Hierarchical Reconstruction of Sparse Signals

Ming Zhong

mzhong1@umd.edu

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

End of Year Presentation

May 7\textsuperscript{th}, 2013
Outline

1. Introduction and Background
   Signal Processing
   $\ell_p$ Minimizations

2. Single Scale Reconstruction
   Approximation at A Given Scale
   Theoretical Bounds

3. Multi-scale Construction - Hierarchical Reconstruction
   Introduction
   Implementation

4. Numerics and Summary
   Test Results
   Summary
Introduction and Background
Signal Processing
\(\ell_p\) Minimizations

Single Scale Reconstruction
Approximation at A Given Scale
Theoretical Bounds

Multi-scale Construction - Hierarchical Reconstruction
Introduction
Implementation

Numerics and Summary
Test Results
Summary
Example (Compressed Sensing)

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

- $x \in \mathbb{R}^n$ 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 $Ax \in \mathbb{R}^m$
- In particular, $x$ has $\ell$ 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 $\ell_p$ minimization:

**Problem**

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

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

(1)

Then what would be a suitable $p$?
Outline

1. **Introduction and Background**
   - Signal Processing
   - $\ell_p$ Minimizations

2. **Single Scale Reconstruction**
   - Approximation at A Given Scale
   - Theoretical Bounds

3. **Multi-scale Construction - Hierarchical Reconstruction**
   - Introduction
   - Implementation

4. **Numerics and Summary**
   - Test Results
   - Summary
The Constrained Minimal $\ell_p$-Norm

$\ell_2$, $\ell_0$, and $\ell_1$

**Problem**

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

- $p = 2$, $x = A^T(AA^T)^{-1}b$, 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 $\ell_1$-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 $\ell_2$ and $\ell_1$ Minimizers

The $\ell_1$ problem gives a sparse solution, while the $\ell_2$ one does not.
1. Introduction and Background
   Signal Processing
   $\ell_p$ Minimizations

2. Single Scale Reconstruction
   Approximation at A Given Scale
   Theoretical Bounds

3. Multi-scale Construction - Hierarchical Reconstruction
   Introduction
   Implementation

4. Numerics and Summary
   Test Results
   Summary
With the $\ell_1$ problem possibly being ill-posed, we can add Tikhonov Regularization$^3$ to (2) ($p = 1$):

**Problem (Tikhonov Regularization)**

$$\min_{x \in \mathbb{R}^n} \left\{ ||x||_1 + \frac{\lambda}{2} ||b - Ax||_2^2 \right\}$$ (3)

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

---

$^3$Different from Lagrange Multiplier
It is proven\(^4\) that \(x\) being a solution of (3) is equivalent to then \(x\) and \(r(x) = b - Ax\) satisfying the following:

**Theorem (Validation Principles)**

\[
\langle x, A^T r(x) \rangle = \|x\|_1 \|A^T r(x)\|_\infty \tag{4}
\]

\[
\|A^T r(x)\|_\infty = \frac{1}{\lambda} \tag{5}
\]

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

\[
\frac{1}{\|A^T b\|_\infty} \leq \lambda \tag{6}
\]

\(^4\)Y. Meyer; E. Tadmor, et al, 04 and 08
The Signum Equation

The sub-gradient of (3) is:

\[ T(x) = \text{sign}(x) + \lambda A^T(Ax - b) \]  \hspace{1cm} (7)

- \( 0 \in T(x_{opt}) \iff x_{opt} = \arg \min_{x \in \mathbb{R}^n} \{ ||x||_1 + \frac{\lambda}{2} ||Ax - b||_2^2 \} \)
- \( T(x) \) is a maximal monotone operator\(^5\).
- We can split \( T(x) \) by letting \( T_2(x) = A^T(Ax - b) \) and \( T_1(x) = \frac{1}{\lambda} \text{sign}(x) \), 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

1. Introduction and Background
   Signal Processing
   $\ell_p$ Minimizations

2. Single Scale Reconstruction
   Approximation at A Given Scale
   Theoretical Bounds

3. Multi-scale Construction - Hierarchical Reconstruction
   Introduction
   Implementation

4. Numerics and Summary
   Test Results
   Summary
Relationship between (2) and (3)

From (7), we can derive the following:

**Theorem**

*Given that $A$ has the Null Space Property*, 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 $\mathcal{O} \left( \frac{1}{\lambda} \right)$.
- Then we show that $\|x_c\|_1 - \|x_*\|_1$ is bounded by $\mathcal{O} \left( \frac{1}{\lambda} \right)$.
- Null Space Property ensures that (2) has unique minimizer.
Convergence of the Unconstrained Minimizer

We looked at the difference, \( \| x_c \|_1 - \| x^* \|_1 \), and obtained the following:

<table>
<thead>
<tr>
<th>( \lambda )</th>
<th>( | x_c |_1 - | x^* |_1 )</th>
<th>ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0869e + 000</td>
<td>1.5700e + 002</td>
<td></td>
</tr>
<tr>
<td>4.1738e + 000</td>
<td>1.3911e + 002</td>
<td>1.1286e + 000</td>
</tr>
<tr>
<td>8.3476e + 000</td>
<td>8.3440e + 001</td>
<td>1.6672e + 000</td>
</tr>
<tr>
<td>1.6695e + 001</td>
<td>4.1722e + 001</td>
<td>1.9999e + 000</td>
</tr>
<tr>
<td>3.3390e + 001</td>
<td>2.0861e + 001</td>
<td>2.0000e + 000</td>
</tr>
<tr>
<td>6.6781e + 001</td>
<td>1.0430e + 001</td>
<td>2.0000e + 000</td>
</tr>
<tr>
<td>1.3356e + 002</td>
<td>5.2152e + 000</td>
<td>2.0000e + 000</td>
</tr>
</tbody>
</table>

**Table:** Convergence Rate Using GPSR Basic
Outline

1. Introduction and Background
   Signal Processing
   $\ell_p$ Minimizations

2. Single Scale Reconstruction
   Approximation at A Given Scale
   Theoretical Bounds

3. Multi-scale Construction - Hierarchical Reconstruction
   Introduction
   Implementation

4. Numerics and Summary
   Test Results
   Summary
Using similar ideas from Image Processing\textsuperscript{6}, we start out by letting \((x_{\lambda}, r_{\lambda})\) be an extremal pair, that is:

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

We can extract useful signal from \(r_{\lambda}\) on a refined scale, say \(2\lambda\):

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

We end up with a better two-scale approximation:

\[
b = A(x_{\lambda} + x_{2\lambda}) + r_{2\lambda} \approx A(x_{\lambda} + x_{2\lambda}). \text{ We can keep on extracting, . . .}
\]

\textsuperscript{6}E. Tadmor, et al, 04 and 08
Hierarchical Reconstruction
The Algorithm

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

\[ x_j := \arg \min_{x \in \mathbb{R}^n} \{||x||_1 + \frac{\lambda_j}{2}||r_j - Ax||_2^2\}; \]

\[ r_{j+1} = r_j - Ax_j; \]

\[ \lambda_{j+1} = 2 \ast \lambda_j; \]

\[ x_{HRSS} = x_{HRSS} + x_j; \]

\[ j = j + 1; \]
end

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

\[ b = Ax_{HRSS} + r_{J+1} \text{ and } ||A^T r_{J+1}||_\infty = \frac{1}{\lambda_{J+1}} \rightarrow 0 \text{ as } \lambda_{J+1} \rightarrow \infty. \]
Some Theoretical Bounds

Using (7), we can show that:

$$||A^T Ax_k||_\infty \leq \frac{3}{2\lambda_k}$$

(8)

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

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

(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 want to pick a $\lambda_J$ such that $\frac{1}{\lambda_J} \text{sign}(x_J) - A^T \epsilon$ is small.
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 \( A^T \varepsilon - \frac{1}{\lambda_j} \text{sign}(x_J) \) is small.
- It has built-in de-biasing step: decreasing the residual through the unconstrained minimization and also try to keep the \( \ell_1 \) term small, it is better than de-biasing.
Since the residual at $k^{th}$ iterate satisfies (7), we found that it is bounded above by $O\left(\frac{1}{\lambda}\right)$:

$$||r = b - Ax_{HRSS}||_2$$

| $||r = b - Ax_{HRSS}||_2$ | 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 $\sigma = 0$
The convergence rate should not be affected by noise:

\[ |r = b - Ax_{HRSS}|_2 \]

<table>
<thead>
<tr>
<th>ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5511e + 000</td>
</tr>
<tr>
<td>1.8440e + 000</td>
</tr>
<tr>
<td>1.9148e + 000</td>
</tr>
<tr>
<td>1.9520e + 000</td>
</tr>
<tr>
<td>1.9665e + 000</td>
</tr>
<tr>
<td>1.9743e + 000</td>
</tr>
<tr>
<td>1.9832e + 000</td>
</tr>
</tbody>
</table>

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

1. Introduction and Background
   Signal Processing
   $\ell_p$ Minimizations

2. Single Scale Reconstruction
   Approximation at A Given Scale
   Theoretical Bounds

3. Multi-scale Construction - Hierarchical Reconstruction
   Introduction
   Implementation

4. Numerics and Summary
   Test Results
   Summary
We test 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 \( \pm 1 \)'s.
- \( b = Ax + \epsilon \), where \( \epsilon \) is a white noise with variance \( \sigma^2 = 10^{-4} \).
- The error is measured in \( \text{MSE} = \left( \frac{1}{n} \right) \| x - x_{\text{true}} \|_2^2 \).
We obtain the following results for HRSS:

- Original signal (n = 4096, number of nonzeros = 160)
- Minimum Norm Solution (MSE = 8.93e−001)
And compare HRSS solutions among 3 different solvers:

- **GPSR Basic** (m = 1024, lambda = 4.17e+000, MSE = 9.65e−005)

- **GPSR Barzilai Borwein** (m = 1024, lambda = 4.17e+000, MSE = 9.75e−005)

- **FPC Method** (m = 1024, lambda = 4.17e+000, MSE = 1.01e−004)
Test Results II
Reconstruction Process with no noise

Approx. at 2-th iterate., noise = 0.0e+000

Approx. at 4-th iterate., noise = 0.0e+000

Approx. at 6-th iterate., noise = 0.0e+000

Approx. at 8-th iterate., noise = 0.0e+000

n, position of the pikes
$x_B$, the approx. signal
Test Results III
Reconstruction Process with some noise

Approximation at A Given Scale
Theoretical Bounds

Multi-scale Construction - Hierarchical Reconstruction
Introduction
Implementation
Numerics and Summary
Test Results
Summary

M. Zhong (UMD)
Test Results IV
Reconstruction Process with a lot of noise

Approximation at A Given Scale
Theoretical Bounds
Multi-scale Construction - Hierarchical Reconstruction
Introduction Implementation
Numerics and Summary
Test Results Summary
Outline

1. Introduction and Background
   Signal Processing
   $\ell_p$ Minimizations

2. Single Scale Reconstruction
   Approximation at A Given Scale
   Theoretical Bounds

3. Multi-scale Construction - Hierarchical Reconstruction
   Introduction
   Implementation

4. Numerics and Summary
   Test Results
   Summary
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.
- Validation 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!