Banded Approximation to Inverses of Bandedly Approximable Covariance Matrices with Applications

Hartigan Workshop 05/2009

Peter Bickel

(Joint work with Marko Lindner)

May 2009

(With assistance of Jing Lei)

Outline

1. Statistical motivation for a correct conjecture.
2. Banded matrix approximations to p.d. Hermitian matrices and their inverses.
3. General results and a theorem of Wiener.
4. Regularity and strong mixing of non-stationary Gaussian processes.
5. Applications to statistics.

Statistical Motivation

- Given $\mathbf{X}_1, \ldots, \mathbf{X}_n$: $p \times 1$, i.i.d. $\mathbb{E}\mathbf{X}_1 = 0$, $Var\mathbf{X}_1 = \Sigma$.
- $\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i^T$.
- For any matrix $M = ||m_{ij}||$,

$$B_k(M) \equiv ||m_{ij} 1(|i - j| \leq k)||.$$

- $||M|| = \max^{\frac{1}{2}}\{\mathbf{X}^T M^T M \mathbf{X} : |\mathbf{X}| = 1\} \equiv$ Operator norm.

Basic Property of $||\cdot||$

- $||AB|| \leq ||A||||B||$.
- Given A_n, B_n symmetric $||A_n - B_n|| \to 0$,

  suppose $\lambda_1(B_n) > \lambda_2(B_n) > \cdots > \lambda_k(B_n) > \lambda_{k+1}(B_n)$,

  and define $\lambda_j(A_n)$ analogously.

  Suppose $\lambda_{j+1}(B_n) < \lambda_j(B_n) - \triangle,\ 1 \leq j \leq k$.

  Dimension B_n arbitrary, $k, \triangle > 0$ fixed.

  Then,

  a) $|\lambda_j(A_n) - \lambda_j(B_n)| = O(\triangle^{-(j-1)}||A_n - B_n||)$;

  b) If E_{jA} respectively E_{jB} is projection operator onto eigenspace corresponding to λ_j, then

$$||E_{jA} - E_{jB}|| = O(\triangle^{-j}||A_n - B_n||).$$

B-L (2006) Main Result I

Banded estimator :

$$\hat{\Sigma}_{k,p}(i,j) = \hat{\Sigma}_p(i,j) \cdot \mathbf{1}(|i-j| \leq k).$$

Let

$$\begin{aligned} \mathcal{U}(\epsilon_0, \alpha, C) &= \{\Sigma : 0 < \varepsilon_0 \leq \lambda_{\min}(\Sigma) \leq \lambda_{\max}(\Sigma) \leq 1/\varepsilon_0, \\ &\quad \max_j \sum_i \{|\sigma_{ij}| : |i-j| > k\} \leq Ck^{-\alpha} \ \text{ for all } k \geq 0\}. \end{aligned}$$

Theorem 1

If $\mathbf{X}$ is Gaussian and $k_n \asymp (n^{-1} \log p)^{-\frac{1}{2(\alpha+1)}}$, then, uniformly on $\Sigma \in \mathcal{U}(\varepsilon_0, \alpha, C)$,

$$\|\hat{\Sigma}_{k_n,p} - \Sigma_p\| = O_P\left((n^{-1} \log p)^{\frac{\alpha}{2(\alpha+1)}}\right) = \|\hat{\Sigma}_{k_n,p}^{-1} - \Sigma_p^{-1}\|.$$

The banded estimator and its inverse are consistent if $\frac{\log p}{n} \to 0$.

Remark

- $\lambda_{min} \geq \varepsilon_0$ not needed if only convergence in $||\cdot||$ to Σ is needed.
- Cai, Zhang and Zhou (2008, *Ann. Stat.* submission) show that rates can be improved if banding is replaced by tapering.

A conjecture

Theorem 1 holds equally well if $\mathcal{U}$ is replaced by

$$\mathcal{U}^{-1} \equiv \Big\{ \Sigma : 0 \le \varepsilon_0 \le \lambda_{\min}(\Sigma) \le \lambda_{\max}(\Sigma) \le \varepsilon_0^{-1}, \max_j \Big\{ \sum_j |\sigma^{ij}| : |i-j| \ge k \Big\} \le Ck^{-\alpha}, \forall k \Big\},$$

where

$$\Sigma^{-1} = ||\sigma^{ij}||.$$

Conjecture

Approximation of Σ by banded matrices and approximation of Σ^{-1} by banded matrices are equivalent.

Operator formulation

- Identify linear operator $T : \ell_2 \to \ell_2$ by

$$T \equiv ||t_{ij}||_{i \geq 1, j \geq 1},$$
$$Tx = \left\{ \sum_j t_{ij} x_j : i \geq 1 \right\}, \quad x \in \ell_2.$$

- <u>Assume</u>: T is bounded in operator norm.

  $BO_k \equiv \{ T : t_{ij} = 0, \forall |i - j| > k \}$, (banded operator of order k);

  $BO = \bigcup_k BO_k$;

  $BDO \equiv$ Closure of BO in operator norm.

Conjecture Resolved

Theorem

(Since BDO is a $\mathcal{C}^*$ algebra) BDO is closed under inversion. If $A \in BDO$ and A^{-1} is bounded then $A^{-1} \in BDO$.

(M. Lindner: Infinite Matrices and Their Finite Sections, *Frontiers in Mathematics*, Birkhäuser (2006))

Approximation Theorems (B and Lindner)

Theorem 1

Let $A \in BO_k$ be Hermitian positive definite, $m^{-1} = ||A^{-1}||$, $M = ||A||$. Then, for all n

$$\mathrm{dist}\big(A^{-1}, BO_{nk}\big) \leq \frac{1}{m}\Big(\frac{M-m}{M+m}\Big)^{n+1} = \frac{1}{m}\Big(\frac{K-1}{K+1}\Big)^{n+1},$$

where

$$K \equiv \frac{M}{m},$$

$$\mathrm{dist}(M, L) \equiv \inf\big\{||M - M'|| : M' \in L\big\}.$$

Theorem 2

For arbitrary $A \in BDO$, let $||A||^{-1} = m^{-1} < \infty$.

If $\delta_k \equiv \mathrm{dist}(A, BO_k)$, $\delta_k < \frac{m}{2}$, then

$$\mathrm{dist}\left(A^{-1}, BO_{3nk}\right) \leq \frac{2\delta_k \alpha_k}{m^2} + \frac{\left(K_k^+\right)^2}{M + 2\delta_k} \left(\frac{\left(K_k^+\right)^2 - 1}{\left(K_k^-\right)^2 + 1} \right)^{n+1},$$

where $\alpha_k \equiv \frac{m}{m - 2\delta_k}$, $K_k^{\pm} = \alpha_k \left(K \pm \frac{2\delta_k}{m} \right)$.

Remark

Theorem 2 $\Rightarrow$ conjecture theorem.

Idea of Proof

i) If A is Hermitian, p.d., $\gamma = \frac{2}{M+m}$, then $||\gamma A - I|| < 1$;

ii) $A^{-1} = \gamma[\gamma A]^{-1} = \gamma \sum_{k=0}^{\infty}(I - \gamma A)^k$;

iii) $A^{-1} = [A^T A]^{-1} A^T$.

AND

- BDO is closed under $A \to A^T$, $(A, B) \to AB$, $(A, B) \to A + B$.

Similar Results for Wiener Norm

$||M||_W \equiv \sum_{k=1}^{\infty} |m_k|, \quad m_k \equiv \max\{|m_{ij}| : |i-j| = k\}.$

Theorem (Lindner, 2006)

If W is the closure of BO in Wiener norm, then W is closed under inversion.

This generalizes:

Theorem (N. Wiener)

If $f = \sum_{j=-\infty}^{\infty} f_j e^{2\pi i j\omega}, \quad \sum |f_j| < \infty, \quad f \neq 0,$

then $\frac{1}{f} = \sum_{j=-\infty}^{\infty} d_j e^{2\pi i j\omega}, \quad \sum |d_j| < \infty.$

Connection with Probability Theory

If $\Sigma : \ell_2 \to \ell_2$ symmetric p.d., $\exists X_1, X_2, \ldots, X_n, \ldots$ Gaussian

$\mathbb{E}X_j = 0$, $j \geq 1$, $\Sigma = ||cov(X_i, X_j)||$.

Definition

X is regular iff

$$\bigcap_j \mathcal{B}(X_{j+1}, X_{j+2}, \ldots) = \{\emptyset, \Omega\}.$$

Theorem (Ibragimov and Linnik)

Suppose $\mathbf{X}$ is stationary: $\Sigma = ||\rho(i-j)||$, ρ symmetric.

$f(t) = \sum_k \rho(k) e^{2\pi ikt}$. Suppose $|f|_\infty \leq M < \infty$,

$\left|\frac{1}{f}\right|_\infty \leq m^{-1} < \infty$, then $\mathbf{X}$ (Gaussian) is regular.

$|g|_\infty \equiv \operatorname{ess\,sup}_{|t|\leq 1} |g(t)|$.

Connection with Operator Theory, Wiener

If $\Sigma = ||\rho(i-j)||$ is an operator from ℓ_2 to ℓ_2;

let $f(t) = \sum_k \rho(k) e^{2\pi ikt}$, then

$||\Sigma|| = |f|_\infty$.

Generalization to Non-stationary $\mathbf{X}$

Theorem 3 (B-Lindner) (Generalization of I-L)

For Gaussian $\mathbf{X}$, if $||\Sigma|| = M < \infty$, $||\Sigma^{-1}|| = m^{-1} < \infty$, and $\Sigma \in BDO$, then $\mathbf{X}$ is regular.

Probabilistic Context and Idea of Proof

Regularity $\equiv$ Linear regularity

$$\equiv \mathbb{E}(X_{i+p}|X_1, \ldots, X_i) \xrightarrow{L_2} 0, \quad \text{for all fixed } i, \quad \text{as } p \to \infty.$$

Strong Mixing

- *Mixing coefficient* $\beta(p)$,

  $\beta(p) \equiv \sup\{|\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{B}(X_1, \ldots, X_i), B \in \mathcal{B}(X_{i+p}, X_{i+p+1}, \ldots), \text{ all } i\}.$

- $\mathbf{X}$ strong mixing iff

  $\beta(p) \to 0$ as $p \to \infty$.

Consequences of Strong Mixing

LLN and Central Limit Theorems for $\sum_{i=1}^n g(X_i)$, and more.

Theorem (Kolmogorov/Rozanov)

If $\mathbf{X}$ Gaussian stationary has $|f|_\infty < \infty$, $|\frac{1}{f}|_\infty < \infty$ <u>and</u> f is continuous, then $\mathbf{X}$ is strong mixing.

Partial Generalization to Non-stationary $\mathbf{X}$

Theorem 4 (B-Lindner)

For Gaussian $\mathbf{X}$, suppose $\Sigma \equiv ||\rho_{ij}||$, $||\Sigma|| = M < \infty$, $||\Sigma^{-1}|| = m^{-1} < \infty$. Then, $\mathbf{X}$ is strong mixing if

$$\sup\left\{\sum_{i=1}^{\infty}\sum_{j=k+p}^{\infty}\rho_{ij}^2 : k \geq 1\right\} \rightarrow 0,$$

as $p \rightarrow \infty$.

Non-stationary $\mathbf{X}$

- A sufficient condition: $\sum_{i=1}^{\infty} \sum_{j=i+k}^{\infty} \rho_{ij}^2 < \infty$ for some k.
- Stationary $\mathbf{X}$: Sufficient condition equivalent to

  $\sum_{k=1}^{\infty} k\rho^2(k) < \infty$.

  (1/2 a derivative)

Idea of Proof

- Compute

  $\mathcal{A}(f_m, g_{m+p,k}) \equiv$ Hellinger affinity, where

  $\mathcal{A}(f, g) = \int \sqrt{fg}\, d\mu$, if f, g are densities with respect to μ.

  $f_m \equiv$ joint density of $\mathbf{X}_1, \ldots, \mathbf{X}_m$;

  $g_{m+p,k} =$ joint density of $\mathbf{X}_{m+p+1}, \ldots, \mathbf{X}_{m+p+k}$;

- Show

$$\sup \left\{ |\mathcal{A}(f_m, g_{m+p,k}) - 1| : m, k \right\} \to 0.$$

Generalizations of Banding

Let ρ be a metric on $I = \{1, \ldots, p\}$. E.g.:

i) $\rho(i, j) = |i - j|$.

ii) $I = \tau(S)$, where

τ is a $1 - 1$ map,

$S \subset \underbrace{J \times \ldots \times J}_{r}$, and $J = \{1, \ldots, k\}$ for some k,

$\rho(i, j) = |\tau^{-1}(i) - \tau^{-1}(j)|$,

$|\cdot|$ is the Euclidean norm on S.

Generalizations of Banding

$GB_k(M) \equiv \big|\big| m_{ij} 1\big(\rho(i,j) \leq k\big) \big|\big|$,

$GBD_k \equiv \big\{ GB_k(M) : M$ bounded linear $\ell_2 \mapsto \ell_2 \big\}$,

$GBO \equiv \bigcup_k GBD_k$,

$GBDO \equiv \overline{GBO}$.

Then $GBDO$ closed under inversion.

Pf: GBO is closed under *, multiplication and addition.

Banding up to a Permutation

- Given π a permutation (Matrix obtained by permuting $(1, 0, \ldots, 0)$, $(0, 1, 0, \ldots), \ldots)$.
- $BO^{(\pi)} \equiv \{A : \pi A \pi^* \in BO\}$.
  $BDO^{(\pi)} \equiv \overline{BO}^{(\pi)}$.
- Closure under inversion holds.

Banding up to a Permutation

- $HBO^{(\pi)} \equiv \left\{\Sigma : \Sigma \in BO^{(\pi)}, \Sigma \text{ symmetric, p.d.}\right\}$,

  $HBDO \equiv \overline{\bigcup_{\pi} HBO^{(\pi)}}$,
- Then $HBDO$ is closed under inversion.
- Note:
  1. $HBO^{(\pi)}$ is a cone closed under inversion since $BO^{(\pi)}$ is.
  2. $HBDO$ is NOT closed under multiplication.

Applications to Statistics

$$\hat{\Sigma}_p \equiv \text{empirical covariance} \leftrightarrow \Sigma_p;$$

$$\Sigma_p = \text{upper } p \times p \text{ block of } \Sigma, \text{ bounded linear } \ell_2 \mapsto \ell_2;$$

$$\delta_k(\Sigma) = \min\left\{||B - \Sigma|| : B \in BO_k\right\}$$

$$\tilde{B}_k(\Sigma) \equiv \arg\min \delta_k(\Sigma).$$

Applications to Statistics

Suppose for $k_n \uparrow \infty$,

(i) $||B_{k_n}(\Sigma) - \Sigma|| \to 0$;

(ii) $||\tilde{B}_{k_n}(\Sigma) - B_{k_n}(\Sigma)|| \to 0$.

Can show

(iii) $\left|\left|B_{k_n}(\hat{\Sigma}_p) - B_{k_n}(\Sigma_p)\right|\right| \xrightarrow{\mathbb{P}} 0$, if $\Sigma \in BDO$.

Applications to Statistics

- (i),(ii) $\Rightarrow \big|\big|B_{k_n}(\hat{\Sigma}_p) - \Sigma_p\big|\big| \xrightarrow{\mathbb{P}} 0$;
- Since $||\Sigma_p^{-1}|| = m^{-1}$,

  $\Longrightarrow \big|\big|B_{k_n}^{-1}(\hat{\Sigma}_p) - \Sigma_p^{-1}\big|\big| \xrightarrow{\mathbb{P}} 0.$
- Note that $B_k^{-1}(\Sigma) \notin BO$ generally (unless Σ is diagonal).

Applications to Statistics

- Suppose we assume $\Sigma^{-1} \in BO_k$.
- Then $\Sigma \in BDO \Rightarrow (iii)$.
- How do we approximate $\tilde{B}_k(\Sigma^{-1})$?

Applications to Statistics

Method:

i) Compute $B_{k_n}(\hat{\Sigma}) \xrightarrow{\mathbb{P}} \Sigma$.

ii) Estimate $\hat{m}_k$, $\hat{M}_k$, upper and lower eigenvalues of $B_{k_n}(\hat{\Sigma})$.

iii) Form $\hat{\gamma} = 2\left(\hat{M}_k + \hat{m}_k\right)^{-1}$.

iv) Estimate Σ^{-1} by $\hat{\gamma} \sum_{\ell=0}^{n} \left(I - \hat{\gamma} B_k \left(\hat{\Sigma}_p\right)\right)^{\ell} \in BO_{kn}$.

A Remaining Statistical Question

- Given $\pi : \{1, \ldots, p\} \mapsto \{1, \ldots, p\}$ a permutation.
- Let $S_\pi = \{j : \pi(j) \neq j\}$, $T_m = \{\pi : |S_\pi| \leq m\}$.
- Given Σ_p there exists $k \leq p$, $m \leq p$, such that

$$\pi \Sigma_p \pi^* \in BO_k, \quad \text{for } \pi \in T_m. \tag{1}$$

- Let

$$k(m) \equiv \min\{k : \pi \Sigma_p \pi^* \in BO_k,\ \pi \in T_m\},$$

$$m(\Sigma_p) \equiv \min\{m :\ (\textbf{??}) \text{ holds for some } k,\ \pi\}.$$

A Remaining Statistical Question

- Since $k(m) \downarrow$,

  $m(\Sigma_p)$, $k\big(m(\Sigma_p)\big)$ uniquely defined.

I: CAN WE ESTIMATE π if m is large and $k(m)$ is moderate?

II: What can we say about rates for estimating Σ uniformly over

$\big\{\Sigma : m(\Sigma) \leq m_0,\ k\big(m(\Sigma)\big) \leq k_0\big\}$?