Hierarchical Reconstruction of Sparse Signals

Ming Zhong

mzhong1@umd.edu

Advisor: Dr. Eitan Tadmor
tadmor@cscamm.umd.edu

End of Year Presentation

May 7^{th}, 2013

Outline

Outline

Example (Compressed Sensing)

Can one recover a sparse signal with the fewest possible number of linear measurements?

- $x \in \mathbf{R}^n$ is our target signal.
- A is a linear measurement matrix:
  - A is a given matrix (DCT, etc).
  - A is constructed with certain properties.
- We only know $Ax \in \mathbf{R}^m$
- In particular, x has ℓ non-zero entries, we do not know where they are, and what the values are.

Can we recover x with $m \ll n$? If so, how?

Sampling Principle

Yes for sparse x ($\ell < m \ll n$):

Compressive Sensing Principle

Sparse signal statistics can be recovered from a relatively small number of non-adaptive linear measurements.

Then how? We can find it through the following ℓ_p minimization:

Problem

Given A and b, we want to find the sparest x, such that $Ax = b$. This leads to:

$$\min_{x \in \mathbf{R}^n} \{ ||x||_{\ell_p} \mid Ax = b \} \tag{1}$$

Then what would be a suitable p?

Outline

The Constrained Minimal ℓ_p-Norm

ℓ_2, ℓ_0, and ℓ_1

Problem

$$\min_{x \in \mathbf{R}^n} \{||x||_p \mid Ax = b\} \tag{2}$$

- $p = 2$, $x = A^T(AA^T)^{-1}b$, not sparse!!
- $0 \leq p \leq 1$, it enforces sparsity.
- $p = 0$, $m = \ell + 1$, it's NP hard[1].
- $p = 1$, $m = C\ell log(n)$, it is a convex problem.[2].

But why is the ℓ_1-norm more appropriate?

---

[1] $\ell_0(\cdot)$ measures the number of non-zero entries; and proof done in B.K.Natarajan, 95

[2] D. Dohono, 04; E.J.Candes & T.Tao, 04

2-Dimensional Example

Dense Vs. Sparse

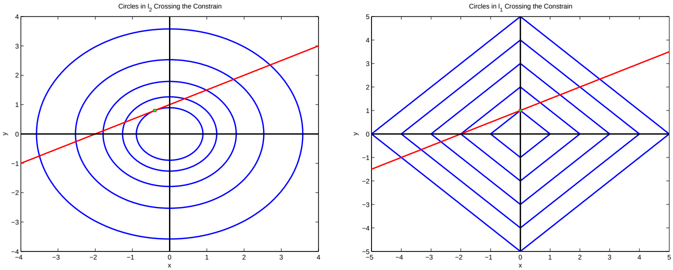


Figure: the ℓ_2 and ℓ_1 Minimizers

The ℓ_1 problem gives a sparse solution, while the ℓ_2 one does not.

Outline

Tikhonov Regularization

With the ℓ_1 problem possibly being ill-posed, we can add Tikhonov Regularization[3] to (2) ($p = 1$):

Problem (Tikhonov Regularization)

$$\min_{x \in \mathbf{R}^n} \{||x||_1 + \frac{\lambda}{2}||b - Ax||_2^2\} \tag{3}$$

- (3) becomes an unconstrained minimization.
- The minimizer depends on the regularization parameter λ (scale).
- Small λ leads to $x = \mathbf{0}$; larger λ leads to the minimizer of (2). So we need large enough λ.
- Our goal is to find a suitable range for λ.

[3] Different from Lagrange Multiplier

Tikhonov Regularizations, Cont.
An Extremal Pair

It is proven[4] that x being a solution of (3) it equivalent to then x and $r(x) = b - Ax$ satisfying the following:

Theorem (Validation Principles)

$$\langle x, A^T r(x) \rangle = ||x||_1 ||A^T r(x)||_\infty \quad (4)$$

$$||A^T r(x)||_\infty = \frac{1}{\lambda} \quad (5)$$

x and $r(x)$ are called an extremal pair. The validation principles are achieved only when λ is sufficiently large,

$$\frac{1}{||A^T b||_\infty} \leq \lambda \quad (6)$$

[4] Y. Meyer; E. Tadmor, et al, 04 and 08

The Signum Equation

The sub-gradient of (3) is:

$$T(x) = sign(x) + \lambda A^T(Ax - b) \quad (7)$$

- $0 \in T(x_{opt}) \Leftrightarrow x_{opt} = \arg\min_{x \in R^n} \{||x||_1 + \frac{\lambda}{2}||Ax - b||_2^2\}$
- $T(x)$ is a maximal monotone operator[5].
- We can split $T(x)$ by letting $T_2(x) = A^T(Ax - b)$ and $T_1(x) = \frac{1}{\lambda} sign(x)$, also making sure $I + \tau T_1$ is invertible.
- A fixed point formula: $x = (I + \tau T_1)^{-1}(I - \tau T_2)x$

[5] R. Rockafellar, Convex Analysis

Outline

Relationship between (2) and (3)

From (7), we can derive the following:

Theorem

Given that A has the Null Space Property[a]*, the minimizer* x_* *of* (3) *converges to the minimizer* x_c *of* (2).

[a] R. Gribonval, 2002

We sketch the proof as the following:

- We show that $||Ax - b||_p$ is bounded by $\mathcal{O}(\frac{1}{\lambda})$.
- Then we show that $\left| ||x_c||_1 - ||x_*||_1 \right|$ is bounded by $\mathcal{O}(\frac{1}{\lambda})$.
- Null Space Property ensures that (2) has unique minimzier.

Convergence of the Unconstrained Minimizer

We looked at the difference, $\left| ||x_c||_1 - ||x_*||_1 \right|$, and obtained the following:

λ	$\left| ||x_c||_1 - ||x_*||_1 \right|$	ratio
$2.0869e+000$	$1.5700e+002$	
$4.1738e+000$	$1.3911e+002$	$1.1286e+000$
$8.3476e+000$	$8.3440e+001$	$1.6672e+000$
$1.6695e+001$	$4.1722e+001$	$1.9999e+000$
$3.3390e+001$	$2.0861e+001$	$2.0000e+000$
$6.6781e+001$	$1.0430e+001$	$2.0000e+000$
$1.3356e+002$	$5.2152e+000$	$2.0000e+000$

Table: Convergence Rate Using GPSR Basic

Outline

Motivation

Using similar ideas from Image Processing[6], we start out by letting (x_λ, r_λ) be an extremal pair, that is:

$$b = Ax_\lambda + r_\lambda, \quad [x_\lambda, r_\lambda] = \underset{Ax+r=b}{\arg\min} \{||x||_1 + \frac{\lambda}{2}||r||_2^2\}$$

We can extract useful signal from r_λ on a refined scale, say 2λ:

$$r_\lambda = Ax_{2\lambda} + r_{2\lambda}, \quad [x_{2\lambda}, r_{2\lambda}] = \underset{Ax+r=r_\lambda}{\arg\min} \{||x||_1 + \frac{2\lambda}{2}||r||_2^2\}$$

We end up with a better two-scale approximation: $b = A(x_\lambda + x_{2\lambda}) + r_{2\lambda} \approx A(x_\lambda + x_{2\lambda})$. We can keep on extracting, ...

[6] E. Tadmor, et al, 04 and 08

Data: A and b, pick $\lambda_0(from(6))$
Initialize: $r_0 = b$, $x_{HRSS} = 0$, and $j = 0$;
while $j \leq J$ **do**

$$x_j := \arg\min_{x \in \mathbf{R}^n} \{||x||_1 + \frac{\lambda_j}{2}||r_j - Ax||_2^2\};$$
$$r_{j+1} = r_j - Ax_j;$$
$$\lambda_{j+1} = 2 * \lambda_j;$$
$$x_{HRSS} = x_{HRSS} + x_j;$$
$$j = j + 1;$$

end

Result: $x = \sum_{j=0}^{J} x_j$

- $b = Ax_{HRSS} + r_{J+1}$ and $||A^T r_{J+1}||_\infty = \frac{1}{\lambda_{J+1}} \to 0$ as $\lambda_{J+1} \to \infty$.

Some Theoretical Bounds

Using (7), we can show that:

$$||A^T A x_k||_\infty \leq \frac{3}{2\lambda_k} \tag{8}$$

Hence $Ax_k \to textNull(A)$ as $\lambda_k \to \infty$.. And we also have

$$A^T(b - Ax_{HRSS}) = \frac{1}{\lambda_J} sign(x_J) \tag{9}$$

If b is noise free, that is $b = Ax_c$, then $||A^T A(x_c - x_{HRSS})||_\infty \leq \frac{1}{\lambda_J}$. If $b = Ax_c + \epsilon$, then we to want pick a λ_J such that $\frac{1}{\lambda_J} sign(x_J) - A^T \epsilon$ is small.

Outline

Numerical Advantages

- The Hierarchical Reconstruction needs only a one scale solver (GPSRs or FPC).
- When there is no noise, we will stop the algorithm using small update and small residual.
- When there is some noise, we want to stop the algorithm when $A^T\epsilon - \frac{1}{\lambda_J} sign(x_J)$ is small.
- It has built-in de-biasing step: decreasing the residual through the unconstrained minimization and and also try to keep the ℓ_1 term small, it is better than de-biasing.

Validation Results I

Since the residual at k^{th} iterate satisfies (7), we found that it is bounded above by $\mathcal{O}(\frac{1}{\lambda})$:

$\|\|r = b - Ax_{HRSS}\|\|_2$	ratio
$5.3806e+000$	
$1.5936e+000$	$3.3763e+000$
$8.1145e-001$	$1.9639e+000$
$4.1502e-001$	$1.9552e+000$
$2.2065e-001$	$1.8809e+000$
$1.2048e-001$	$1.8314e+000$
$6.6032e-002$	$1.8246e+000$
$3.5953e-002$	$1.8366e+000$

Table: Convergence Rate of Residual with Noise Level $\sigma = 0$

Validation Results II

The convergence rate should not be affected by noise:

$\|\|r = b - Ax_{HRSS}\|\|_2$	ratio
$6.3408e+000$	
$2.4855e+000$	$2.5511e+000$
$1.3479e+000$	$1.8440e+000$
$7.0396e-001$	$1.9148e+000$
$3.6064e-001$	$1.9520e+000$
$1.8339e-001$	$1.9665e+000$
$9.2890e-002$	$1.9743e+000$
$4.6838e-002$	$1.9832e+000$

Table: Convergence Rate of Residual with Noise Level $\sigma = 0.1$

Outline

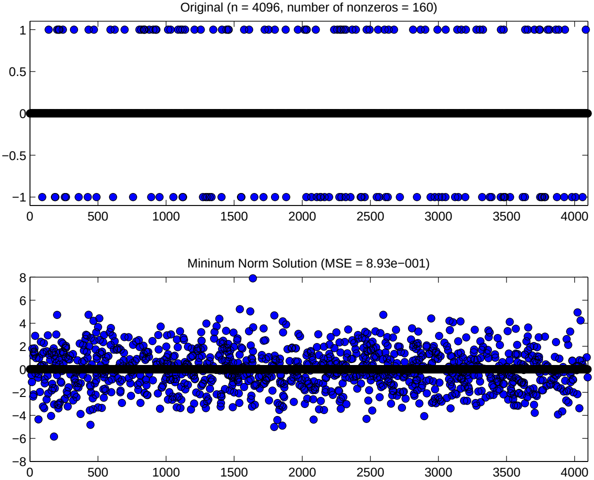
We tests the HRSS algorithm with the following case:

- $m = 1024$, $n = 4096$, and A is obtained by first filling it with independent samples of a standard Gaussian distribution and then orthonormalizing the rows.
- The original signal has only $k = 160$ non-zeros, and they are ± 1's.
- $b = Ax + \epsilon$, where ϵ is a white noise with variance $\sigma^2 = 10^{-4}$.
- The error is measured in $\text{MSE} = (\frac{1}{n})||x - x_{true}||_2^2$.

Test Results 0

The Original Signal And Minimum Norm Solution

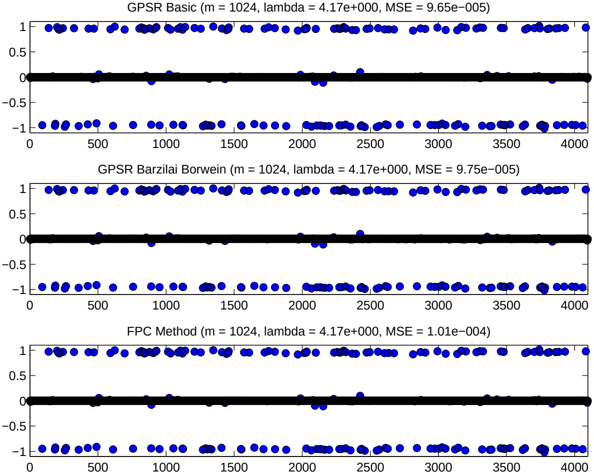
We obtain the following results for HRSS:



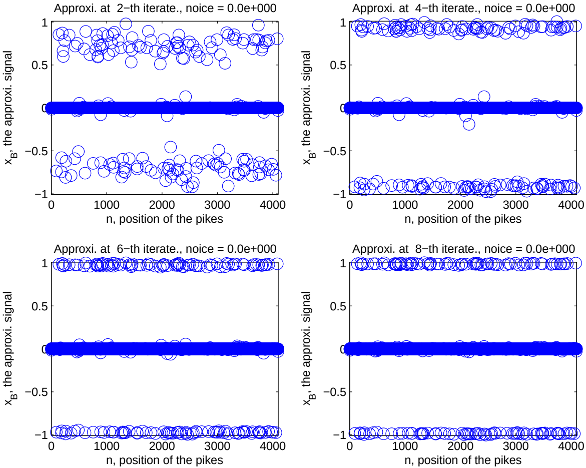
Test Results I

HRSS with 3 different solvers

And compare HRSS solutions among 3 different solvers:



Test Results II
Reconstruction Process with no noise

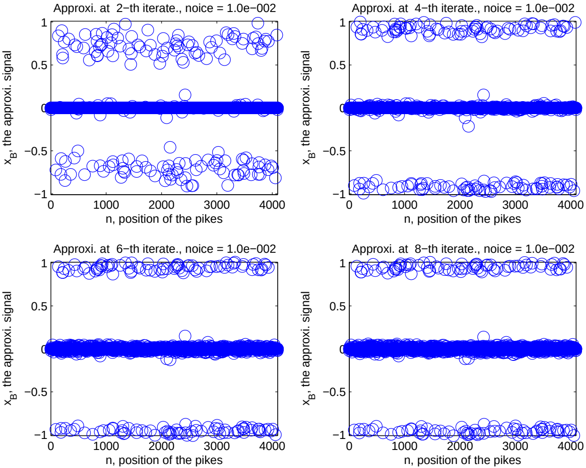


Test Results III
Reconstruction Process with some noise

HRoSS

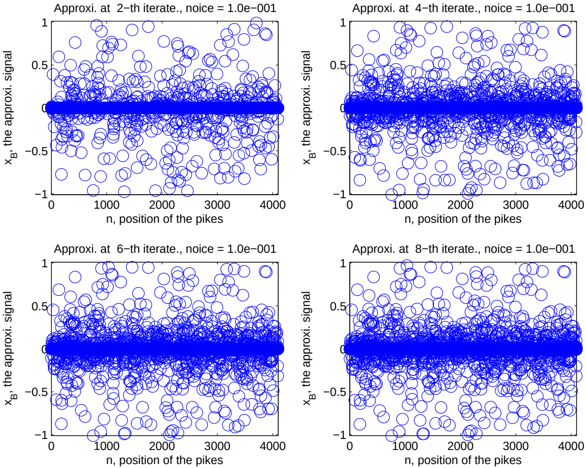
M. Zhong

Introduction and Background
Signal Processing
ℓ_p Minimizations
Single Scale Reconstruction
Approximation at A Given Scale
Theoretical Bounds
Multi-scale Construction - Hierarchical Reconstruction
Introduction
Implementation
Numerics and Summary
Test Results
Summary



Test Results IV

Reconstruction Process with a lot of noise



Outline

Milestones

- Project Background Research started on 08/29/2013.
- Presentation given on 10/02/2012 and Project Proposal written on 10/05/2012.
- Implementation of the GPSR algorithm finished and debugged on 11/05/2012, validation finished on 11/21/2012.
- Preparation for mid-year report and presentation started on 11/22/2012, FPC implementation started.

Milestones, Cont.

- Implementation of FPC done by 12/21/2012, debugged and validated by 01/22/2013.
- Implementation of HRSS finished by 02/22/2013, Near-Completion Presentation on 03/07/2013.
- Validatin of HRSS done by 03/22/2013, theoretical results obtained by 04/22/2013.
- More tests done by 04/30/2013, End-of-year Presentation on 05/07/2013.

Deliverables

- Whole Matlab Package for GPSR, FPC, and HRSS
- Test results and graphs.
- Proposal, mid-year, mid-spring, and end-of-year presentation slides.
- Complete project document.

Thank You Note

Thank you!