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

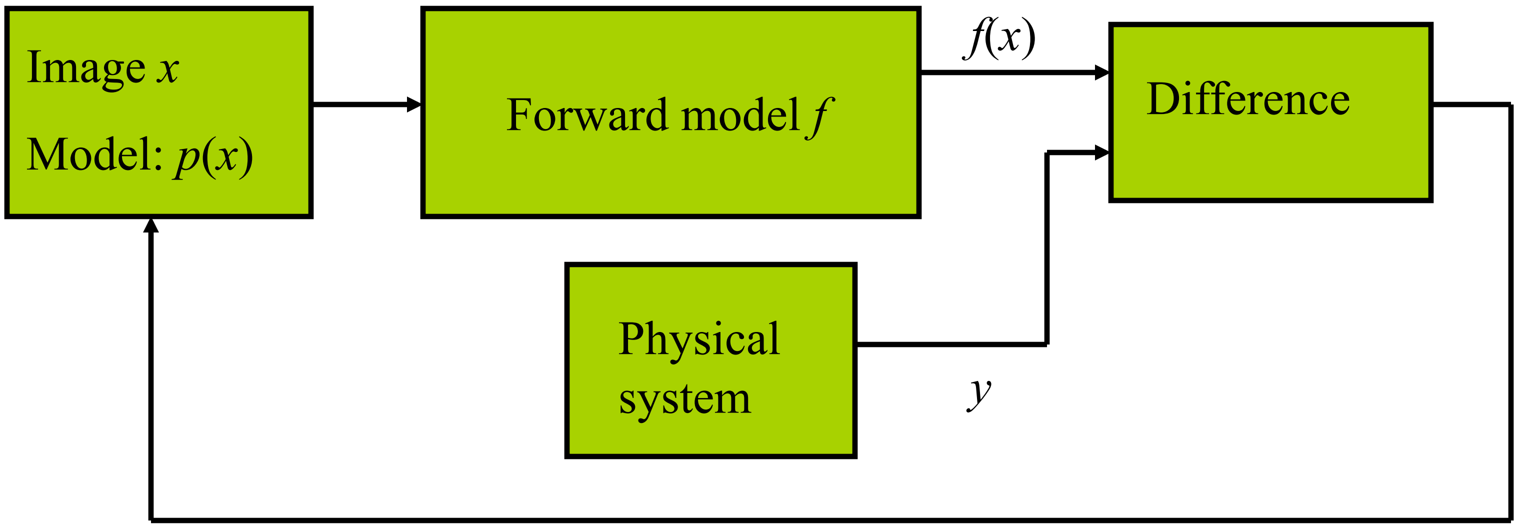
October 5, 2009

*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



x – Unknown image

y – Known surface measurement

- Objective: reconstruct unknown *x* from known *y*

Model-based Iterative Reconstruction

- MAP estimate of x:

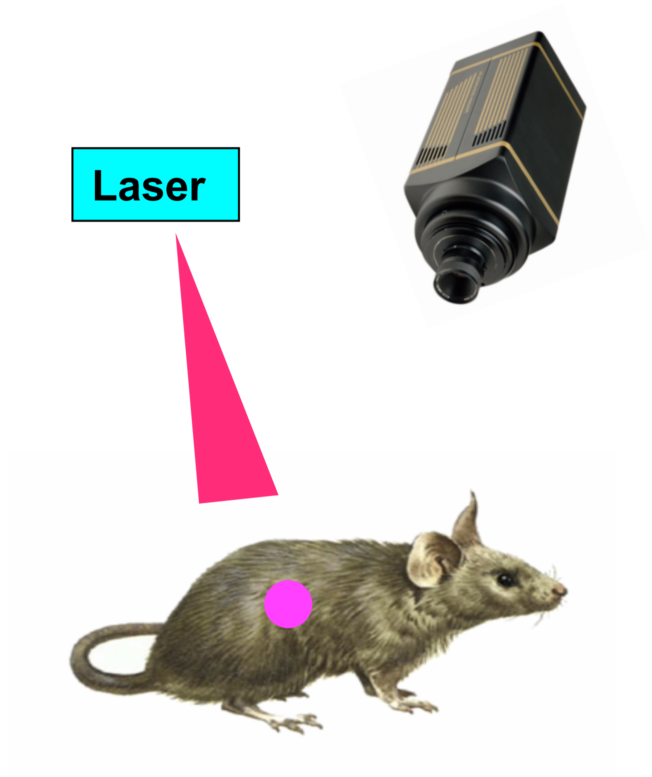
$$\hat{x} = \arg\max_{x \geq 0} \left\{ \log p(y \mid x) + \log p(x) \right\}$$
$$= \arg\max_{x \geq 0} \left\{ \left\| y - f(x) \right\|_{\Lambda}^{2} + \log p(x) \right\}$$

Λ – 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:

  ϕ – photon flux density (W/cm)

  D – diffusion coefficient (cm)

  μ_a – absorption coefficient (cm^{-1})

  c – speed of light in medium (cm/sec)

- System description:

$$y = f(x)$$

  x – Unknown images of optical parameters, e.g. (μ_a, D)

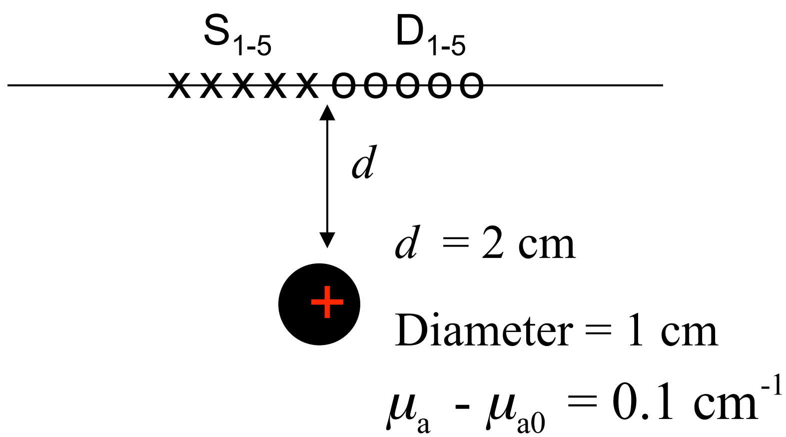
  y – Known surface measurement, e.g. ϕ

Approach 1: Fast Maximum Likelihood Localization

- Maximum likelihood (ML) estimate of inhomogeneity location:

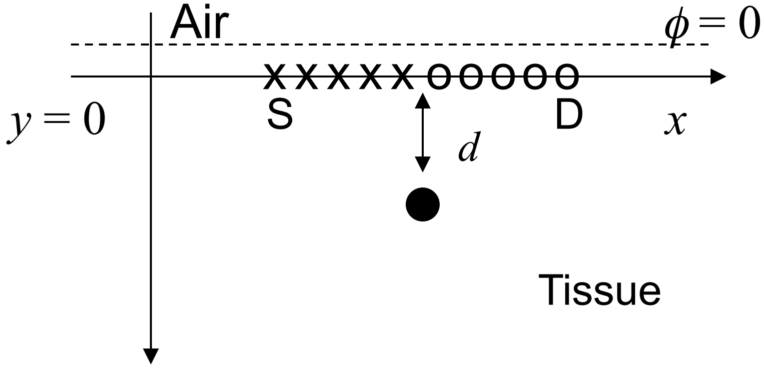
$$\hat{r} = \arg\min_{r \in \Omega} \| y - f(x_r) \|_{\Lambda}^2$$

- Parameterization of inhomogeneity? Size, contrast (μ_a - μ_{a0})...
- Inhomogeneity can be effectively modeled as a point: $\delta u(r)$



Localization versus Reconstruction

Reflectance measurement geometry



Optical parameters

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

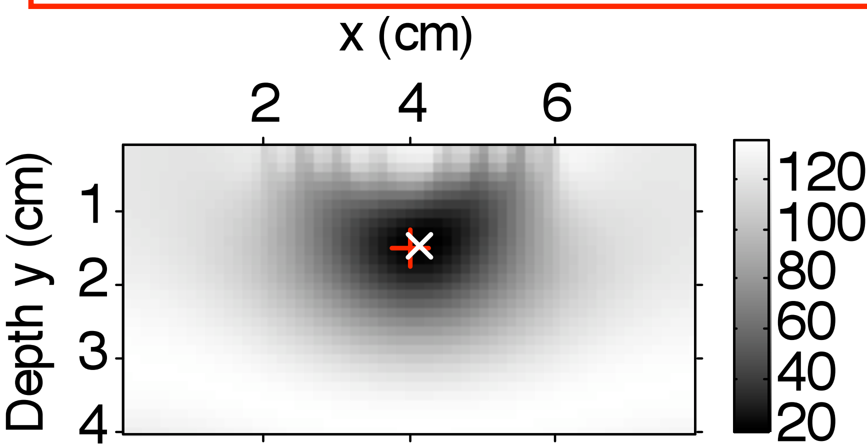
$D_0 = 0.03$ cm

Inhomogeneity: $d = 1.5$ cm

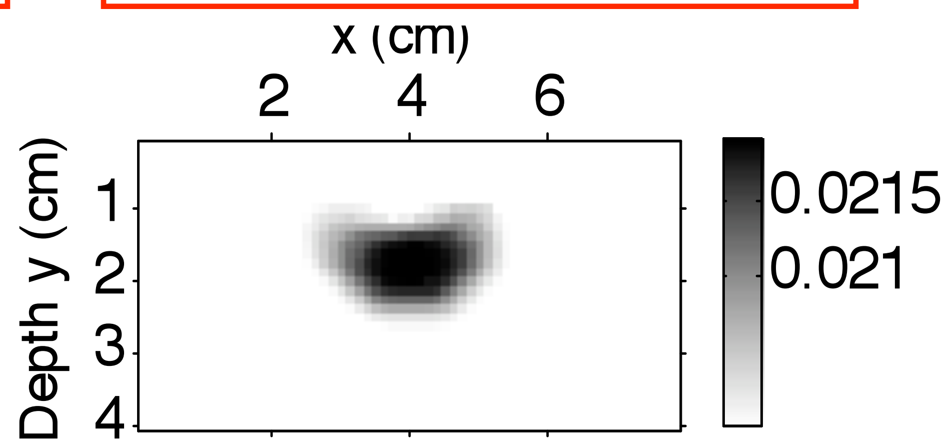
(Diameter = 6.25 mm) $\mu_a = 0.12$ cm^{-1}

Average SNR ~ 40 dB

Localization – *negative log likelihood*

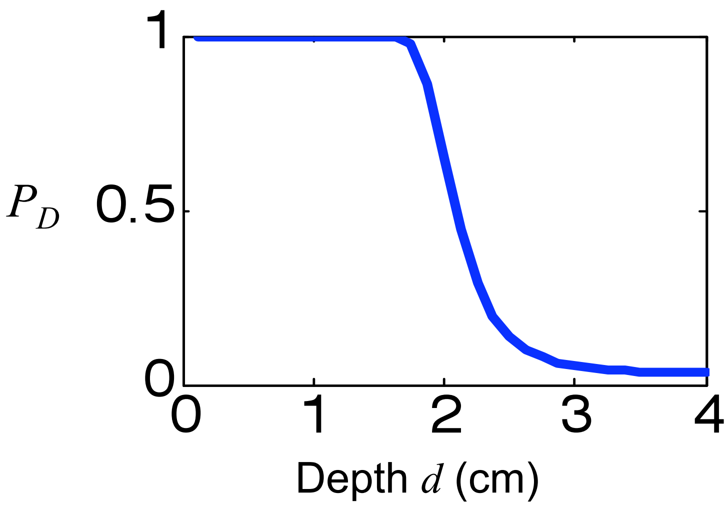


Reconstruction – $\hat{\mu}_a$

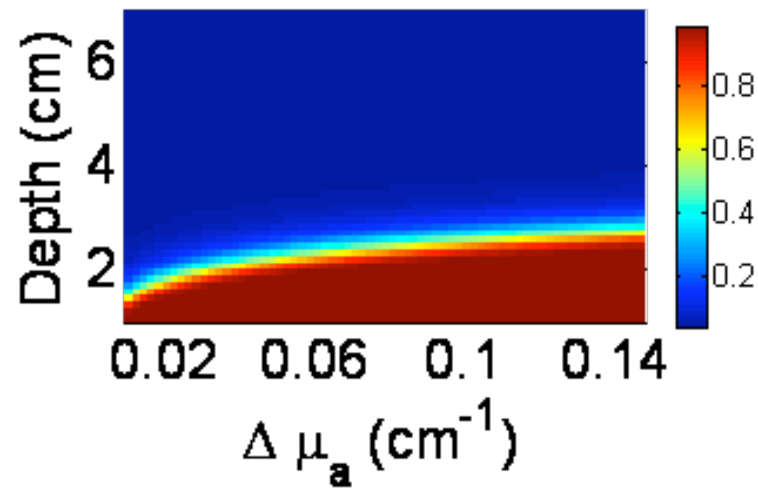


Probability of Detection Under Various Conditions

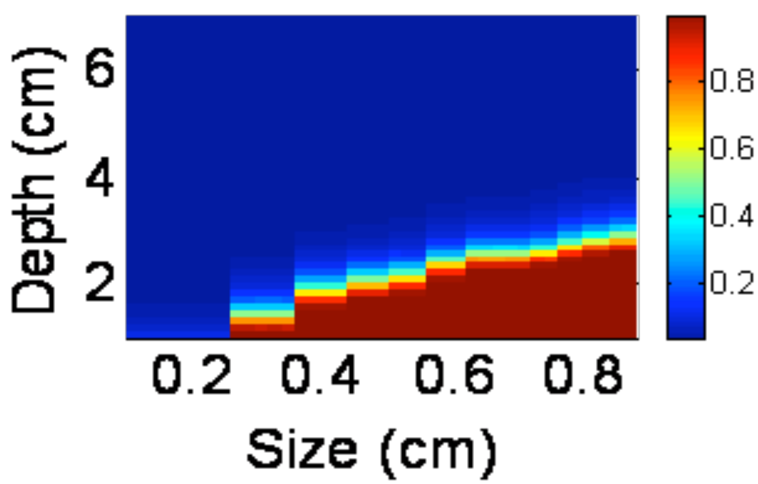
SNR=40dB, size=6.25mm, $\Delta\mu_a$=0.1cm^{-1}



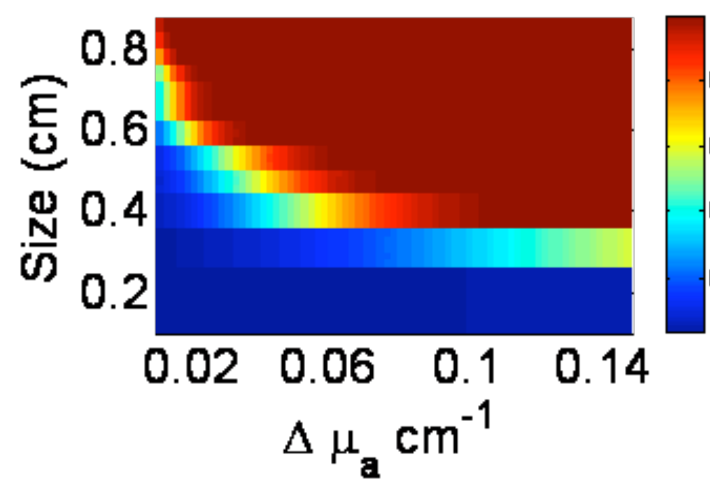
Absorption & depth, size = 6.25mm



Size & depth, $\Delta\ \mu_a = 0.1\ \text{cm}^{-1}$



Absorption & size, depth = 1.5 cm



Approach 2: Non-Iterative MAP Reconstruction

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

- 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:

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

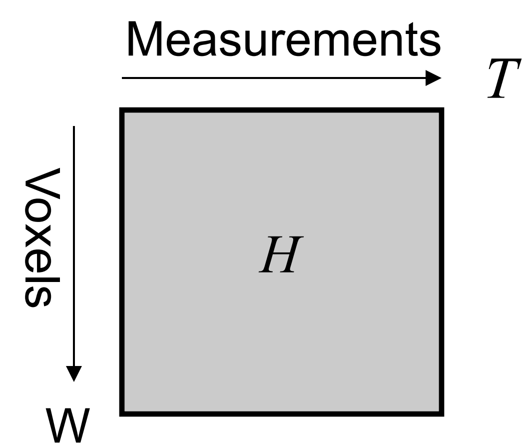
where $R_y = \mathrm{E}[yy^T]$.

- 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

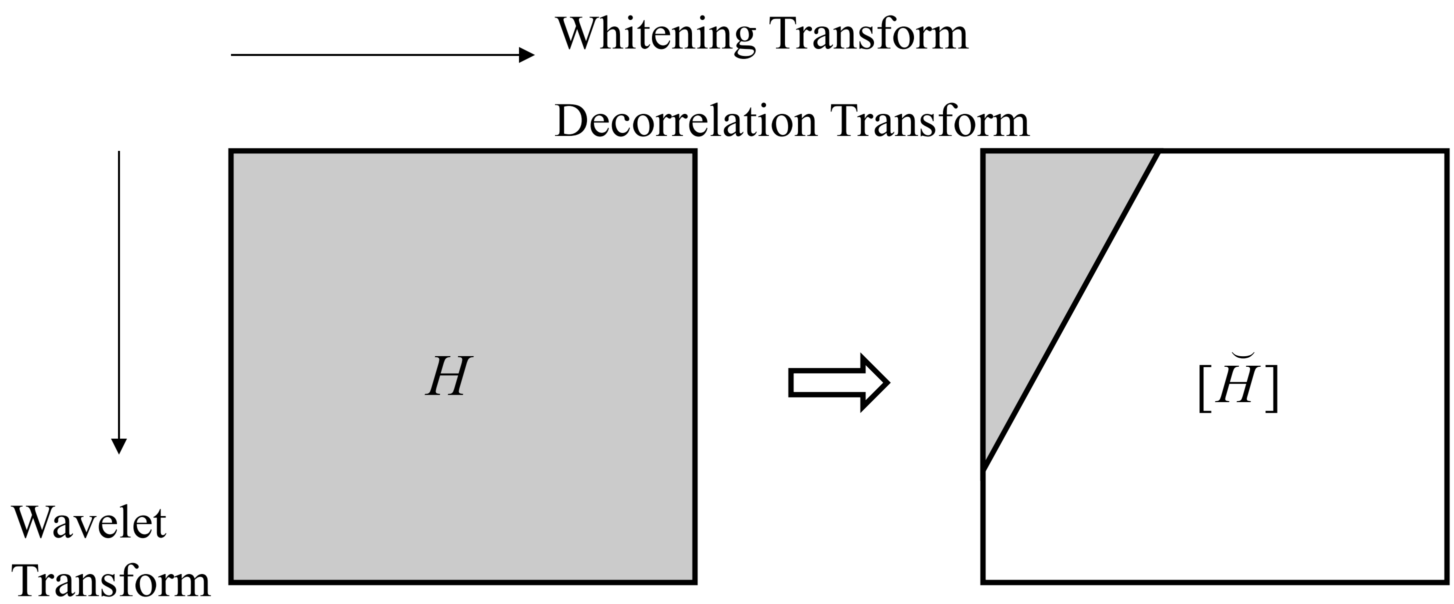
$$\Rightarrow \quad \breve{H} = W H T^{-1}, \quad \breve{y} = T y$$



- **Conclusion:**

$$\hat{x} = W^T \breve{H} \, \breve{y}$$

Reconstruction from Compressed Transform

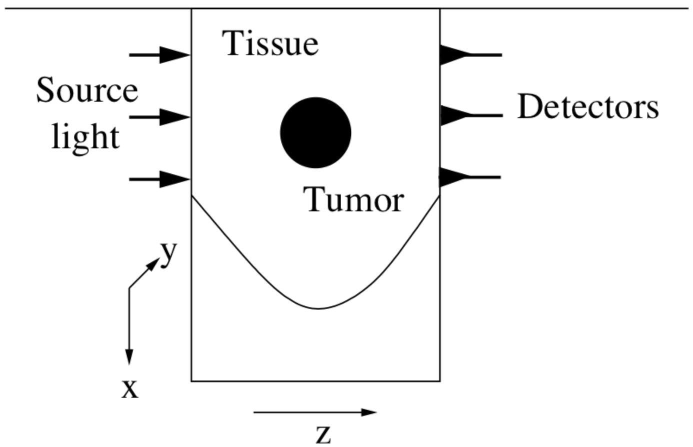


$$\hat{x} = W^T [\breve{H}]\, \breve{y}$$

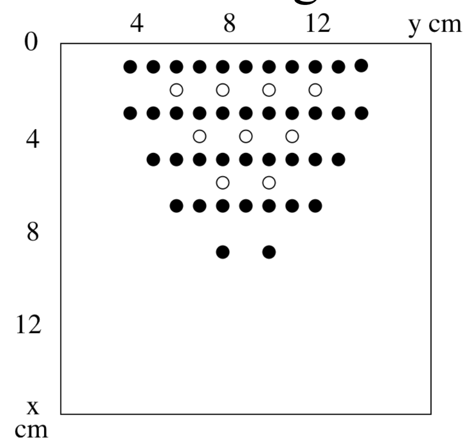
- Quantization:
  - Reduces data storage
  - Reduces computation for reconstruction
- Can use a variety of encoding methods for lossy coding of matrix columns

Numerical Example: ODT for Female Breast Imaging

Measurement System



Probe Configuration



- Bulk optical parameters:

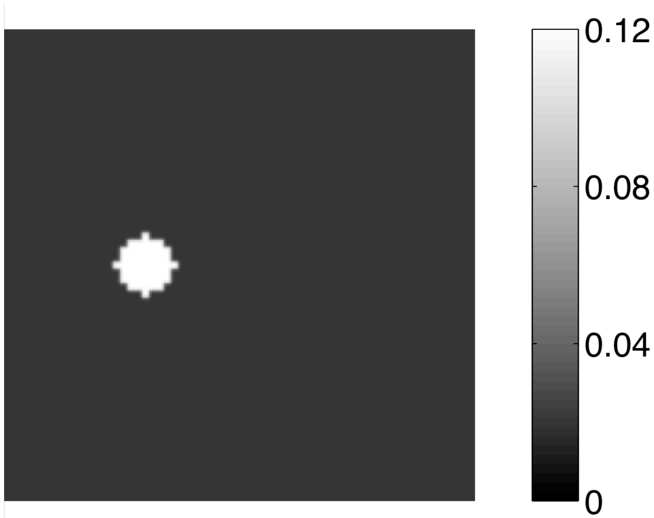
$$\mu_a = 0.02 \text{ cm}^{-1} \text{ and } D = 0.03 \text{ cm}$$

- Heterogeneity optical parameters (Radius: 1 cm, depth: 3 cm):

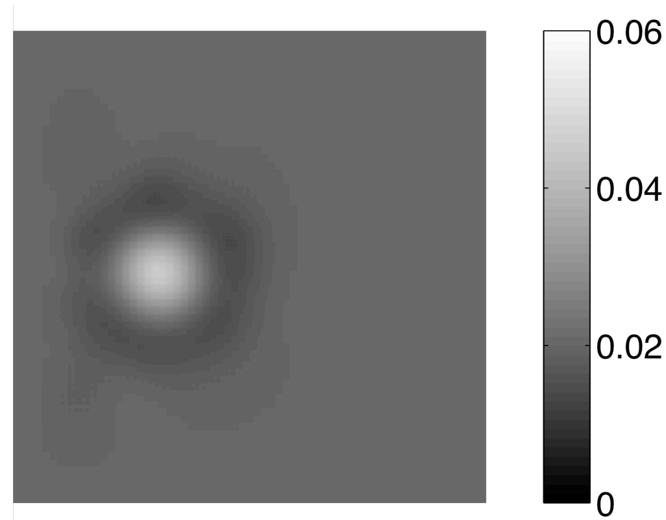
$$\mu_a = 0.12 \text{ cm}^{-1} \text{ and } D = 0.03 \text{ cm}$$

- Modulation frequency: 70 MHz
- Noise: shot noise with an average SNR of 35.8 dB

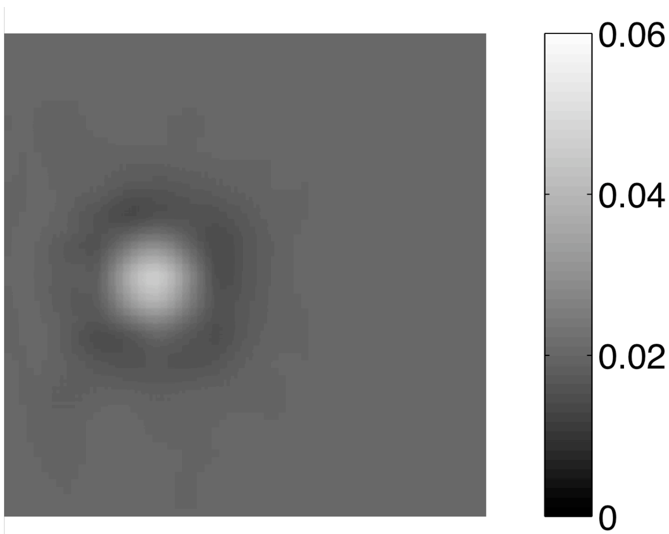
Reconstruction Results for Absorption Distribution



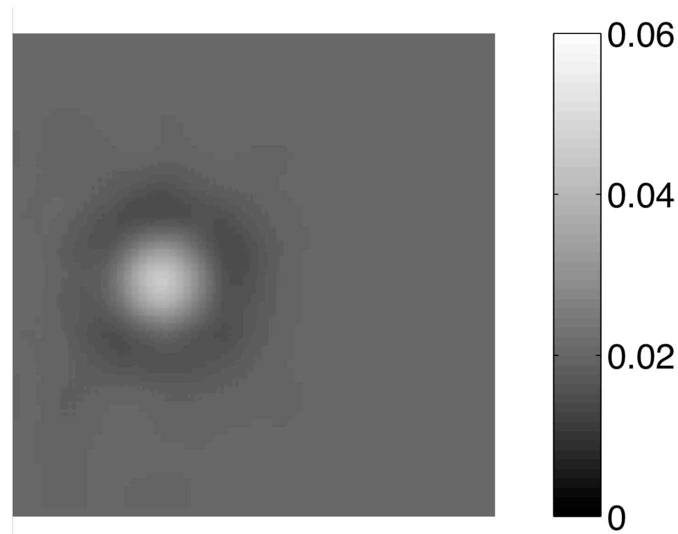
Original Image



Uncompressed

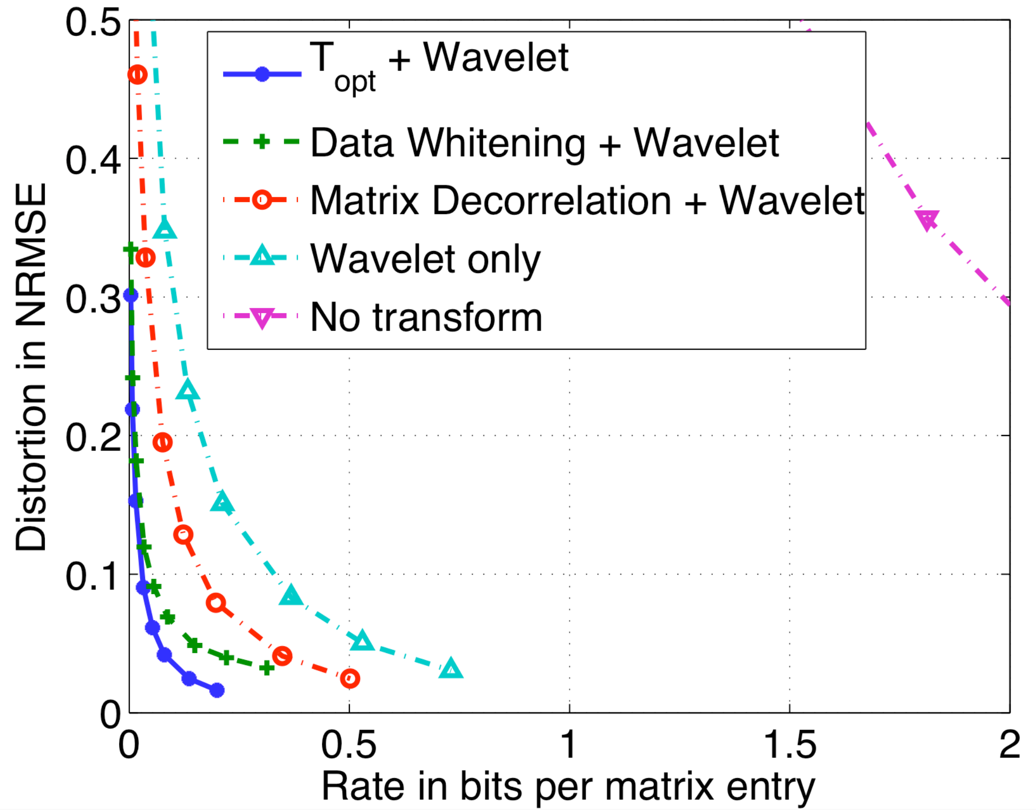


0.015 bpme, NRMSE=16.98%



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 cm^3.
- M=720 – number of measurements, N=65x65x33 – number of pixels, I =30 - number of iterations, c = 1808:1 - compression ratio (NRMSE≈10%)

	Computation		Storage	
	Order	Seconds	Order	Mbytes
Conjugate Grad.	MNI	*126*	MN	*766*
Uncompressed	NM	*0.89*	NM	*766*
Compressed with KLT	$NM/c+N+M^2$	*0.03*	$NM/c+M^2$	*4.4 = 0.4 + 4.0*

Problem: Storage and Computation of Orthonormal Transforms

- Reconstruction of x:

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

But

$$\breve{y} = Ty$$

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

- As # of measurements becomes large $\Longrightarrow$ 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:

$$\underbrace{Y = \left[y_1, y_2, \cdots, y_n \right]}_{\text{Observed Data}} \quad \underbrace{S \triangleq \tfrac{1}{n} Y Y^t}_{\text{Sample Covariance}} \quad \underbrace{R = E\left[S \right]}_{\text{True Covariance}}$$

- Likelihood of Y given R:

$$p_R(Y) = \frac{1}{(2\pi)^{np/2}} \left| R \right|^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2} tr\{ Y^t R^{-1} Y \} \right\}, \text{ with } R = E \Lambda E^t .$$

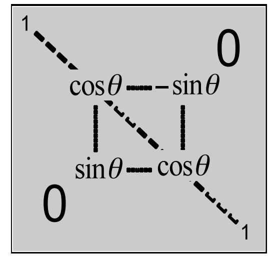
- ML estimate of eigenvectors and eigenvalues is given by

$$\begin{cases} \hat{E} = \arg\min \left\{ \left| diag(E^t S E) \right| \right\} \\ \hat{\Lambda} = diag(\hat{E}^t S \hat{E}) \end{cases}$$

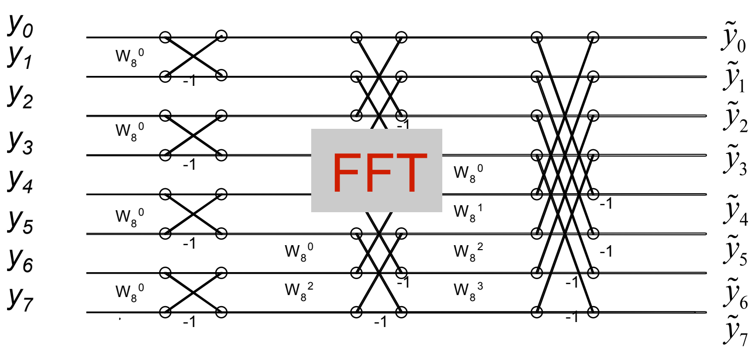
Sparse Matrix Transform (SMT)

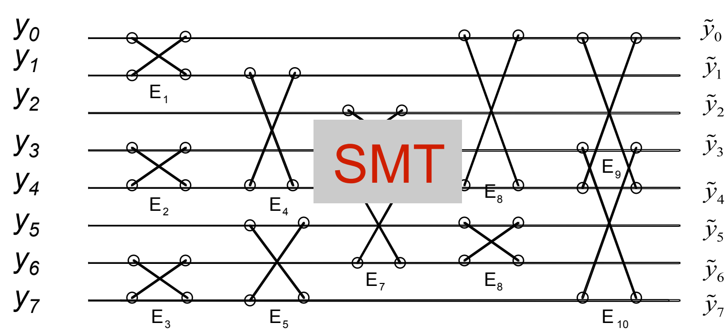
- SMT is product of Givens rotations:

$$E = E_1\, E_2 \cdots E_k \text{ where } E_k =$$



- 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*) $\in$ Ω (Sparse Matrix Transform)

- The constrained ML estimation of eigenvectors and eigenvalues:

$$\begin{cases} \hat{E} = \arg\min_{E \in \Omega_K} \left\{ \left| diag(E^t S E) \right| \right\} \\ \hat{\Lambda} = diag(\hat{E}^t S \hat{E}) \end{cases}$$

- 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:

$$\hat{E} = \arg \min_{E = E_1 E_2 \cdots E_K} \left\{ \left| diag(E^t S E) \right| \right\}$$

- A greedy algorithm is used for optimization
- The algorithm:

For $k = 1$ to K {

Select most correlated coordinate pair

Decorrelate the coordinate pair with rotation E_k

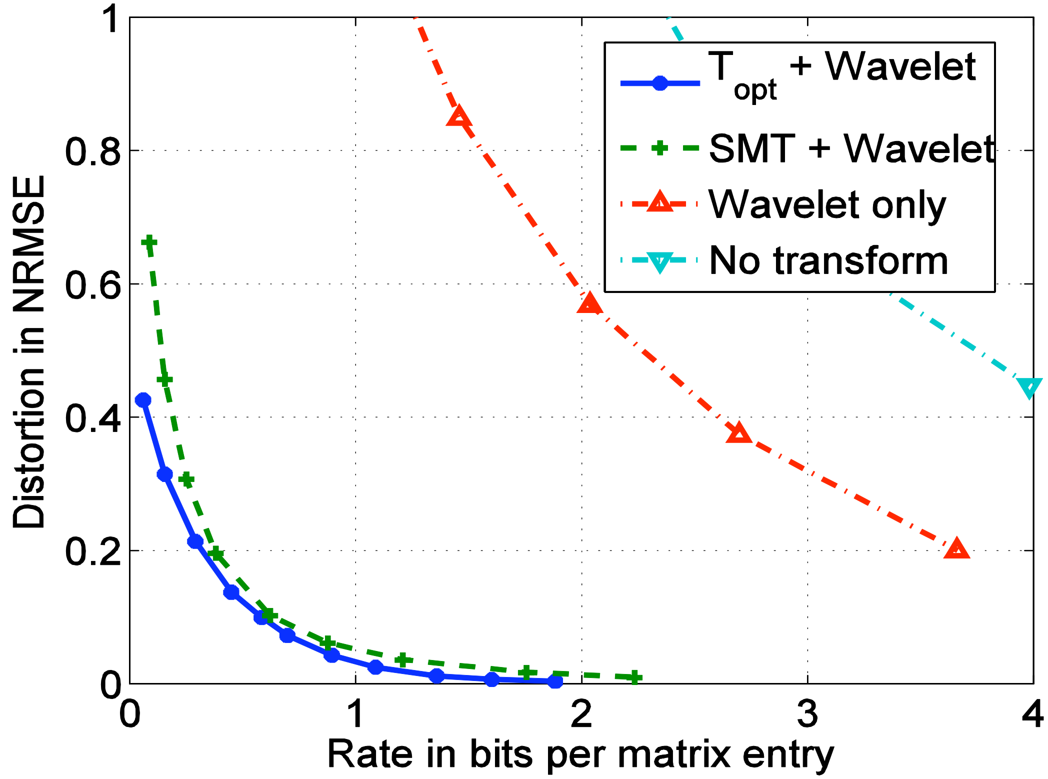
}

$E \leftarrow E_1 E_2 \cdots E_K$

- Comments: 1) K is chosen to maximize cross-validated likelihood

  2) Does not depend on ordering of vector or stationary assumption

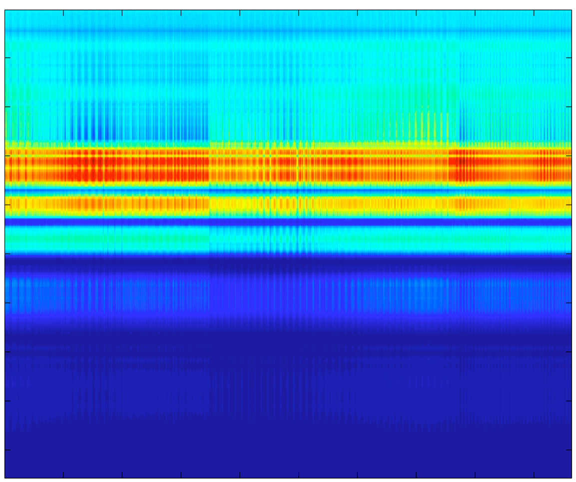
Distortion versus Rate – SMT for Optical Tomography



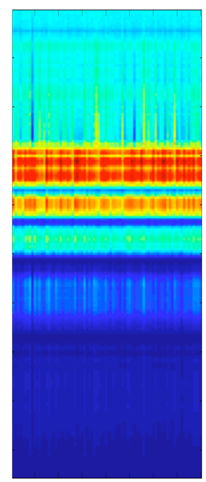
- The number of SMT butterflies used is $p \log_2(p)$
- Comment: SMT both reduces the data storage and computation required to compute Ty

Covariance Estimation for Hyperspectral Data: Grass Class

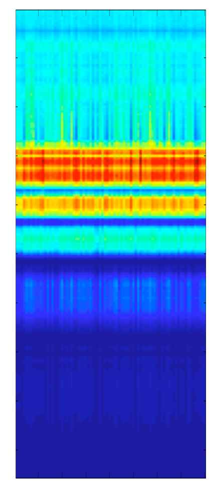
- # of hyperspectral bands: $p = 191$, # of samples: $n = 80$



Sample Data (1928)



Y (Gaussian)



Y (non-Gaussian)

Estimators Compared

- Shrinkage estimator:

$$\hat{R} = (1-\alpha)\cdot S + \alpha \cdot diag(S)$$

- Graphic lasso (glasso) covariance estimator:

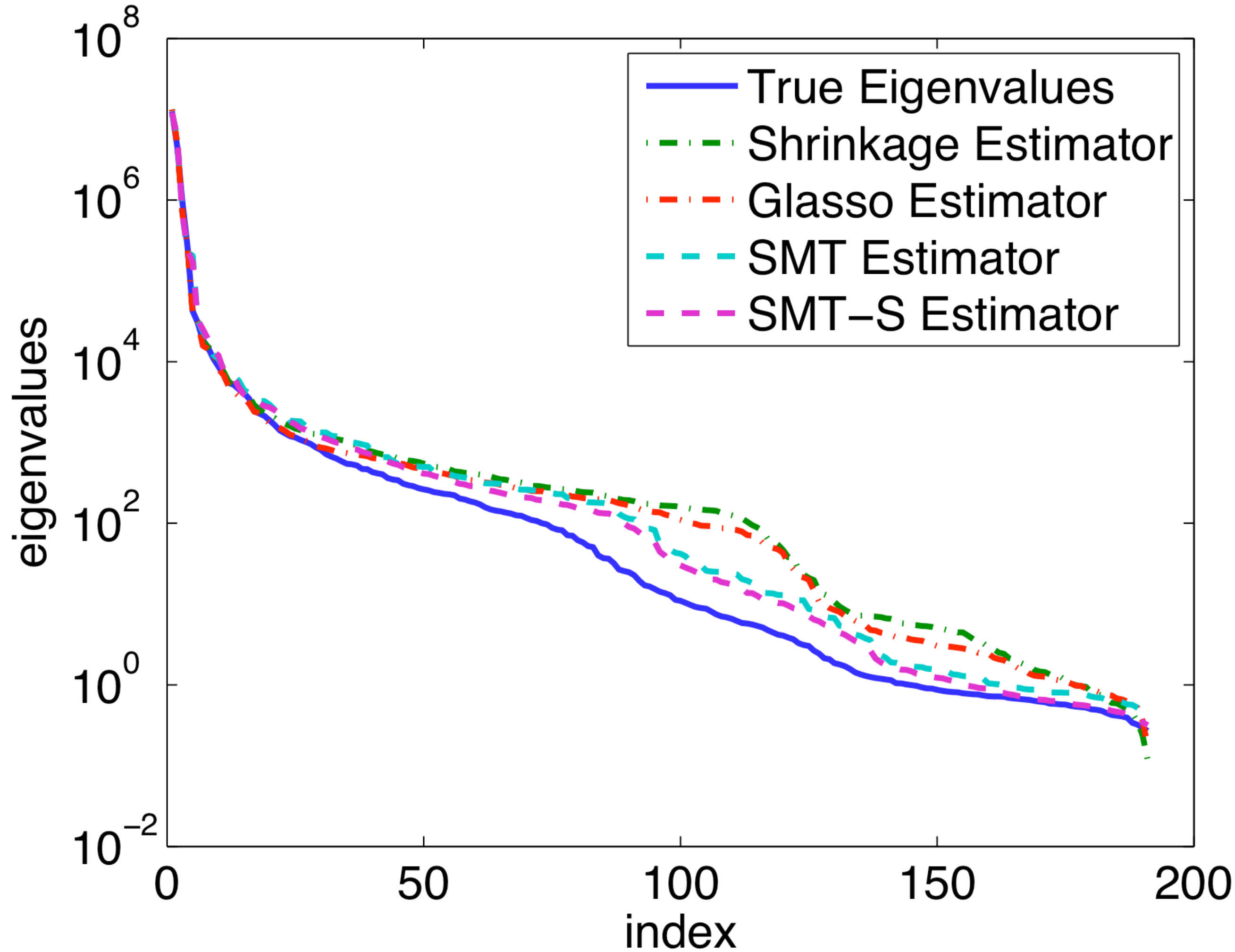
$$\hat{R} = \arg\max\left\{\log(Y|R) - \alpha\left\|R^{-1}\right\|_1\right\}$$

- SMT estimator
- SMT-shrinkage (SMT-S) estimator:

$$\hat{R} = (1-\alpha)\cdot S + \alpha \cdot \hat{R}_{SMT}$$

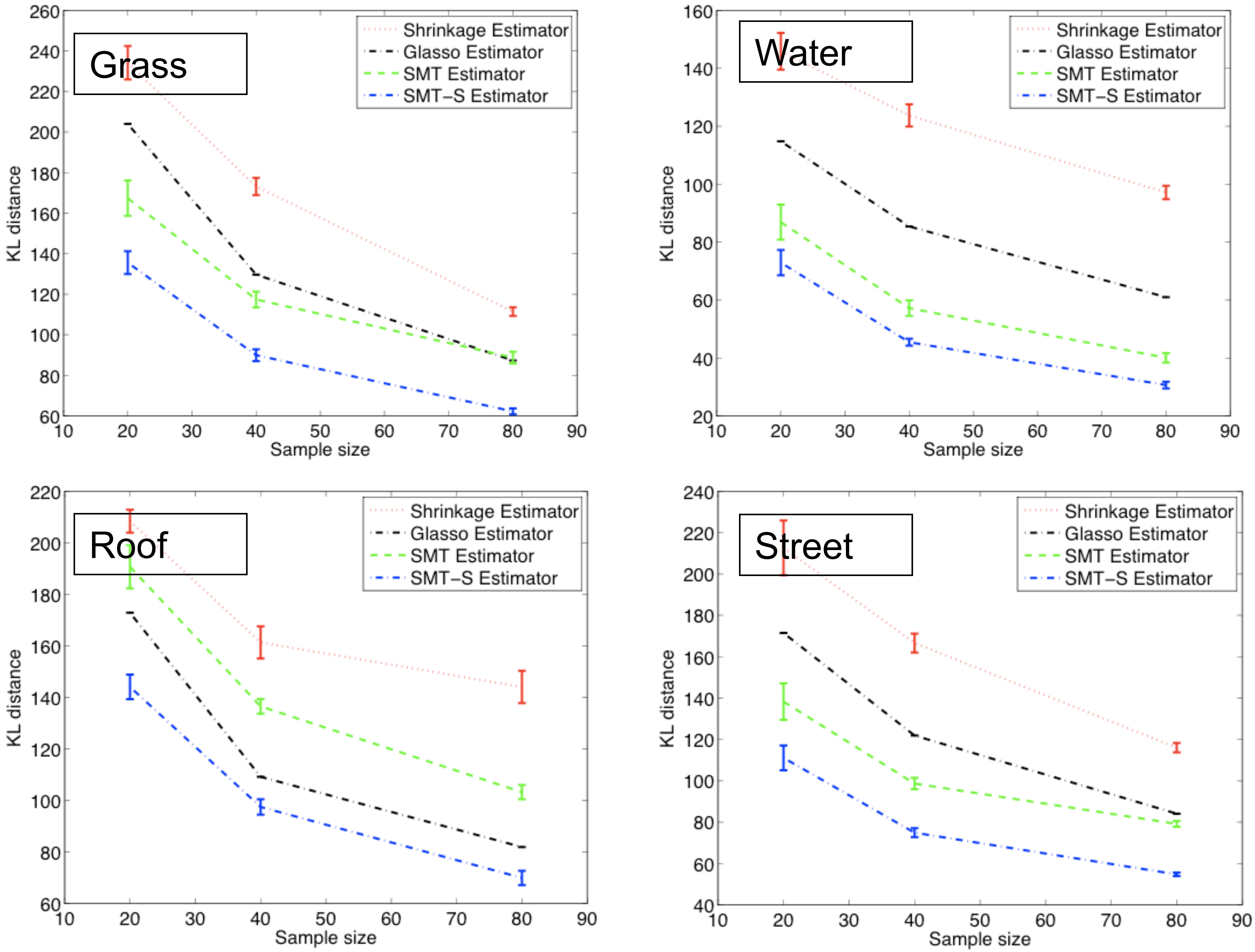
- Comment: α is estimated using cross-validation

Eigenvalue Estimation: Gaussian Case



$$\hat{\Lambda} = diag(\hat{E}^t S \hat{E})$$

Results in Kullback-Leibler Distance: Gaussian Case



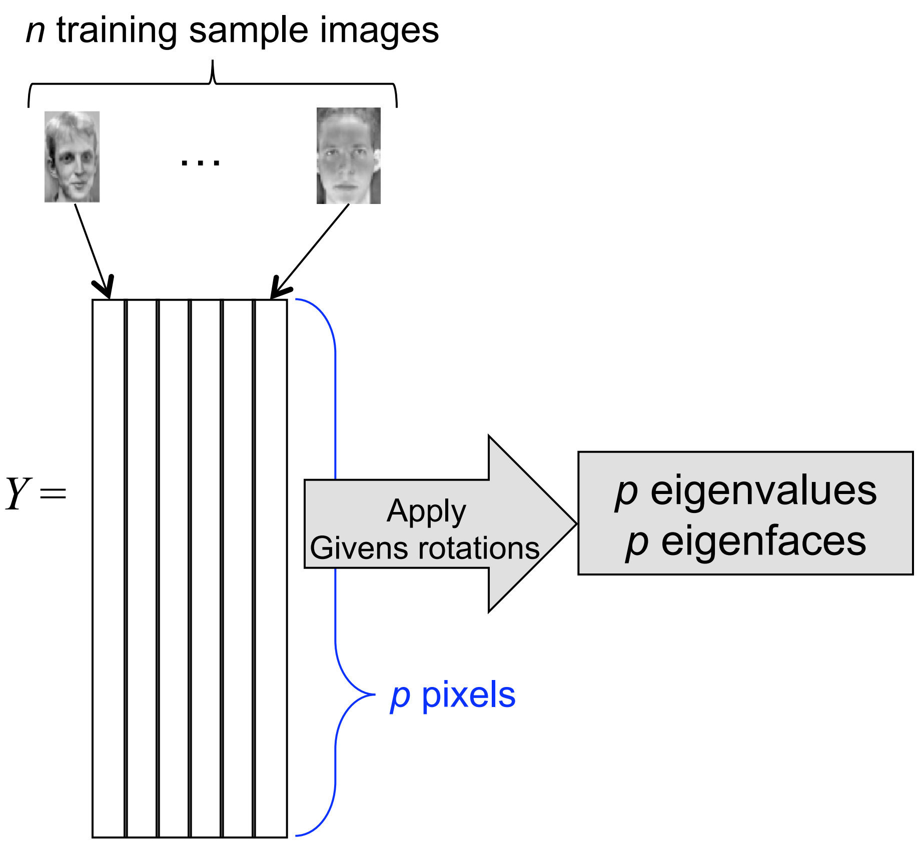
Computational Complexity

	Complexity (without c.v.)	CPU time (seconds)	Model order
Shrinkage	p	8.6 (with c.v.)	1
glasso	$p^3 I$	422.6 (without c.v.)	4939
SMT	$p^2 + Kp$	6.5 (with c.v.)	495
SMT-S	$p^2 + Kp$	7.2 (with c.v.)	496

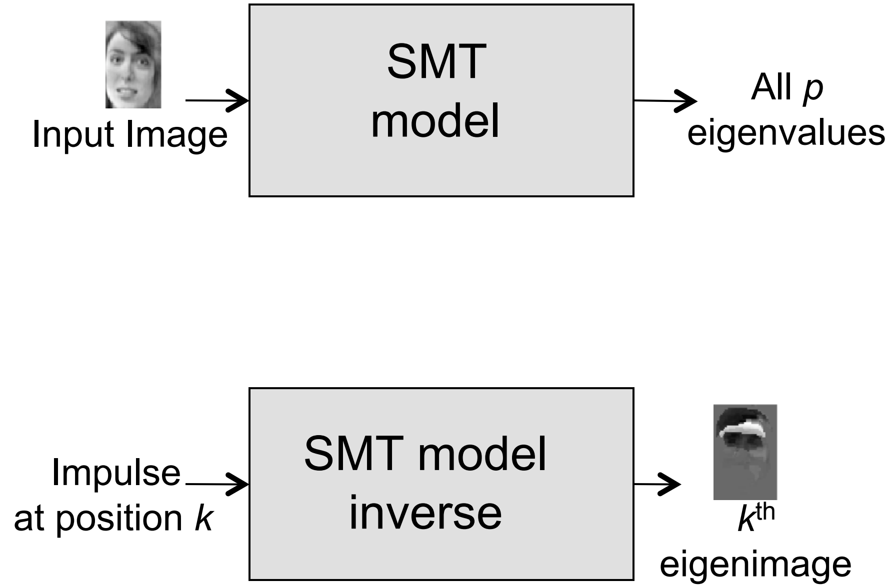
- I - cycles used in glasso, K - number of Givens rotations in SMT
- Numerical results are based on the grass Guassian case with $n = 80$
- c.v.: cross-validation

Computing Eigen-Images Using SMT

SMT Model Training



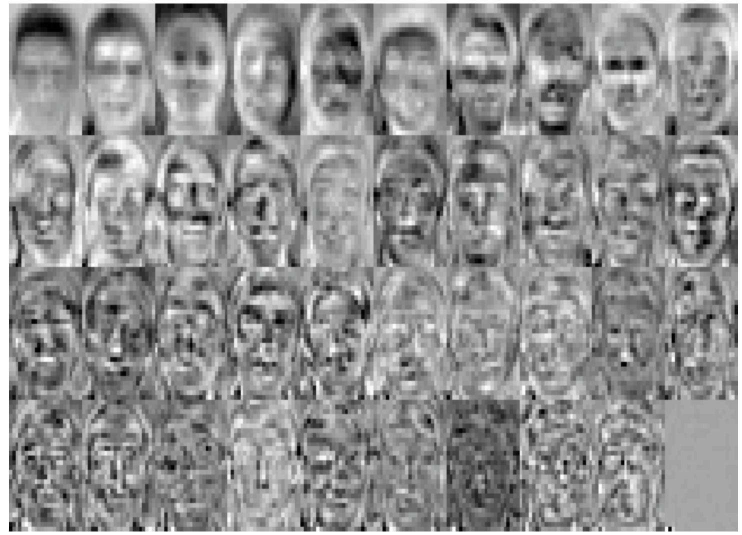
SMT Eigenimage Analysis



SMT versus Traditional Eigen-faces

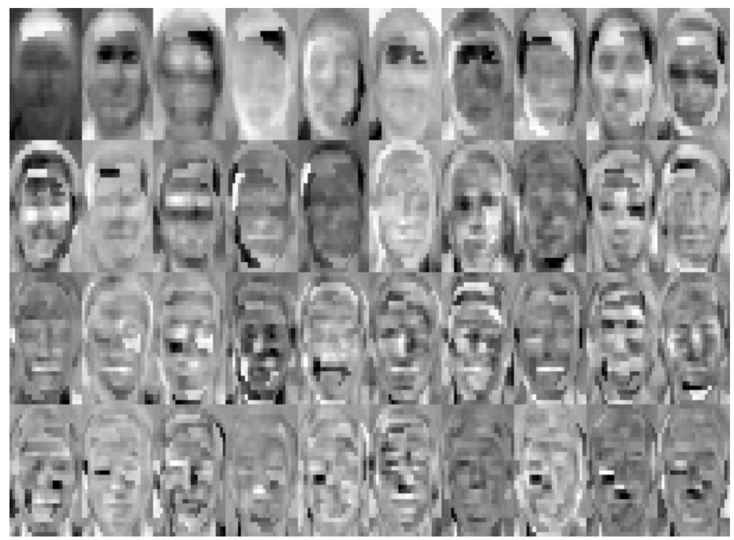
Face dataset:

PCA + Shrinkage:



SMT:

SMT-S:



Comparison of Traditional and SMT Eigenimages

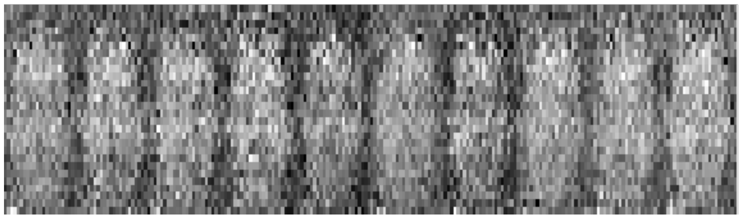
- Eigenimage experiment
  - Dataset: face image
  - Number of samples (n) = 80
  - Dimensions (p) = 644
- Use cross-validation to compute expected log likelihood
  - Δloglikelihood=167.3

Method	Maximum log-likelihood	Δ	K_{max}
PCA+Shrinkage	-2863.6	0	-
glasso	-2699.1	164.5	-
SMT	-2797.2l	113.6	952
SMT-S	-2696.3	167.3	952
Diagonal	-3213.3	-349.7	-

- SMT produces much better fit to image data
- SMT can produce *all* eigenimages

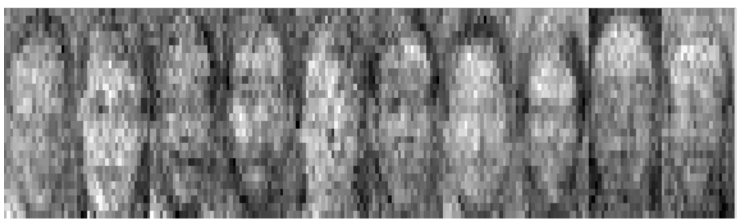
Samples Generated Using Estimated Covariance

Diagonal:

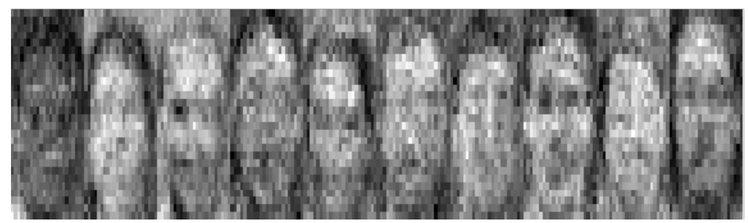


$$y \sim N(\bar{y}, \hat{R})$$

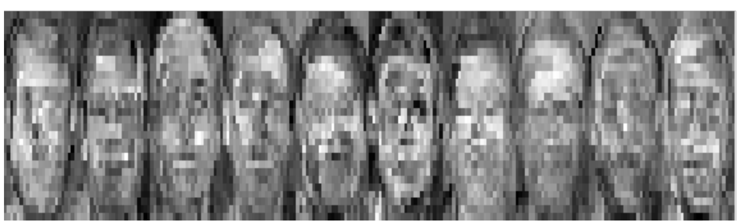
PCA + Shrinkage:



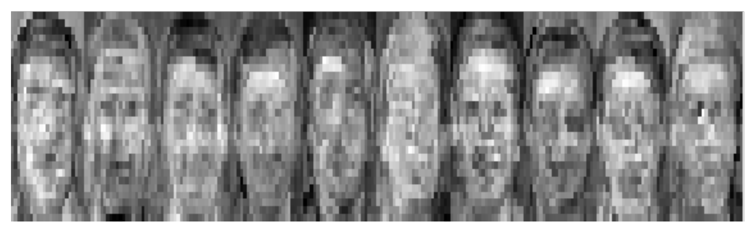
Glasso:



SMT:



SMT-S:



Signal Detection in Hyper-Spectral Remote Sensing

- Formulation of signal detection :

  $\mathrm{H}_0 : r = x$

  $\mathrm{H}_1 : r = x + \lambda \cdot t$,

  where: x is a background with covariance R

  t is a target signature

  λ is a scalar signal strength

  r is the observed pixel

- A filter q can be used for signal detection:

$$\boxed{q^t \cdot r \geq \delta ?}$$

- The signal-to-clutter ratio (SCR) for a filter q is given by

$$SCR = \frac{(q^t t)^2}{\mathrm{E}[(q^t x)^2]} = \frac{(q^t t)^2}{q^t R q}$$

  - The matched filter $q = R^{-1} t$ optimizes the SCR

Matched Filter Performance: SCRR

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

$$q = \hat{R}^{-1} t$$

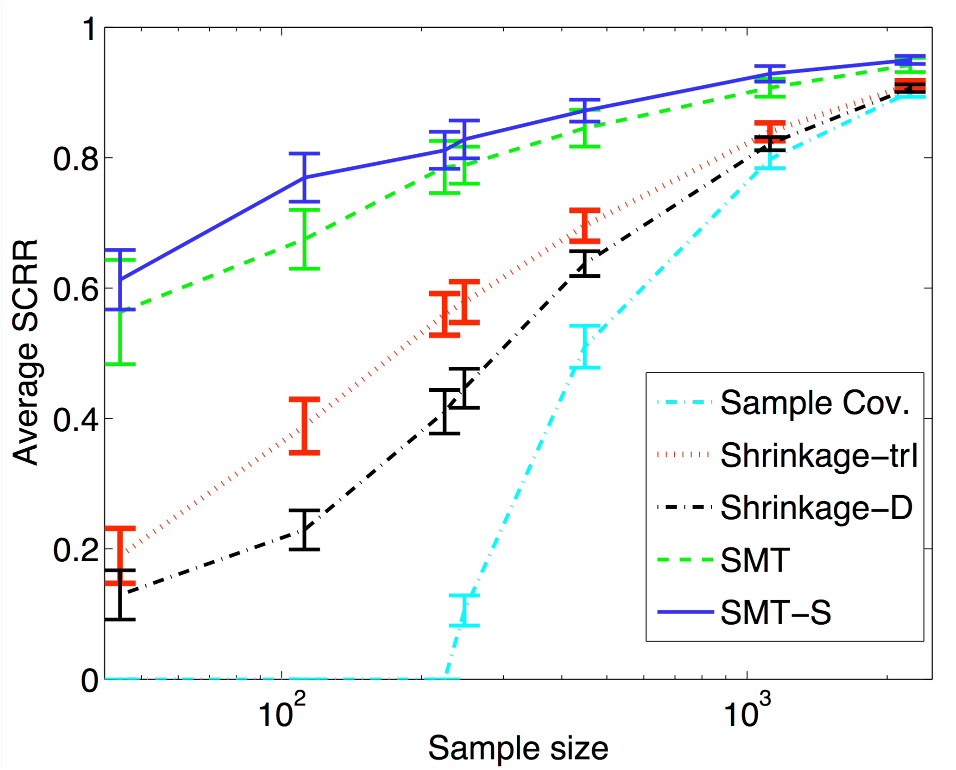
- The SCR ratio (SCRR) of the matched filter using an estimated $\hat{R}$:

$$SCRR = \frac{SCR(\hat{R})}{SCR(R)} = \frac{SCR \text{ using estmated } \hat{R}}{\text{optimal } SCR}$$

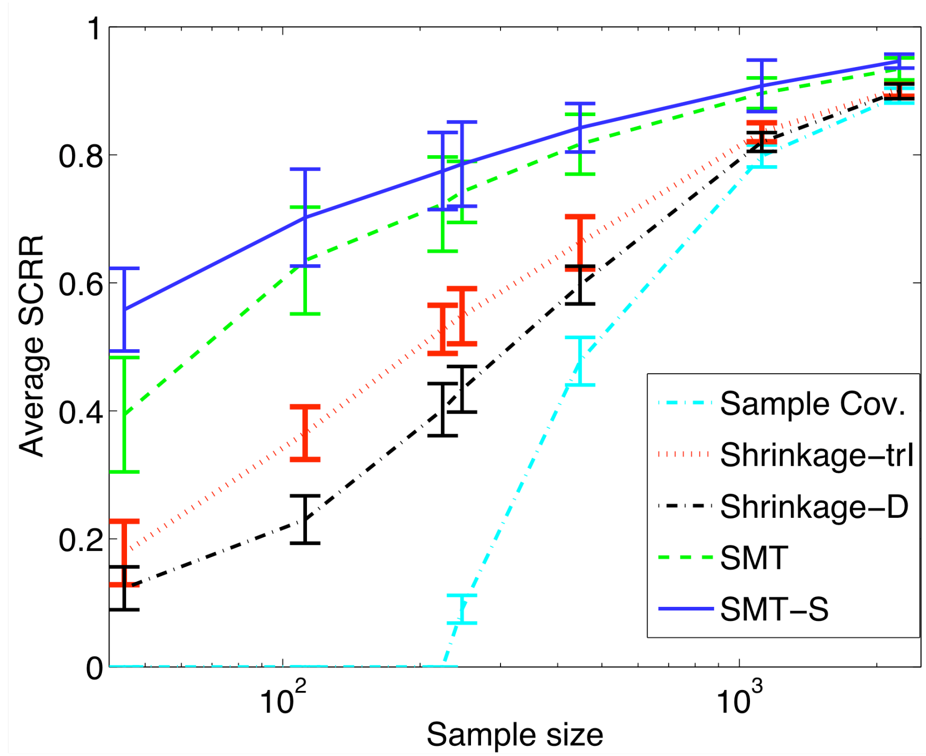
- Numerical experiments on AVIRIS Florida data (p=224):

Detection Results for AVIRIS Florida Coastline Image

$$SCRR = \frac{SCR(\hat{R})}{SCR(R)} = \frac{SCR \text{ using estmated } \hat{R}}{\text{opitmal } SCR}$$ *larger* $\Rightarrow$ *better!*

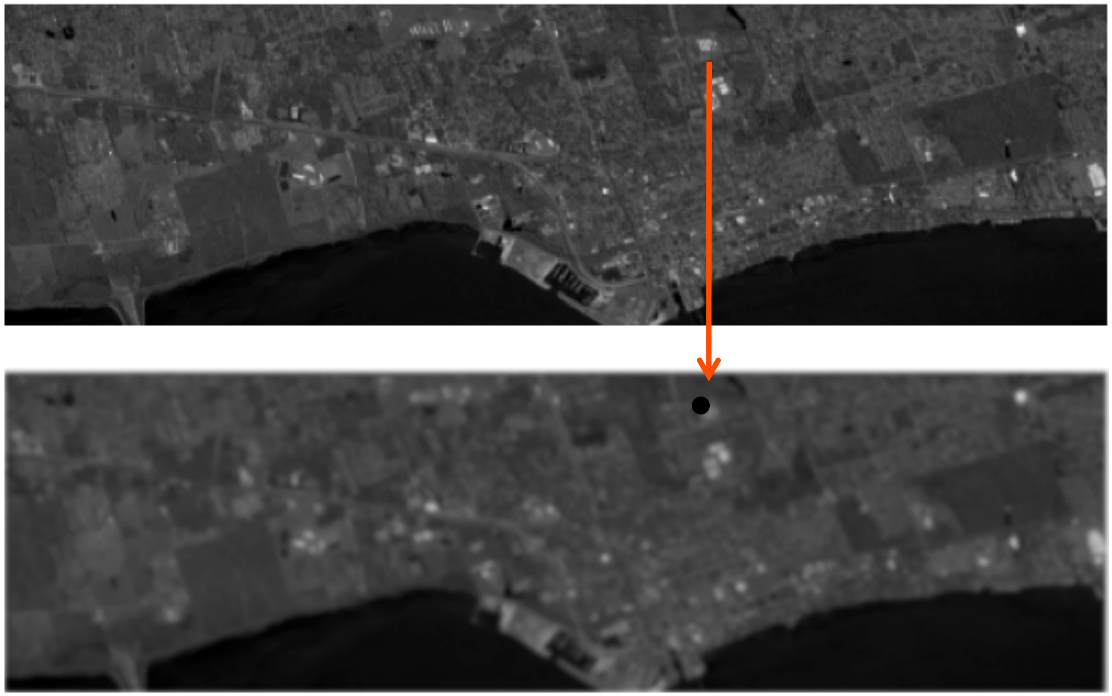


Gaussian Case



Non-Gaussian Case

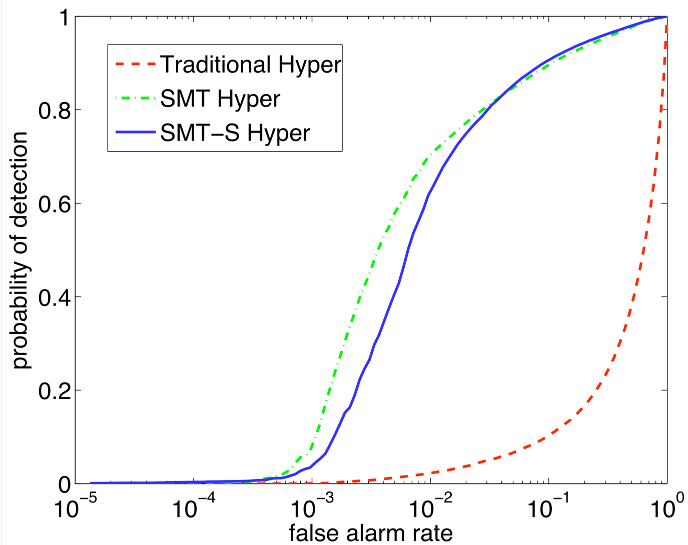
Anomaly Change Detection (ACD)



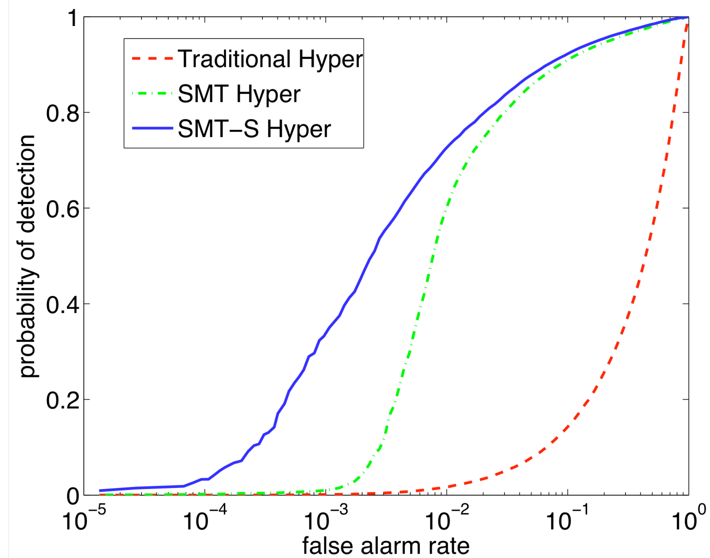
- Hyperbolic anomaly change detection:

$$\frac{P(x,y)}{P_x(x)P_y(y)} > \delta ?$$

SMT for Anomaly Change Detecion --- HACD

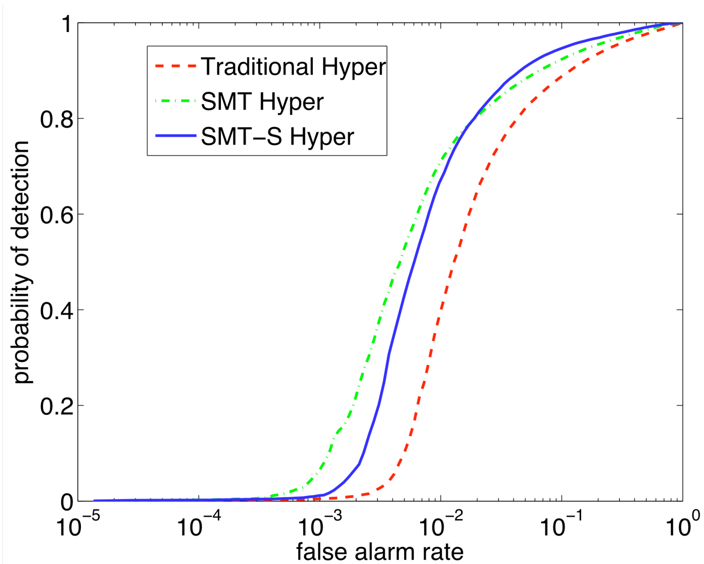


n=1p

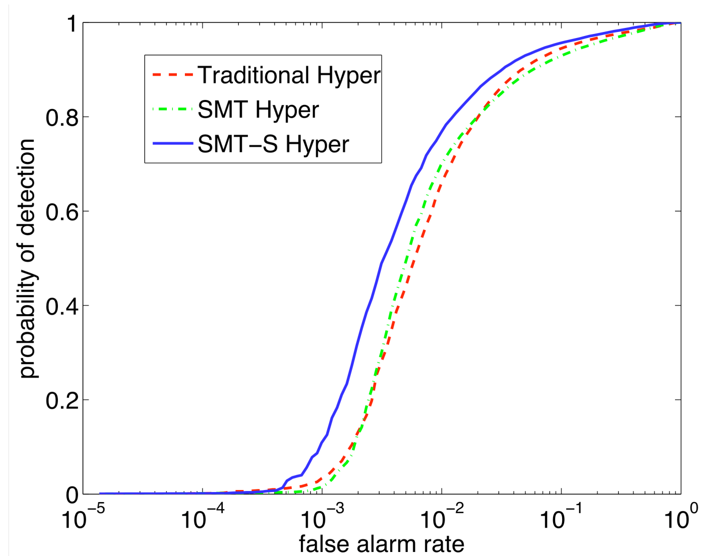


n=2p

p = 224
"Smooth"

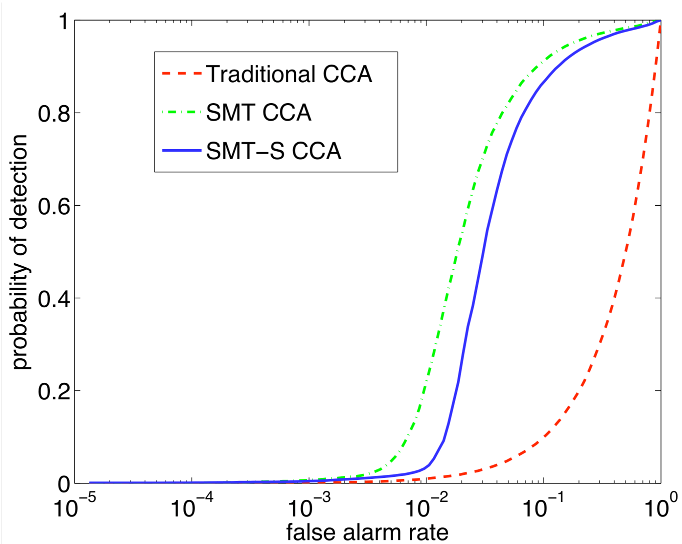


n=5p

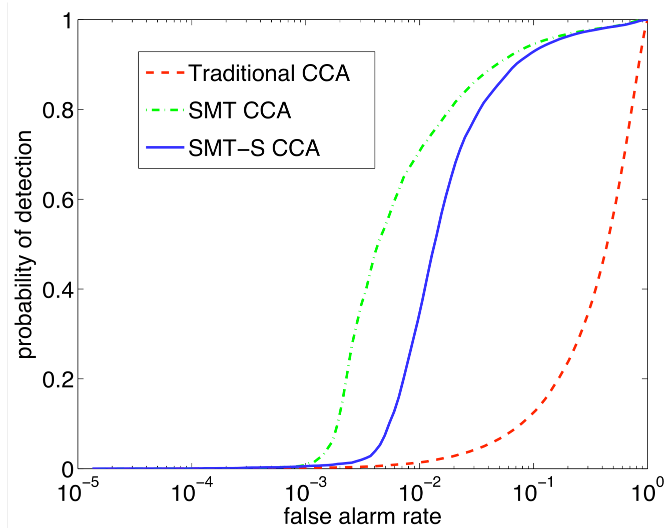


n=10p

SMT for Anomaly Change Detection --- Dimension Reduction

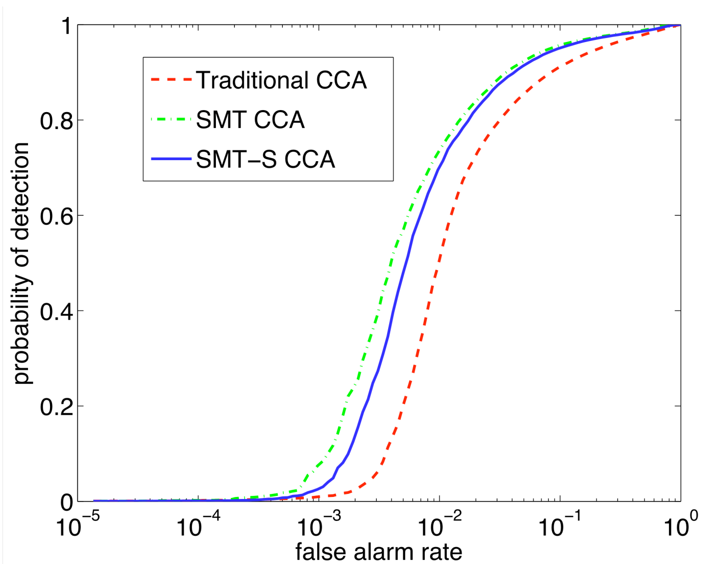


n=1p

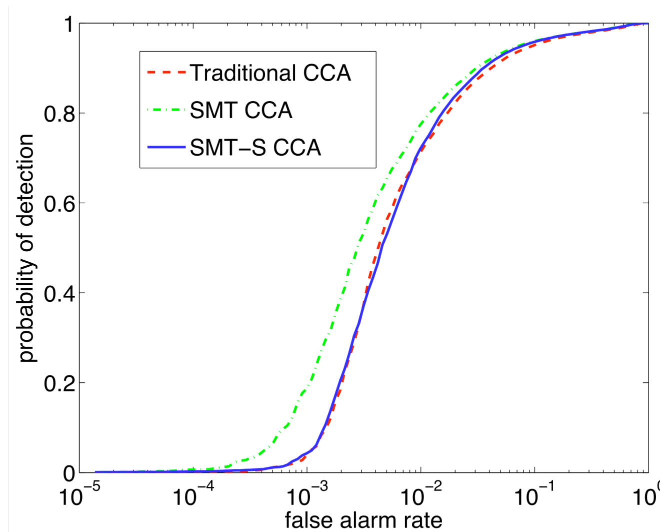


n=2p

HACD detection
d = 30
“Smooth”



n=5p



n=10p

Part III. Regression from High Dimensional Vectors

- Conventional regression model

$$y = Xb + W$$

  - y is nx1 vector to be predicted; X is nxp matrix of observations
  - b is px1 vector of prediction weights

- MMSE estimate of b: $\hat{b} = \underbrace{\mathrm{E}[X^t X]}_{R_x}^{-1} \cdot \underbrace{\mathrm{E}[X^t y]}_{\rho_{xy}}$

- 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

$$\hat{R} = \hat{E}\hat{\Lambda}\hat{E}^t \cong \frac{1}{n} X^t X$$

- Decorrelate and whiten X

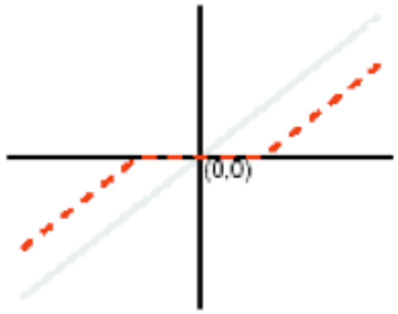
$$\tilde{X} = \hat{\Lambda}^{-1/2} \hat{E}^t X$$

- Compute correlation between Y against X

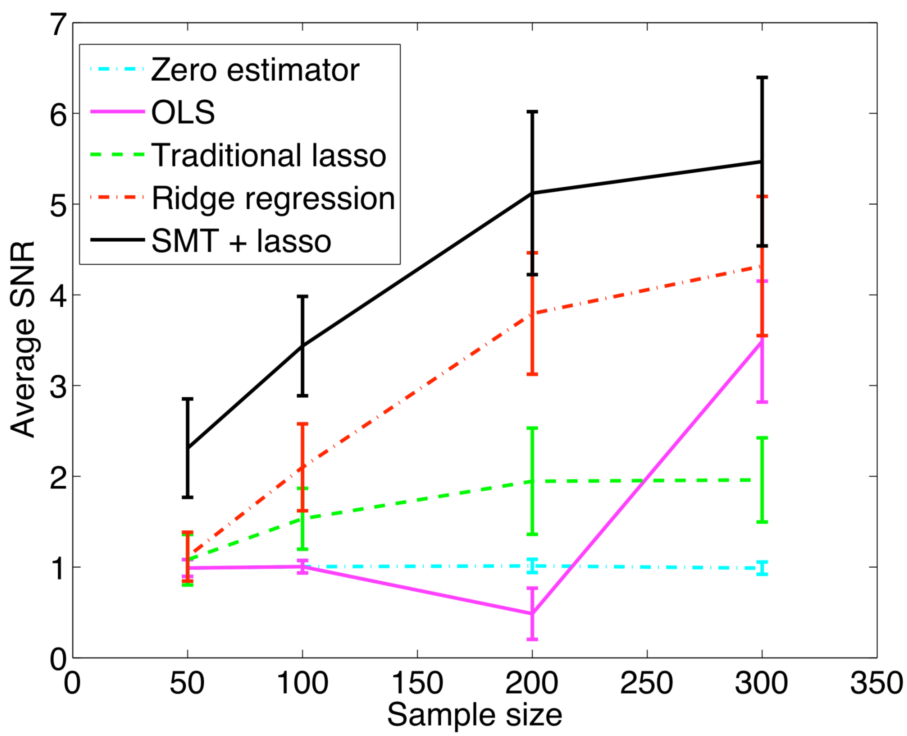
$$\hat{b} = y^t \tilde{X} / n$$

- Shrink coefficients using lasso, shrinkage or subset selection

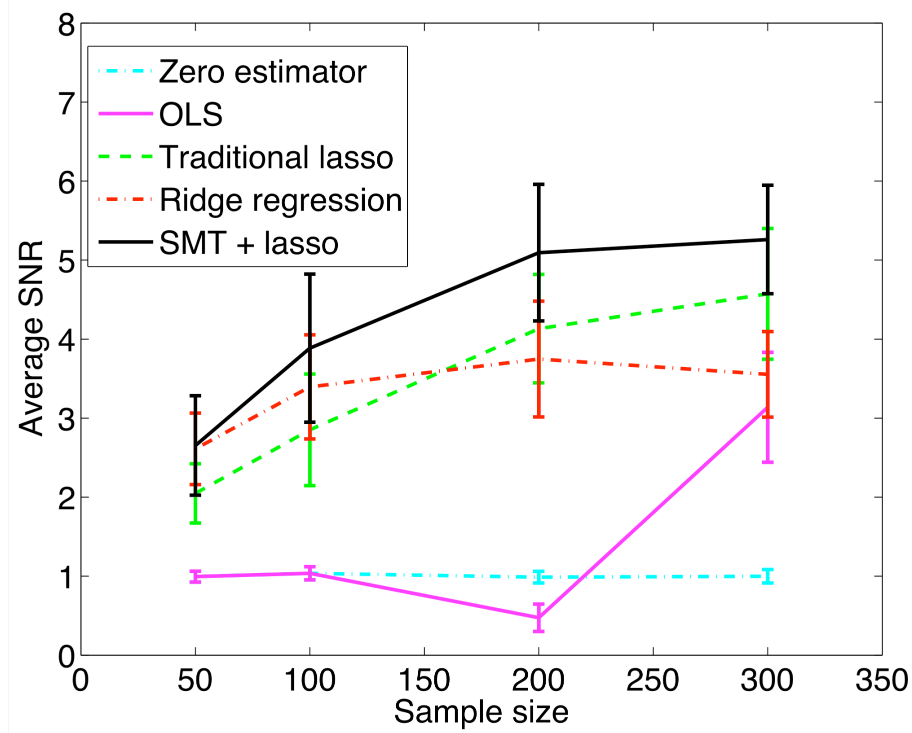
Shrinkage: $\hat{b}_\gamma = sign(\hat{b})(\hat{b} - \gamma)_+$



Numerical Results



Grass



Water

- 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?