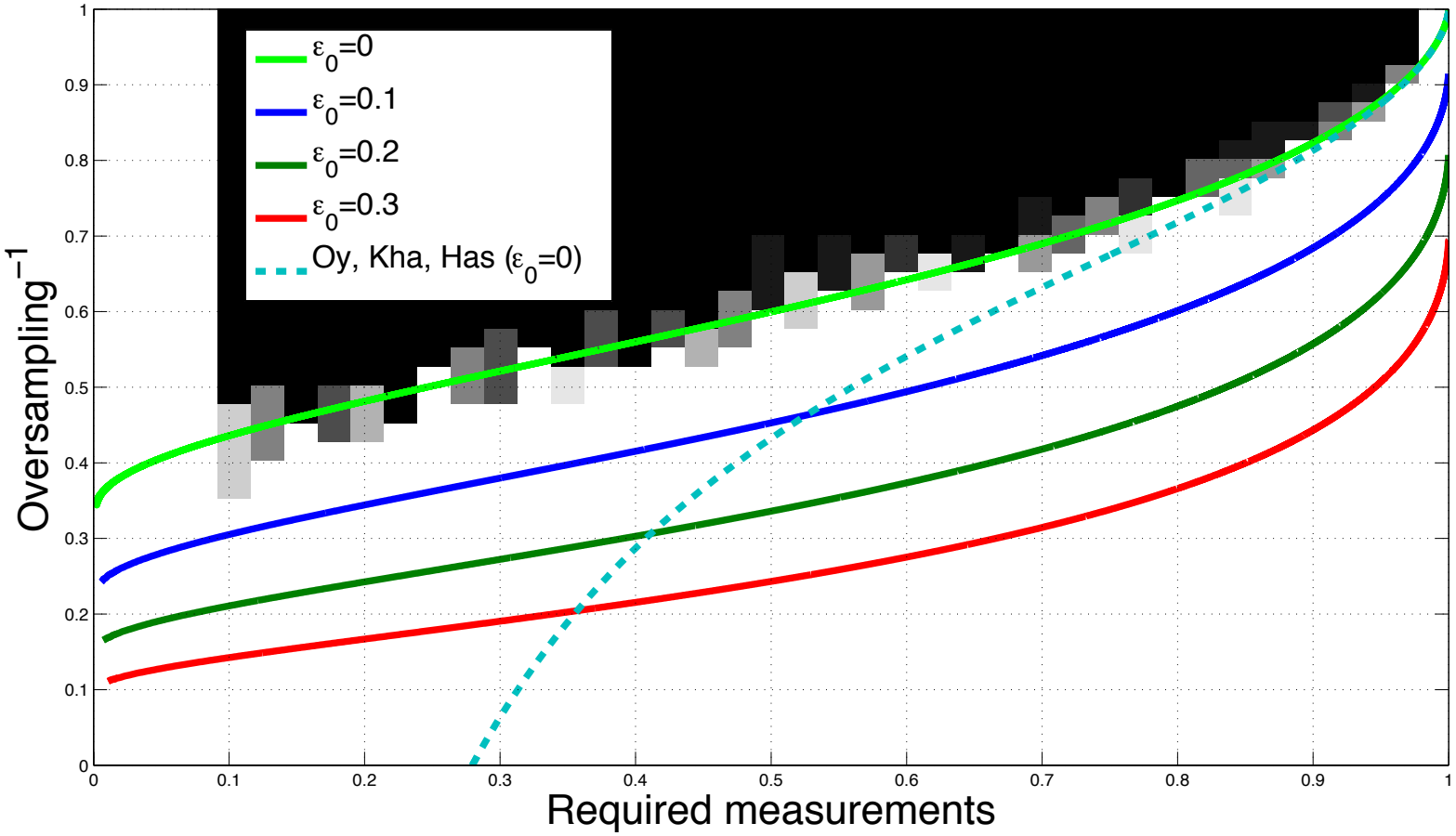
Definition (Weak robustness)

Given a fixed M, $\mathcal{A}(\cdot)$ is weakly robust with ϵ_0, r if the following holds:

$$\|M - M^*\|_\star < \frac{2\|M - M^r\|_\star}{\epsilon_0}$$

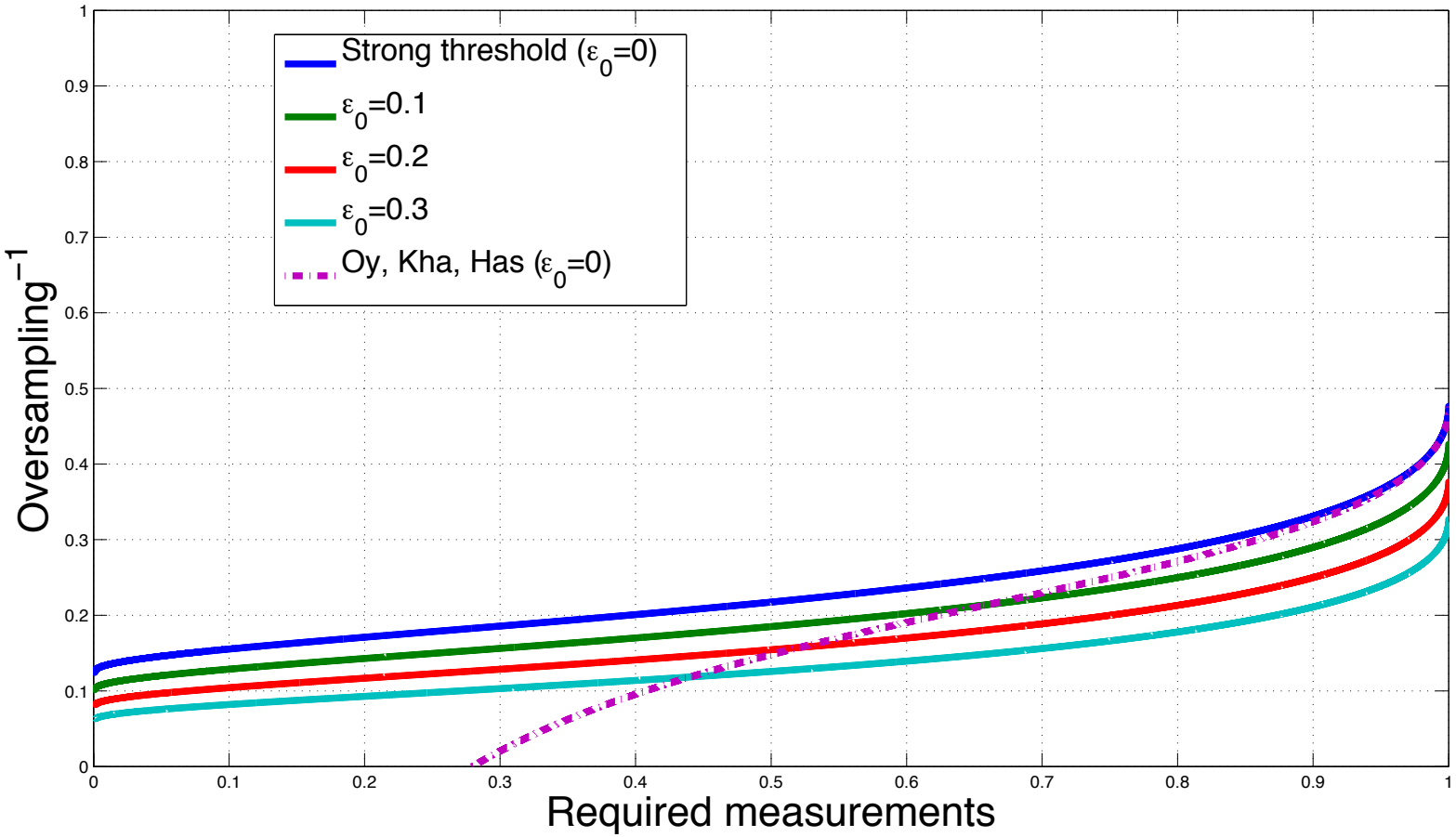
Weak recovery curves

Oversampling: How much measurement per degrees of freedom?



DOF $= r(2n - r)$

Strong recovery curves



Closed Form Results

Theorem (Closed Form Bounds)

Let dimensions be $n_1 \times n_2$, maximum rank be r (no tail). Then

$$\begin{aligned} &\textit{for strong recovery:} \quad m \geq 4(\sqrt{n_1} + \sqrt{n_2})^2 r \\ &\textit{for weak recovery:} \quad m \geq 2(n_1 + \sqrt{n_1 n_2} + n_2) r \end{aligned}$$

is sufficient.

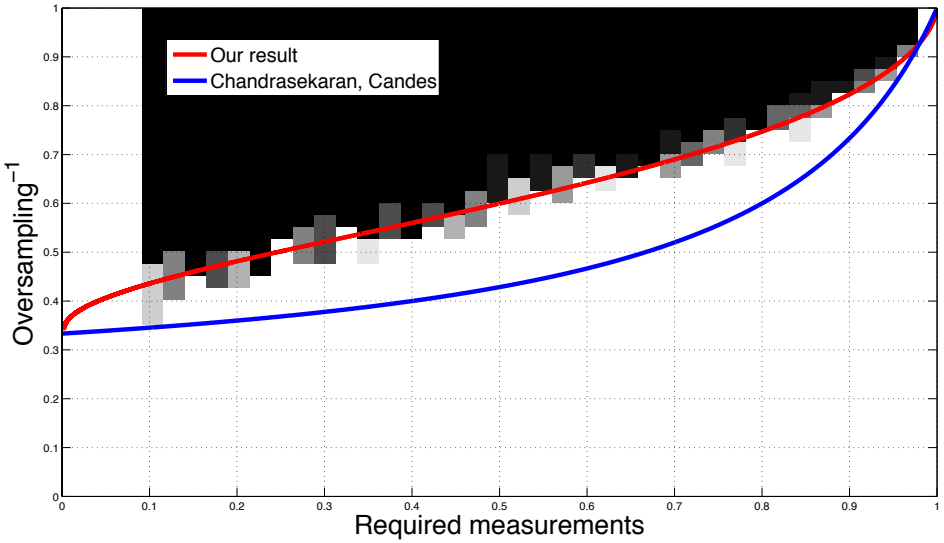
Theorem (Closed Form Bounds)

In the square case:

$$\begin{aligned} &\textit{for strong recovery:} \quad m \geq 16nr \\ &\textit{for weak recovery:} \quad m \geq 6nr \end{aligned}$$

Recent and Simultaneous Work

- Chandrasekaran, Recht, Parrilo, and Willsky "The convex geometry of linear inverse problems" (2010)
- Candes, Recht "Simple bounds for low complexity model construction" (2011)



Some Other Examples

- n dimensional vectors that are k-sparse:

$$f(x) = \|x\|_1 \quad \text{and} \quad \omega^2 \leq 2k \log \frac{2k}{n}$$

- $n \times n$ matrices that are rank r:

$$f(x) = \|x\|_\star \quad \text{and} \quad \omega^2 \leq 3r(2n - r)$$

- qb dimensional vectors that have k non-zero blocks of size b:

$$f(x) = \|x\|_{1,2} \quad \text{and} \quad \omega^2 \leq 4k \left(b + log \frac{q}{k}\right)$$

Relation Between Compressed Recovery and Denoising

- **CS problem:** Recover a signal from underdetermined linear observations.

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 \in \mathbb{R}^m, \ \mathbf{x}_0 \in \mathbb{R}^n, \ \mathbf{A} \sim \text{i.i.d. } \mathcal{N}(0,1)$$

Aim: Recover structured $\mathbf{x}_0$ when $m \ll n$.

Relation Between Compressed Recovery and Denoising

- **CS problem:** Recover a signal from underdetermined linear observations.

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 \in \mathbb{R}^m, \ \mathbf{x}_0 \in \mathbb{R}^n, \ \mathbf{A} \sim \text{i.i.d. } \mathcal{N}(0,1)$$

  Aim: Recover structured $\mathbf{x}_0$ when $m \ll n$.
- Recovery method:

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x}_0 \qquad \text{(BP)}$$

Relation Between Compressed Recovery and Denoising

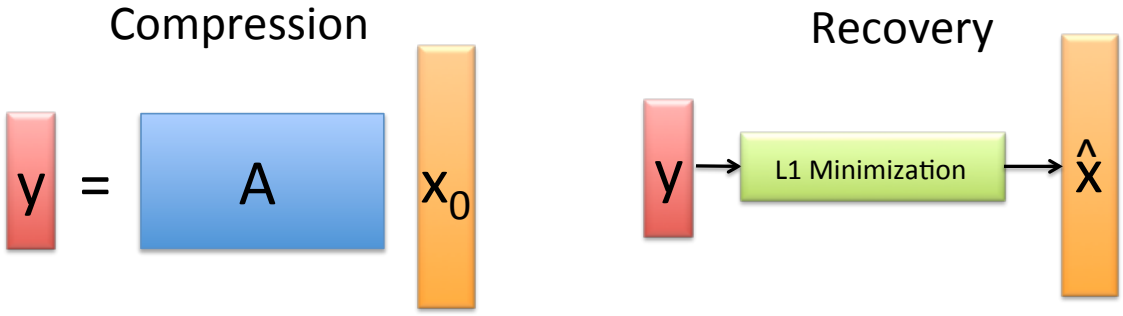
- **CS problem:** Recover a signal from underdetermined linear observations.

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 \in \mathbb{R}^m, \ \mathbf{x}_0 \in \mathbb{R}^n, \ \mathbf{A} \sim \text{i.i.d. } \mathcal{N}(0,1)$$

  Aim: Recover structured $\mathbf{x}_0$ when $m \ll n$.
- Recovery method:

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x}_0 \qquad \text{(BP)}$$

- Choose $f(\cdot)$ to exploit the structure. Eg: $\mathbf{x}_0$ is sparse vector and $f(\cdot) = \|\cdot\|_{\ell_1}$



- We will focus on convex $f(\cdot)$

Denoising Problem

- **Denoising problem:** Estimate a signal corrupted by additive noise.

$$\mathbf{y} = \mathbf{x}_0 + \mathbf{z}, \ \mathbf{x}_0 \in \mathbb{R}^n, \ \mathbf{z} \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2)$$

Aim: Estimate structured $\mathbf{x}_0$

Denoising Problem

- **Denoising problem:** Estimate a signal corrupted by additive noise.

$$\mathbf{y} = \mathbf{x}_0 + \mathbf{z}, \ \mathbf{x}_0 \in \mathbb{R}^n, \ \mathbf{z} \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2)$$

  Aim: Estimate structured $\mathbf{x}_0$
- Denoising method: proximity operator

$$\min_{\mathbf{x}} \lambda f(\mathbf{x}) + \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 \qquad \text{(LASSO)}$$

Denoising Problem

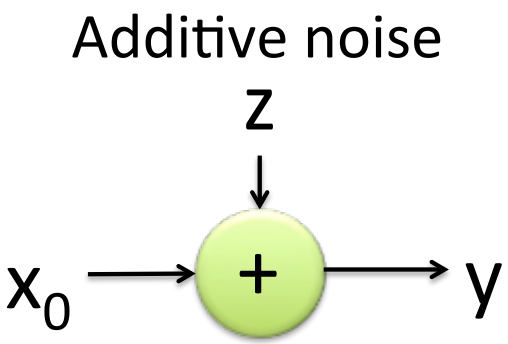
- **Denoising problem:** Estimate a signal corrupted by additive noise.

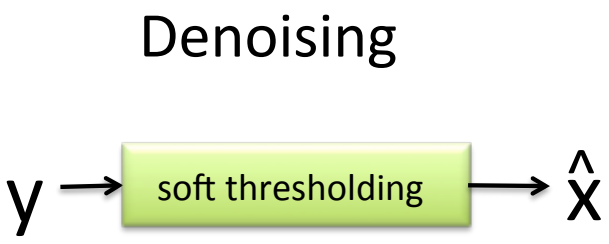
$$\mathbf{y} = \mathbf{x}_0 + \mathbf{z}, \ \mathbf{x}_0 \in \mathbb{R}^n, \ \mathbf{z} \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2)$$

  Aim: Estimate structured $\mathbf{x}_0$
- Denoising method: proximity operator

$$\min_{\mathbf{x}} \lambda f(\mathbf{x}) + \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 \qquad \text{(LASSO)}$$

- Choose $f(\cdot)$ to induce the structure. Eg: $\mathbf{x}_0$ is sparse vector and $f(\cdot) = \|\cdot\|_{\ell_1}$.





Performance Criteria

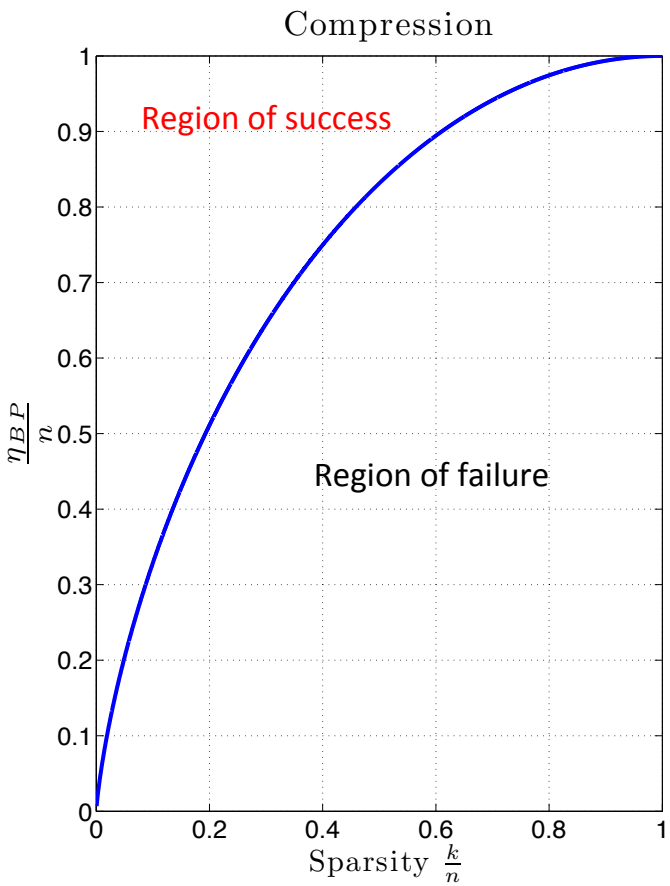
- **CS problem:** The smallest number $\eta_{BP}(\mathbf{x}_0)$ s.t. $m = \eta_{BP}(\mathbf{x}_0)$ measurements are sufficient for recovery of $\mathbf{x}_0$ via (BP) w.h.p. where $\mathbf{y} = \mathbf{A}\mathbf{x}_0 \in \mathbb{R}^m$.
- **Denoising problem:** Tune λ optimally to minimize normalized estimation error.

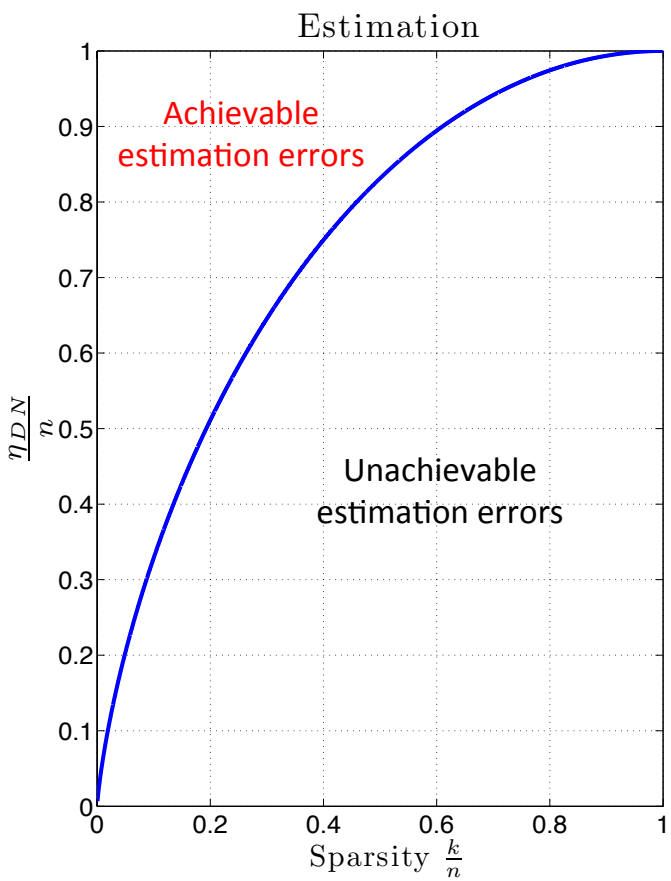
$$\eta_{DN}(\mathbf{x}_0) = \lim_{\sigma \to 0} \inf_{\lambda \geq 0} \frac{\mathbb{E}[\|\mathbf{x}_0 - \mathbf{x}^*(\lambda, \mathbf{z})\|_2^2]}{\sigma^2} \tag{1.1}$$

where $\mathbf{x}^*(\lambda, \mathbf{z})$ is the minimizer of LASSO:

$$\mathbf{x}^*(\lambda, \mathbf{z}) = \arg\min_{\mathbf{x}} \lambda f(\mathbf{x}) + \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 \tag{1.2}$$

ℓ_1 Phase Transitions





A General Relation

- Bayati, Donoho and Montanari have proven $\eta_{BP} = \eta_{DN}$ for ℓ_1 optimization

A General Relation

- Bayati, Donoho and Montanari have proven $\eta_{BP} = \eta_{DN}$ for ℓ_1 optimization
- Recently, based on extensive empirical evidence, Donoho, Montanari and Johnstone have proposed that

$$\eta_{BP} = \eta_{DN}$$

for a general convex $f(\cdot)$.

  - "Accurate Prediction of Phase Transitions in Compressed Sensing ..." (2011)

A General Relation

- Bayati, Donoho and Montanari have proven $\eta_{BP} = \eta_{DN}$ for ℓ_1 optimization
- Recently, based on extensive empirical evidence, Donoho, Montanari and Johnstone have proposed that

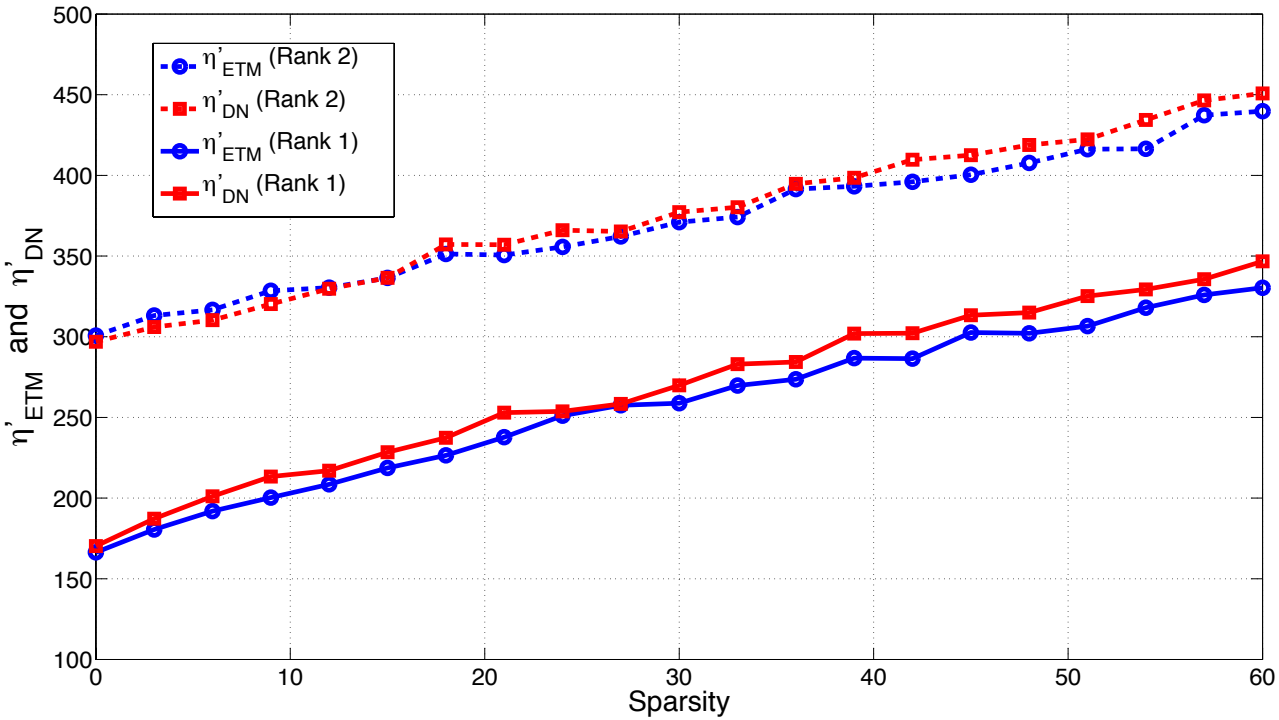
$$\eta_{BP} = \eta_{DN}$$

  for a general convex $f(\cdot)$.
  - "Accurate Prediction of Phase Transitions in Compressed Sensing ..." (2011)
- In Oymak-H using the escape-through-mesh framework we show that $\eta_{BP} = \eta_{DN} = \omega^2$, the statistical dimension, for a general convex $f(\cdot)$.

Low Rank plus Sparse Signals

$$\mathbf{X} \text{ is LPS and } f(\mathbf{X}) = \min_{\mathbf{L}} \|\mathbf{L}\|_\star + \Theta\|\mathbf{X} - \mathbf{L}\|_1$$

sparsity: $0 \to 60$, rank: $1 \to 2$



Observation: $\eta_{DN} \approx \eta_{ETM}$.

Noiseless CS – Simple Denoising – LASSO

$\mathbf{x}_0 \in \mathbf{R}^n$: *structured* signal of interest (sparse, low-rank, etc.)

Noiseless CS – Simple Denoising – LASSO

$\mathbf{x}_0 \in \mathbf{R}^n$: *structured* signal of interest (sparse, low-rank, etc.)
$\mathbf{A} \in \mathbf{R}^{m \times n}$: measurement matrix ,

Noiseless CS – Simple Denoising – LASSO

$\mathbf{x}_0 \in \mathbf{R}^n$: *structured* signal of interest (sparse, low-rank, etc.)
$\mathbf{A} \in \mathbf{R}^{m \times n}$: measurement matrix , $\mathbf{z} \in \mathbf{R}^m$: noise

Noiseless CS – Simple Denoising – LASSO

$\mathbf{x}_0 \in \mathbf{R}^n$: *structured* signal of interest (sparse, low-rank, etc.)
$\mathbf{A} \in \mathbf{R}^{m \times n}$: measurement matrix , $\mathbf{z} \in \mathbf{R}^m$: noise
$f(\cdot)$: *structure inducing* convex function ($\|\cdot\|_1$, $\|\cdot\|_*$, etc.)

Noiseless CS – Simple Denoising – LASSO

$\mathbf{x}_0 \in \mathbf{R}^n$: *structured* signal of interest (sparse, low-rank, etc.)
$\mathbf{A} \in \mathbf{R}^{m \times n}$: measurement matrix , $\mathbf{z} \in \mathbf{R}^m$: noise
$f(\cdot)$: *structure inducing* convex function ($\|\cdot\|_1$, $\|\cdot\|_*$, etc.)

1. Noiseless Compressed Sensing,

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{y} = \mathbf{A}\mathbf{x}$$

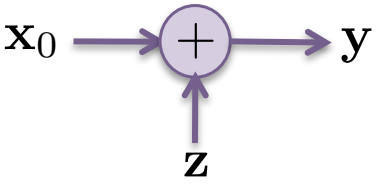
Noiseless CS – Simple Denoising – LASSO

$\mathbf{x}_0 \in \mathbf{R}^n$: *structured* signal of interest (sparse, low-rank, etc.)
$\mathbf{A} \in \mathbf{R}^{m \times n}$: measurement matrix , $\mathbf{z} \in \mathbf{R}^m$: noise
$f(\cdot)$: *structure inducing* convex function ($\|\cdot\|_1$, $\|\cdot\|_*$, etc.)

1. Noiseless Compressed Sensing,

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{y} = \mathbf{A}\mathbf{x}$$

2. Simple Denoising:



$$\min_{\mathbf{x}} \left\{ \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda f(\mathbf{x}) \right\}$$

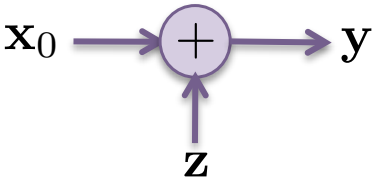
Noiseless CS – Simple Denoising – LASSO

$\mathbf{x}_0 \in \mathbf{R}^n$: *structured* signal of interest (sparse, low-rank, etc.)
$\mathbf{A} \in \mathbf{R}^{m \times n}$: measurement matrix , $\mathbf{z} \in \mathbf{R}^m$: noise
$f(\cdot)$: *structure inducing* convex function ($\|\cdot\|_1$, $\|\cdot\|_*$, etc.)

1. Noiseless Compressed Sensing,

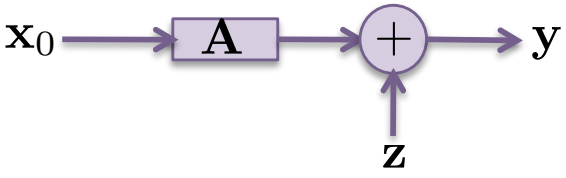
$$\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{y} = \mathbf{A}\mathbf{x}$$

2. Simple Denoising:



$$\min_{\mathbf{x}} \left\{ \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda f(\mathbf{x}) \right\}$$

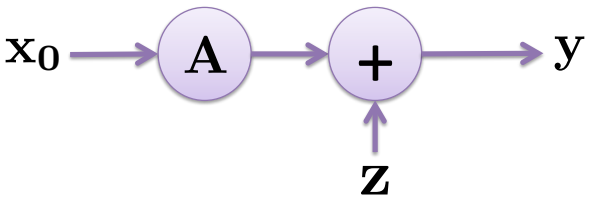
3. Noisy CS (LASSO) can be seen as a "merger" of 1 and 2:



$$\min_{\mathbf{x}} \left\{ \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda f(\mathbf{x}) \right\}$$

The Generalized LASSO

Observe



Solve $\min_{\mathbf{x}} \lambda f(\mathbf{x}) + \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2$

- $f(\mathbf{x}) = \|\mathbf{x}\|_1$: R. Tibshirani "Regression shrinkage and selection via the lasso", '96., Chen, Donoho, Saunders'98
- $f(\mathbf{X}) = \|\mathbf{X}\|_*$: Koltchinskii'10, Negahban'12, Candes'11 and more.

✓ $f(\mathbf{x})$ = any *convex* function of $\mathbf{x}$: **"Generalized LASSO"**

The Squared Error of Generalized LASSO

Consider the generalized LASSO:

$$\min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda f(\mathbf{x}) \right\},$$

then we have been able to show:

The Squared Error of Generalized LASSO

Consider the generalized LASSO:

$$\min_{\mathbf{x}} \left\{ \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda f(\mathbf{x}) \right\},$$

then we have been able to show:

- for z iid $N(0, \frac{1}{m})$ Gaussian

$$\min_{\lambda} \frac{\|x_0 - \hat{x}\|_2^2}{\|z\|_2^2} \rightarrow \frac{\omega^2}{m - \omega^2}$$

- for z arbitrary, but independent of A:

$$\min_{\lambda} \frac{\|x_0 - \hat{x}\|_2^2}{\|z\|_2^2} \rightarrow \leq \left(\frac{\omega}{\sqrt{m} - \omega} \right)^2$$

- for the worst-case z, chosen with knowledge of A:

$$\min_{\lambda} \frac{\|x_0 - \hat{x}\|_2^2}{\|z\|_2^2} \rightarrow \left(\frac{\sqrt{m}}{\sqrt{m} - \omega} \right)^2$$

The Squared Error of Generalized LASSO

These are the same formulae obtained for standard least-squares, except that the *ambient dimension* , n, has been replaced by the *statistical dimension*, ω^2.

The Squared Error of Generalized LASSO

These are the same formulae obtained for standard least-squares, except that the *ambient dimension* , n, has been replaced by the *statistical dimension*, ω^2.
These results are also true for the constrained LASSO

$$\min_x \|y - Ax\|_2^2 \quad \text{subject to } f(x) \leq f(x_0).$$

For example,

- for n-dimensional k-sparse signals and ℓ_1 minimization, we have the bounds

$$\frac{2k \log \frac{2n}{k}}{m - 2k \log \frac{2n}{k}} \quad , \quad \left(\frac{\sqrt{2k \log \frac{2n}{k}}}{\sqrt{m} - \sqrt{2k \log \frac{2n}{k}}} \right)^2 \quad , \quad \left(\frac{\sqrt{m}}{\sqrt{m} - \sqrt{2k \log \frac{2n}{k}}} \right)^2$$

The Squared Error of Generalized LASSO

- for $n \times n$-dimensional rank r matrices and nuclear norm minimization, we have the bounds

$$\frac{3r(2n-r)}{m-3r(2n-r)}\,,\quad \left(\frac{\sqrt{3r(2n-r)}}{\sqrt{m}-\sqrt{3r(2n-r)}}\right)^2\,,\quad \left(\frac{\sqrt{m}}{\sqrt{m}-\sqrt{3r(2n-r)}}\right)^2$$

The Squared Error of Generalized LASSO

- for $n \times n$-dimensional rank r matrices and nuclear norm minimization, we have the bounds

$$\frac{3r(2n-r)}{m-3r(2n-r)}\,,\quad \left(\frac{\sqrt{3r(2n-r)}}{\sqrt{m}-\sqrt{3r(2n-r)}}\right)^2\,,\quad \left(\frac{\sqrt{m}}{\sqrt{m}-\sqrt{3r(2n-r)}}\right)^2$$

- for qb-dimensional k-block sparse vectors and mixed ℓ_1/ℓ_2 minimization we have the bounds

$$\frac{4k(b+\log\frac{q}{k})}{m-4k(b+\log\frac{q}{k})}\,,\quad \left(\frac{\sqrt{4k(b+\log\frac{q}{k})}}{\sqrt{m}-\sqrt{4k(b+\log\frac{q}{k})}}\right)^2\,,\quad \left(\frac{\sqrt{m}}{\sqrt{m}-\sqrt{4k(b+\log\frac{q}{k})}}\right)$$

The Squared Error of Generalized LASSO

- These results required either knowledge of the *optimal* λ, or of $f(x_0)$.

The Squared Error of Generalized LASSO

- These results required either knowledge of the *optimal* λ, or of $f(x_0)$.
- But what if we do not have these?

The Squared Error of Generalized LASSO

- These results required either knowledge of the *optimal* λ, or of $f(x_0)$.
- But what if we do not have these?
- Can we still give formulae for an arbitrary $\lambda \geq 0$?

Three Versions of the LASSO Problem

Observe $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}$, where $\mathbf{A} \sim \mathcal{N}(0, \frac{1}{m}\mathbf{I}_{m\times n})$ $\mathbf{z} \sim \mathcal{N}(0, \sigma^2\mathbf{I}_m)$

Three Versions of the LASSO Problem

Observe $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}$, where $\mathbf{A} \sim \mathcal{N}(0, \frac{1}{m}\mathbf{I}_{m\times n})$ $\mathbf{z} \sim \mathcal{N}(0, \sigma^2\mathbf{I}_m)$

1. Constrained LASSO

$$\mathbf{x}_c^* = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \quad \text{subject to} \quad f(\mathbf{x}) \leq f(\mathbf{x}_0)$$

Three Versions of the LASSO Problem

Observe $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}$, where $\mathbf{A} \sim \mathcal{N}(0, \frac{1}{m}\mathbf{I}_{m\times n})$ $\mathbf{z} \sim \mathcal{N}(0, \sigma^2\mathbf{I}_m)$

1. Constrained LASSO

$$\mathbf{x}_c^* = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \quad \text{subject to} \quad f(\mathbf{x}) \leq f(\mathbf{x}_0)$$

2. ℓ_2-LASSO

$$\mathbf{x}_{\ell_2}^* = \arg\min_{\mathbf{x}} \{\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 + \lambda f(\mathbf{x})\}$$

Three Versions of the LASSO Problem

Observe $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}$, where $\mathbf{A} \sim \mathcal{N}(0, \frac{1}{m}\mathbf{I}_{m\times n})$ $\mathbf{z} \sim \mathcal{N}(0, \sigma^2\mathbf{I}_m)$

1. Constrained LASSO

$$\mathbf{x}_c^* = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \quad \text{subject to} \quad f(\mathbf{x}) \leq f(\mathbf{x}_0)$$

2. ℓ_2-LASSO

$$\mathbf{x}_{\ell_2}^* = \arg\min_{\mathbf{x}} \{\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 + \lambda f(\mathbf{x})\}$$

3. ℓ_2^2-LASSO

$$\mathbf{x}_{\ell_2^2}^* = \arg\min_{\mathbf{x}} \left\{\frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \sigma\tau f(\mathbf{x})\right\}$$

Three Versions of the LASSO Problem

Observe $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}$, where $\mathbf{A} \sim \mathcal{N}(0, \frac{1}{m}\mathbf{I}_{m\times n})$ $\mathbf{z} \sim \mathcal{N}(0, \sigma^2\mathbf{I}_m)$

1. Constrained LASSO

$$\mathbf{x}_c^* = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \quad \text{subject to} \quad f(\mathbf{x}) \leq f(\mathbf{x}_0)$$

2. ℓ_2-LASSO

$$\mathbf{x}_{\ell_2}^* = \arg\min_{\mathbf{x}} \{\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 + \lambda f(\mathbf{x})\}$$

3. ℓ_2^2-LASSO

$$\mathbf{x}_{\ell_2^2}^* = \arg\min_{\mathbf{x}} \left\{\frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \sigma\tau f(\mathbf{x})\right\}$$

Quantity of Interest:

$$\boxed{\text{Normalized Squared Error} := \frac{\|\mathbf{x}_{\text{LASSO}}^* - \mathbf{x}_0\|_2^2}{\|z\|_2^2}}$$

Relevant Literature

- Bayati & Montanari, '11: Exact Characterization of the squared error of ℓ_2^2-LASSO for $f(\mathbf{x}) = \|\mathbf{x}\|_1$

Relevant Literature

- Bayati & Montanari, '11: Exact Characterization of the squared error of ℓ_2^2-LASSO for $f(\mathbf{x}) = \|\mathbf{x}\|_1$
  - for z iid $N(0, \sigma^2)$

Relevant Literature

- Bayati & Montanari, '11: Exact Characterization of the squared error of ℓ_2^2-LASSO for $f(\mathbf{x}) = \|\mathbf{x}\|_1$
  - for z iid $N(0, \sigma^2)$
  - through equivalence to "Approximate Message Passing" (AMP) algorithm

Relevant Literature

- Bayati & Montanari, '11: Exact Characterization of the squared error of ℓ_2^2-LASSO for $f(\mathbf{x}) = \|\mathbf{x}\|_1$
  - for z iid $N(0, \sigma^2)$
  - through equivalence to "Approximate Message Passing" (AMP) algorithm
- Stojnic, '13:

Relevant Literature

- Bayati & Montanari, '11: Exact Characterization of the squared error of ℓ_2^2-LASSO for $f(\mathbf{x}) = \|\mathbf{x}\|_1$
  - for z iid $N(0, \sigma^2)$
  - through equivalence to "Approximate Message Passing" (AMP) algorithm
- Stojnic, '13:
  - for z iid $N(0, \sigma^2)$

Relevant Literature

- Bayati & Montanari, '11: Exact Characterization of the squared error of ℓ_2^2-LASSO for $f(\mathbf{x}) = \|\mathbf{x}\|_1$
  - for z iid $N(0, \sigma^2)$
  - through equivalence to "Approximate Message Passing" (AMP) algorithm
- Stojnic, '13:
  - for z iid $N(0, \sigma^2)$
  - precise bounds for the ℓ_1-constrained LASSO

Relevant Literature

- Bayati & Montanari, '11: Exact Characterization of the squared error of ℓ_2^2-LASSO for $f(\mathbf{x}) = \|\mathbf{x}\|_1$
  - for z iid $N(0, \sigma^2)$
  - through equivalence to "Approximate Message Passing" (AMP) algorithm
- Stojnic, '13:
  - for z iid $N(0, \sigma^2)$
  - precise bounds for the ℓ_1-constrained LASSO
  - developed framework for analysis based on Gordon's lemma that compares Gaussian processes ['88]

Relevant Literature

- Bayati & Montanari, '11: Exact Characterization of the squared error of ℓ_2^2-LASSO for $f(\mathbf{x}) = \|\mathbf{x}\|_1$
  - for z iid $N(0, \sigma^2)$
  - through equivalence to "Approximate Message Passing" (AMP) algorithm
- Stojnic, '13:
  - for z iid $N(0, \sigma^2)$
  - precise bounds for the ℓ_1-constrained LASSO
  - developed framework for analysis based on Gordon's lemma that compares Gaussian processes ['88]
  - *we rely on this framework*

What's New?

- **Simplify** the powerful framework proposed by Stojnic '13

What's New?

- **Simplify** the powerful framework proposed by Stojnic '13
- **Generalize** results on the constrained LASSO for *arbitrary* convex functions

What's New?

- **Simplify** the powerful framework proposed by Stojnic '13
- **Generalize** results on the constrained LASSO for *arbitrary* convex functions
- **Extend** analysis to the ℓ_2-LASSO; precise bounds as functions of the regularizer parameter λ

What's New?

- **Simplify** the powerful framework proposed by Stojnic '13
- **Generalize** results on the constrained LASSO for *arbitrary* convex functions
- **Extend** analysis to the ℓ_2-LASSO; precise bounds as functions of the regularizer parameter λ
- Establish **connection** of the ℓ_2-LASSO to the ℓ_2^2-LASSO; propose a formula for calculating the squared estimation error of the latter

What's New?

- **Simplify** the powerful framework proposed by Stojnic '13
- **Generalize** results on the constrained LASSO for *arbitrary* convex functions
- **Extend** analysis to the ℓ_2-LASSO; precise bounds as functions of the regularizer parameter λ
- Establish **connection** of the ℓ_2-LASSO to the ℓ_2^2-LASSO; propose a formula for calculating the squared estimation error of the latter
- Simple recipe for **optimal tuning** of the regularization parameter λ

What's New?

- **Simplify** the powerful framework proposed by Stojnic '13
- **Generalize** results on the constrained LASSO for *arbitrary* convex functions
- **Extend** analysis to the ℓ_2-LASSO; precise bounds as functions of the regularizer parameter λ
- Establish **connection** of the ℓ_2-LASSO to the ℓ_2^2-LASSO; propose a formula for calculating the squared estimation error of the latter
- Simple recipe for **optimal tuning** of the regularization parameter λ
- **Converse results**: When does robust estimation of $\mathbf{x}_0$ fail? Connection to phase transitions of noiseless CS, to statistical dimension, and to similar formulae in standard least-squares.

What's New?

- **Simplify** the powerful framework proposed by Stojnic '13
- **Generalize** results on the constrained LASSO for *arbitrary* convex functions
- **Extend** analysis to the ℓ_2-LASSO; precise bounds as functions of the regularizer parameter λ
- Establish **connection** of the ℓ_2-LASSO to the ℓ_2^2-LASSO; propose a formula for calculating the squared estimation error of the latter
- Simple recipe for **optimal tuning** of the regularization parameter λ
- **Converse results**: When does robust estimation of $\mathbf{x}_0$ fail? Connection to phase transitions of noiseless CS, to statistical dimension, and to similar formulae in standard least-squares.
- **Extend** the results from z iid $N(0, \sigma^2)$ to **arbitrary** z and **worst-case** z.

First-Order Approximation

$$\min_{\mathbf{x}} \left\{ \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 + \lambda f(\mathbf{x}) \right\}$$

First-Order Approximation

$$\min_{\mathbf{x}} \{\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 + \lambda f(\mathbf{x})\}$$

- $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(0, \sigma \mathbf{I}_m)$

First-Order Approximation

$$\min_{\mathbf{x}} \{\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 + \lambda f(\mathbf{x})\}$$

- $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(0, \sigma \mathbf{I}_m)$
- $f(\mathbf{x}_0 + \mathbf{w}) \gtrsim f(\mathbf{x}_0) + \sup_{\mathbf{s} \in \partial f(\mathbf{x}_0)} \mathbf{s}^T \mathbf{w}$

First-Order Approximation

$$\min_{\mathbf{x}} \{\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 + \lambda f(\mathbf{x})\}$$

- $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(0, \sigma \mathbf{I}_m)$
- $f(\mathbf{x}_0 + \mathbf{w}) \gtrsim f(\mathbf{x}_0) + \sup_{\mathbf{s} \in \partial f(\mathbf{x}_0)} \mathbf{s}^T \mathbf{w}$
- Lower bound becomes **tight** when $\sigma \to 0$, since $\|\mathbf{w}^*\|_2$ becomes small

First-Order Approximation

$$\min_{\mathbf{x}} \{\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 + \lambda f(\mathbf{x})\}$$

- $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(0, \sigma \mathbf{I}_m)$
- $f(\mathbf{x}_0 + \mathbf{w}) \gtrsim f(\mathbf{x}_0) + \sup_{\mathbf{s} \in \partial f(\mathbf{x}_0)} \mathbf{s}^T \mathbf{w}$
- Lower bound becomes **tight** when $\sigma \to 0$, since $\|\mathbf{w}^*\|_2$ becomes small

Approximated LASSO Problem:

$$\min_{\mathbf{w}} \left\{ \|\mathbf{A}\mathbf{w} - \mathbf{z}\|_2 + \sup_{\mathbf{s} \in \lambda \partial f(\mathbf{x}_0)} \mathbf{s}^T \mathbf{w} \right\}$$

First-Order Approximation

$$\min_{\mathbf{x}} \{\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 + \lambda f(\mathbf{x})\}$$

- $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(0, \sigma \mathbf{I}_m)$
- $f(\mathbf{x}_0 + \mathbf{w}) \gtrsim f(\mathbf{x}_0) + \sup_{\mathbf{s} \in \partial f(\mathbf{x}_0)} \mathbf{s}^T \mathbf{w}$
- Lower bound becomes **tight** when $\sigma \to 0$, since $\|\mathbf{w}^*\|_2$ becomes small

Approximated LASSO Problem:

$$\min_{\mathbf{w}} \left\{ \|\mathbf{A}\mathbf{w} - \mathbf{z}\|_2 + \sup_{\mathbf{s} \in \lambda \partial f(\mathbf{x}_0)} \mathbf{s}^T \mathbf{w} \right\}$$

- For the Constrained Problem, just replace $\lambda \partial f(\mathbf{x}_0)$ with $\text{cone}(\partial f(\mathbf{x}_0))$.

Why $\sigma \to 0$?

- **Precise Formulas**
  - First-order approximation is tight $\implies$ Perform the analysis on the more tractable Approximate LASSO, instead.
  - Precise analysis for arbitrary σ impossible with f.o. approximation.

Why $\sigma \to 0$?

- **Precise Formulas**
  - First-order approximation is tight $\implies$ Perform the analysis on the more tractable Approximate LASSO, instead.
  - Precise analysis for arbitrary σ impossible with f.o. approximation.
- **Worst-case error scenario**
  - Worst case NSE of the LASSO is achieved for $\sigma \to 0$.
  - Our formulae *upper bound* the NSE for *arbitrary* values of the noise variance.

Gordon's Lemma

- Introduced by Gordon in '88. Compares two centered Gaussian processes:

Lemma (Gordon)

Let $\mathbf{G} \in \mathbf{R}^{\mathbf{m}\times\mathbf{n}}, g \in \mathbf{R}, \mathbf{g} \in \mathbf{R}^{\mathbf{m}}, \mathbf{h} \in \mathbf{R}^{\mathbf{n}}$, *all having entries i.i.d.* $\mathcal{N}(0,1)$ *and being independent of each other. Also, let* $\mathcal{S} \subset \mathbb{R}^n$ *be arbitrary set and* $\psi : \mathcal{S} \to \mathbb{R}$ *be arbitrary function. Then, for all choices of* $c \in \mathbb{R}$,

$$\mathbb{P}\left(\min_{\mathbf{x}\in\mathcal{S}} \{\|\mathbf{G}\mathbf{x}\|_2 + \|\mathbf{x}\|_2 g - \psi(\mathbf{x})\} \geq c\right) \geq \mathbb{P}\left(\min_{\mathbf{x}\in\mathcal{S}} \left\{\|\mathbf{x}\|_2\|\mathbf{g}\|_2 - \mathbf{h}^T\mathbf{x} - \psi(\mathbf{x})\right\} \geq c\right)$$

Gordon's Lemma

- Introduced by Gordon in '88. Compares two centered Gaussian processes:

Lemma (Gordon)

Let $\mathbf{G} \in \mathbf{R}^{\mathbf{m}\times\mathbf{n}}, g \in \mathbf{R}, \mathbf{g} \in \mathbf{R}^{\mathbf{m}}, \mathbf{h} \in \mathbf{R}^{\mathbf{n}}$, *all having entries i.i.d.* $\mathcal{N}(0,1)$ *and being independent of each other. Also, let* $\mathcal{S} \subset \mathbb{R}^n$ *be arbitrary set and* $\psi : \mathcal{S} \to \mathbb{R}$ *be arbitrary function. Then, for all choices of* $c \in \mathbb{R}$,

$$\mathbb{P}\left(\min_{\mathbf{x}\in\mathcal{S}} \left\{\|\mathbf{G}\mathbf{x}\|_2 + \|\mathbf{x}\|_2 g - \psi(\mathbf{x})\right\} \geq c\right) \geq \mathbb{P}\left(\min_{\mathbf{x}\in\mathcal{S}} \left\{\|\mathbf{x}\|_2\|\mathbf{g}\|_2 - \mathbf{h}^T\mathbf{x} - \psi(\mathbf{x})\right\} \geq c\right)$$

- "Escape through mesh" is a Corollary of this Lemma.

Gordon's Lemma

- Introduced by Gordon in '88. Compares two centered Gaussian processes:

Lemma (Gordon)

Let $\mathbf{G} \in \mathbf{R}^{\mathbf{m}\times\mathbf{n}}, g \in \mathbf{R}, \mathbf{g} \in \mathbf{R}^{\mathbf{m}}, \mathbf{h} \in \mathbf{R}^{\mathbf{n}}$, *all having entries i.i.d.* $\mathcal{N}(0,1)$ *and being independent of each other. Also, let* $\mathcal{S} \subset \mathbb{R}^n$ *be arbitrary set and* $\psi : \mathcal{S} \to \mathbb{R}$ *be arbitrary function. Then, for all choices of* $c \in \mathbb{R}$,

$$\mathbb{P}\left(\min_{\mathbf{x}\in\mathcal{S}} \{\|\mathbf{G}\mathbf{x}\|_2 + \|\mathbf{x}\|_2 g - \psi(\mathbf{x})\} \geq c\right) \geq \mathbb{P}\left(\min_{\mathbf{x}\in\mathcal{S}} \left\{\|\mathbf{x}\|_2\|\mathbf{g}\|_2 - \mathbf{h}^T\mathbf{x} - \psi(\mathbf{x})\right\} \geq c\right)$$

- "Escape through mesh" is a Corollary of this Lemma.

Apply (a slight modification) to Approximated LASSO:

$$\mathbb{P}\left(\min_{\mathbf{w}} \left\{\|\mathbf{A}\mathbf{w} - \mathbf{z}\|_2 + \sup_{\mathbf{s}\in\lambda\partial f(\mathbf{x}_0)} \mathbf{s}^T\mathbf{w}\right\} \geq c\right)$$

$$\geq 2\,\mathbb{P}\left(\min_{\mathbf{w}} \left\{\sqrt{\|\mathbf{w}\|_2^2 + \sigma^2}\|\mathbf{g}\|_2 - \mathbf{h}^T\mathbf{w} + \sup_{\mathbf{s}\in\lambda\partial f(\mathbf{x}_0)} \mathbf{s}^T\mathbf{w}\right\} \geq c\right) - 1$$

After Gordon's Lemma: Deterministic Analysis

Key Optimization

$$\mathcal{L}(\mathbf{g},\mathbf{h}) = \min_{\mathbf{w}} \left\{ \sqrt{\|\mathbf{w}\|_2^2 + \sigma^2}\|\mathbf{g}\|_2 - \mathbf{h}^T\mathbf{w} + \sup_{\mathbf{s}\in\lambda\partial f(\mathbf{x}_0)} \mathbf{s}^T\mathbf{w} \right\}$$

After Gordon's Lemma: Deterministic Analysis

Key Optimization

$$\mathcal{L}(\mathbf{g},\mathbf{h}) = \min_{\mathbf{w}} \left\{ \sqrt{\|\mathbf{w}\|_2^2 + \sigma^2}\|\mathbf{g}\|_2 - \mathbf{h}^T\mathbf{w} + \sup_{\mathbf{s}\in\lambda\partial f(\mathbf{x}_0)} \mathbf{s}^T\mathbf{w} \right\}$$

- Reduce to *scalar* optimization:

$$\mathcal{L}(\mathbf{g},\mathbf{h}) = \min_{\alpha} \left\{ \sqrt{\alpha^2 + \sigma^2}\|\mathbf{g}\|_2 - \underbrace{\max_{\|\mathbf{w}\|_2=\alpha} \min_{\mathbf{s}\in\lambda\partial f(\mathbf{x}_0)} (\mathbf{h}-\mathbf{s})^T\mathbf{w}}_{=\alpha\cdot\mathrm{dist}(\mathbf{h},\lambda\partial f(\mathbf{x}_0))} \right\}$$

After Gordon's Lemma: Deterministic Analysis

Key Optimization

$$\mathcal{L}(\mathbf{g},\mathbf{h}) = \min_{\mathbf{w}} \left\{ \sqrt{\|\mathbf{w}\|_2^2 + \sigma^2}\|\mathbf{g}\|_2 - \mathbf{h}^T\mathbf{w} + \sup_{\mathbf{s}\in\lambda\partial f(\mathbf{x}_0)} \mathbf{s}^T\mathbf{w} \right\}$$

- Reduce to *scalar* optimization:

$$\mathcal{L}(\mathbf{g},\mathbf{h}) = \min_{\alpha} \left\{ \sqrt{\alpha^2+\sigma^2}\|\mathbf{g}\|_2 - \underbrace{\max_{\|\mathbf{w}\|_2=\alpha} \min_{\mathbf{s}\in\lambda\partial f(\mathbf{x}_0)} (\mathbf{h}-\mathbf{s})^T\mathbf{w}}_{=\alpha\cdot\mathrm{dist}(\mathbf{h},\lambda\partial f(\mathbf{x}_0))} \right\}$$

- Solve:

$$\|\mathbf{w}^*\|_2^2 = (\alpha^*)^2 = \sigma^2 \frac{\mathrm{dist}^2(\mathbf{h},\lambda\partial f(\mathbf{x}_0))}{\|\mathbf{g}\|_2^2 - \mathrm{dist}^2(\mathbf{h},\lambda\partial f(\mathbf{x}_0))} \quad , \quad \mathcal{L}(\mathbf{g},\mathbf{h}) = \sigma\sqrt{\|\mathbf{g}\|_2^2 - \mathrm{dist}^2(\mathbf{h},\lambda\partial f(\mathbf{x}_0))}$$

After Gordon's Lemma: Probabilistic Analysis

Key Optimization

$$\mathcal{L}(\mathbf{g},\mathbf{h}) = \min_{\mathbf{w}} \left\{ \sqrt{\|\mathbf{w}\|_2^2 + \sigma^2}\|\mathbf{g}\|_2 - \mathbf{h}^T\mathbf{w} + \sup_{\mathbf{s}\in\lambda\partial f(\mathbf{x}_0)} \mathbf{s}^T\mathbf{w} \right\}$$

After Gordon's Lemma: Probabilistic Analysis

Key Optimization

$$\mathcal{L}(\mathbf{g},\mathbf{h}) = \min_{\mathbf{w}} \left\{ \sqrt{\|\mathbf{w}\|_2^2 + \sigma^2}\|\mathbf{g}\|_2 - \mathbf{h}^T\mathbf{w} + \sup_{\mathbf{s}\in\lambda\partial f(\mathbf{x}_0)} \mathbf{s}^T\mathbf{w} \right\}$$

- *Easy*: $\mathbf{h} \sim \mathcal{N}(0,\mathbf{I}_n)$, $\mathbf{g} \sim \mathcal{N}(0,\mathbf{I}_m)$ and independent!
- Basic *Concentration* Arguments:

$$\text{dist}^2(\mathbf{h},\lambda\partial f(\mathbf{x}_0)) \approx \underbrace{\mathbb{E}_{\mathbf{h}}\left[\text{dist}^2(\mathbf{h},\lambda\partial f(\mathbf{x}_0))\right] := \mathbf{D}_f(\mathbf{x}_0,\lambda)}_{\text{"Gaussian Squared Distance"}}$$

After Gordon's Lemma: Probabilistic Analysis

Key Optimization

$$\mathcal{L}(\mathbf{g},\mathbf{h}) = \min_{\mathbf{w}} \left\{ \sqrt{\|\mathbf{w}\|_2^2 + \sigma^2}\|\mathbf{g}\|_2 - \mathbf{h}^T\mathbf{w} + \sup_{\mathbf{s}\in\lambda\partial f(\mathbf{x}_0)} \mathbf{s}^T\mathbf{w} \right\}$$

- *Easy*: $\mathbf{h} \sim \mathcal{N}(0,\mathbf{I}_n)$, $\mathbf{g} \sim \mathcal{N}(0,\mathbf{I}_m)$ and independent!
- Basic *Concentration* Arguments:

$$\text{dist}^2(\mathbf{h},\lambda\partial f(\mathbf{x}_0)) \approx \underbrace{\mathbb{E}_{\mathbf{h}}\left[\text{dist}^2(\mathbf{h},\lambda\partial f(\mathbf{x}_0))\right] := \mathbf{D}_f(\mathbf{x}_0,\lambda)}_{\text{"Gaussian Squared Distance"}}$$

- Apply to Deterministic results:

$$\mathcal{L}(\mathbf{g},\mathbf{h}) = \sigma\sqrt{\|\mathbf{g}\|_2^2 - \text{dist}^2(\mathbf{h},\lambda\partial f(\mathbf{x}_0))} \approx \boxed{\sigma\sqrt{m - \mathbf{D}_f(\mathbf{x}_0,\lambda)}} \implies$$

High-Probability Lower Bound for LASSO cost!

$$\frac{\|\mathbf{w}^*\|_2^2}{\|z\|_2^2} = \frac{\text{dist}^2(\mathbf{h},\lambda\partial f(\mathbf{x}_0))}{\|\mathbf{g}\|_2^2 - \text{dist}^2(\mathbf{h},\lambda\partial f(\mathbf{x}_0))} \approx \boxed{\frac{\mathbf{D}_f(\mathbf{x}_0,\lambda)}{m - \mathbf{D}_f(\mathbf{x}_0,\lambda)}} \xrightarrow{???}$$

Normalized Square Error of the LASSO !

Synopsis of the Technical Framework

$$\mathcal{F}_{\ell_2}(\mathbf{A},\mathbf{z}) = \min_{\mathbf{w}} \left\{ \|\mathbf{A}\mathbf{w}-\mathbf{z}\|_2 + \sup_{\mathbf{s}\in\lambda\partial f(\mathbf{x}_0)} \mathbf{s}^T\mathbf{w} \right\}$$

1. Gordon's Lemma to $\mathcal{F}_{\ell_2}(\mathbf{A},\mathbf{z})$ to find a *high-probability* **lower bound** for it. We showed that!
2. Gordon's Lemma to the *dual* of $\mathcal{F}_{\ell_2}(\mathbf{A},\mathbf{z})$ to find a *high-probability* **upper bound** for it
3. Lower and Upper Bounds **match** $\Longrightarrow$ $\boxed{\mathcal{F}_{\ell_2}(\mathbf{A},\mathbf{z}) \approx \sigma\sqrt{m-\mathbf{D}_f(\mathbf{x}_0,\lambda)}}$
4. Assume $\|\mathbf{w}^*_{\ell_2}\|_2^2$ deviates from $\sigma^2\frac{\mathbf{D}_f(\mathbf{x}_0,\lambda)}{m-\mathbf{D}_f(\mathbf{x}_0,\lambda)}$. Then, Gordon's Lemma shows that optimization cost is *strictly larger* than $\sigma\sqrt{m-\mathbf{D}_f(\mathbf{x}_0,\lambda)}$.
5. Conclude $\boxed{\frac{\|\mathbf{w}^*_{\ell_2}\|_2^2}{\|z\|_2^2} \approx \frac{\mathbf{D}_f(\mathbf{x}_0,\lambda)}{m-\mathbf{D}_f(\mathbf{x}_0,\lambda)}}$

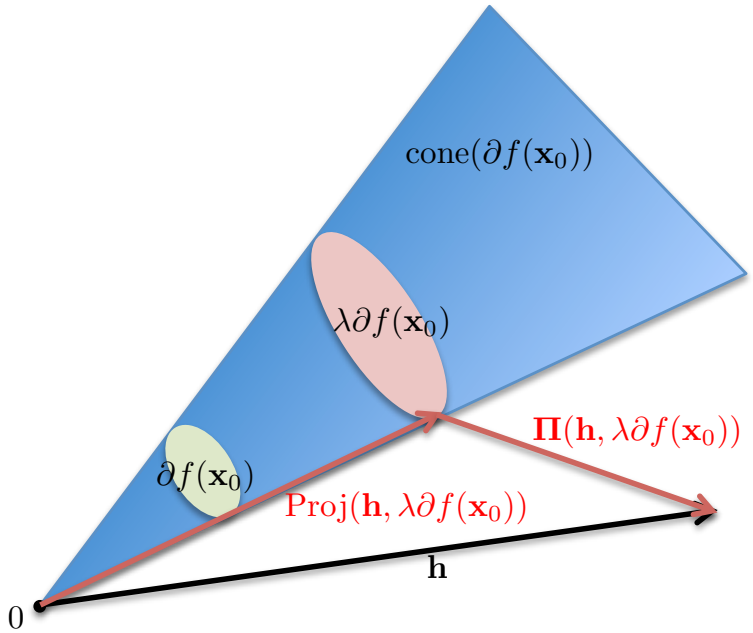
Gaussian Squared Distance and Related Quantities

$\text{dist}(\mathbf{h}, \lambda\partial f(\mathbf{x}_0)) := \|\Pi(\mathbf{h}, \lambda\partial f(\mathbf{x}_0))\|_2$

$\mathbf{D}_f(\mathbf{x}_0, \lambda) := \mathbb{E}_{\mathbf{h}}\left[\text{dist}^2(\mathbf{h}, \lambda\partial f(\mathbf{x}_0))\right]$

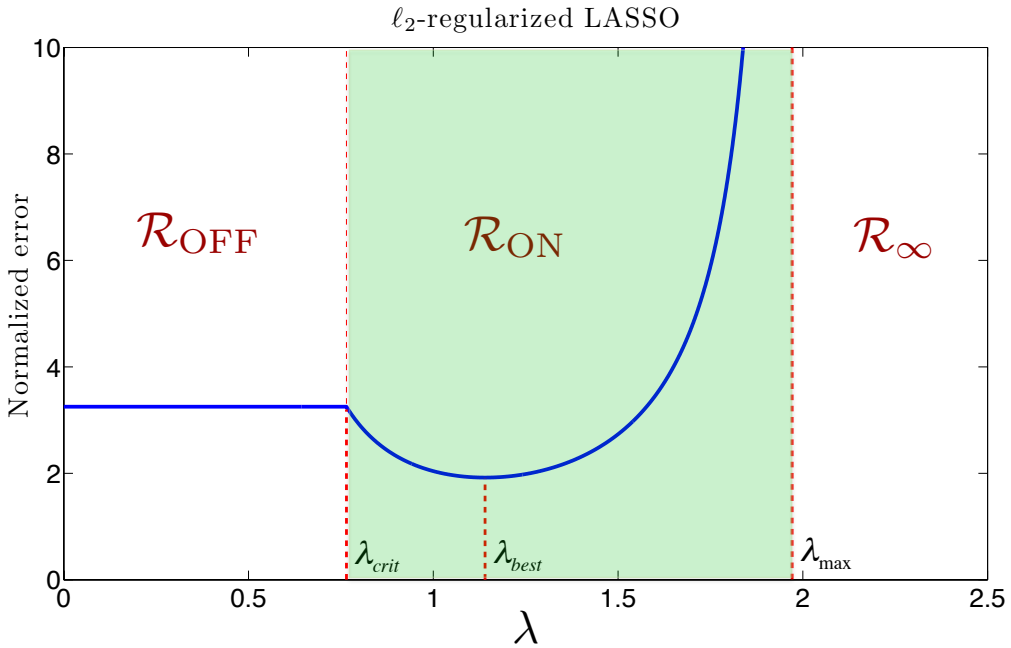
$\mathbf{\Delta}_f(\mathbf{x}_0) = \omega^2 := \mathbb{E}_{\mathbf{h}}\left[\text{dist}^2(\mathbf{h}, \text{cone}(\partial f(\mathbf{x}_0)))\right]$

$\mathbf{C}_f(\mathbf{x}_0, \lambda) := \mathbb{E}_{\mathbf{h}}\left[\langle \text{Proj}(\mathbf{h}, \lambda\partial f(\mathbf{x}_0)), \Pi(\mathbf{h}, \lambda\partial f(\mathbf{x}_0))\rangle\right]$



- $\mathbf{\Delta}_f(\mathbf{x}_0) = \omega^2$: minimum number of measurements required to success of the *Noiseless CS* (*Gaussian width* or *statistical dimension*)
- $\mathbf{D}_f(\mathbf{x}_0, \lambda)$: upper bounds the normalized squared error of the *simple denoiser*
- $\min_{\lambda \geq 0} \mathbf{D}_f(\mathbf{x}_0, \lambda) = \mathbf{\Delta}_f(\mathbf{x}_0)$
- $\mathbf{C}_f(\mathbf{x}_0, \lambda)$: Appears when upper bounding LASSO cost in Step 2

ℓ_2-LASSO: Regions of Operation



R_{OFF}

$$m - \mathbf{D}_f(\mathbf{x}_0, \lambda) \leq \mathbf{C}_f(\mathbf{x}_0, \lambda)$$

$$\mathbf{A}\mathbf{x}^* = \mathbf{y}$$

R_{ON}

$$m \geq \mathbf{D}_f(\mathbf{x}_0, \lambda)$$

$$m - \mathbf{D}_f(\mathbf{x}_0, \lambda) \geq \mathbf{C}_f(\mathbf{x}_0, \lambda)$$

Precise Bounds!

R_{∞}

$$m \leq \mathbf{D}_f(\mathbf{x}_0, \lambda)$$

Not Robust Recovery

ℓ_2-LASSO: R_{ON}

Theorem (ℓ_2-LASSO)

Assume,

- $\frac{\mathbf{D}_f(\mathbf{x}_0,\lambda)}{\epsilon_L} > m > (1+\epsilon_L)\mathbf{D}_f(\mathbf{x}_0,\lambda)$, *for some constant* ϵ_L,
- m *is sufficiently large,*
- $\lambda \in \mathcal{R}_{ON}$.

For any $\epsilon > 0$, *there exists a constant* $C = C(\epsilon, \epsilon_L)$ *and* $\sigma_0 = \sigma_0(\epsilon, \epsilon_L, n)$ *such that, whenever* $\sigma \leq \sigma_0$, *with probability* $1 - \exp(-C^2 \min\{m, \frac{m^2}{n}\})$, *we have,*

$$\left| \frac{\|\mathbf{x}^*_{\ell_2} - \mathbf{x}_0\|_2^2}{\|z\|_2^2} - \frac{\mathbf{D}_f(\mathbf{x}_0,\lambda)}{m - \mathbf{D}_f(\mathbf{x}_0,\lambda)} \right| < \epsilon$$

Optimal Regularizer Parameter:

$$\boxed{\lambda_{\text{best}} = \arg\min_{\lambda \geq 0} \mathbf{D}_f(\mathbf{x}_0,\lambda)}$$

Example

$\mathbf{X}_0 \in \mathbb{R}^{n\times n}$ is rank r. Observe, $\mathbf{y} = \mathcal{A}(\mathbf{X}_0) + \mathbf{z}$, solve the Matrix LASSO,

$$\min_{\mathbf{X}} \{\|\mathbf{y} - \mathcal{A}(\mathbf{X})\|_2 + \lambda\|\mathbf{X}\|_\star\}$$

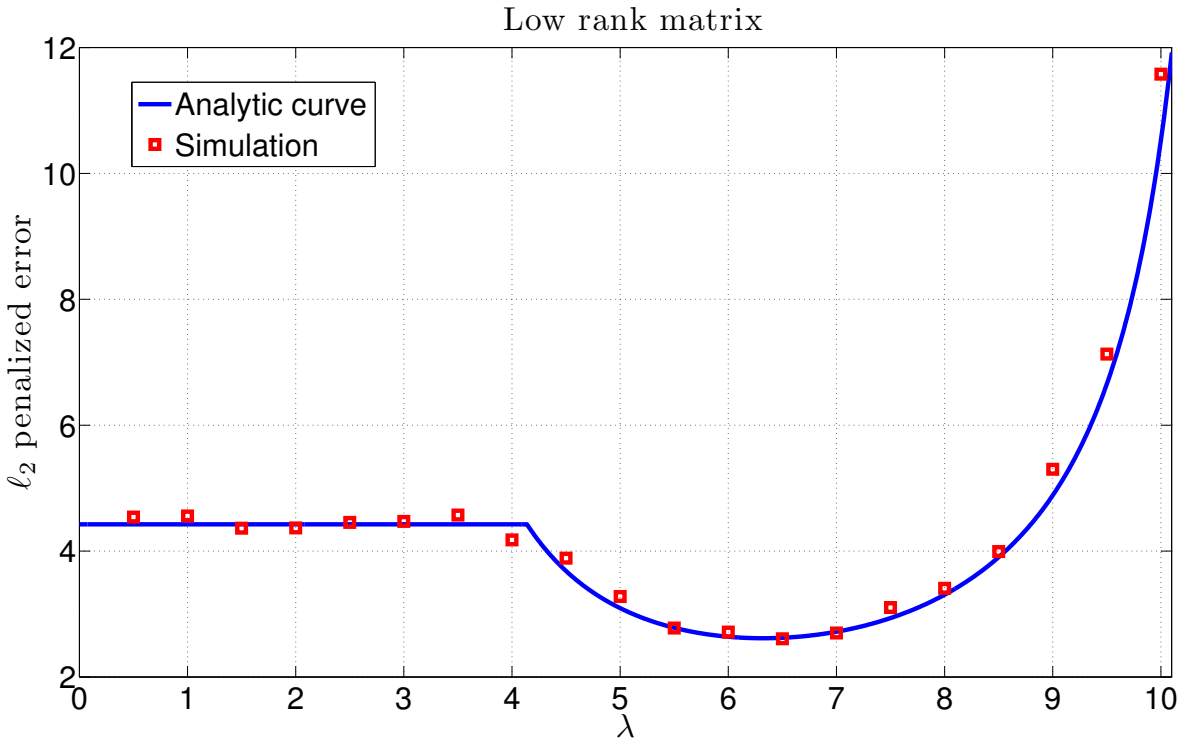


Figure : $n = 45$, $r = 6$, measurements $m = 0.6n^2$.

ℓ_2^2-LASSO: Connection to ℓ_2-LASSO

$$\mathbf{x}^*_{\ell_2} = \arg\min_{\mathbf{x}} \{\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 + \lambda f(\mathbf{x})\}$$

$$\boxed{\frac{\mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x}^*_{\ell_2})}{\|\mathbf{y} - \mathbf{A}\mathbf{x}^*_{\ell_2}\|_2} \in \lambda \partial f(\mathbf{x}^*_{\ell_2})}$$

$$\mathbf{x}^*_{\ell_2^2} = \arg\min_{\mathbf{x}} \{\tfrac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \sigma\tau f(\mathbf{x}$$

$$\boxed{\mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x}^*_{\ell_2^2}) \in \sigma\tau \partial f(\mathbf{x}^*_{\ell_2^2})}$$

Connection

Fix any $\lambda \geq 0$. Suppose $\mathbf{x}^*_{\ell_2}$ is optimal for ℓ_2-LASSO for that λ and $\|\mathbf{y} - \mathbf{A}\mathbf{x}^*_{\ell_2}\|_2 \neq 0$. Then, it is also optimal for the ℓ_2^2-LASSO for

$$\sigma\tau = \|\mathbf{y} - \mathbf{A}\mathbf{x}^*_{\ell_2}\|\lambda.$$

ℓ_2-ℓ_2^2 calibration

Formula

Define,

$$calib(\lambda) = \frac{m - \mathbf{D}_f(\mathbf{x}_0, \lambda) - \mathbf{C}_f(\mathbf{x}_0, \lambda)}{\sqrt{m - \mathbf{D}_f(\mathbf{x}_0, \lambda)}}, \qquad map(\lambda) = \lambda \cdot calib(\lambda)$$

Then, choosing,

$$\tau = map(\lambda)$$

connects ℓ_2 to ℓ_2^2 LASSO.

Intuition

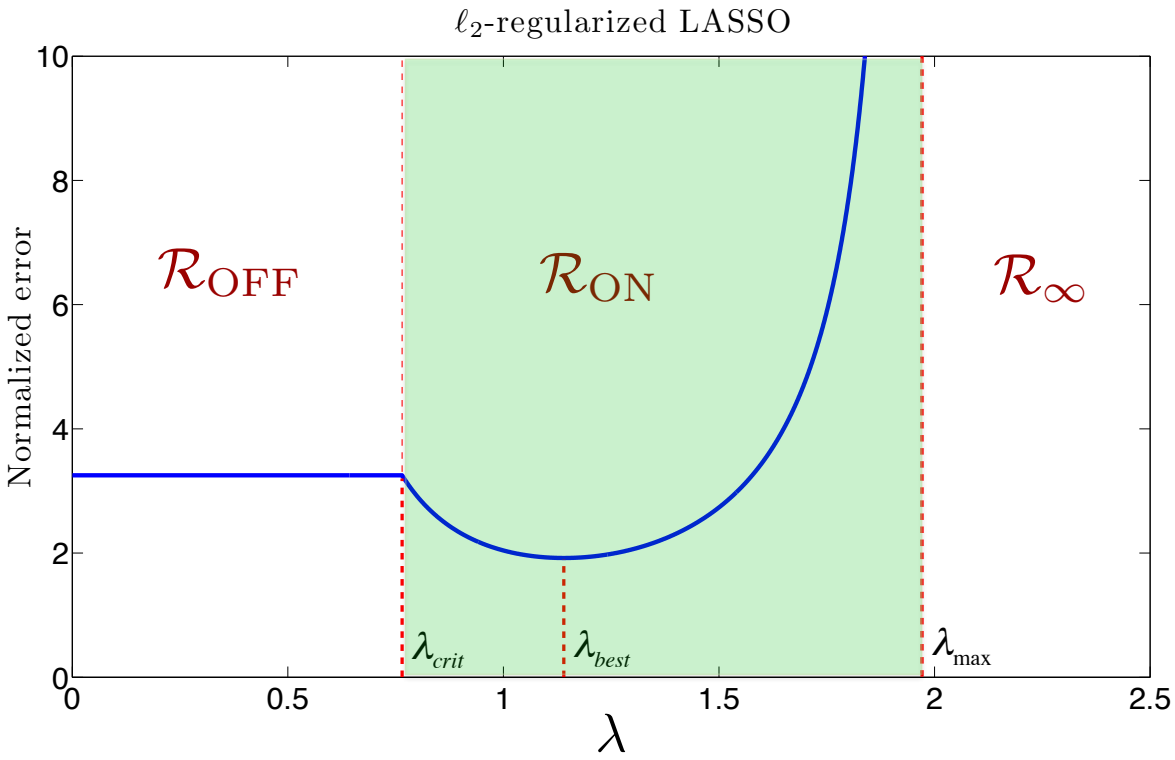
Recall: $\left\{ \|\mathbf{A}\mathbf{w}_{\ell_2}^* - \mathbf{z}\|_2 + \max_{\mathbf{s} \in \lambda \partial f(\mathbf{x}_0)} \mathbf{s}^T \mathbf{w}_{\ell_2}^* \right\} \approx \sigma \sqrt{m - \mathbf{D}_f(\mathbf{x}_0, \lambda)}$.

Conjecture: $\max_{\mathbf{s} \in \lambda \partial f(\mathbf{x}_0)} \mathbf{s}^T \mathbf{w}_{\ell_2}^* \approx \sigma \frac{\mathbf{C}_f(\mathbf{x}_0, \lambda)}{\sqrt{m - \mathbf{D}_f(\mathbf{x}_0, \lambda)}}$.

Proof based on same Framework?

Properties of the mapping

- map(λ) **takes** $\mathcal{R}_{\mathbf{ON}}$ **to** $\mathbb{R}^+$
  - map(λ) is *bijective* and *strictly increasing*.
  - map(λ_{crit}) $= 0$, map(λ_{max}) $= \infty$.

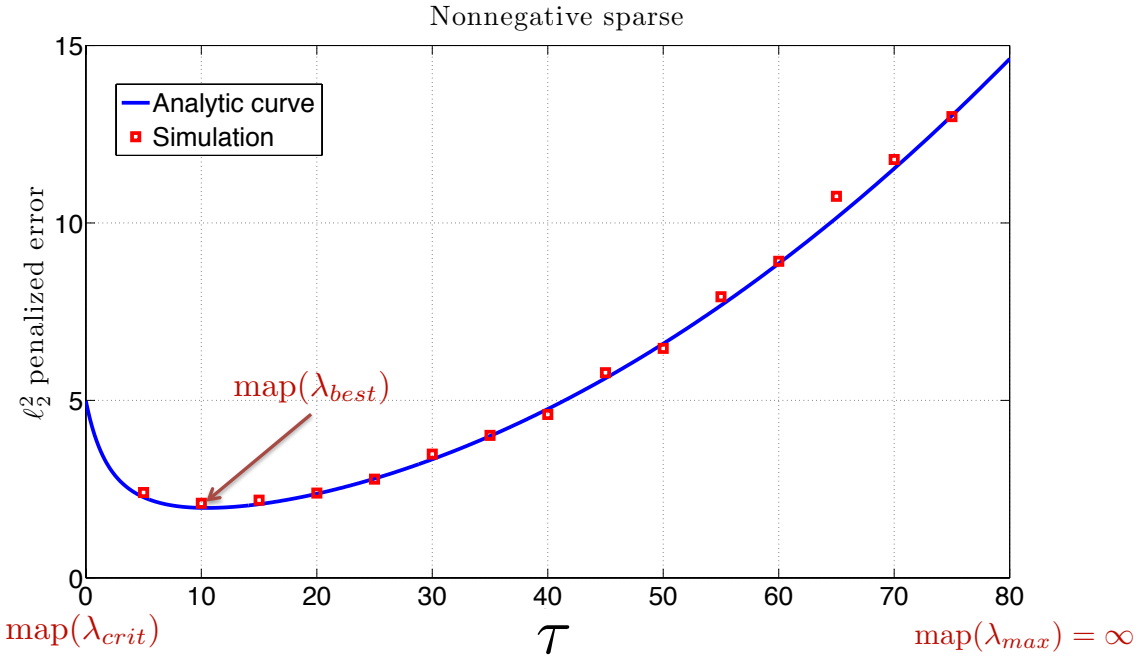


$\Longrightarrow$ R_{ON} is indeed the most important region of ℓ_2-LASSO.

Main Result on ℓ_2^2-LASSO

Formula (ℓ_2^2-LASSO error)

$$\frac{\|\mathbf{x}^*_{\ell_2^2} - \mathbf{x}_0\|}{\|z\|_2^2} \approx \frac{\mathbf{D}_f(\mathbf{x}_0, map^{-1}(\tau))}{m - \mathbf{D}_f(\mathbf{x}_0, map^{-1}(\tau))}$$



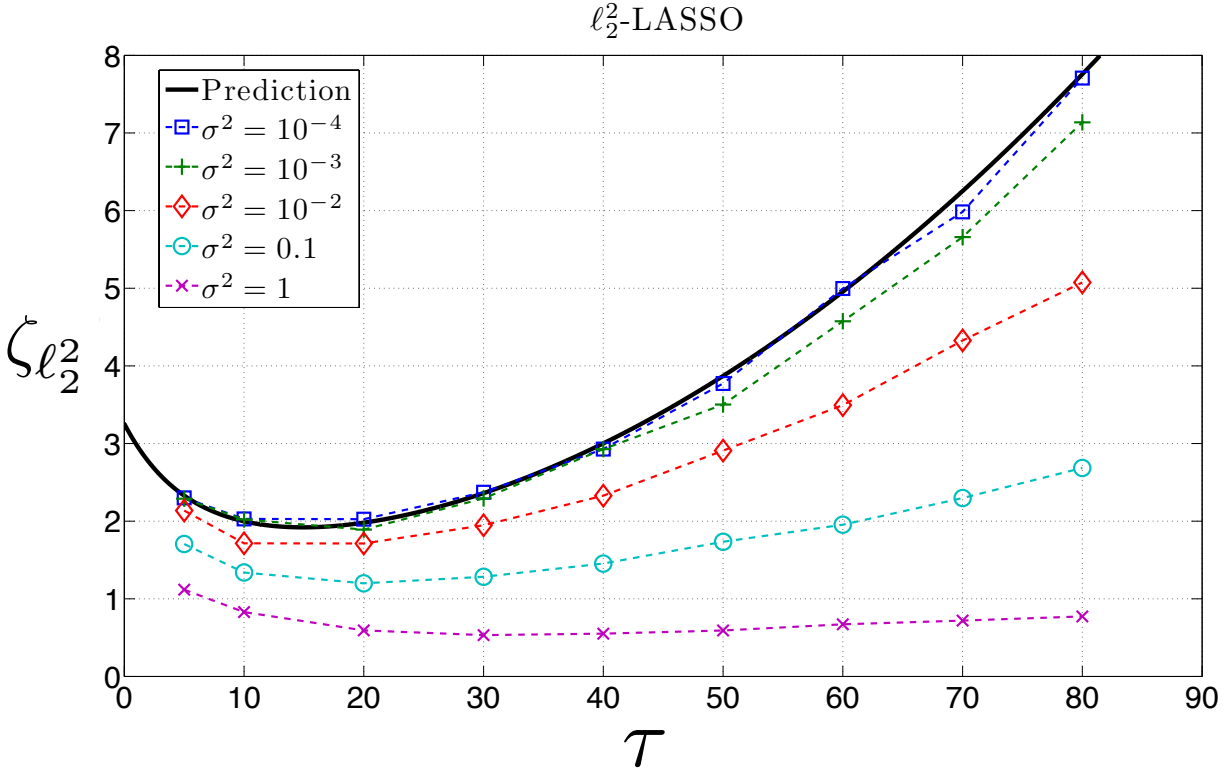
$\frac{k}{n} = 0.1$. $\frac{m}{n} = 0.3$, $n = 1000$, $\sigma^2 = 10^{-5}$.

Long Story Short

	Normalized Squared Error	Optimal Tuning
C-LASSO	$\frac{\mathbf{\Delta}_f(\mathbf{x}_0)}{m-\mathbf{\Delta}_f(\mathbf{x}_0)}$	
ℓ_2-LASSO	$\frac{\mathbf{D}_f(\mathbf{x}_0,\lambda)}{m-\mathbf{D}_f(\mathbf{x}_0,\lambda)}$ for $\lambda \in \mathcal{R}_{\text{ON}}$	$\lambda^* = \arg\min \mathbf{D}_f(\mathbf{x}_0, \lambda)$
ℓ_2^2-LASSO	$\frac{\mathbf{D}_f(\mathbf{x}_0,\text{map}^{-1}(\tau))}{m-\mathbf{D}_f(\mathbf{x}_0,\text{map}^{-1}(\tau))}$ for $\tau \in \mathbb{R}^+$	$\tau^* = \lambda^* \sqrt{m - \mathbf{D}_f(\mathbf{x}_0, \lambda^*)}$

- Our expressions are precise for sufficiently small noise level σ.
- They are upper bounds for arbitrary values of σ. Worst case error happens as $\sigma \to 0$.
- Converse results: When $m < \mathbf{\Delta}_f(\mathbf{x}_0)$, *not* robust recovery.
- Similar formulae for z arbitrary and z worst-case.

Example: Combining ℓ_2^2 and variance predictions



$n = 1000,\ \frac{k}{n} = 0.1,\ \frac{m}{n} = 0.5$

Summary and Conclusion

- Studied the recovery of a structured signal from noisy measurements using generalized LASSO

Summary and Conclusion

- Studied the recovery of a structured signal from noisy measurements using generalized LASSO
- Found formulae for the squared error that are counterparts of those encountered in standard least-squares

Summary and Conclusion

- Studied the recovery of a structured signal from noisy measurements using generalized LASSO
- Found formulae for the squared error that are counterparts of those encountered in standard least-squares
  - formulae hold for arbitrary convex regularizer $f(\cdot)$ and arbitrary regularization parameter $\lambda \geq 0$

Summary and Conclusion

- Studied the recovery of a structured signal from noisy measurements using generalized LASSO
- Found formulae for the squared error that are counterparts of those encountered in standard least-squares
  - formulae hold for arbitrary convex regularizer $f(\cdot)$ and arbitrary regularization parameter $\lambda \geq 0$
  - obtained formulae for z iid $N(0, \sigma^2)$, z arbitrary, and z worst-case

Summary and Conclusion

- Studied the recovery of a structured signal from noisy measurements using generalized LASSO
- Found formulae for the squared error that are counterparts of those encountered in standard least-squares
  - formulae hold for arbitrary convex regularizer $f(\cdot)$ and arbitrary regularization parameter $\lambda \geq 0$
  - obtained formulae for z iid $N(0, \sigma^2)$, z arbitrary, and z worst-case
  - three different versions: constrained LASSO, ℓ_2 LASSO, ℓ_2^2 LASSO

Summary and Conclusion

- Studied the recovery of a structured signal from noisy measurements using generalized LASSO
- Found formulae for the squared error that are counterparts of those encountered in standard least-squares
  - formulae hold for arbitrary convex regularizer $f(\cdot)$ and arbitrary regularization parameter $\lambda \geq 0$
  - obtained formulae for z iid $N(0, \sigma^2)$, z arbitrary, and z worst-case
  - three different versions: constrained LASSO, ℓ_2 LASSO, ℓ_2^2 LASSO
- *Ambient dimension* of the signal is replaced by its *statistical dimension*

Summary and Conclusion

- Studied the recovery of a structured signal from noisy measurements using generalized LASSO
- Found formulae for the squared error that are counterparts of those encountered in standard least-squares
  - formulae hold for arbitrary convex regularizer $f(\cdot)$ and arbitrary regularization parameter $\lambda \geq 0$
  - obtained formulae for z iid $N(0, \sigma^2)$, z arbitrary, and z worst-case
  - three different versions: constrained LASSO, ℓ_2 LASSO, ℓ_2^2 LASSO
- *Ambient dimension* of the signal is replaced by its *statistical dimension*
  - for the optimal λ this is the expected squared distance of a Gaussian vector to the *subgradient cone*

Summary and Conclusion

- Studied the recovery of a structured signal from noisy measurements using generalized LASSO
- Found formulae for the squared error that are counterparts of those encountered in standard least-squares
  - formulae hold for arbitrary convex regularizer $f(\cdot)$ and arbitrary regularization parameter $\lambda \geq 0$
  - obtained formulae for z iid $N(0, \sigma^2)$, z arbitrary, and z worst-case
  - three different versions: constrained LASSO, ℓ_2 LASSO, ℓ_2^2 LASSO
- *Ambient dimension* of the signal is replaced by its *statistical dimension*
  - for the optimal λ this is the expected squared distance of a Gaussian vector to the *subgradient cone*
  - for an arbitrary λ, it is the expected squared distance of a Gaussian vector to the *scaled subgradient set*

Summary and Conclusion

- Studied the recovery of a structured signal from noisy measurements using generalized LASSO
- Found formulae for the squared error that are counterparts of those encountered in standard least-squares
  - formulae hold for arbitrary convex regularizer $f(\cdot)$ and arbitrary regularization parameter $\lambda \geq 0$
  - obtained formulae for z iid $N(0, \sigma^2)$, z arbitrary, and z worst-case
  - three different versions: constrained LASSO, ℓ_2 LASSO, ℓ_2^2 LASSO
- *Ambient dimension* of the signal is replaced by its *statistical dimension*
  - for the optimal λ this is the expected squared distance of a Gaussian vector to the *subgradient cone*
  - for an arbitrary λ, it is the expected squared distance of a Gaussian vector to the *scaled subgradient set*
- Results allow one to compute the optimal value of λ and predict the error behavior of generalized LASSO

Summary and Conclusion

- Studied the recovery of a structured signal from noisy measurements using generalized LASSO
- Found formulae for the squared error that are counterparts of those encountered in standard least-squares
  - formulae hold for arbitrary convex regularizer $f(\cdot)$ and arbitrary regularization parameter $\lambda \geq 0$
  - obtained formulae for z iid $N(0, \sigma^2)$, z arbitrary, and z worst-case
  - three different versions: constrained LASSO, ℓ_2 LASSO, ℓ_2^2 LASSO
- *Ambient dimension* of the signal is replaced by its *statistical dimension*
  - for the optimal λ this is the expected squared distance of a Gaussian vector to the *subgradient cone*
  - for an arbitrary λ, it is the expected squared distance of a Gaussian vector to the *scaled subgradient set*
- Results allow one to compute the optimal value of λ and predict the error behavior of generalized LASSO
- All results in noiseless structured signal recovery (compressed sensing, etc.) follow as special cases