Hierarchical Reconstruction of Sparse Signals

Ming Zhong

mzhong1@umd.edu

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

End-of-Semester Presentation

December 11^{th}, 2012

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

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

Sampling Principle

No for general x[1]. Yes for sparse x:

Compressive Sensing Principle

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

Then how? We can find it with this minimization problem:

Problem

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

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

but which p?

[1] $Ax = b$ is ill-posed

The Constrained Minimal ℓ_p-Norm

ℓ_0, ℓ_2, and ℓ_1

Problem

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

- $p = 0$, $m = l + 1$, it's NP hard[2].
- $p = 2$, $x = A^*(AA^*)^{-1} b$, not sparse!!
- $p = 1$, $m \simeq l * log(n)$, it leads to sparse solution![3].

But why is the ℓ_1-norm more appropriate?

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

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

2-Dimensional Example
Dense Vs. Sparse

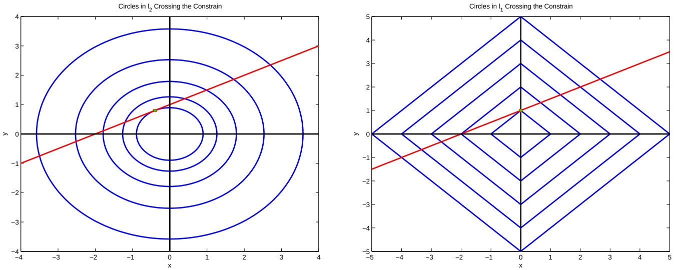


Figure: Circles in ℓ_2 and ℓ_1 Spaces

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 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 unconstrained.
- The minimizer depends on the regularization parameter λ (scale).
- Small λ leads to $x = \mathbf{0}$.
- The bigger λ is, the more emphasis is put on the least square part, and pushing for a true recovery.
- Our goal is to find a suitable range for λ's.

Tikhonov Regularizations, Cont.

An Extremal Pair

It is proven[4] that if x solves (3), then x and $r(x) = b - Ax$ satisfy the following:

Theorem (Validation Principles)

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

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

x and $r(x)$ are called an extremal pair. We will also pick an initial λ satisfying:

$$\frac{1}{||A^* b||_\infty} \leq \lambda \leq \frac{2}{||A^* b||_\infty} \quad (6)$$

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

Outline

Gradient Projection for Sparse Reconstruction

The Setup

We consider:

$$\min_{x \in \mathbf{R}^n} \{||x||_1 + \frac{\lambda}{2}||r(x)||_2^2\} \rightarrow \min_{x \in \mathbf{R}^n} \{\tau||x||_1 + \frac{1}{2}||r(x)||_2^2\} \quad (7)$$

Define $(x)_+ = \max\{0, x\}$, then we simply (7) by using:

- $u = (x)_+$, $v = (-x)_+$, and $z = \begin{bmatrix} u \\ v \end{bmatrix}$.
- $y = A^*b$, $c = \tau\mathbf{1}_{2n} + \begin{bmatrix} -y \\ y \end{bmatrix}$ and $B = \begin{bmatrix} A^*A & -A^*A \\ -A^*A & A^*A \end{bmatrix}$.

Then (7) becomes:

$$\min_{z \in \mathbf{R}^{2n}} \{F(z) \equiv c^*z + \frac{1}{2}z^*Bz \mid z \geq 0\} \quad (8)$$

Two Gradient Projection Algorithms

We will present two gradient projection type algorithms: GPSR Basic and GPSR Barzilai Borwein. They both pick the descent direction as:

$$d^{(k)} = (z^{(k)} - \alpha^{(k)}\nabla F(z^{(k)}))_+ - z^{(k)} \tag{9}$$

and update the next iterative $z^{(k+1)}$ as:

$$z^{(k+1)} = z^{(k)} + \nu^{(k)} d^{(k)} \tag{10}$$

They differ by choosing different $\alpha^{(k)}$ and $\nu^{(k)}$.

Algorithm Description
GPSR Basic

Data: A, b, τ, pick $\alpha_{\min}$, $\alpha_{\max}$, $z^{(0)}$, $\beta \in (0,1)$, and $\mu \in (0, 1/2)$;

Initialize: $k = 0$;

while *A convergence test is not satisfied* **do**

Compute $\alpha_0 = arg \min_{\alpha} F(z^{(k)} - \alpha g^{(k)})$;

Let $\alpha^{(k)} \in [\alpha_{\min}, \alpha_{\max}]$ be the first in $\alpha_0, \beta\alpha_0, \beta^2\alpha_0, \ldots,$ such that $F((z^{(k)} - \alpha^{(k)}\nabla F(z^{(k)}))_+) \leq F(z^{(k)}) + \mu \nabla F(z^{(k)})^*((z^{(k)} - \alpha^{(k)}\nabla F(z^{(k)}))_+ - z^{(k)})$;

Set $z^{(k+1)} = (z^{(k)} - \alpha^{(k)}\nabla F(z^{(k)}))_+$;

$k = k + 1$;

end

Data: A, b, τ, pick $\alpha_{\min}$, $\alpha_{\max}$, $z^{(0)}$, and $\alpha^{(0)} \in [\alpha_{\min}, \alpha_{\max}]$;
Initialize: $k = 0$;
while *A convergence test is not satisfied* **do**

  Compute $d^{(k)} = (z^{(k)} - \alpha^{(k)} \nabla F(z^{(k)}))_+ - z^{(k)}$;
  Compute $\omega^{(k)} = (d^{(k)})^* B d^{(k)}$;
  if $\omega^{(k)} = 0$ **then**
    Set $\alpha^{(k+1)} = \alpha_{\max}$ and $\nu^{(k)} = 1$;
  else
    Compute $\alpha^{(k+1)} = \text{mid}\left\{\alpha_{\min}, \frac{\|d^{(k)}\|_2^2}{\omega^{(k)}}, \alpha_{\max}\right\}$ and
    $\nu^{(k)} = \text{mid}\left\{0, -\frac{(d^{(k)})^* \nabla F(z^{(k)})}{(d^{(k)})^* B d^{(k)}}, 1\right\}$;
  end
  Set $z^{(k+1)} = z^{(k)} + \nu^{(k)} d^{(k)}$;
  $k = k + 1$;

end

Stopping Criterion

We have picked the following stopping criterion for the codes:

- $||z - (z - \bar{\alpha}\nabla F(z))_+||_2 \leq tolP$; if z is optimal then the left hand side is 0.
- Perturbation Results from Linear Complementarity Problems (CLP): $dist(z, \mathcal{S}) \leq C_{CLP}|| \min(z, \nabla F(z))||_2$ for some constant C_{CLP} and $\mathcal{S}$ is the solution set of (8). So we take: $|| \min(z, \nabla F(z))||_2 \leq tolP$, where $\min(\cdot)$ is taken component wise.

Stopping Criterion, Cont.

HRoSS

M. Zhong

The dual[5] problem of (7) is:

$$\max_s \{-\frac{1}{2}s^*s - b^*s \,|\, -\tau\mathbf{1}_n \leq A^*s \leq \tau\mathbf{1}_n\} \tag{11}$$

If s is a solution for (11) and x is a solution for (7), then $\frac{1}{2}||r(x)||_2^2 + \tau||x||_1 + \frac{1}{2}s^*s + b^*s = 0$. Hence we have the following stopping criteria:

- $\frac{1}{2}||r(x)||_2^2 + \tau||x||_1 + \frac{1}{2}s^*s + b^*s \leq tolP$.

Where $s = \tau\frac{-r(x)}{||A^*(-r(x))||_\infty}$.

[5] S. Kim, et al, 07

Stopping Criterion, Cont.

We want to track how the non-zero indices of z are changing in recent iteration. Let us define:

$$\begin{aligned} \mathcal{I}_k &= \{i \big| z_i^{(k)} \neq 0\} \\ \mathcal{C}_k &= \{i \big| i \in \mathcal{I}_k \oplus \mathcal{I}_{k-1}\} \end{aligned}$$

And stop the iteration if:

- $\frac{|\mathcal{C}_k|}{|\mathcal{I}|_k} \leq tolP$.

De-Biasing for GPSR

After the approximation $z = [u^*, v^*]^*$ is computed, we convert it to $x_{GPSR} = u - v$. Then we might choose to do a debiasing step:

- The zero entries of x_{GPSR} is kept zero.
- The least square objective $||r(x)||_2^2$ is minimized using a (restricted) Conjugate Gradient algorithm.
- Terminates when $||r(x)||_2^2 \leq \text{tolD} * ||r(x_{GPSR})||_2^2$
- We are fine-tuning the non-zero entries of x_{GPSR}, it might not give a satisfactory result.

- The convergence of the algorithm is proved.[6]
- $F(z) = c^* z + \frac{1}{2} z^* B z$, and $\nabla F(z) = c + Bz$.
- $c^* z = \tau \mathbf{1}_n^* (u + v) - y^* (u - v)$ and $z^* B z = ||A(u - v)||_2$.
- $c = \begin{bmatrix} \tau \mathbf{1}_n - y \\ \tau \mathbf{1}_n + y \end{bmatrix}$ and $Bz = \begin{bmatrix} A^* A(u - v) \\ -A^* A(u - v) \end{bmatrix}$.
- We do $A(u - v)$ first, then $A^*(A(u - v))$.
- Ax and $A^* x$ can be defined as function calls instead of direct matrix-vector multiplication.

[6] M.A.T. Figueiredo, et al, 2007

Outline

Test Setup

We validated tested the code by the following setting:

- $m = 1024$, $n = 4096$, and A is obtained by first filling it with independent samples of a standard Gaussian distribution and then orthonormalizing the rows.
- $\beta = 0.5$, $\mu = 0.1$, $\tau = 0.1||A^*b||_\infty$, $b = Ax + \epsilon$, where ϵ is a white noise with variance $\sigma^2 = 10^{-4}$.
- The original signal has only $k = 160$ non-zeros, and they are ± 1's.
- The error is measured in $\text{MSE} = (\frac{1}{n})||x - x_{true}||_2^2$.

Validation Results

In order to validate the codes, we hope to show that the approximation which we find satisfies (4) and (5), so we define the following quantities:

$$\begin{aligned} \text{diff}_1 &= \langle x, A^* r(x) \rangle - ||x||_1 ||A^* r(x)||_\infty \\ \text{diff}_2 &= ||A^* r(x)||_\infty - \frac{1}{\lambda} \\ J(x) &= \tau ||x||_1 + \frac{1}{2} ||r(x)||_2^2 \end{aligned}$$

We want to show that the diff_1 and diff_2 go down with the preset tolerance.

Validation Results, Cont.

We obtain the following results for GPSR:

tol	diff_1	diff_2	Num. of Iter.	$J(x)$
10^{-4}	$-2.4644e-003$	$2.0224e-005$	27	6.7937
10^{-5}	$-2.2962e-004$	$1.8994e-006$	33	6.793661
10^{-6}	$-2.1692e-005$	$1.8028e-007$	39	6.7937
10^{-7}	$-2.0270e-006$	$1.6894e-008$	45	6.7937

Table: Result with $\sigma = 0$ for Basic

tol	diff_1	diff_2	Num. of Iter.	$J(x)$
10^{-4}	$-2.3277e-003$	$1.9723e-005$	31	6.7937
10^{-5}	$-2.0638e-004$	$1.8571e-006$	37	6.793661
10^{-6}	$-2.5418e-005$	$2.1726e-007$	43	6.7937
10^{-7}	$-2.2379e-006$	$2.0232e-008$	49	6.7937

Table: Result with $\sigma = 0$ for Barzilai Borwein

Validation Results, Cont.

The original signal has: $J(x) = 1.4910e + 001$:

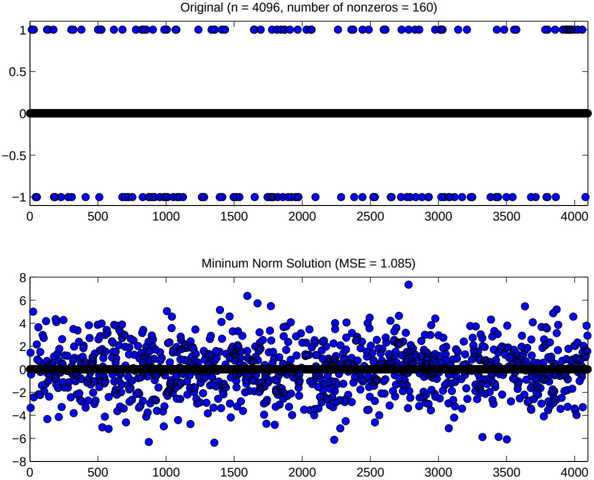
tol	$diff_1$	$diff_2$	Num. of Iter.	$J(x)$
10^{-4}	$-2.5181e-003$	$1.3017e-005$	65	$1.1393e+001$
10^{-5}	$-2.4369e-004$	$1.2570e-006$	93	$1.1393e+001$
10^{-6}	$-2.6812e-005$	$1.3817e-007$	121	$1.1393e+001$
10^{-7}	$-2.6666e-006$	$1.3707e-008$	151	$1.1393e+001$

Table: Result with $\sigma = 10^{-1}$ for Basic

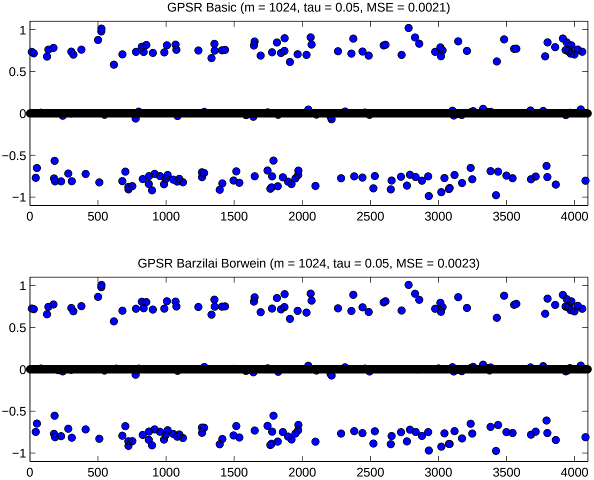
tol	$diff_1$	$diff_2$	Num. of Iter.	$J(x)$
10^{-4}	$-1.6146e-003$	$1.3495e-005$	92	$1.1393e+001$
10^{-5}	$-1.3726e-004$	$1.1382e-006$	136	$1.1393e+001$
10^{-6}	$-1.5475e-005$	$1.2528e-007$	180	$1.1393e+001$
10^{-7}	$-1.7440e-006$	$1.3996e-008$	224	$1.1393e+001$

Table: Result with $\sigma = 10^{-1}$ for Barzilai Borwein

We obtain the following results for GPSR:



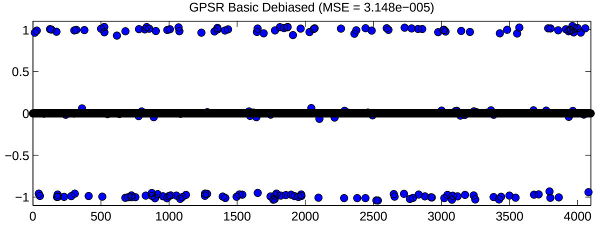
We obtain the following results for GPSR:

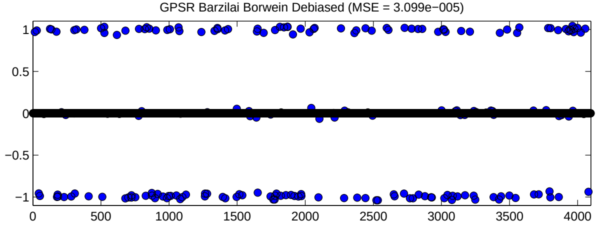


Test Results

Basic Vs. Barzilai Borwein (Debiased)

We obtain the following results for GPSR:





Outline

Let (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: $b = A(x_\lambda + x_{2\lambda}) + r_{2\lambda} \approx A(x_\lambda + x_{2\lambda})$. And we can keep on extracting, . . .

Hierarchical Reconstruction
The Algorithm

HRoSS
M. Zhong
Single Scale Reconstruction
Introduction
Approximation on a given Scale
Implementation
Validations and Tests
Hierarchical Reconstruction
Introduction
Summary

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

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

end

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

- $b = Ax + r_J$ and $||A^* r_J||_\infty = \frac{1}{\lambda_J} \to 0$ as $\lambda_J \to \infty$.

Implementation

- The Hierarchical Reconstruction needs a one scale solver.
- GPRS and FPC will be the built-in solver inside the j-loop.
- Validation will be done at each j iterate.
- Convergence test will be tried.

Outline

Milestones

- 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.
- Finish implementation of FPC by 12/21/2012, debug and validate FPC codes by 01/22/2013.

Deliverables

- Whole Matlab Package for GPSR.
- Test results and graphs.
- Mid year presentation slides.
- Mid year project document.

Thank You Note

Thank you!