Constrained L1-norm Optimization for Sparse Signal Reconstruction

Computer Cognitive and Security Lab
Research Seminar
Jan.22 2010
Chuan Li
ECE Department
The University of Alabama
Optimization Problem

\[ b = Au \]

\( b \): observed value
\( A \): Linear Transform
\( u \): unknown vector

When \( u \) is sufficiently sparse and \( b \) is long enough, exact reconstruction is possible (both in theory and in practical
Different Optimization Methods

L2-Regression suppresses over-fitting
L2-Regression does not add too much complexity
to existing problems -> easy to calculate  not accurate

L0-Calculation is NP hard
L1-Regression creates sparse answers, and
better approximations in relevant cases
L1-Regression problems are not differentiable
need other ways of solving problem (using
convex optimization techniques, iterative
approaches, etc.)
Usage:

• Signal Processing
  Basic Pursuit
  Compressed Sensing
  Signal Recovery
  Wavelet Thresholding

• Statistics
  Lasso Algorithm
  Fused Lasso

• Others
  Decoding Linear Codes
  Geophysics Problems
  Maximum Likelihood Estimation
Different Optimization Methods

- Unconstrained
  \[
  \min_u \mu \|u\|_1 + \frac{1}{2} \|Au - b\|_2^2
  \]

- Basis pursuit
  \[
  \min_u \|u\|_1, \text{ s.t. } \|Au - b\|_2 \leq \delta
  \]

- LASSO
  \[
  \min_u \|Au - b\|_2, \text{ s.t. } \|u\|_1 \leq \gamma
  \]

- Constrained
  \[
  \min_u \|u\|_1, \text{ s.t. } Au = b
  \]
When is L1-reconstruction successful

Def:
• n=dim(u) signal size
• k=\|u\|_0 signal sparsity
• m=Number of row(A) sample size

• When A is subgaussian (e.g. Gaussian, Bernoulli, etc.) (Mendelson et al)
  • m \approx O(k \log (n/k))

• When A is Fourier submatrix (Candes, Romberg and Tao, 2004)
  • m \approx O(k \log (n))

• m can be even less using \|u\|_p for p<1 minimization (Candes-Wakin-Boyd, Chartrand, and Chartrand-Y)
When is $u$ is not sparse?

- $u$ is not sparse in most cases, but it needs to be sparse in some sense
- Sparse under basis (e.g., Fourier, wavelet, contourlet)
- Sparse under linear transforms (overcomplete basis, not invertible)
- Sparse under nonlinear transforms (e.g., total variation)
  \[ TV(u) = \sum_{i,j} \| (\nabla u)_{i,j} \|_2 \]
- Has a low rank
- Has a low dimension embedding
Recent Algorithms?

• Path-following: LARS, etc.
  – Start from an easy problem
  – Gradually transform the easy problem to the original one
  – Solution path is piece-wise linear
  – Solve one (small) system of linear equations for each breakpoint
• Specialized interior-point method: l1_lS
  – More accurate than first-order methods
  – Use truncated Newton’s method and preconditioned conjugate gradient
• Operator splitting:
  – Cheap per-iteration cost, more iterations
  – Accelerated by line search and continuation
  – Obtain optimal support quickly
• Gradient projection: GPSR, similar to operator splitting
• Bregman method:
  – Originally for the constrained problem, finite convergence
L1-Norm Optimization Compare:

<table>
<thead>
<tr>
<th>Optimization Method</th>
<th>Approx Objective</th>
<th>Sub-Gradient</th>
<th>Explicit Constraints</th>
<th>Convergence Ranking</th>
<th>Iteration Speed Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gauss-Seidel [16]</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Shooting [15]</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Grafting [6]</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Sub-Gradient</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>Log(norm(x))</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>EM [4]</td>
<td>Y*</td>
<td>Y***</td>
<td>N</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Log-Barrier [14]</td>
<td>Y*</td>
<td>N</td>
<td>Y</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>SmoothL1 [ThisPaper]</td>
<td>Y*</td>
<td>N</td>
<td>N</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>SQP [11]</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>ProjectionL1 [ThisPaper]</td>
<td>N</td>
<td>Y***</td>
<td>Y</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Interior Point [5]</td>
<td>Y**</td>
<td>N</td>
<td>Y</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

*: method improves the approximation between iterations.

**: method uses a constrained objective that is improved between iterations.

***: methods use the correct gradient, but only for the working set (other sub-gradient methods also restrict to the working set).
Reconstruction Example

Least Square Reconstruction

Simplex Method
## Experimental Data

<table>
<thead>
<tr>
<th></th>
<th><img src="image1.png" alt="Image" /></th>
<th><img src="image2.png" alt="Image" /></th>
<th><img src="image3.png" alt="Image" /></th>
<th><img src="image4.png" alt="Image" /></th>
<th><img src="image5.png" alt="Image" /></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>6 subjects</strong></td>
<td>water drinking starts</td>
<td>water drinking ends</td>
<td>neural activities start</td>
<td>neural activities over</td>
<td>after all activities</td>
</tr>
<tr>
<td><strong>6 subjects</strong></td>
<td>glucose intake starts</td>
<td>glucose intake ends</td>
<td>neural activities start</td>
<td>neural activities over</td>
<td>after all activities</td>
</tr>
</tbody>
</table>

- **Sampling rate:** 4 frames/min
- **Water drinking time:** 54”, 45”, 1’40” 1’ 57” 1’36”
- **Glucose intaking time:** 1’28”, 1’, 1’, 50”, 1’2”, 1’10”
Proposed Solutions: Constrained L1 Optimization I

The impulse response model: \[ h(t; \tau_1, \tau_2, \delta_1, \delta_2, c) = \left( \frac{t}{\tau_1} \right) \delta_1 e^{-\left( \delta_1 / \tau_1 \right)(t-\tau_1)} - c \left( \frac{t}{\tau_2} \right) \delta_2 e^{-\left( \delta_2 / \tau_2 \right)(t-\tau_2)} \]

\( \tau_1, \delta_1, c \) determine the shape of the IR model.
Proposed Solutions: Optimization Approach
Constrained L1 Optimization II

Multi-layer neural activity model:
S: fMRI time course data
x: sparse signal sources at different layer
A: hemodynamic model.

\[
S = Ax = \begin{bmatrix} A_1 & A_1 & A_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}
\]

\[
A_i = \begin{bmatrix}
1 & 0 & \ldots & 0 \\
1 & 1 & \ldots & 0 \\
\ldots & \ldots & \ldots & \ldots \\
1 & 1 & \ldots & 1 \\
\end{bmatrix}
\begin{bmatrix}
h_0^i \\
h_1^i \\
h_0^i \\
h_n^i \\
h_{n-1}^i \\
h_0^i \\
\end{bmatrix}
\]

Goal: \( \min_x \|x\|_1 \) subject to \( S = Ax \)
Constrained L1 Optimization I
Model Detection

- In each E-step, the ternary neural activity signals are estimated using the given impulse response model.
- In each M-step, the impulse response model parameters are optimized to maximize the correlation between the predicted and measured fMRI data.
Part II: Constrain L1 Optimization
Signal Reconstruction
Future work II: ICA Enhanced L1 Optimization

- ICA generates prior for L1 optimization, hence improve the robustness of the L1

- Region based ICA incorporates the spatial correlation of the fMRI
Thank You So Much

Questions or Comments?