Covariance Operator Estimation for Random Variables Observed on a Lattice Graph

Ming Yuan

Department of Statistics

University of Wisconsin-Madison

myuan@stat.wisc.edu

http://www.stat.wisc.edu/~myuan

(Based on joint work with T.T. Cai)

Covariance Operator for Discrete Stochastic Process

- Discrete stochastic process $X : \mathcal{T} \mapsto \mathbb{R}$
- Covariance operator

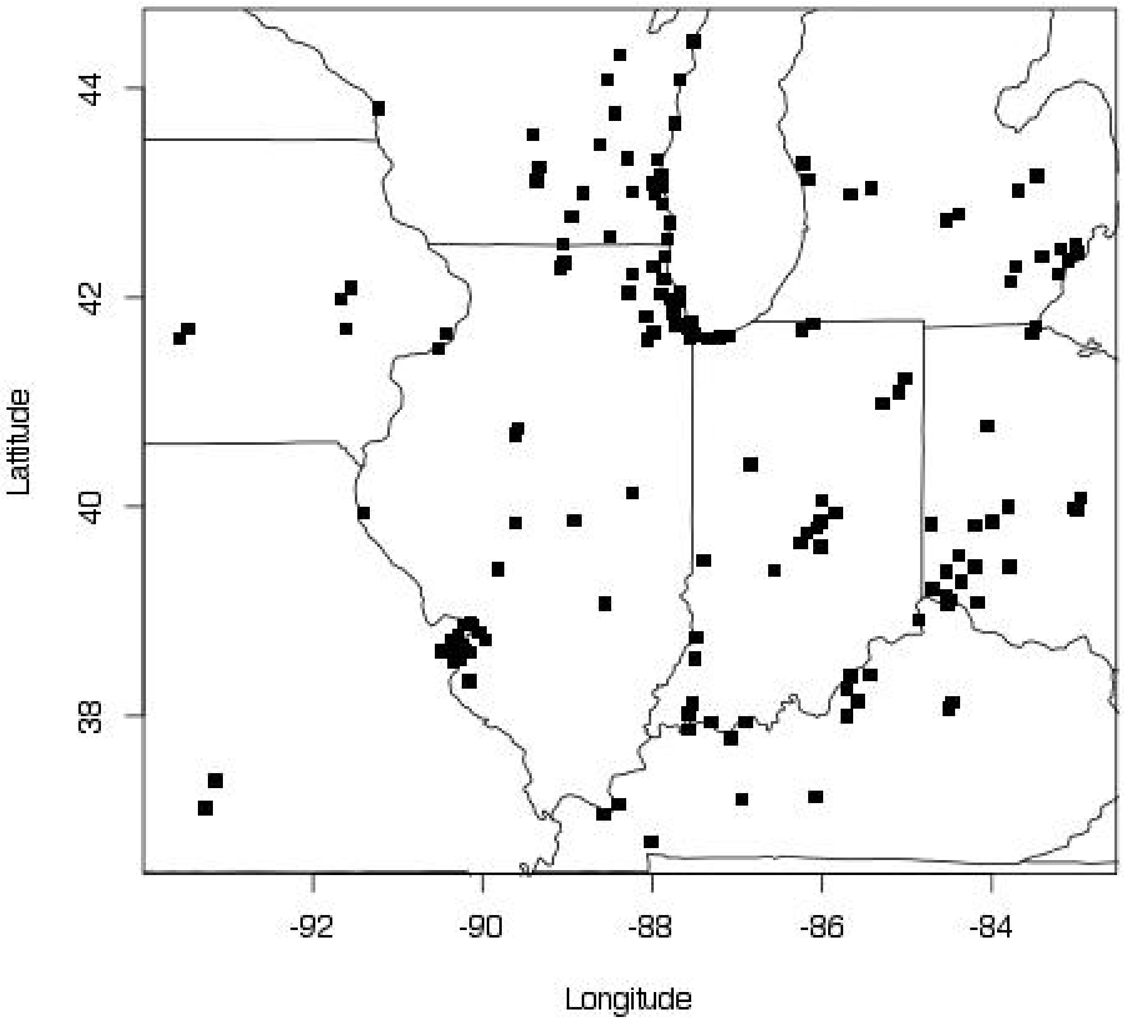
$$\Sigma = (\text{cov}(X(s), X(t)))_{s,t \in \mathcal{T}} : \ell_2(\mathcal{T}) \mapsto \ell_2(\mathcal{T})$$

- Estimating covariance operator

$$X_1, \ldots, X_n \mapsto \hat{\Sigma}$$

- Topological structure of domain – $(\mathcal{T}, D)$

SPATIAL STATISTICS

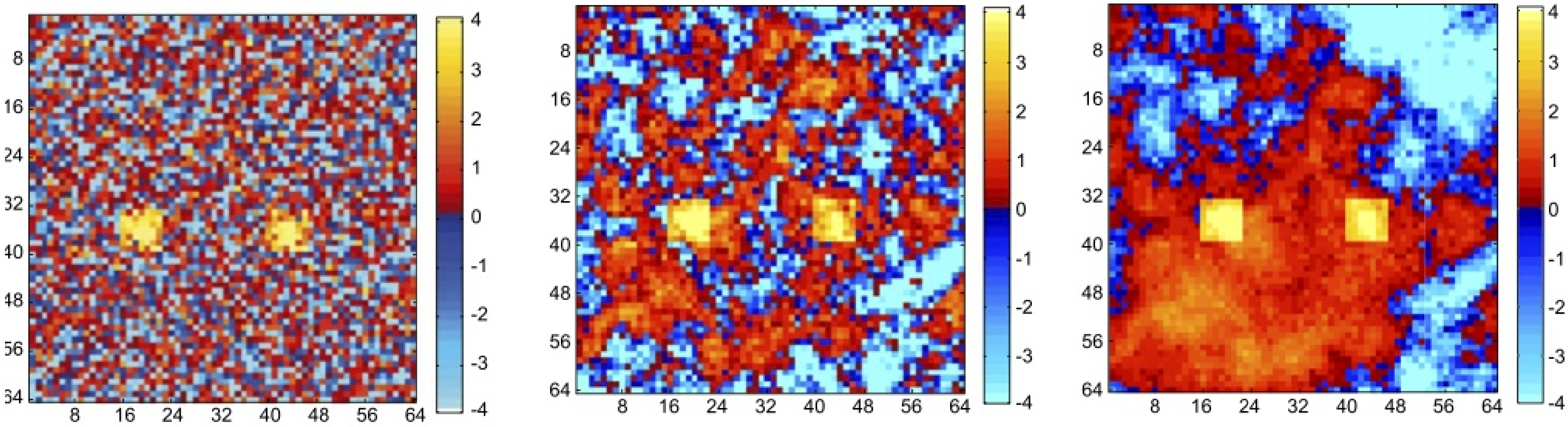


FACE RECOGNITION

[AT&T Laboratory]

$$\text{Face Images} : X_1, \ldots, X_n \Longrightarrow \Sigma \Longrightarrow \text{Eigenfaces}$$

EFFECT OF CORRELATION IN MULTIPLE TESTING

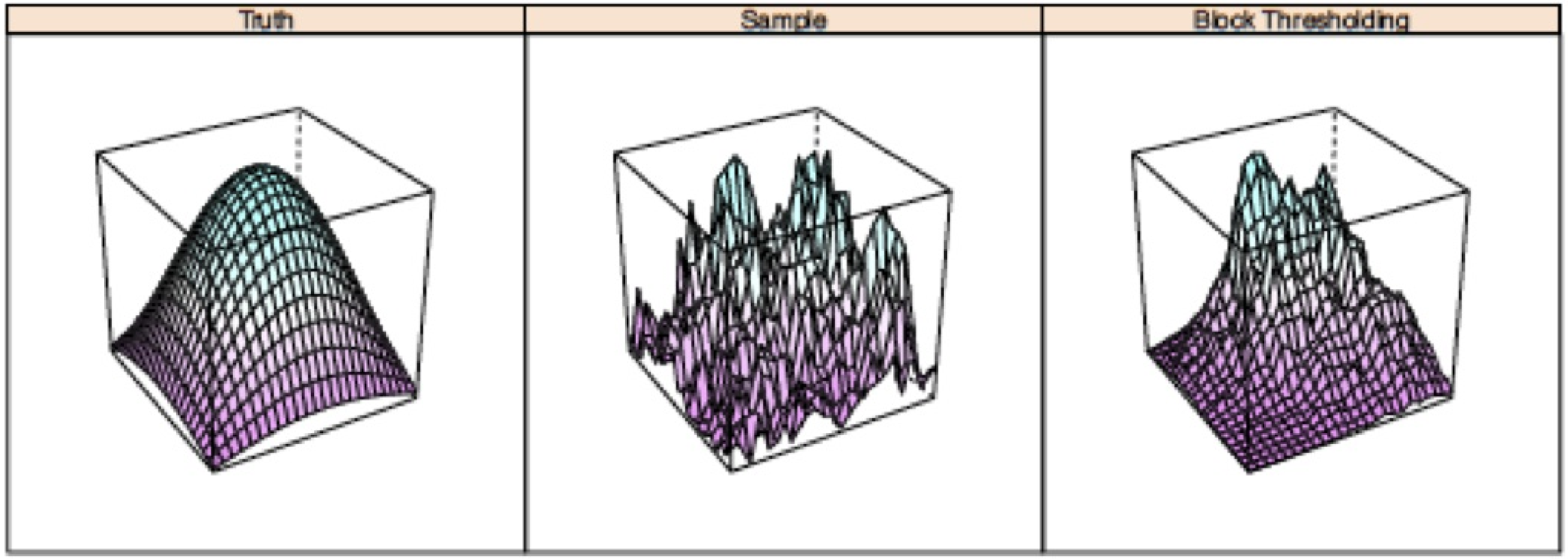


[Logan and Rowe, 2004]

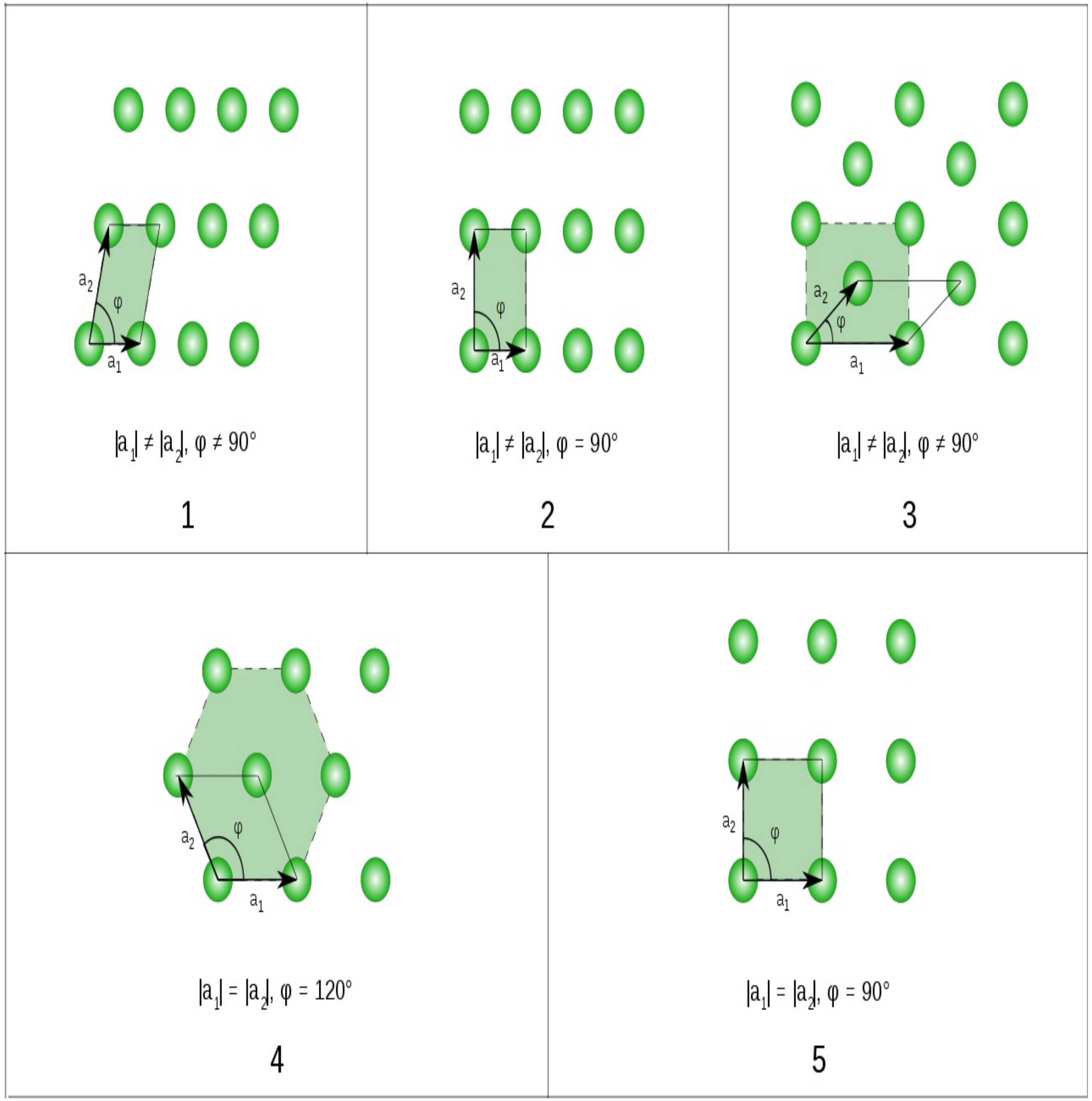
$$\text{Test Statistics}: \tilde{X} \Longrightarrow \Sigma \Longrightarrow \text{Region of interests}$$

Importance of Accounting for Lattice Structure

- Eigenimage for 2-d lattice
- Markov random field of order two
- $n = 400$ images of resolution 25×25



Covariance Operator for Lattice Graphs



- Lattice – $\mathcal{G}(q_1, \ldots, q_d) = \{1, 2, \ldots, q_1\} \times \cdots \times \{1, 2, \ldots, q_d\}$
- Stochastic process on the lattice

$$X = (X(t) : t \in \mathcal{G}_d)$$

- Covariance operator

$$\Sigma = (\mathrm{cov}(X(s), X(t)))_{s,t \in \mathcal{G}_d}$$

  - Operator

$$\Sigma : \ell_2(\mathcal{G}_d) \mapsto \ell_2(\mathcal{G}_d)$$

  - Operator norm – $\|\Sigma\|$
- Goal – $X_1, \ldots, X_n \to \Sigma$

Exponentially Decaying Covariance Operators

- Lattice – $\mathcal{G}(q_1, \ldots, q_d) = \{1, 2, \ldots, q_1\} \times \cdots \times \{1, 2, \ldots, q_d\}$
  - "Dimension" of lattice – d
  - Lengths – $q_1, \ldots, q_d$
  - Number of variables – $p = q_1 \times q_2 \times \ldots q_d$
- Distance on lattice – Taxicab distance/ℓ_1 distance
- Covariance decays exponentially with distance:

$$\mathcal{F}_d^*(\alpha_0, \alpha; M_0, M) = \left\{\Sigma : \ \Sigma \succ 0, \ \|\Sigma\| \le M_0, \sum_{s:D(s,t)\ge k} |\sigma(s,t)| \le M \exp(-\alpha_0 k^\alpha)\right\}.$$

  - Markov random field

How well can we estimate Σ

- Minimax optimal rate of convergence

$$\inf_{\tilde\Sigma(\text{data})} \sup_{\mathcal{F}_d^*(\alpha_0,\alpha;M_0,M)} \mathbb{E}\|\tilde\Sigma - \Sigma\|^2 \asymp \frac{\log p}{n} + \frac{1}{n}\prod_{k=1}^{d}(\min\{q_k, (\log n)^{1/\alpha}\}).$$

- In particular – "squared" lattice $q_1 = \ldots = q_d = q$

$$\inf_{\tilde\Sigma(\text{data})} \sup_{\Sigma\in\mathcal{F}_d^*(\alpha_0,\alpha;M_0,M)} \mathbb{E}\|\tilde\Sigma - \Sigma\|^2 \asymp \min\left\{\frac{(\log n)^{d/\alpha}}{n} + \frac{\log p}{n}, \frac{p}{n}\right\}.$$

  - If $\log p \ll (\log n)^{d/\alpha}$

$$\inf_{\tilde\Sigma(\text{data})} \sup_{\Sigma\in\mathcal{F}_d^*(\alpha_0,\alpha;M_0,M)} \mathbb{E}\|\tilde\Sigma - \Sigma\|^2 \asymp \min\left\{\frac{(\log n)^{d/\alpha}}{n}, \frac{p}{n}\right\}.$$

  - If $\log p \gg (\log n)^{d/\alpha}$

$$\inf_{\tilde\Sigma(\text{data})} \sup_{\Sigma\in\mathcal{F}_1^*(\alpha_0,\alpha;M_0,M)} \mathbb{E}\|\tilde\Sigma - \Sigma\|^2 \asymp \frac{\log p}{n}.$$

POLYNOMIALLY DECAYING COVARIANCE OPERATORS

$$\mathcal{F}_d^*(\alpha; M_0, M) = \left\{ \Sigma : \ \Sigma \succ 0, \ \|\Sigma\| \le M_0, \sum_{s:D(s,t)\ge k} |\sigma(s,t)| \le a_k = Mk^{-\alpha} \right\}.$$

- Minimax optimal rate of convergence

$$\inf_{\tilde\Sigma(\text{data})} \sup_{\mathcal{F}_d(\alpha;M_0,M)} \mathbb{E}\|\tilde\Sigma - \Sigma\|^2 \asymp \frac{\log p}{n} + \min\left\{ \left(n^{-1}\prod_{l=0}^{k} q_{[l]}\right)^{\frac{2\alpha}{2\alpha+d-k}} : 0 \le k \le d \right\}.$$

- In particular – 1-d lattice or path graph (see also Cai, Zhang and Zhou, 2010)

$$\inf_{\tilde\Sigma(\text{data})} \sup_{\Sigma\in\mathcal{F}_1(\alpha;M_0,M)} \mathbb{E}\|\tilde\Sigma - \Sigma\|^2 \asymp \min\left\{ n^{-\frac{2\alpha}{2\alpha+1}} + \frac{\log p}{n}, \frac{p}{n} \right\}.$$

- Squared Lattices

$$\inf_{\tilde{\Sigma}(\text{data})} \sup_{\Sigma\in\mathcal{F}_d(\alpha;M_0,M)} \mathbb{E}\|\tilde{\Sigma}-\Sigma\|^2 \asymp \min\left\{n^{-\frac{2\alpha}{2\alpha+d}} + \frac{\log p}{n},\ \frac{p}{n}\right\}$$

  - Role of p – mostly as $\log p$ factor
  - Role of d – mostly in exponent: curse of dimensionality

- 2-d Rectangular Lattice

$$\inf_{\tilde{\Sigma}(\text{data})} \sup_{\Sigma\in\mathcal{F}_2(\alpha;M_0,M)} \mathbb{E}\|\tilde{\Sigma}-\Sigma\|^2 \asymp \frac{\log p}{n} + \min\left\{n^{-\frac{\alpha}{\alpha+1}},\ \left(\frac{q_{[1]}}{n}\right)^{\frac{2\alpha}{2\alpha+1}},\ \frac{p}{n}\right\}.$$

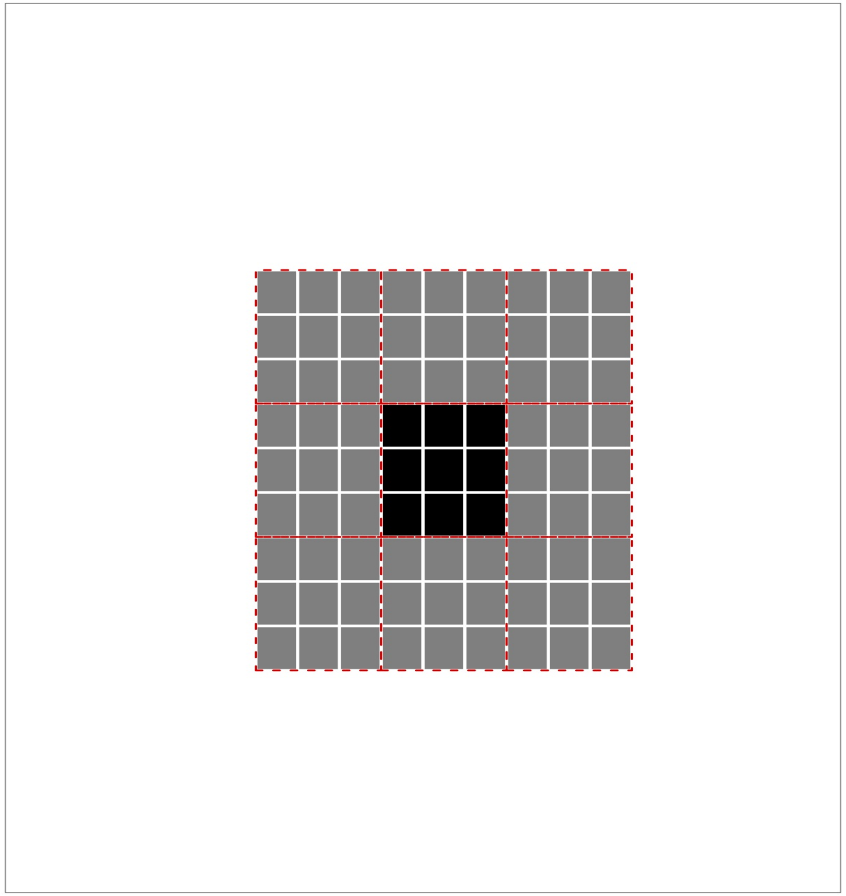
  - If $q_{[2]} \gg (n/q_{[1]})^{1/(2\alpha+1)}$

$$\inf_{\tilde{\Sigma}(\text{data})} \sup_{\Sigma\in\mathcal{F}_2(\alpha;M_0,M)} \mathbb{E}\|\tilde{\Sigma}-\Sigma\|^2 \asymp \frac{\log p}{n} + \min\left\{n^{-\frac{\alpha}{\alpha+1}},\ \left(\frac{q_{[1]}}{n}\right)^{\frac{2\alpha}{2\alpha+1}}\right\}.$$

  - If $q_{[1]} \gg n^{1/(2\alpha+2)}$

$$\inf_{\tilde{\Sigma}(\text{data})} \sup_{\Sigma\in\mathcal{F}_2(\alpha;M_0,M)} \mathbb{E}\|\tilde{\Sigma}-\Sigma\|^2 \asymp \frac{\log p}{n} + n^{-\frac{\alpha}{\alpha+1}}$$

Blockwise Banding Estimator



- Dividing the lattice $\mathcal{G}_d$ into blocks

$$I_j^{(l)} = \{(j-1)b+1, (j-1)b+2, \dots, jb\}$$

- Define a "block"

$$B_{\mathbf{j}} = I_{j_1}^{(1)} \times I_{j_2}^{(2)} \times \dots \times I_{j_d}^{(d)},$$

- Block of an operator

$$A_{\mathbf{j}\mathbf{j}'} = (A(s,t))_{s\in B_{\mathbf{j}}, t\in B_{\mathbf{j}'}}$$

- Blockwise banding

$$\hat{\Sigma}_{\mathbf{j}\mathbf{j}'} = \begin{cases} S_{\mathbf{j}\mathbf{j}'} & \text{if } \|\mathbf{j}-\mathbf{j}'\|_\infty \le 1 \\ 0 & \text{otherwise} \end{cases}.$$

OPTIMALITY

- Polynomially decaying covariance operators
  - Block size

$$b \sim \min_k \left\{ \left(n^{-1} \prod_{l=0}^{k} q_l \right)^{\frac{2\alpha}{2\alpha+d-k}} : 0 \le k \le d \right\}.$$

  - Rate of convergence

$$\sup_{\Sigma \in \mathcal{F}_d(\alpha; M_0, M)} \mathbb{E}\|\hat{\Sigma} - \Sigma\|^2 \lesssim \frac{\log p}{n} + \min \left\{ \left(n^{-1} \prod_{l=0}^{k} q_l \right)^{\frac{2\alpha}{2\alpha+d-k}} : 0 \le k \le d \right\}$$

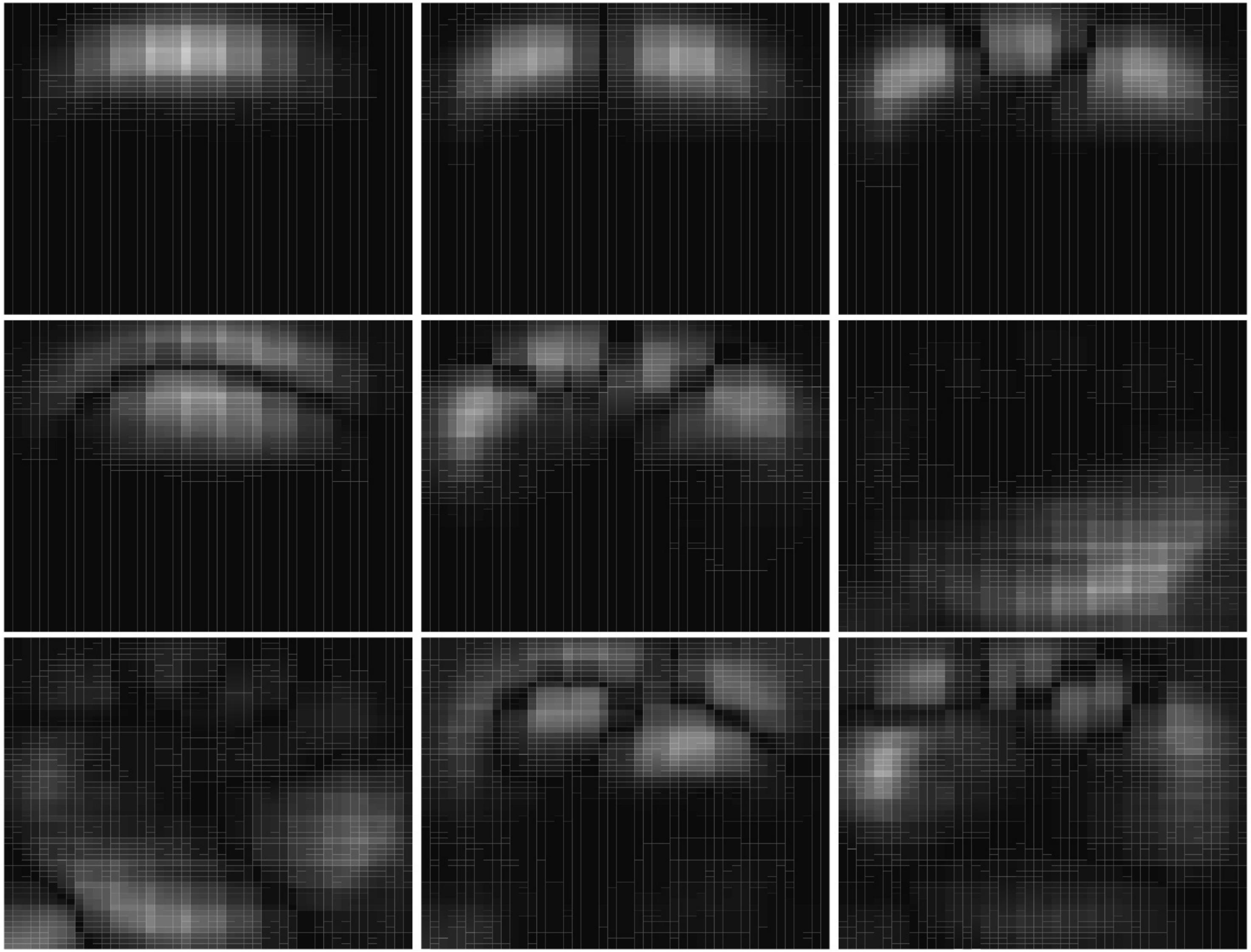
- Exponentially decaying covariance operators
  - Block size

$$b \sim \text{ solution of } \log n + d \log x = 2\alpha_0 x^{\alpha}$$

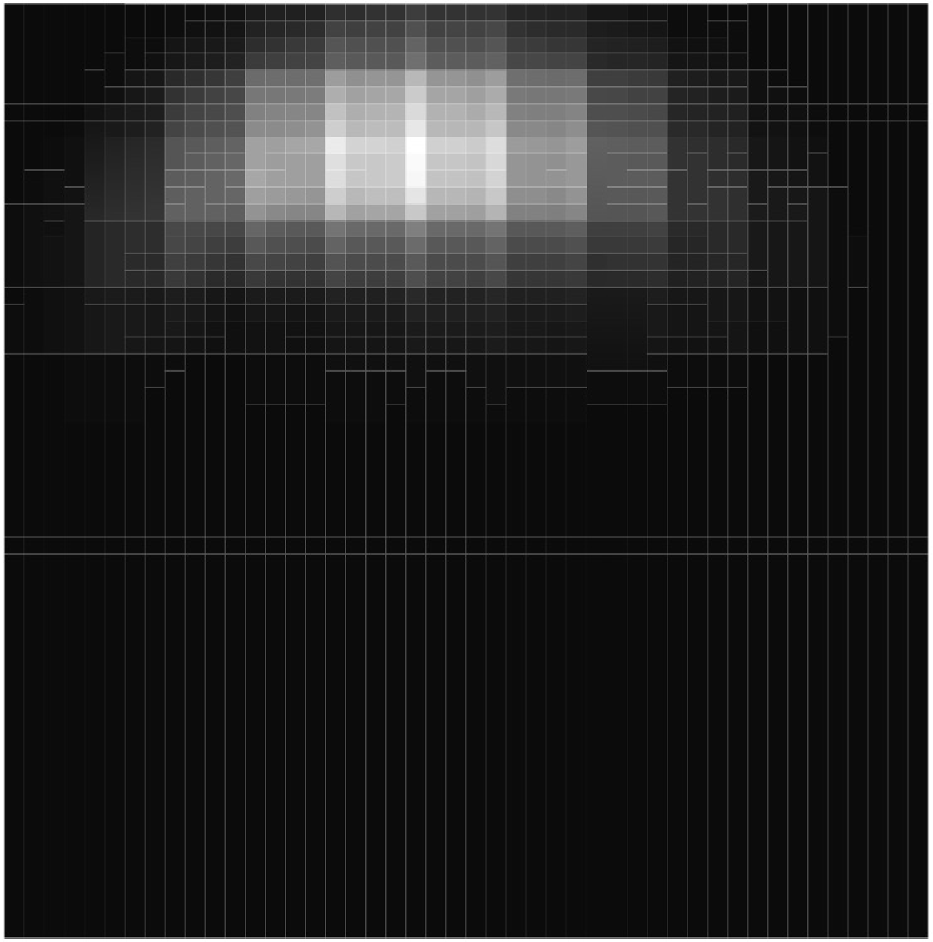
  - Rate of convergence

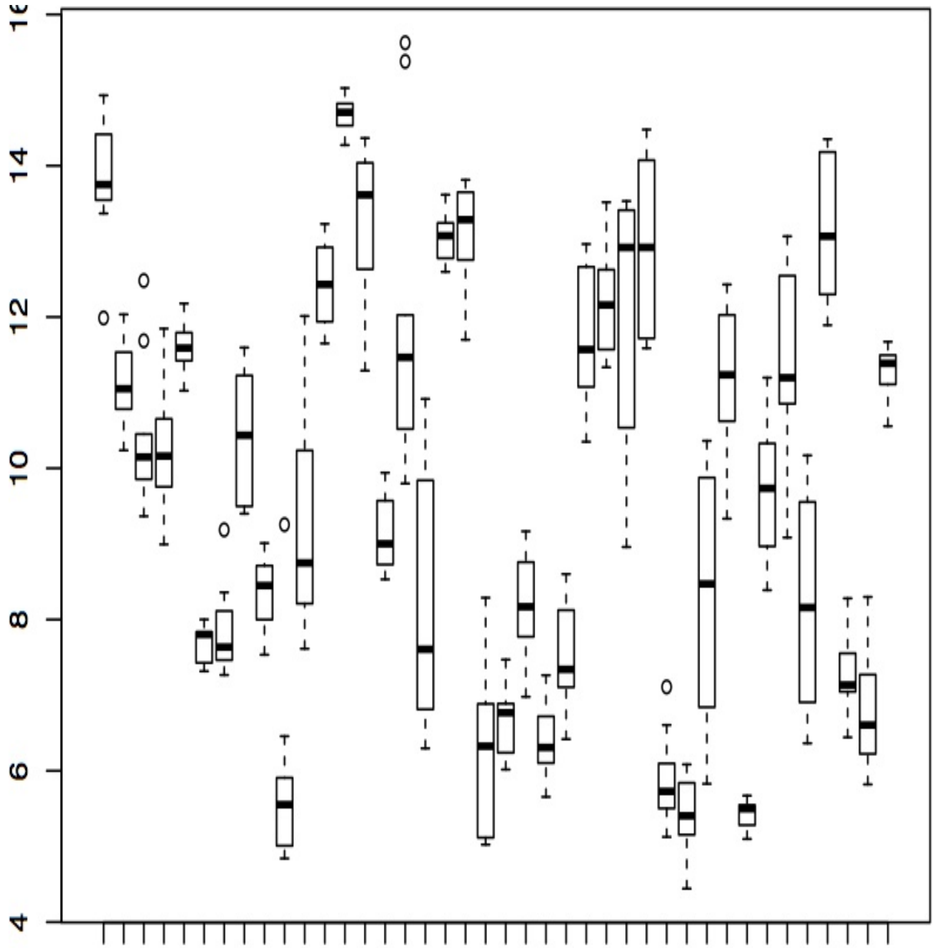
$$\sup_{\Sigma \in \mathcal{F}_d^*(\alpha_0, \alpha; M_0, M)} \mathbb{E}\|\hat{\Sigma} - \Sigma\|^2 \lesssim \frac{\log p}{n} + \frac{1}{n} \prod_{k=1}^{d} \left(\min\{q_k, (\log n)^{1/\alpha}\} \right)$$

EIGENFACES



Face Recognition





MORE GENERALLY...

- General decaying rate – $\sum_{s:D(s,t)\geq k} |\sigma(s,t)| \leq a_k$ where $a_k \downarrow 0$
  - Key quantity
$$k(q) = \min\{1 \leq k \leq q : a_k \leq n^{-1/2} k^{d/2-1}\}$$
  - Optimal rate of convergence
$$\inf_{\tilde{\Sigma}(\text{data})} \sup_{\Sigma \in \mathcal{F}_d(\{a_k\};M)} \mathbb{E}\|\tilde{\Sigma} - \Sigma\|^2 \asymp \frac{[k(q)]^d + \log p}{n}.$$
- General domains – rate of convergence depends on the "volume" of balls
$$N(r) = \max_{t \in \mathcal{T}} \text{card}\{s \in \mathcal{T} : D(s,t) \leq r\},$$
  - Large volume $\Longrightarrow$ slower rate
  - Small volume $\Longrightarrow$ faster rate