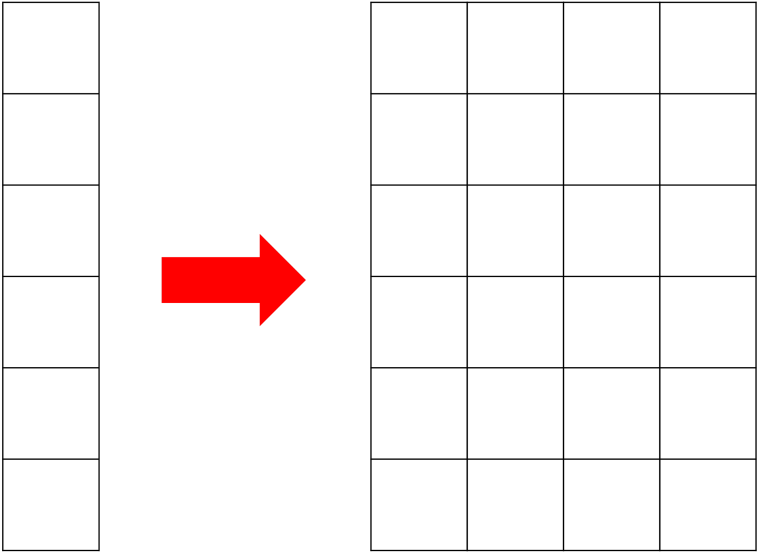
Adapting to arbitrary quadratic loss via singular value shrinkage

Takeru Matsuda

RIKEN Center for Brain Science

June 4, 2022 @ EcoSta 2022

Shrinkage estimation: from vector to matrix



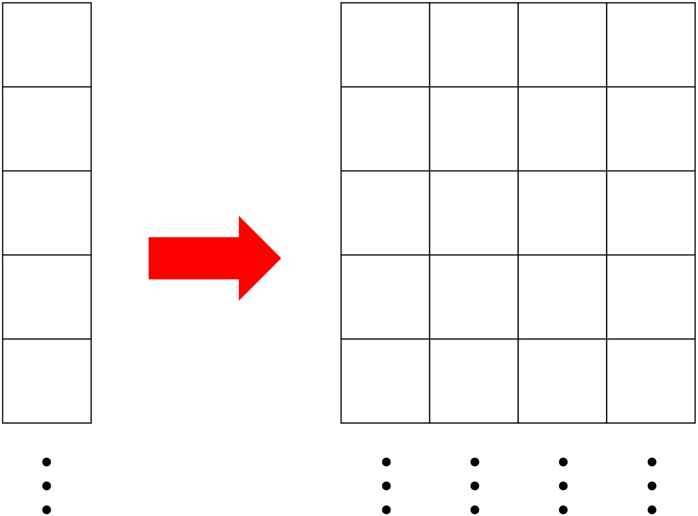
Shrinkage estimation: from vector to matrix

- Efron and Morris (1972): extension of James–Stein to matrix
- M. and Komaki (2015): Bayesian counterpart of Efron–Morris
- M. and Strawderman (2022): minimaxity of Efron–Morris under matrix quadratic loss

vector	matrix
James–Stein estimator $\hat{\mu}_{\mathrm{JS}} = \left(1 - \frac{n-2}{\|x\|^2}\right) x$	Efron–Morris estimator $\hat{M}_{\mathrm{EM}} = X\left(I_p - (n-p-1)(X^\top X)^{-1}\right)$
Stein's prior $\pi_{\mathrm{S}}(\mu) = \|\mu\|^{-(n-2)}$	singular value shrinkage prior $\pi_{\mathrm{SVS}}(M) = \det(M^\top M)^{-(n-p-1)/2}$

- How about nonparametric (infinite-dimensional) estimation ??

Sequence estimation: from univariate to multivariate



Abstract

- Efromovich and Pinsker (1984): The blockwise James–Stein estimator is adaptive minimax over the Sobolev ellipsoids.
  - "crowning result for linear estimation" (Johnstone, 2012)

$$\downarrow$$

- This study: The blockwise Efron–Morris estimator is adaptive minimax over the multivariate Sobolev ellipsoids.
  - adaptation not only to unknown smoothness and scale but also to arbitrary quadratic loss

	parametric	nonparametric
vector	James–Stein estimator	blockwise JS
matrix	Efron–Morris estimator	blockwise EM

- arXiv:2205.13840

Background:
Adaptive minimaxity of blockwise James–Stein estimator

Problem setting

Gaussian sequence model

$$y_i = \theta_i + \varepsilon \xi_i, \quad \xi_i \sim \mathrm{N}(0, 1), \quad i = 1, 2, \cdots$$

- estimation of $\theta = (\theta_i)$ from $y = (y_i)$ under the quadratic loss:

$$L(\theta, \hat{\theta}) = \|\hat{\theta} - \theta\|^2 = \sum_i (\hat{\theta}_i - \theta_i)^2$$

- risk function

$$R(\theta, \hat{\theta}) = \mathrm{E}_\theta [L(\theta, \hat{\theta})]$$

- canonical model in nonparametric estimation
  - nonparametric regression, density estimation, ...

Sobolev ellipsoid

Sobolev ellipsoid ($\beta > 0, Q > 0$)

$$\Theta(\beta, Q) = \left\{ \theta = (\theta_i) \,\middle|\, \sum_{i=1}^{\infty} a_{\beta,i}^2 \theta_i^2 \le Q \right\}$$

$$a_{\beta,i} = \begin{cases} i^{\,\beta} & (i : \text{even}) \\ (i-1)^{\,\beta} & (i : \text{odd}) \end{cases}$$

- A function $f : [0,1] \to \mathbb{R}$ belongs to the Sobolev class iff its Fourier coefficients belong to $\Theta(\beta, Q)$

Notions of minimaxity

- We consider the asymptotics $\varepsilon \to 0$ and write $a \sim b$ if $a/b \to 1$

Definition

$\hat{\theta}_*$: asymptotically minimax on Θ

$$\Leftrightarrow \quad \sup_{\theta \in \Theta} R(\theta, \hat{\theta}_*) \sim \inf_{\hat{\theta}} \sup_{\theta \in \Theta} R(\theta, \hat{\theta})$$

Definition

$\hat{\theta}_*$: adaptive minimax over $C = \{\Theta\}$
$\Leftrightarrow$ $\hat{\theta}_*$: asymptotically minimax on every $\Theta \in C$

- note: We focus on exact minimaxity in this study
    - stronger than rate-minimaxity

Pinsker's theorem

Theorem (Pinsker, 1980)

For a constant $\kappa = \kappa(\varepsilon, \beta, Q)$, the linear estimator

$$\hat{\theta}_{\mathrm{P},i} = (1 - \kappa a_i)_+ y_i$$

is asymptotically minimax on $\Theta(\beta, Q)$:

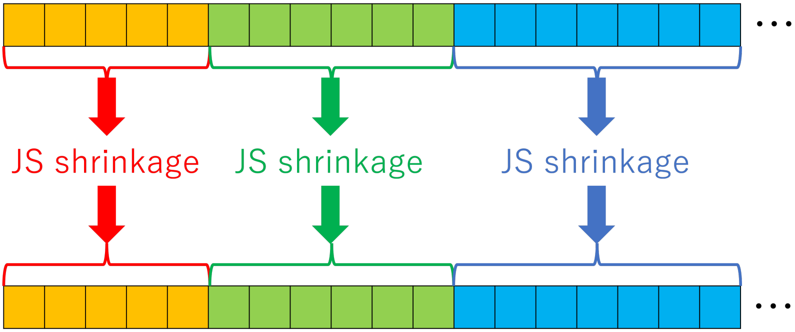
$$\sup_{\theta \in \Theta(\beta,Q)} R(\theta, \hat{\theta}_{\mathrm{P}}) \sim \inf_{\hat{\theta}} \sup_{\theta \in \Theta(\beta,Q)} R(\theta, \hat{\theta}) \sim P(\beta, Q)\varepsilon^{4\beta/(2\beta+1)}$$

- Pinsker estimator requires the knowledge of β and Q...

Blockwise James–Stein estimator

Blockwise James–Stein estimator $\hat{\theta}_{\mathrm{BJS}}$

- Take sufficiently large N (e.g. ε^{-2})
- Partition $\{1, 2, \ldots, N\}$ into consecutive blocks
- apply James–Stein shrinkage to each block



Adaptive minimaxity of blockwise JS

Theorem (Efromovich and Pinsker, 1984)

The blockwise James–Stein estimator $\hat{\theta}_{\mathrm{BJS}}$ with the weakly geometric blocks is adaptive minimax over the Sobolev ellipsoids:

$$\sup_{\theta \in \Theta(\beta, Q)} R(\theta, \hat{\theta}_{\mathrm{BJS}}) \sim \inf_{\hat{\theta}} \sup_{\theta \in \Theta(\beta, Q)} R(\theta, \hat{\theta}) \sim P(\beta, Q) \varepsilon^{4\beta/(2\beta+1)}$$

for every β and Q.

- proof: use Pinsker's theorem and oracle inequality for James–Stein estimator
    - Pinsker estimator is linear & JS attains almost the same risk with linear estimators
- Blockwise JS does not require the knowledge of β and Q!!

Efron–Morris estimator and its oracle inequality

Efron–Morris estimator

$$X \sim \mathrm{N}_{n,p}(M, I_n, I_p) \quad \Leftrightarrow \quad X_{ai} \sim \mathrm{N}(M_{ai}, 1)$$

- estimation of M from X under the Frobenius loss:

$$\|\hat{M} - M\|_{\mathrm{F}}^2 = \sum_{a=1}^{n} \sum_{i=1}^{p} (\hat{M}_{ai} - M_{ai})^2$$

- Efron–Morris estimator (= James–Stein estimator when $p = 1$)

$$\hat{M}_{\mathrm{EM}}(X) = X \left(I_p - (n-p-1)(X^\top X)^{-1} \right)$$

Theorem (Efron and Morris, 1972)

When $n \geq p + 2$, $\hat{M}_{\mathrm{EM}}$ is minimax and dominates $\hat{M}_{\mathrm{MLE}}(X) = X$.

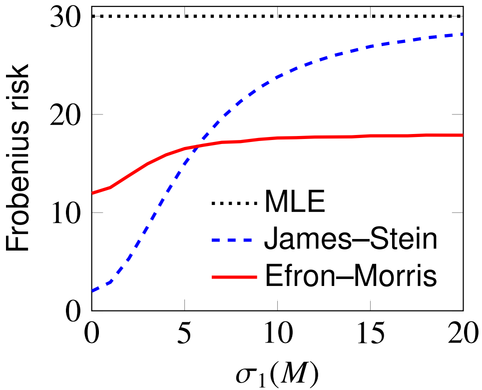
Singular value shrinkage of Efron–Morris

- Stein (1974): $\hat{M}_{\mathrm{EM}}$ shrinks singular values towards zero.

$$\sigma_i(\hat{M}_{\mathrm{EM}}) = \left(1 - \frac{n-p-1}{\sigma_i(X)^2}\right)\sigma_i(X)$$

$\to \hat{M}_{\mathrm{EM}}$ works well when M is close to low-rank!!

$$n = 10,\ p = 3,\ \sigma_2(M) = \sigma_3(M) = 0$$



Minimaxity of EM under matrix quadratic loss

- How about matrix quadratic loss ?

$$L(M, \hat{M}) = (\hat{M} - M)^\top (\hat{M} - M) \in \mathbb{R}^{p \times p}$$

Theorem (M. and Strawderman, *Biometrika* 2022)

When $n \geq p + 2$, $\hat{M}_{\mathrm{EM}}$ is minimax and dominates MLE:

$$\mathrm{E}_M[(\hat{M}_{\mathrm{EM}} - M)^\top (\hat{M}_{\mathrm{EM}} - M)] \preceq n I_p$$

- Therefore, $\hat{M}_{\mathrm{EM}}$ is minimax under arbitrary quadratic loss:

$$\mathrm{E}_M[\mathrm{tr}(\hat{M}_{\mathrm{EM}} - M) Q (\hat{M}_{\mathrm{EM}} - M)^\top] \leq n \mathrm{tr}(Q), \quad Q \succ O$$

  - $Q = I_p$: Frobenius loss
  - $Q = cc^\top$: $\|\hat{M}c - Mc\|^2$ (c = column weights)

Oracle inequality for Efron–Morris estimator

- $\hat{M}_C = XC$: linear estimator
- For fixed M,

$$\mathrm{E}_M[(\hat{M}_{C_*} - M)^\top(\hat{M}_{C_*} - M)] \preceq \mathrm{E}_M[(\hat{M}_C - M)^\top(\hat{M}_C - M)]$$

for every C, where $C_* = C_*(M) := (M^\top M + nI_p)^{-1} M^\top M$.

$\to \hat{M}_{C_*}$: linear oracle

Theorem

$$\mathrm{E}_M[(\hat{M}_{\mathrm{EM}} - M)^\top(\hat{M}_{\mathrm{EM}} - M)] \preceq \mathrm{E}_M[(\hat{M}_{C_*} - M)^\top(\hat{M}_{C_*} - M)] + 2(p+1)I_p$$

- $\hat{M}_{\mathrm{EM}}$ attains almost the same risk with linear oracle!!

Oracle inequality for Efron–Morris estimator

Corollary

$$\mathrm{E}_M[(\hat{M}_{\mathrm{EM}} - M)Q(\hat{M}_{\mathrm{EM}} - M)^\top] \leq \mathrm{E}_M[(\hat{M}_{C_*} - M)Q(\hat{M}_{C_*} - M)^\top] + 2(p+1)\mathrm{tr}(Q)$$

- $\hat{M}_{\mathrm{EM}}$ attains almost the same risk with linear oracle under arbitrary quadratic loss!!
  - key to adaptive minimaxity of blockwise EM

Adaptive minimaxity of blockwise Efron–Morris estimator

Problem setting

Multivariate Gaussian sequence model ($p \geq 2$)

$$y_i = \theta_i + \varepsilon \xi_i, \quad \xi_i \sim \mathrm{N}_p(0, I_p), \quad i = 1, 2, \cdots$$

- estimation of $\theta = (\theta_i)$ from $y = (y_i)$ under Q-quadratic loss
  - cf. weighted L^2 loss in function estimation

$$L_Q(\theta, \hat{\theta}) = \sum_i (\hat{\theta}_i - \theta_i)^\top Q(\hat{\theta}_i - \theta_i), \quad Q \succ O$$

- risk function

$$R_Q(\theta, \hat{\theta}) = \mathrm{E}_\theta[L_Q(\theta, \hat{\theta})]$$

Multivariate Sobolev ellipsoid

Multivariate Sobolev class ($\beta > 0, L > O$)

$$W(\beta, L) = \left\{ f : [0,1] \to \mathbb{R}^p \;\middle|\; \int_0^1 f^{(\beta)}(x)^\top L^{-2} f^{(\beta)}(x) \mathrm{d}x \le 1 \right\}$$

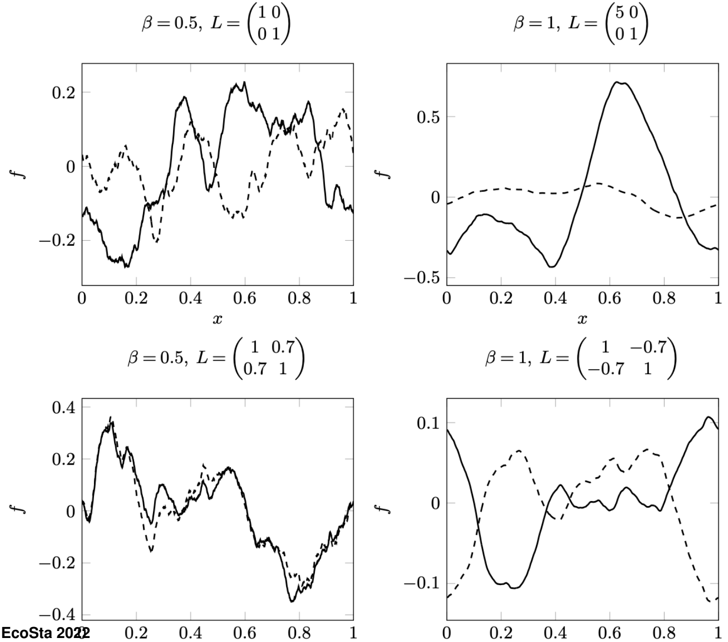
- β: smoothness of f
- L: scale & correlation between $f_1, \ldots, f_p$

Multivariate Sobolev ellipsoid ($\beta > 0, Q > O$)

$$\Theta(\beta, Q) = \left\{ \theta = (\theta_1, \theta_2, \cdots) \;\middle|\; \sum_{i=1}^{\infty} a_{\beta,i}^2 \theta_i^\top Q^{-1} \theta_i \le 1 \right\}$$

- $f \in W(\beta, L) \Leftrightarrow \theta \in \Theta(\beta, L^2/\pi^{2\beta})$ (Fourier coefficients)

Functions in multivariate Sobolev class



Multivariate extension of Pinsker's theorem

- S_+: projection of $S \in \mathbb{R}^{p \times p}$ onto the positive semidefinite cone
  - truncation of negative eigenvalues

Theorem

For a constant $\kappa = \kappa(\varepsilon, \beta, Q)$, the linear estimator

$$\hat{\theta}_{\mathrm{P},i} = (I_p - a_{\beta,i} \kappa Q^{-1})_+ y_i$$

is asymptotically minimax under L_Q on $\Theta(\beta, Q)$:

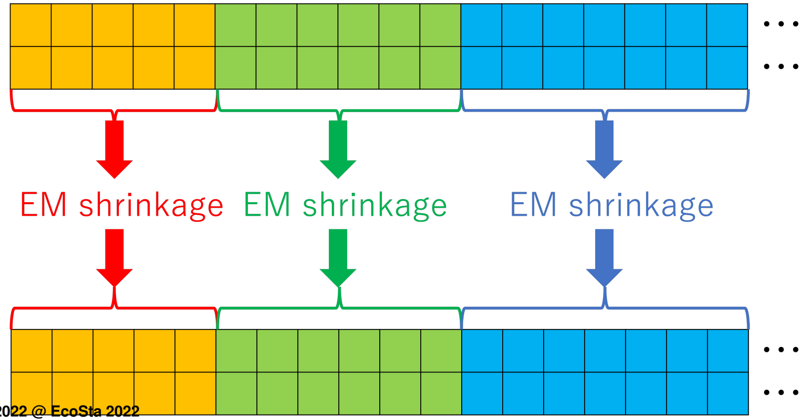
$$\sup_{\theta \in \Theta(\beta,Q)} R_Q(\theta, \hat{\theta}_{\mathrm{P}}) \sim \inf_{\hat{\theta}} \sup_{\theta \in \Theta(\beta,Q)} R_Q(\theta, \hat{\theta}) \sim P(\beta, Q) \varepsilon^{4\beta/(2\beta+1)}.$$

- generalization of Pinsker's theorem to $p \geq 2$

Blockwise Efron–Morris estimator

Blockwise Efron–Morris estimator $\hat{\theta}_{\mathrm{BEM}}$

- Take sufficiently large N (e.g. ε^{-2})
- Partition $\{1, 2, \ldots, N\}$ into consecutive blocks
- apply Efron–Morris shrinkage to each block matrix



Adaptive minimaxity of blockwise EM

Theorem

The blockwise Efron–Morris estimator $\hat{\theta}_{\mathrm{BEM}}$ is adaptive minimax over the multivariate Sobolev ellipsoids:

$$\sup_{\theta \in \Theta(\beta, Q)} R_Q(\theta, \hat{\theta}_{\mathrm{BEM}}) \sim \inf_{\hat{\theta}} \sup_{\theta \in \Theta(\beta, Q)} R_Q(\theta, \hat{\theta}) \sim P(\beta, Q) \varepsilon^{4\beta/(2\beta+1)}$$

for every β and Q.

- generalization of Efromovich and Pinsker (1984) to $p \geq 2$
- proof: use oracle inequality for Efron–Morris estimator
- adaptation not only to smoothness and scale but also to arbitrary quadratic loss!!
    - due to singular value shrinkage of Efron–Morris estimator

Summary

- Efromovich and Pinsker (1984): The blockwise James–Stein estimator is adaptive minimax over the Sobolev ellipsoids.
  - "crowning result for linear estimation" (Johnstone, 2012)

$$\downarrow$$

- This study: The blockwise Efron–Morris estimator is adaptive minimax over the multivariate Sobolev ellipsoids.
  - adaptation not only to unknown smoothness and scale but also to arbitrary quadratic loss

	parametric	nonparametric
vector	James–Stein estimator	blockwise JS
matrix	Efron–Morris estimator	blockwise EM

- arXiv:2205.13840