Recent Developments in Compressed Sensing

M. Vidyasagar

Cecil & Ida Green Chair
The University of Texas at Dallas
M.Vidyasagar@utdallas.edu
www.utdallas.edu/~m.vidyasagar

Purdue University, 11 October 2013
Outline

1 Introduction

2 Some Known Results
   - Exact Measurements
   - Noisy Measurements

3 Recent Advances
   - A Modified Elastic Net Algorithm
   - Sparse Group LASSO
   - Compression with $\| \cdot \|_\mu$ Sparsity Index

4 References
Outline

1 Introduction

2 Some Known Results
   - Exact Measurements
   - Noisy Measurements

3 Recent Advances
   - A Modified Elastic Net Algorithm
   - Sparse Group LASSO
   - Compression with $\| \cdot \|_\mu$ Sparsity Index

4 References
Compressed Sensing: Rough Formulation

Simple Version:

Knowing that an $n$-dimensional vector $x$ has very few nonzero components (say $k$), but not knowing the locations of the nonzero components,

- Is it possible to recover $x$ exactly by making $m \ll n$ noise-free linear measurements?
- Is it possible to recover $x$ approximately by making $m \ll n$ noisy linear measurements?
Precise Formulation

Define the set of \textit{k-sparse} vectors in $\mathbb{R}^n$:

$$\Sigma_k = \{ x \in \mathbb{R}^n : |\text{supp}(x)| \leq k \},$$

where $\text{supp}(x) = \{ i : x_i \neq 0 \}$ is the \textbf{support} of $x$.

Is it possible to choose (a) an integer $m \ll n$, (ii) a matrix $A \in \mathbb{R}^{m \times n}$, and (iii) a “demodulation” map $\Delta : \mathbb{R}^m \rightarrow \mathbb{R}^n$, such that

- $\Delta(Ax) = x \ \forall x \in \Sigma_k$?
- $\| \Delta(Ax + \eta) - x \| \leq c\|\eta\| \ \forall x \in \Sigma_k$, where $c$ is a “universal” constant that does not depend on $x$ or $\eta$?

\textbf{Note:} Measurements are linear, but demodulation can be highly nonlinear.
Suppose \( x \in \mathbb{R}^n \) is “nearly \( k \)-sparse,” though not exactly so. Suppose we have \( m \ll n \) exact or noisy linear measurements of \( x \). Is it possible to recover a \( k \)-sparse approximation of \( x \)?

**Signal Compression Interpretation:** Suppose \( x \) represents the Fourier coefficients of a periodic signal, and only \( k \) coefficients are “significant.” Can we construct a good approximation of \( x \) **without** knowing **which** Fourier coefficients are significant?
Define the $k$-sparsity index of $x$ in the norm $\| \cdot \|$.

$$\sigma_k(x, \| \cdot \|) = \inf \{ \| x - z \| : z \in \Sigma_k \}.$$  

Note: $\sigma_k(x, \| \cdot \|)$ depends on the norm $\| \cdot \|$.  

**Question:** Is it possible to choose an integer $m \ll n$, a matrix $A \in \mathbb{R}^{m \times n}$, $m \ll n$, and a “demodulation” map $\Delta : \mathbb{R}^m \to \mathbb{R}^n$, such that

$$\| \Delta(Ax) - x \|_2 \leq C_0 \sigma_k(x, \| \cdot \|_1) \forall x \in \mathbb{R}^n?$$

$$\| \Delta(Ax + \eta) - x \|_2 \leq C_0 \sigma_k(x, \| \cdot \|_1) + C_2 \| \eta \|_2?$$

for “universal” constants $C_0$ and $C_2$?

*Note mixture of $\ell_1$- and $\ell_2$-norms!* More on this later.
1 Introduction

2 Some Known Results
   - Exact Measurements
   - Noisy Measurements

3 Recent Advances
   - A Modified Elastic Net Algorithm
   - Sparse Group LASSO
   - Compression with $\| \cdot \|_\mu$ Sparsity Index

4 References
Outline

1. Introduction

2. Some Known Results
   - Exact Measurements
   - Noisy Measurements

3. Recent Advances
   - A Modified Elastic Net Algorithm
   - Sparse Group LASSO
   - Compression with $\| \cdot \|_\mu$ Sparsity Index

4. References
Restricted Isometry Property (RIP)

**Note:** *Not* most general result, but easy to state!

A matrix $A \in \mathbb{R}^{m \times n}$ is said to satisfy the RIP (Restricted Isometry Property) of order $k$ with constant $\delta_k$ if

$$(1 - \delta_k) \| u \|_2^2 \leq \| Au \|_2^2 \leq (1 + \delta_k) \| u \|_2^2, \quad \forall u \in \Sigma_k.$$ 

**Interpretation:** Every set of $k$ or fewer columns of $A$ is “nearly orthonormal.”

Precisely, if we take columns of $A$ from the set $J \subseteq \{1, \ldots, n\}$, call the submatrix $A_J$, then all eigenvalues of $A_J^t A_J$ lie in the interval $[1 - \delta_k, 1 + \delta_k]$ whenever $|J| \leq k$. 
Candès-Tao Result on $\ell_1$-Norm Minimization

**Theorem:** (Candès-Tao (2005); see also Donoho (2006)). Suppose $A \in \mathbb{R}^{m \times n}$ satisfies the RIP of order $2k$ with constant $\delta_{2k} < \sqrt{2} - 1$, and that $y = Ax$ for some $x \in \Sigma_k$. Define the “demodulation map” $\Delta$ by

$$
\Delta(y) = \hat{x} = \arg\min_z \|z\|_1 \text{ s.t. } y = Az.
$$

Then $\hat{x} = x$.

**Note:** Problem at hand is a linear programming problem.

*Exact recovery of sparse vectors*, if only we can design a matrix $A$ that satisfies RIP.
Designing Matrices with RIP

(Candès-Tao (2005)): Choose columns of $A$ to be realizations of $m$-dimensional zero-mean Gaussians. Then with “high probability” (which can be computed), $A$ satisfies RIP.

**Difficulty:** Resulting $A$ matrix has all nonzero entries with probability one – *implementation issues!*

(Achlioptas (2003)): Choose columns of $A$ to be realizations of i.i.d. (independent and identically distributed) random process $\{X_t\}$ assuming values in $\{-1, 0, +1\}$, with

$$\Pr\{X_t = -1\} = \Pr\{X_t = +1\} = \epsilon, \Pr\{X_t = 0\} = 1 - 2\epsilon.$$ 

**Benefit:** Resulting $A$ matrix is very sparse, *no implementation issues*, and also satisfies RIP with “high probability.”
Outline

1. Introduction

2. Some Known Results
   - Exact Measurements
   - Noisy Measurements

3. Recent Advances
   - A Modified Elastic Net Algorithm
   - Sparse Group LASSO
   - Compression with $\| \cdot \|_\mu$ Sparsity Index

4. References
Defining Near Ideal Behavior

Suppose \( x \in \Sigma_k \), and we measure \( y = A x + \eta \), where \( \| \eta \|_2 \leq \epsilon \), where \( \epsilon \) is known. An “oracle” would know the support set \( J \) of \( x \), and then (in obvious notation)

\[
y = A_J x_J + \eta.
\]

So estimate and estimation error of the oracle are

\[
\hat{x} = (A_J^t A_J)^{-1} A_J^t y,
\]

\[
\hat{x} - x = (A_J^t A_J)^{-1} A_J^t \eta,
\]

\[
\| \hat{x} - x \|_2 \leq \text{const.} \epsilon.
\]

An algorithm is **near ideal** if, *without knowing the support of \( x \)*, it achieves an error proportional to \( \epsilon \), for all \( x \in \Sigma_k \).
A General Theorem

**Theorem:** (Candès-Plan (2009); see also DDEK (2012)). Suppose $A \in \mathbb{R}^{m \times n}$ satisfies the RIP of order $2k$ with constant $\delta_{2k} < \sqrt{2} - 1$, and that $y = Ax + \eta$ for some $x \in \mathbb{R}^n$ and $\eta \in \mathbb{R}^m$ with $\|\eta\|_2 \leq \epsilon$.

$$\hat{x} = \arg\min_z \|z\|_1 \text{ s.t. } \|y - Az\|_2 \leq \epsilon.$$ 

Then

$$\|\hat{x} - x\|_2 \leq C_0 \frac{\sigma_k(x, \|\cdot\|_1)}{\sqrt{k}} + C_2 \epsilon,$$

where

$$C_0 = 2 \frac{1 + (\sqrt{2} - 1)\delta_{2k}}{1 - (\sqrt{2} + 1)\delta_{2k}}, \quad C_2 = \frac{4\sqrt{1 + \delta_{2k}}}{1 - (\sqrt{2} + 1)\delta_{2k}}.$$
Corollary: Suppose $A \in \mathbb{R}^{m \times n}$ satisfies the RIP of order $2k$ with constant $\delta_{2k} < \sqrt{2} - 1$, and that $y = Ax + \eta$ for some $x \in \Sigma_k$ and $\eta \in \mathbb{R}^m$ with $\|\eta\|_2 \leq \epsilon$.

$$\hat{x} = \arg\min_{z} \|z\|_1 \text{ s.t. } \|y - Az\|_2 \leq \epsilon.$$ 

Then

$$\|\hat{x} - x\|_2 \leq C_2 \epsilon,$$

where

$$C_2 = \frac{4\sqrt{1 + \delta_{2k}}}{1 - (\sqrt{2} + 1)\delta_{2k}}.$$
Candès-Plan Result on Near Ideal Behavior of LASSO – 2

Original Problem:

$$\min_{z} \|z\|_1 \text{ s.t. } \|y - Az\|_2 \leq \epsilon.$$ 

Lagrangian Formulation:

$$\min_{z} \|z\|_1 + \lambda(\|y - Az\|_2 - \epsilon)^2.$$ 

“Reverse” Lagrangian Formulation:

$$\min_{z} \|y - Az\|_2^2 \text{ s.t. } \|z\|_1 \leq c.$$ 

This is known as “LASSO” (Tibshirani (1996)). Therefore LASSO has “near ideal behavior.”
Outline

1. Introduction

2. Some Known Results
   - Exact Measurements
   - Noisy Measurements

3. Recent Advances
   - A Modified Elastic Net Algorithm
   - Sparse Group LASSO
   - Compression with $\| \cdot \|_\mu$ Sparsity Index

4. References
Introduction

Some Known Results
- Exact Measurements
- Noisy Measurements

Recent Advances
- A Modified Elastic Net Algorithm
- Sparse Group LASSO
- Compression with $\| \cdot \|_\mu$ Sparsity Index

References
Elastic Net (EN) Algorithm

LASSO (Tibshirani (1996)):

$$\min_z \|y - Az\|_2^2 \text{ s.t. } \|z\|_1 \leq c.$$  

Elastic Net (EN) (Zou-Hastie (2005)):

$$\min_z \|y - Az\|_2^2 \text{ s.t. } [(1 - \mu)\|z\|_1 + \mu\|z\|_2^2] \leq c.$$  

“Reverse” Lagrangian Formulation:

$$\min_z [(1 - \mu)\|z\|_1 + \mu\|z\|_2^2] \text{ s.t. } \|y - Az\|_2 \leq \epsilon.$$  

EN contains LASSO as a special case (take $\mu = 0$). So does EN have nearly ideal behavior?

Difficulty: $(1 - \mu)\|z\|_1 + \mu\|z\|_2^2$ is not a norm!
A Modified Elastic Net (MEN) Algorithm

Modified Elastic Net (MEN) algorithm:

\[
\min \left[ (1 - \mu) \|z\|_1 + \mu \|z\|_2 \right] \text{ s.t. } \|y - Az\|_2 \leq \epsilon.
\]

Equivalently, define

\[
\|z\|_\mu := (1 - \mu) \|z\|_1 + \mu \|z\|_2,
\]

and study the problem

\[
\min_{z} \|z\|_\mu \text{ s.t. } \|y - Az\|_2 \leq \epsilon.
\]

Note that, since \(\| \cdot \|\) is strictly convex, and the feasible region is convex, this problem has a unique solution.
**Question:** Is this algorithm nearly ideal?

In other words, if \( x \in \Sigma_k \) and \( y = Ax + \eta \), and

\[
\hat{x} = \arg\min_z \|z\|_\mu \quad \text{s.t.} \quad \|y - Az\|_2 \leq \epsilon,
\]

is \( \|\hat{x} - x\|_2 \) bounded by a constant times \( \epsilon \)? **YES!**
Near Ideal Behavior of MEN Algorithm

**Theorem (MV CDC 2013):** Suppose $A \in \mathbb{R}^{m \times n}$ satisfies the RIP of order $2k$ with constant $\delta_{2k} < \sqrt{2} - 1$, and that $y = Ax + \eta$ for some $x \in \mathbb{R}^n$ and $\eta \in \mathbb{R}^m$ with $\|\eta\|_2 \leq \epsilon$. Define

$$\hat{x}_{\text{MEN}} := \arg\min_z \|z\|_\mu \text{ s.t. } \|y - Az\|_2 \leq \epsilon.$$  

Suppose in addition that

$$\frac{\mu}{\sqrt{k} (1 - \mu)} < \frac{1 - (\sqrt{2} - 1) \delta_{2k}}{1 + (\sqrt{2} + 1) \delta_{2k}}.$$  

If $\mu = 0$, this is always true, hence true for sufficiently small $\mu$, depending on $\delta_{2k}$. The smaller $\delta_{2k}$, the larger $\mu$ can be.
Near Ideal Behavior of MEN Algorithm – 2

Then there exist constants $C_{0,\mu}$ and $C_{2,\mu}$ such that

$$\|\hat{x}_{\text{MEN}} - x\|_2 \leq C_{0,\mu} \frac{\sigma_k(x, \|\cdot\|_1)}{\sqrt{k}} + C_{2,\mu}\epsilon.$$ 

Moreover when $\mu = 0$ these reduce to earlier constants.

$$C_0 = 2 \frac{1 + (\sqrt{2} - 1)\delta_{2k}}{1 - (\sqrt{2} + 1)\delta_{2k}}, \quad C_2 = \frac{4\sqrt{1 + \delta_{2k}}}{1 - (\sqrt{2} + 1)\delta_{2k}}.$$
Theorem: Suppose $A \in \mathbb{R}^{m \times n}$ satisfies the RIP of order $2k$ with constant $\delta_{2k} < \sqrt{2} - 1$, and that $y = Ax$ for some $x \in \Sigma_k$. Define

$$\hat{x} = \arg\min_z \|z\|_\mu \text{ s.t. } y = Az.$$  

Then $\hat{x} = x$ provided $\mu$ is sufficiently small.

In short, there are *infinitely many norms* $\| \cdot \|_\mu$ that permit exact recovery of sparse signals.
Advantages of MEN Algorithm

Minimizing $\| \cdot \|_1$ is a quadratic program. What are the advantages of minimizing $\| \cdot \|_\mu$?

EN has better numerical behavior than LASSO.

- LASSO uses fewer features.
- EN produces lower errors.

Does MEN outperform LASSO?

No theoretical results as yet, but on lung and ovarian cancer data, MEN combines *accuracy* of EN with *sparsity* of LASSO.
Introduction

Some Known Results

- Exact Measurements
- Noisy Measurements

Recent Advances

- A Modified Elastic Net Algorithm
- Sparse Group LASSO
- Compression with $\| \cdot \|_{\mu}$ Sparsity Index

References
Recall LASSO in primal form:

$$\min \|y - Az\|_2 \text{ s.t. } \|z\|_1 \leq \text{const.}$$

Idea: LASSO chooses very nonzero few components of $z$.

**Group LASSO:** Partition $\{1, \ldots, n\}$ into $p$ sets $I_1, \ldots, I_p$. Let $z_j$ denote projection of $z$ onto $I_j$.

$$\min \|y - Az\|_2 \text{ s.t. } \sum_{j=1}^{p} \|z_j\|_2 \leq \text{const.}$$

If every set is a singleton, then Group LASSO becomes LASSO. In general, Group LASSO tries to choose from very few *distinct subsets* $I_j$ as possible.
Sparse Group LASSO

Partition \{1, \ldots, n\} into \( p \) sets \( I_1, \ldots, I_p \). Let \( z_j \) denote projection of \( z \) onto \( I_j \).

\[
\min \| y - Az \|_2 \text{ s.t. } \sum_{j=1}^{p} [(1 - \mu) \| z \|_1 + \mu \| z_j \|_2] \leq \text{const}.
\]

Sparse Group LASSO simultaneously chooses from very few groups and also very few features from within each group.

If \( p = 1 \) (everything in one group), then Sparse Group LASSO is MEN.

So does Sparse Group LASSO have near ideal behavior? **YES!**
Near Ideal Behavior of Sparse Group LASSO

**Theorem (MV CDC 2013):** Suppose $A \in \mathbb{R}^{m \times n}$ satisfies the RIP of order $2k$ with constant $\delta_{2k} < \sqrt{2} - 1$, and that $y = Ax + \eta$ for some $x \in \mathbb{R}^{n}$ and $\eta \in \mathbb{R}^{m}$ with $\|\eta\|_2 \leq \epsilon$. Partition $\{1, \ldots, n\}$ into sets $I_1, \ldots, I_p$. Define (note “reverse” or “dual” Lagrangian formulation):

$$\hat{x}_{\text{MEN}} := \arg\min_z \sum_{j=1}^{p} [(1 - \mu)\|z\|_1 + \mu\|z_j\|_2] \text{ s.t. } \|y - Az\|_2 \leq \epsilon.$$

Suppose in addition that

$$\frac{\mu}{\sqrt{k}(1 - \mu)} 2^{(\lfloor \log p \rfloor)/2} < \frac{1 - (\sqrt{2} - 1)\delta_{2k}}{1 + (\sqrt{2} + 1)\delta_{2k}}.$$ 

M. Vidyasagar

Recent Developments in Compressed Sensing
Then there exist constants \( C_{0,\mu,p} \) and \( C_{2,\mu,p} \) such that

\[
\| \hat{x}_{\text{MEN}} - x \|_2 \leq C_{0,\mu,p} \frac{\sigma_k(x, \| \cdot \|_1)}{\sqrt{k}} + C_{2,\mu,p} \epsilon.
\]

Moreover when \( \mu = 0 \) and \( p = 1 \), these reduce to earlier constants.

\[
C_0 = 2 \frac{1 + (\sqrt{2} - 1)\delta_{2k}}{1 - (\sqrt{2} + 1)\delta_{2k}}, \quad C_2 = \frac{4\sqrt{1 + \delta_{2k}}}{1 - (\sqrt{2} + 1)\delta_{2k}}.
\]
Outline

1 Introduction

2 Some Known Results
   - Exact Measurements
   - Noisy Measurements

3 Recent Advances
   - A Modified Elastic Net Algorithm
   - Sparse Group LASSO
   - Compression with $\| \cdot \|_\mu$ Sparsity Index

4 References
Recall earlier definition:

$$\| \Delta (Ax + \eta) - x \|_2 \leq C_0 \sigma_k(x, \| \cdot \|_1) + C_2 \| \eta \|_2 ?$$

With exact measurements, earlier conclusion becomes

$$\| \Delta (Ax) - x \|_2 \leq C_0 \sigma_k(x, \| \cdot \|_1).$$

Why this mixture of $\ell_2$- and $\ell_1$-norms?
Theorem: (CDD (2009)) Suppose there exist an integer $m$, a matrix $A \in \mathbb{R}^{m \times n}$ and a function $\Delta : \mathbb{R}^m \rightarrow \mathbb{R}^n$ such that, for some constant $C_0$, we have

$$\|\Delta(Ax) - x\|_2 \leq C_0 \sigma_k(x, \|\cdot\|_2).$$

Then $m \geq C_0^{-2}n$.

No compression is possible using $\ell_2$-norm sparsity index.
Compression with $\| \cdot \|_\mu$ Sparsity Index

Roughly stated theorem (to avoid messy details) Suppose there exist an integer $m$, a matrix $A \in \mathbb{R}^{m \times n}$ and a function $\Delta : \mathbb{R}^m \rightarrow \mathbb{R}^n$ such that, for some constant $C_0$, we have

$$\| \Delta(Ax) - x \|_\mu \leq C_0 \sigma_k(x, \| \cdot \|_2).$$

Then

$$m \geq C(\mu, n)^{-2} \frac{\mu}{1 - \mu} n.$$

On the other hand, compression in $\| \cdot \|_\mu$ is possible for sufficiently small $\mu$.

These results provide a nice continuum between two extreme known results, namely: compressed sensing is possible in $\| \cdot \|_1$, and is not possible in $\| \cdot \|_2$. 
Outline

1 Introduction

2 Some Known Results
   - Exact Measurements
   - Noisy Measurements

3 Recent Advances
   - A Modified Elastic Net Algorithm
   - Sparse Group LASSO
   - Compression with $\| \cdot \|_\mu$ Sparsity Index

4 References
References


D. Donoho, “For most large underdetermined systems of linear equations, the minimal $\ell_1$-norm solution is also the sparsest solution,” *Comm. Pure and Appl. Math.*, 59(6), 797-829, 2006.
References (Cont’d)