Covariance Estimation for High Dimensional Data Vectors

Charles A. Bouman
Purdue University
School of Electrical and Computer Engineering

Co-authored with: Guangzhi Cao and Leonardo R Bachega

PURDUE
UNIVERSITY

Covariance Estimation for High Dimensional Data Vectors

- Let $Y = \left[y_1, y_2, \cdots, y_n \right]$

  where $y_i \sim N(0, R)$ is a $p-$dimensional random vector

- **Objective:** Estimate the eigenvalues and eigenvectors of R

$$R = E \Lambda E^t$$

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

- **Proposed Solution:** Model based estimation
  - Does not depend on ordering of vector or stationarity assumption

Data Model

- Notation:

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

- Likelihood of Y given R:

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

- ML estimate of eigenvectors and eigenvalues is given by

$$\hat{E} = \arg \min_{E \in \text{Prior Model}} \left\{ \left| diag(E^t S E) \right| \right\}$$

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

Prior Model: The Sparse Matrix Transform (SMT)

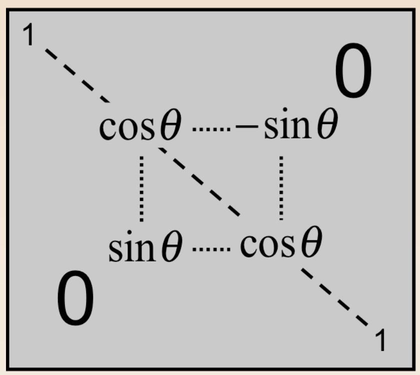
- **Big idea:**

  (Eigenvector matrix *E)* = (Sparse Matrix Transform)

- What is a Sparse Matrix Transform?

$$E = E_1 E_2 \cdots E_k$$

were E_k are Givens rotations



- Each Givens rotation operates on only two coordinates
  - Only 4 multiplies per rotation (really only 2)
  - When $k=p(p-1)/2$, this is any orthonormal transform

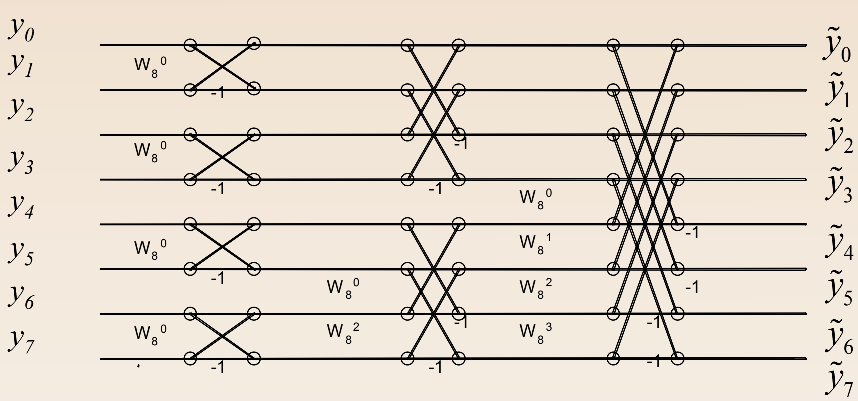
PURDUE UNIVERSITY

SMT is a Generalization of the FFT and Orthonormal Wavelet Transform

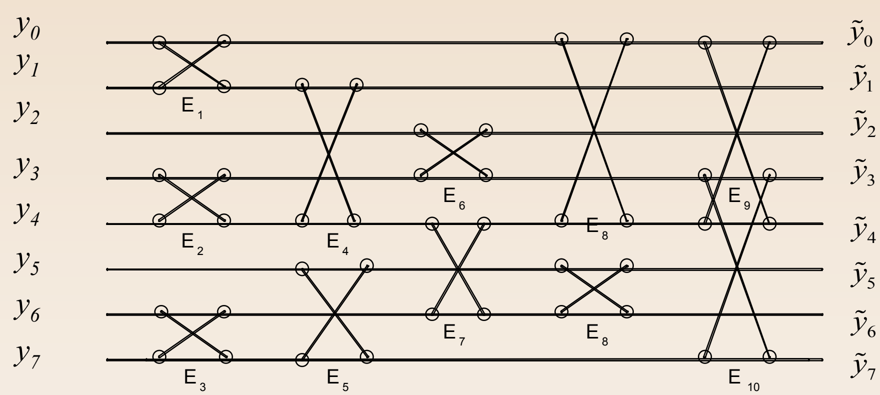
- SMT is product of Givens rotations:

$$E = E_1\, E_2 \cdots E_k \text{ where } E_k = \begin{bmatrix} 1 & & & & 0 \\ & \cos\theta & \cdots & -\sin\theta & \\ & \vdots & \ddots & \vdots & \\ & \sin\theta & \cdots & \cos\theta & \\ 0 & & & & 1 \end{bmatrix}$$

- So the SMT is a generalization of the FFT



FFT



SMT

- SMT is also a generalization of orthonormal (paraunitary) wavelets

Design of SMT using Cost Optimization

- ML estimate of Eigenvectors is

$$\hat{E} = \arg \min_{E \in \text{SMT of order } K} \left\{ \left| diag(E^t S E) \right| \right\}$$

where $E = E_1 E_2 \cdots E_K$ is an SMT transform.

- The algorithm:

For $k = 1$ to K {

Select most correlated coordinate pair

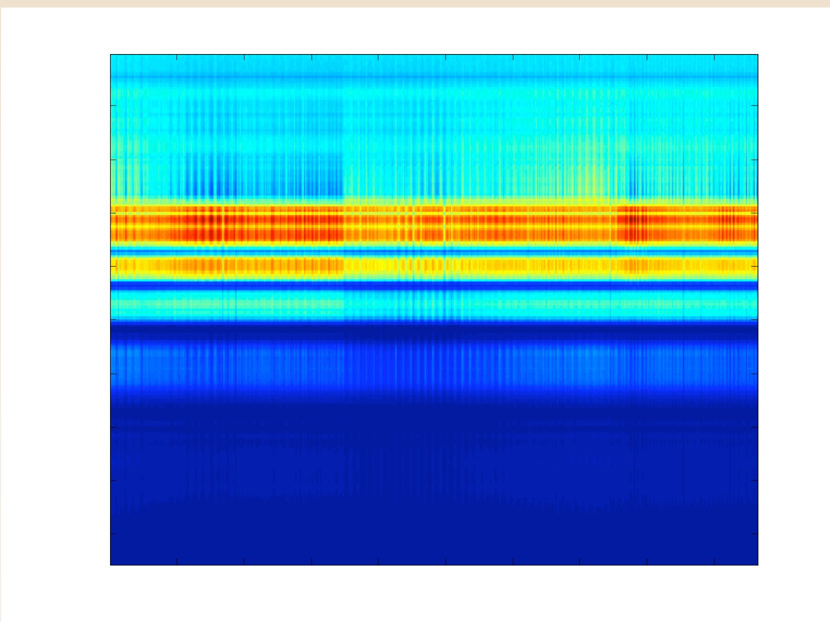
Decorrelate the coordinate pair with rotation E_k

}

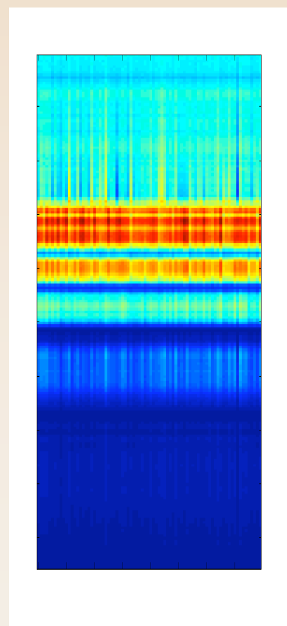
- K is choose to maximize cross-validated likelihood

Covariance Estimation for Hyperspectral Data

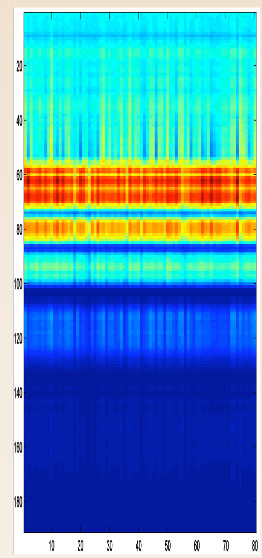
- # of hyperspectral bands: $p = 191$, # of samples: $n = 80$, grass class



Sample Data (1928)



Y (Gaussian)



Y (non-Gaussian)

Estimators Compared

- Shrinkage estimator

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

  - estimate α using cross-validation

- Graphic lasso (glasso) covariance estimator *:

$$\hat{R} = \arg\max_{R:\,P.D.} \left\{ \log(Y \mid R) - \rho \left\| R^{-1} \right\|_{L1} \right\}$$

  - L1 regularized ML estimate

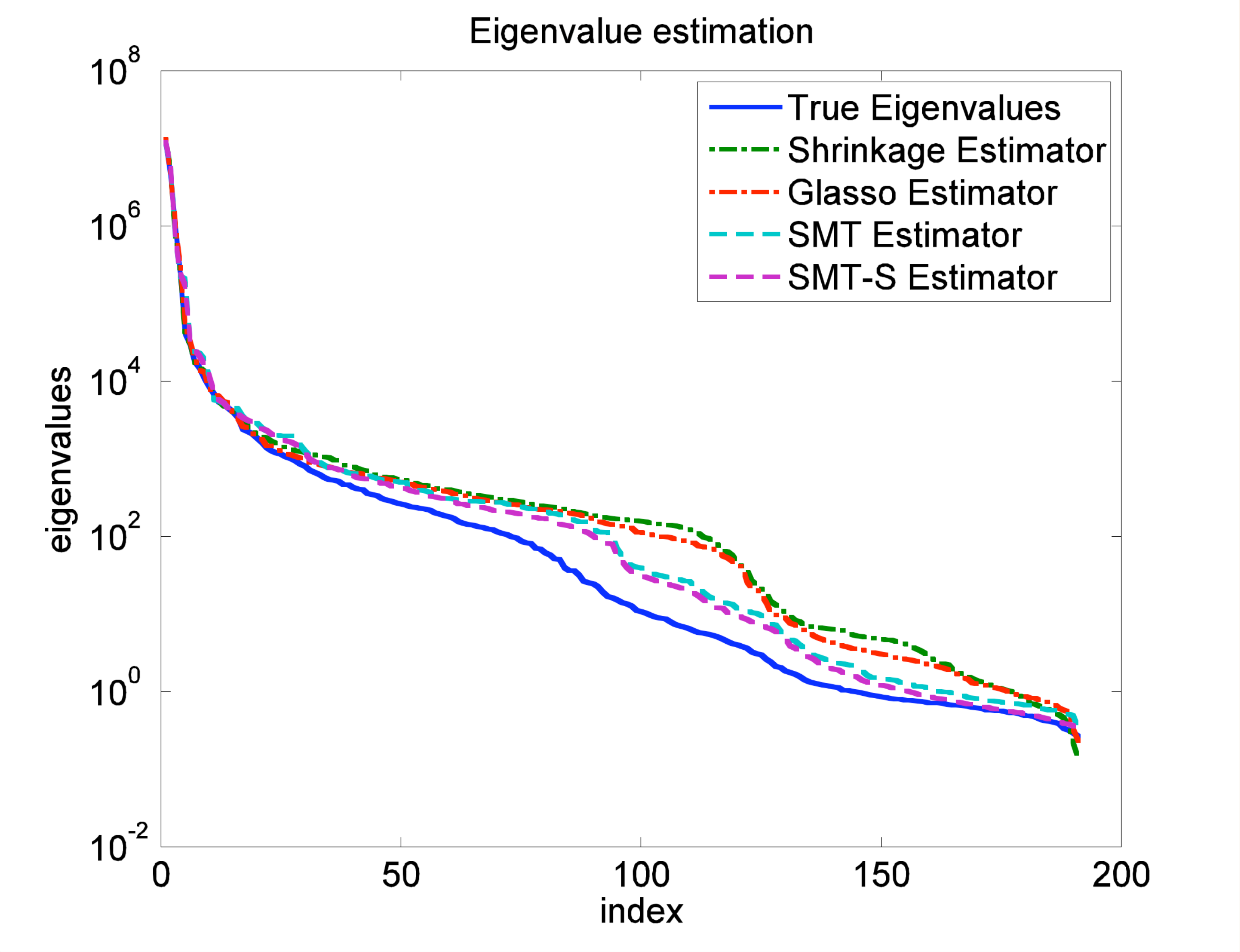
- SMT estimator
  - Use K^{th} order SMT model
  - Estimate k using cross-validation

- SMT-shrinkage (SMT-S) model
  - Combine SMT covariance estimate with shrinkage

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

* See *Sparse inverse covariance estimation with the graphic lasso*, Friedman et al. 2007

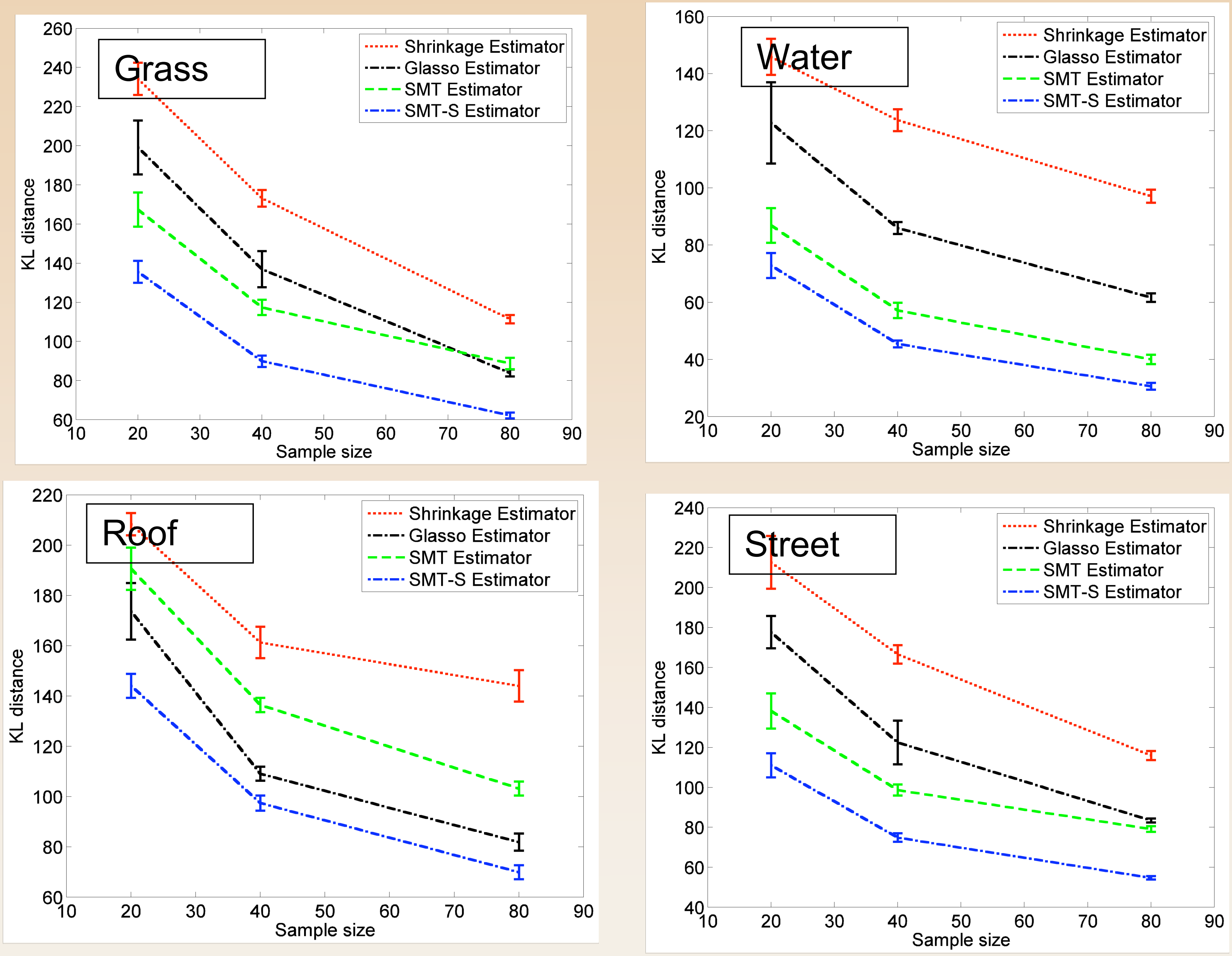
Eigenvalue and Eigenvector Estimation: Gaussian Case



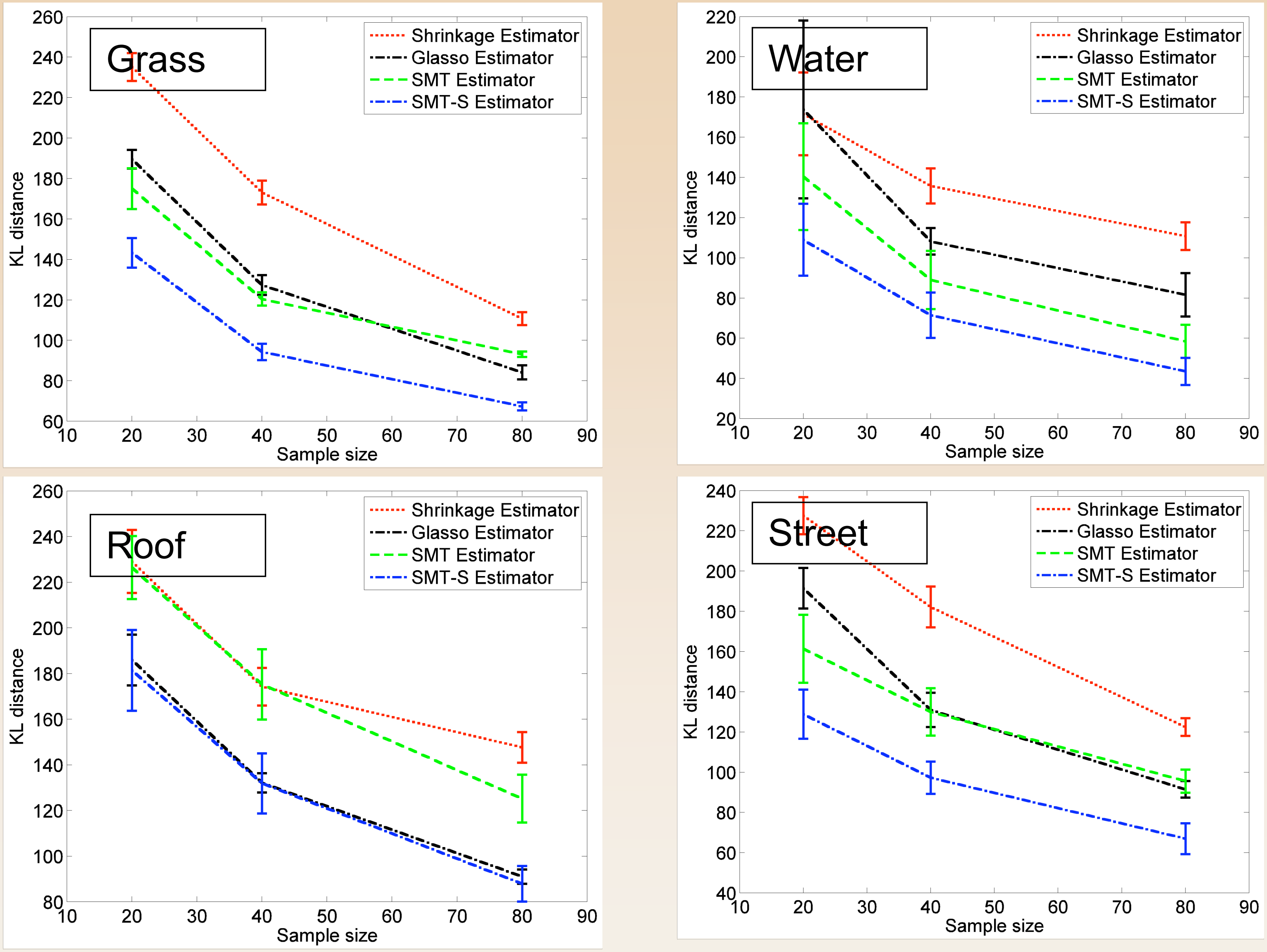
$$\hat{\Lambda} = diag(\hat{E}^t S \hat{E})$$

PURDUE UNIVERSITY

Results in Kullback-Leibler Distance: Gaussian Case



Results in Kullback-Leibler Distance*: non-Gaussian Case



* *KL distance of Gaussian distributions estimated from non-Gaussian samples*

Computational Complexity

	Complexity	CPU time (seconds)	Model order
Shrinkage	p^3	8.6	1
glasso	$p^3 I$	422.6	4939
SMT	Kp	6.5	495
SMT-S	$Kp + p^3$	7.2	496

- I-iterations required for for glasso, K-number of Givens rotations
- Numerical results are based on the Guassian grass case with n = 80

Signal Detection Based on Covariance Estimation

- Formulation of signal detection :

$$H_0 : r = x$$

$$H_1 : r = x + e \cdot t, \text{ with } e \text{ not zero}$$

where: x is a background with covariance R

t is a target signature

e is a scalar signal strength

r is the observed pixel

- The SNR of signal detection:

$$SNR = \frac{(q^t t)^2}{q^t R q}$$

where $q = \hat{R}^{-1} t$ is the linear matching filter used to test for signal

PURDUE UNIVERSITY

Results in Signal Detection

- Simulation:
  - R was chosen to be the true covariance of each class
  - $\hat{R}$ was estimated from the Gaussian sample (n = 80)
  - Dense case:

    t = rand(p,1) has the uniform distribution in [0, 1]
  - Sparse case:

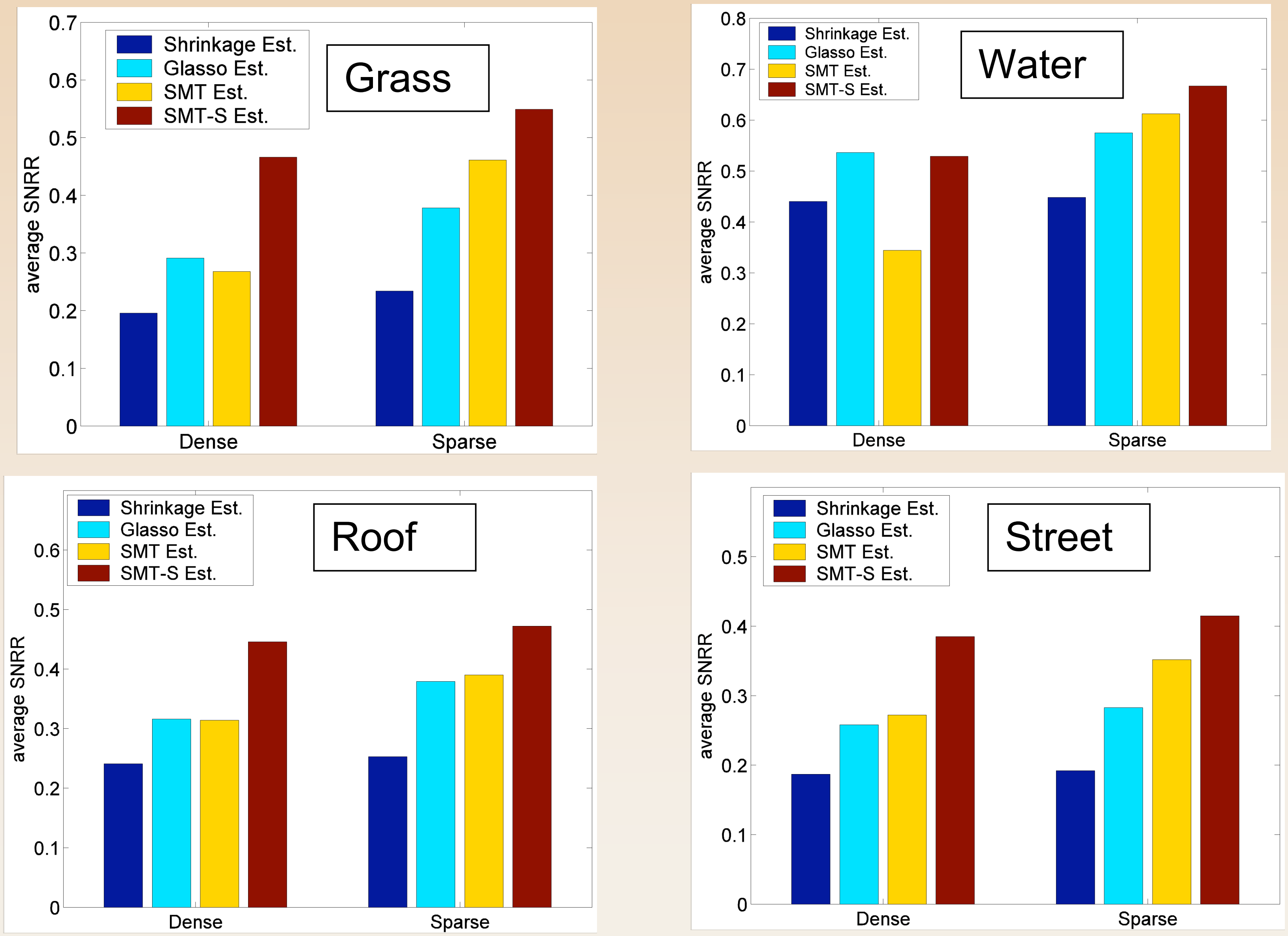
    setting all but the largest values (>0.9) of t to zero
  - Test was run 100 times, and the average SNRR was calculated:

$$SNRR = \frac{SNR(\hat{R})}{SNR(R)} = \frac{SNR \text{ using estmated } \hat{R}}{\text{opitmal } SNR}$$

Results in Signal Detection – Dense Case

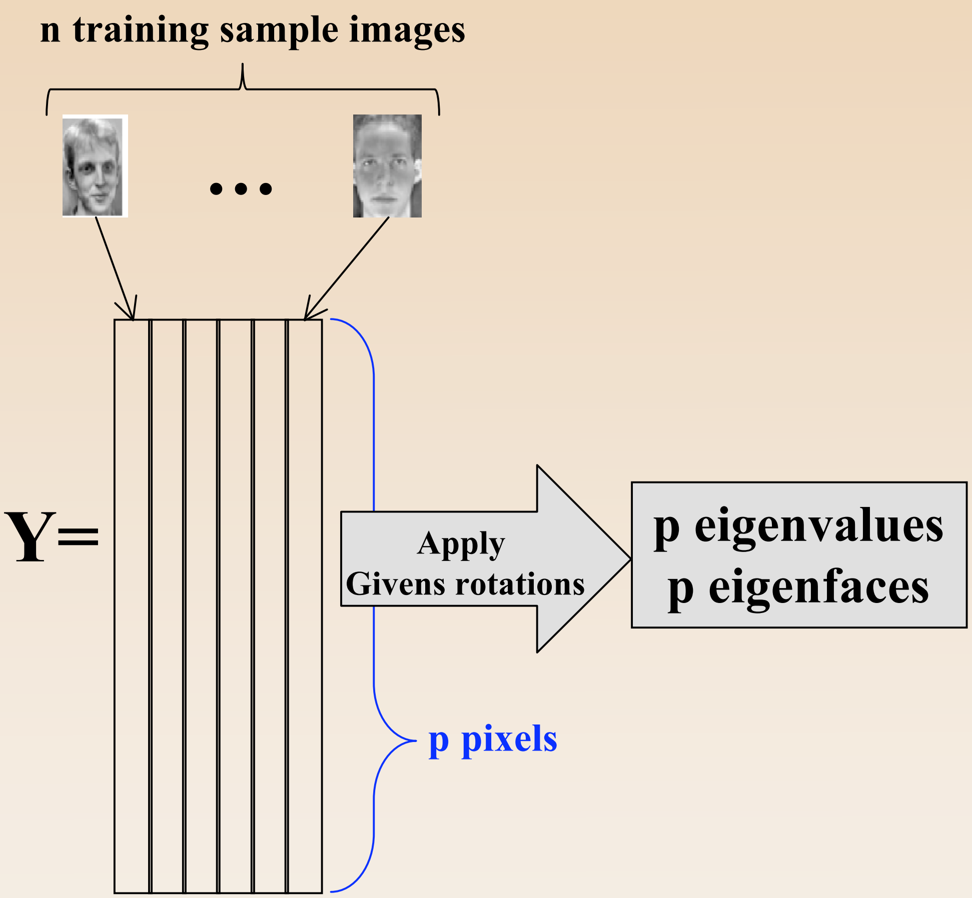
$$SNRR = \frac{SNR(\hat{R})}{SNR(R)} = \frac{SNR \text{ using estmated } \hat{R}}{\text{opitmal } SNR}$$

The higher, the better!

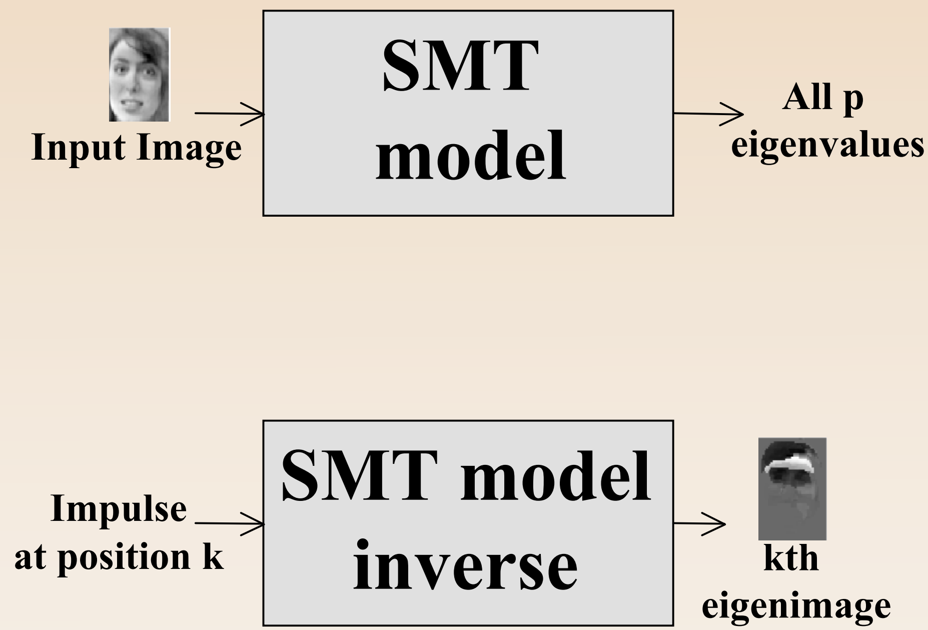


Computing Eigenimages with SMT

SMT Model Training



SMT Eigenimage Analysis

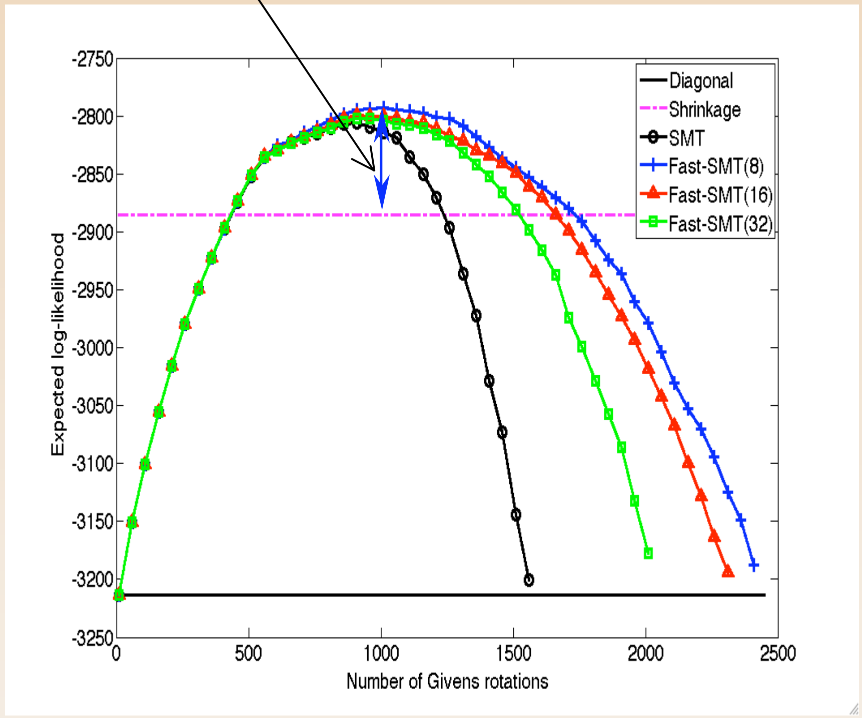


Comparison of Traditional and SMT Eigenimages

- Eigenimage experiment
  - Dataset: face image
  - Number of samples (n) = 40
  - Dimensions (p) = 644
- Use cross-validation to compute expected log likelihood

Method	Maximum log-likelihood	Δ	K_{max}
PCA+Shrinkage	-2885.71	0	-
Full SMT	-2805.81	79.88	910
Fast SMT(8)	**-2793.33**	**92.37**	**1010**
Fast SMT(16)	-2799.03	86.67	910
Fast SMT(32)	-2802.36	83.34	910
Diagonal	-3213.10	-327.40	-

Δlog likelihood= **92.37**



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

SMT versus Traditional Eigenfaces

Face dataset:

Traditional (PCA) eigenfaces:

Full SMT:

SMT(8pt neighborhood):