Modeling and Processing of High Dimensional Signals and Systems Using the Sparse Matrix Transform

Guangzhi Cao
School of Electrical and Computer Engineering
Purdue University

Ph.D. Final Examination
October 5, 2009

*Supported by the National Science Foundation
Outline

- Non-iterative MAP tomographic reconstruction
  - Localization versus reconstruction
  - Framework of non-iterative MAP
  - Matrix source coding theory
  - Experimental results on medical imaging

- Covariance estimation for high dimensional signals
  - SMT framework
  - SMT for covariance estimation
  - Experimental results of SMT for modeling hyper-spectral data and face images

- Regression of high dimensional data
Part I. Iterative Statistical Reconstruction in Tomography

- Image $x$
- Model: $p(x)$
- Forward model $f$
- Physical system
- Difference

$x$ – Unknown image
$y$ – Known surface measurement

- Objective: reconstruct unknown $x$ from known $y$
Model-based Iterative Reconstruction

- MAP estimate of $x$:

$$\hat{x} = \arg \max_{x \geq 0} \left\{ \log p(y \mid x) + \log p(x) \right\}$$

$$= \arg \max_{x \geq 0} \left\{ \| y - f(x) \|_{\Lambda}^2 + \log p(x) \right\}$$

$\Lambda$ – Noise model; $p(x)$ – prior image distribution

- Optimization:
  - Preconditioned conjugate gradient
  - Iterative Coordinate Descent
  - Multigrid
  - All are computationally expensive!
Introduction to Optical Tomography

- Optical tomography: Deep tissue imaging with light

  - Optical Diffusion Tomography (ODT or DOT)
  - Fluorescence Optical Diffusion Tomography (FODT)
  - Bioluminescence Tomography (BLT)
Forward Model: Diffusion Equation

- Transport of light in scattering media:

\[ \nabla \cdot [D(r)\nabla \phi(r, \omega)] - [\mu_a(r) + j\omega/c] \phi(r, \omega) = -S(r, \omega) \]

- Parameters:
  \[ \begin{align*}
  \phi & \rightarrow \text{photon flux density (W/cm)} \\
  D & \rightarrow \text{diffusion coefficient (cm)} \\
  \mu_a & \rightarrow \text{absorption coefficient (cm}^{-1}) \\
  c & \rightarrow \text{speed of light in medium (cm/sec)}
  \end{align*} \]

- System description:

\[ y = f(x) \]

\[ \begin{align*}
  x & \rightarrow \text{Unknown images of optical parameters, e.g. } (\mu_a, D) \\
  y & \rightarrow \text{Known surface measurement, e.g. } \phi
  \end{align*} \]
Approach 1: Fast Maximum Likelihood Localization

- Maximum likelihood (ML) estimate of inhomogeneity location:
  \[ \hat{r} = \arg \min_{r \in \Omega} \| y - f(x_r) \|_2^2 \]

- Parameterization of inhomogeneity? Size, contrast \((\mu_a - \mu_{a0})\)...

- Inhomogeneity can be effectively modeled as a point: \(\delta u(r)\)
Localization versus Reconstruction

Reflectance measurement geometry

Optical parameters

Bulk: $\mu_{a0} = 0.02 \text{ cm}^{-1}$
$D_0 = 0.03 \text{ cm}$

Inhomogeneity: $d = 1.5 \text{ cm}$
(Diameter = 6.25 mm) $\mu_a = 0.12 \text{ cm}^{-1}$

Average SNR ~ 40 dB

Localization – negative log likelihood

Reconstruction – $\hat{\mu}_a$
Probability of Detection Under Various Conditions

- **SNR=40dB, size=6.25mm, \( \Delta \mu_a = 0.1 \text{cm}^{-1} \)**
- **Absorption & depth, size = 6.25mm**

- **Size & depth, \( \Delta \mu_a = 0.1 \text{ cm}^{-1} \)**
- **Absorption & size, depth = 1.5 cm**
Approach 2: Non-Iterative MAP Reconstruction

- MAP reconstruction: \[ \hat{x} = \arg\max_x \left\{ \|y - f(x)\|_A^2 + \log p(x) \right\} \]

- Assume a linear forward model and Gaussian prior:
  \[ y = f(x) = Ax \]
  \[ \log p(x) = x^T S x \]

- Then the MAP estimate has the closed form:
  \[ \hat{x} = \left( A^T \Lambda A + S \right)^{-1} A^T \Lambda y \]
  \[ = Hy \]

- Can we just store \( H \) and directly compute \( \hat{x} \)?
  - But \( H \) is not sparse!
Our Approach to Non-Iterative MAP Reconstruction

- Use lossy source coding strategies to store $H$
  - Makes $H$ sparse
  - Less storage
  - Less computation
- Compute exact $H$ off-line using iterative methods
  \[
  h_i = \arg \max_x \{ \| e_i - f(x) \| + \log p(x) \}
  \]
- Then directly compute
  \[
  \hat{x} = [H]y
  \]
  
  $[H]$ represents quantized version of $H$.
- Questions: How do we code $H$?
Distortion Metric for $H$

- Reconstruction error due to quantization of $H$:

$$\hat{x} + \delta \hat{x} = (H + \delta H)y$$

- Theorem:

$$E \left[ \| \delta \hat{x} \|^2 \mid \delta H \right] = \text{trace} \left\{ \delta H \ R_y \ \delta H^T \right\} = \| \delta H \|^2_{R_y}$$

where $R_y = E[yy^T]$.

- Intuition: Whiten the measurement $y$ using eigen-decomposition
Three Orthonormal Transforms Used to Code $H$

- Step 1. Whiten $y$ using eigen-decomposition

- Step 2. Apply KL transform to decorrelate columns of $H$

- Step 3. Apply wavelet transform to decorrelate rows of $H$

Conclusion:

$$\hat{x} = W^T \tilde{H} \tilde{y}$$
Reconstruction from Compressed Transform

- Wavelet Transform
- Decorrelation Transform
- Whitening Transform

\[ \hat{x} = W^T [\tilde{H}] \tilde{y} \]

- Quantization: • Reduces data storage
  • Reduces computation for reconstruction

- Can use a variety of encoding methods for lossy coding of matrix columns
Numerical Example: ODT for Female Breast Imaging

- **Measurement System**
  - Bulk optical parameters:
    \[
    \mu_a = 0.02 \text{ cm}^{-1} \text{ and } D = 0.03 \text{ cm}
    \]
  - Heterogeneity optical parameters (Radius: 1 cm, depth: 3 cm):
    \[
    \mu_a = 0.12 \text{ cm}^{-1} \text{ and } D = 0.03 \text{ cm}
    \]
  - Modulation frequency: 70 MHz
  - Noise: shot noise with an average SNR of 35.8 dB
Reconstruction Results for Absorption Distribution

Original Image

Uncompressed

0.015 bpme, NRMSE=16.98%

0.032 bpme, NRMSE=10.52%
Distortion versus Rate

- Distortion Defined as: \[ NRMSE = \sqrt{\frac{\| [H]y - Hy \|^2}{\| Hy \|^2}}. \]
- Whitening is more important than decorrelation!
Computation Complexity

- The size of the computational domain is 16 x 16 x 6 cm³.
- \( M=720 \) – number of measurements, \( N=65\times65\times33 \) – number of pixels, \( I=30 \) - number of iterations, \( c = 1808:1 \) - compression ratio (NRMSE≈10%)

<table>
<thead>
<tr>
<th></th>
<th>Computation</th>
<th>Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Order</td>
<td>Seconds</td>
</tr>
<tr>
<td>Conjugate Grad.</td>
<td>MNI</td>
<td>126</td>
</tr>
<tr>
<td>Uncompressed</td>
<td>NM</td>
<td>0.89</td>
</tr>
<tr>
<td>Compressed with KLT</td>
<td>( NM/c+N+M^2 )</td>
<td>0.03</td>
</tr>
</tbody>
</table>
Problem: Storage and Computation of Orthonormal Transforms

- Reconstruction of $x$: 
  $$\hat{x} = W^T[H]\tilde{y}$$

  But 
  $$\tilde{y} = Ty$$

- $T$ is a dense matrix consisting of two orthonormal transforms
  - $T$ must be stored - $(\# \text{ of measurements})^2$
  - $Ty$ must be computed - $(\# \text{ of measurements})^2$

- As # of measurements becomes large $\implies$ Problem!

- More general problem: how to estimate $T$?!
Part II: Covariance Estimation for High Dimensional Signals

- **Setting:** Let $Y = [y_1, y_2, \cdots, y_n]$
  
  where $y_i \sim N(0, R)$ is a $p$-dimensional random vector.

- **Objective:** Estimate the eigenvalues and eigenvectors of $R$
  
  $$R = E \Lambda E^t$$

- **Challenge:** This a classically difficult problem when $n < p$
  
  - Curse of dimensionality

- **Application:** Widely used for
  
  - PCA analysis, Eigenimage and eigensignal analysis
  - Machine learning, pattern recognition…
Data Model

- Notation:
  \[ Y = \begin{bmatrix} y_1, y_2, \cdots, y_n \end{bmatrix} \quad S \triangleq \frac{1}{n} YY^t \quad R = E \mathbb{E} \begin{bmatrix} S \end{bmatrix} \]
  
  Observed Data  
  Sample Covariance  
  True Covariance

- Likelihood of \( Y \) given \( R \):
  \[
p_R(Y) = \frac{1}{(2\pi)^{np/2}} |R|^{-n/2} \exp \left\{ -\frac{1}{2} tr\{Y^t R^{-1}Y\} \right\}, \text{ with } R = E \Lambda E'.
  \]

- ML estimate of eigenvectors and eigenvalues is given by
  \[
  \begin{cases}
  \hat{E} = \text{arg min} \left\{ \| \text{diag}(E^t SE) \| \right\} \\
  \hat{\Lambda} = \text{diag}(\hat{E}^t \hat{S} \hat{E})
  \end{cases}
  \]
Sparse Matrix Transform (SMT)

- SMT is product of Givens rotations:

\[ E = E_1 E_2 \cdots E_k \text{ where } E_k = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \]

- So the SMT is a generalization of the FFT

- SMT is also a generalization of orthonormal wavelet transforms
Proposed Solution: Model-based estimation

- Big idea:
  (Eigenvector matrix $E$) $\in \Omega$ (Sparse Matrix Transform)

- The constrained ML estimation of eigenvectors and eigenvalues:

$$
\hat{E} = \arg \min_{E \in \Omega_K} \left\{ \text{diag}(E^tSE) \right\}
$$

$$
\hat{\Lambda} = \text{diag}(\hat{E}^tS\hat{E})
$$

- Each Givens rotation operates on only two coordinates
  - Only 4 multiplies per rotation (actually, only 2)
  - When $K=p(p-1)/2$, this can be any $p$-dimensional orthonormal transform
Design of SMT

- Design of SMT is formulated as a cost optimization problem:

\[
\hat{E} = \arg \min_{E=E_1E_2\cdots E_K} \left\{ \left\| \text{diag}(E^tSE) \right\| \right\}
\]

- A greedy algorithm is used for optimization

- The algorithm:

```plaintext
For k = 1 to K {
    Select most correlated coordinate pair
    Decorrelate the coordinate pair with rotation \( E_k \)
}
E \leftarrow E_1 E_2 \cdots E_K
```

- Comments: 1) \( K \) is chosen to maximize cross-validated likelihood
  2) Does not depend on ordering of vector or stationary assumption
The number of SMT butterflies used is $p \log_2(p)$

Comment: SMT both reduces the data storage and computation required to compute $T_y$
Covariance Estimation for Hyperspectral Data: Grass Class

- # of hyperspectral bands: \( p = 191 \), # of samples: \( n = 80 \)
Estimators Compared

- Shrinkage estimator:
  \[
  \hat{R} = (1 - \alpha) \cdot S + \alpha \cdot \text{diag}(S)
  \]

- Graphic lasso (glasso) covariance estimator:
  \[
  \hat{R} = \arg\max \left\{ \log(Y|R) - \alpha \|R^{-1}\|_1 \right\}
  \]

- SMT estimator

- SMT-shrinkage (SMT-S) estimator:
  \[
  \hat{R} = (1 - \alpha) \cdot S + \alpha \cdot \hat{R}_{SMT}
  \]

- Comment: \(\alpha\) is estimated using cross-validation
Eigenvalue Estimation: Gaussian Case

\[ \hat{\Lambda} = \text{diag}(\hat{E}' S \hat{E}) \]
Results in Kullback-Leibler Distance: Gaussian Case
## Computational Complexity

<table>
<thead>
<tr>
<th></th>
<th>Complexity (without c.v.)</th>
<th>CPU time (seconds)</th>
<th>Model order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shrinkage</td>
<td>$p$</td>
<td>8.6 (with c.v.)</td>
<td>1</td>
</tr>
<tr>
<td>glasso</td>
<td>$p^3I$</td>
<td>422.6 (without c.v.)</td>
<td>4939</td>
</tr>
<tr>
<td>SMT</td>
<td>$p^2 + Kp$</td>
<td>6.5 (with c.v.)</td>
<td>495</td>
</tr>
<tr>
<td>SMT-S</td>
<td>$p^2 + Kp$</td>
<td>7.2 (with c.v.)</td>
<td>496</td>
</tr>
</tbody>
</table>

- $I$ - cycles used in glasso, $K$ - number of Givens rotations in SMT
- Numerical results are based on the grass Guassian case with $n = 80$
- c.v.: cross-validation
Computing Eigen-Images Using SMT

**SMT Model Training**

\[ Y = \text{apply Givens rotations} \]

- \( n \) training sample images
- \( p \) pixels

**SMT Eigenimage Analysis**

- Input Image
- All \( p \) eigenvalues
- SMT model
- SMT model inverse
- Impulse at position \( k \)
- \( k^{th} \) eigenimage

\[ \text{p eigenvalues} \]
\[ \text{p eigenfaces} \]
SMT versus Traditional Eigen-faces

Face dataset:

PCA + Shrinkage:

SMT:

SMT-S:
Comparison of Traditional and SMT Eigenimages

- Eigenimage experiment
  - Dataset: face image
  - Number of samples \((n) = 80\)
  - Dimensions \((p) = 644\)
- Use cross-validation to compute expected log likelihood
  - \(\Delta \text{loglikelihood}=167.3\)

<table>
<thead>
<tr>
<th>Method</th>
<th>Maximum log-likelihood</th>
<th>(\Delta)</th>
<th>(K_{\text{max}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA+Shrinkage</td>
<td>-2863.6</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>glasso</td>
<td>-2699.1</td>
<td>164.5</td>
<td>-</td>
</tr>
<tr>
<td>SMT</td>
<td>-2797.2</td>
<td>113.6</td>
<td>952</td>
</tr>
<tr>
<td>SMT-S</td>
<td>-2696.3</td>
<td>167.3</td>
<td>952</td>
</tr>
<tr>
<td>Diagonal</td>
<td>-3213.3</td>
<td>-349.7</td>
<td>-</td>
</tr>
</tbody>
</table>

- SMT produces much better fit to image data
- SMT can produce *all* eigenimages
Samples Generated Using Estimated Covariance

Diagonal: $y \sim N(\bar{y}, \hat{R})$

PCA + Shrinkage:

SMT:

Glasso:

SMT-S:
Signal Detection in Hyper-Spectral Remote Sensing

- Formulation of signal detection:
  \[ H_0 : r = x \]
  \[ H_1 : r = x + \lambda t, \]
  where: \( x \) is a background with covariance \( R \)
  \( t \) is a target signature
  \( \lambda \) is a scalar signal strength
  \( r \) is the observed pixel

- A filter \( q \) can be used for signal detection:
  \[ q^t \cdot r \geq \delta? \]

- The signal-to-clutter ratio (SCR) for a filter \( q \) is given by
  \[ SCR = \frac{(q^t t)^2}{E[(q^t x)^2]} \]
  \[ = \frac{(q^t t)^2}{q^t R q} \]

- The matched filter \( q = R^{-1} t \) optimizes the SCR
Matched Filter Performance: SCRR

- Matched filter using an estimated $\hat{R}$:

\[ q = \hat{R}^{-1} t \]

- The SCR ratio (SCRR) of the matched filter using an estimated $\hat{R}$:

\[ SCRR = \frac{SCR(\hat{R})}{SCR(\hat{R})} = \frac{SCR \text{ using estimated } \hat{R}}{\text{optimal } SCR} \]

- Numerical experiments on AVIRIS Florida data ($p=224$):
Detection Results for AVIRIS Florida Coastline Image

\[ SCRR = \frac{SCR(\hat{R})}{SCR(R)} = \frac{SCR \text{ using estimated } \hat{R}}{\text{optimal } SCR} \]

larger $\Rightarrow$ better!

Gaussian Case

Non-Gaussian Case
Anomaly Change Detection (ACD)

- Hyperbolic anomaly change detection:

\[
\frac{P(x,y)}{P_x(x)P_y(y)} > \delta?
\]
SMT for Anomaly Change Detection --- HACD

\[ p = 224 \]

“Smooth”
SMT for Anomaly Change Detection --- Dimension Reduction

HACD detection
\( d = 30 \)
“Smooth”
Part III. Regression from High Dimensional Vectors

- Conventional regression model
  \[ y = Xb + W \]
  - \( y \) is \( n \times 1 \) vector to be predicted; \( X \) is \( n \times p \) matrix of observations
  - \( b \) is \( p \times 1 \) vector of prediction weights

- MMSE estimate of \( b \):
  \[ \hat{b} = \left( \sum_{x_i} X_i^t X_i \right)^{-1} \cdot \sum_{x_i} X_i^t y_i \]

- Traditional solutions
  - Ordinary least squares: \( \hat{b} = (X^t X)^{-1} X^t y \)
  - Ridge regression: \( \hat{b} = \arg\min_b \left\{ ||y - Xb||^2 + \gamma ||b||^2 \right\} \)
  - Lasso regression: \( \hat{b} = \arg\min_b \left\{ ||y - Xb||^2 + \gamma ||b||_1 \right\} \)
SMT Regression

\[ y = Xb + W \]

- Use SMT method to estimate \( p \times p \) covariance of \( X \)
  \[ \hat{R} = \hat{E}\hat{E}' \equiv \frac{1}{n} X'X \]

- Decorrelate and whiten \( X \)
  \[ \tilde{X} = \hat{\Lambda}^{-1/2} \hat{E}'X \]

- Compute correlation between \( Y \) against \( X \)
  \[ \hat{b} = y'\tilde{X}/n \]

- Shrink coefficients using lasso, shrinkage or subset selection

\[ \text{Shrinkage: } \hat{b}_\gamma = \text{sign}(\hat{b})(\hat{b} - \gamma)_+ \]
Numerical Results

- Each experiment was repeated 30 times with a re-generated $t$ and $W$. 
Summary

- We presented
  - A fast non-iterative MAP reconstruction algorithm
    - Localization versus reconstruction
    - Framework of non-iterative MAP
    - Matrix source coding theory
  
  - Covariance estimation for high dimensional data
    - SMT representation
    - SMT covariance estimation

- SMT regression of high dimensional data
Thank you!!!

- Especially,
  - Prof. Bouman and Prof. Webb
  - Prof. Doerschuk, Prof. Low and Doctor Theiler
  - Prof. Allebach and Prof. Bell
  - Vaibhav, Leonardo, Jianing, Zhou, Yandong, Dalton…
  - Finally, my wife Yue!

Questions?