

高次元信号行列に対する自動スパース推定について

筑波大学・数理物質科学研究群 海野 哲也
筑波大学・数理物質系 矢田 和善
筑波大学・数理物質系 青嶋 誠

1 はじめに

本講演では、高次元データに対する低ランク行列の復元問題について考える．未知である $d \times n$ の低ランク行列 $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_n]$ に対し，そのランク数を r とする．今， $d \times n$ のデータ行列 $\mathbf{X}$ が以下で与えられているとする．

$$\mathbf{X} = \sqrt{n}\mathbf{A} + \mathbf{W}. \quad (1)$$

ここで， $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_n]$ はノイズ行列であり， $\mathbf{w}_1, \dots, \mathbf{w}_n$ は独立同分布に， $E(\mathbf{w}_l) = \mathbf{0}$ かつ $\text{Cov}(\mathbf{w}_l) = \Sigma_W$ を持つ分布に従うとする．信号行列とも呼ばれる低ランク行列 $\mathbf{A}$ の推定は，ゲノム科学や画像解析等，高次元データを扱う多くの分野において重要な技術であり，これまでにいくつかの先行研究がある．Shabalin and Nobel [5] は $\mathbf{W}$ に正規性を仮定し， $d/n \rightarrow c > 0$ の枠組みのもとで，ランダム行列理論に基づく $\mathbf{A}$ の推定量を提案した．Negahban and Wainwright [3] 及び Rohde and Tsybakov [4] は，高次元回帰モデルにおける低ランク行列の推定問題を検討した．Yata and Aoshima [8] は $\mathbf{W}$ に正規性を課さない場合における標本特異値・特異ベクトルの高次元漸近的性質を示し，特異値分解に基づく従来の推定量が高次元で巨大なノイズを内包することを示した．その解決策として，Yata and Aoshima [8] は Yata and Aoshima [7] で提案されたノイズ掃き出し法を特異値の推定に応用することで， $d/n \rightarrow \infty$ となる高次元小標本設定においても高精度な特異値の推定を可能とした．しかしながら，高次元小標本設定において d 次元の特異ベクトルを高精度に推定する手法は，未だ確立されていない．

最近，Yata and Aoshima [9] は高次元主成分ベクトルの推定に対し，自動スパース推定と呼ばれる方法論を提案し，これが高次元小標本設定においても主成分ベクトルを高精度に推定できることを示した．さらに，Umino et al. [6] は自動スパース推定を高次元相互共分散行列の推定に応用した．本講演では，高次元特異ベクトルの推定に自動スパース推定を応用することで，次元数が標本数を凌駕する状況においても高精度な信号行列の再構築手法を提案する．

2 高次元における標本特異値・特異ベクトルの漸近的性質

$\mathbf{A}$ の特異値分解を $\mathbf{A} = \sum_{i=1}^r \lambda_{i(A)}^{1/2} \mathbf{u}_{i(A)} \mathbf{v}_{i(A)}^T$ とする．ただし， $\lambda_{1(A)}^{1/2} \geq \dots \geq \lambda_{r(A)}^{1/2} (> 0)$ は $\mathbf{A}$ の特異値であり， $\mathbf{u}_{i(A)} \in \mathbf{R}^d$ ， $\mathbf{v}_{i(A)} \in \mathbf{R}^n$ はそれぞれ $\lambda_{i(A)}^{1/2}$ に対応する左特異ベクトル，右特異ベクトルである．各特異値・特異ベクトルは， d だけでなく n にも依存することに注意する． $r \geq 2$

である場合、各特異値は

$$\liminf_{d,n \rightarrow \infty} \frac{\lambda_{i(A)}}{\lambda_{i'(A)}} > 1, \quad \text{for all } i < i' (\leq r)$$

の意味で区別できるとする。したがって、 $\mathbf{u}_{i(A)}$ と $\mathbf{v}_{i(A)}$ は符号の違いを除いてそれぞれ一意に定まる。また、 Σ_W の固有値を $\lambda_{1(W)} \geq \cdots \geq \lambda_{d(W)} (\geq 0)$ とし、適当な直交行列 $\mathbf{H}_W$ を用いて $\Sigma_W = \mathbf{H}_W \Lambda_W \mathbf{H}_W^T$, $\Lambda_W = \text{diag}(\lambda_{1(W)}, \dots, \lambda_{d(W)})$ と分解する。 Σ_A と Σ_W の構造について、次を仮定する。

$$(A-i) \quad \liminf_{d,n \rightarrow \infty} \frac{\lambda_{r(A)}^2}{\text{tr}(\Sigma_W^2)} > 0 \quad \text{かつ} \quad \frac{\lambda_{1(W)}}{\lambda_{r(A)}} = o(1), \quad d, n \rightarrow \infty.$$

(A-i) の 1 つ目の仮定は、Aoshima and Yata [1] で提案された強スパイク固有値モデルに基づく仮定である。また、2 つ目の仮定は Yata and Aoshima [8] で仮定された、(1) に対するパワースパイクモデルよりも緩い仮定となっている。

データ行列 $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ に対し、 $\mathbf{x}_l = \sqrt{n} \mathbf{a}_l + \mathbf{H}_W \Lambda_W^{1/2} \mathbf{z}_l$, $l = 1, \dots, n$ と表記する。このとき、 $\mathbf{I}_d$ を d 次の単位行列として $E(\mathbf{z}_l) = \mathbf{0}$ かつ $\text{Cov}(\mathbf{z}_l) = \mathbf{I}_d$ となる。 $\mathbf{z}_l = (z_{1l}, \dots, z_{dl})^T$, $l = 1, \dots, n$ とする。Chen and Qin [2] 及び Yata and Aoshima [9] に基づき、次を仮定する。

(A-ii) $\mathbf{z}_l$ の 4 次モーメントが有界かつ、任意の $s_1 \neq s_2, s_3, s_4$ を満たす $s_1, s_2, s_3, s_4 = 1, \dots, d$ に対し、

$$E(z_{s_1 l}^2 z_{s_2 l}^2) = E(z_{s_1 l}^2) E(z_{s_2 l}^2) = 1 \quad \text{かつ} \quad E(z_{s_1 l} z_{s_2 l} z_{s_3 l} z_{s_4 l}) = 0$$

(A-ii) は $\mathbf{W}$ の母集団分布について、正規分布よりも緩い仮定であることに注意する。

標本共分散行列を $\mathbf{S} = n^{-1} \mathbf{X} \mathbf{X}^T$ とする。さらに、双対標本共分散行列を $\mathbf{S}_D = n^{-1} \mathbf{X}^T \mathbf{X}$ とする。 $\mathbf{S}$ と $\mathbf{S}_D$ は非ゼロの固有値を共有し、 $\text{rank}(\mathbf{S}) = \text{rank}(\mathbf{S}_D) \leq \min\{d, n\} (= m)$ となることに注意する。 $\mathbf{S}_D$ の固有値を $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_m (\geq 0)$ とし、 $\mathbf{S}$ と $\mathbf{S}_D$ の固有値分解をそれぞれ $\mathbf{S} = \sum_{i=1}^m \hat{\lambda}_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^T$, $\mathbf{S}_D = \sum_{i=1}^m \hat{\lambda}_i \hat{\mathbf{v}}_i \hat{\mathbf{v}}_i^T$ とする。ただし、 $\hat{\mathbf{u}}_i$, $\hat{\mathbf{v}}_i$ はそれぞれ $\hat{\lambda}_i$ に対応する $\mathbf{S}$, $\mathbf{S}_D$ の固有ベクトルである。このとき、 $\mathbf{X}/\sqrt{n}$ の特異値分解は $\mathbf{X}/\sqrt{n} = \sum_{i=1}^m \hat{\lambda}_i^{1/2} \hat{\mathbf{u}}_i \hat{\mathbf{v}}_i^T$ で与えられる。

各 i に対し $\kappa_i = \text{tr}(\Sigma_W)/(n\lambda_{i(A)})$ とする。標本特異値、標本特異ベクトルについて、Yata and Aoshima [8] より次の結果が得られる。

定理 1 (Theorem 3.1 in Yata and Aoshima [8]). $(A-i), (A-ii)$ のもと、各 $i \leq r$ に対し $d, n \rightarrow \infty$ で以下が成立する。

$$\begin{aligned} \frac{\hat{\lambda}_i}{\lambda_{i(A)}} &= 1 + \kappa_i + o_p(1), \quad \|\hat{\mathbf{u}}_i - \mathbf{u}_{i(A)}\|^2 = 2\{1 - (1 + \kappa_i)^{-1/2}\} + o_p(1) \\ \text{and} \quad \|\hat{\mathbf{v}}_i - \mathbf{v}_{i(A)}\|^2 &= o_p(1). \end{aligned}$$

したがって標本右特異ベクトル $\hat{\mathbf{v}}_i$ は高次元においても精度の良い推定量となるものの、標本特異値 $\hat{\lambda}_i$ 、及び標本左特異ベクトル $\hat{\mathbf{u}}_i$ の性能は κ_i に依存することがわかる。

$\|\cdot\|_F$ をフロベニウスノルムとし、 $d \times n$ 行列 $\hat{\mathbf{A}}$ に対する損失を $L(\hat{\mathbf{A}}|\mathbf{A}) = \|\hat{\mathbf{A}} - \mathbf{A}\|_F^2$ で定める． $\mathbf{X}$ の特異値分解に基づく $\mathbf{A}$ の推定量 $\hat{\mathbf{A}}_r = \sum_{i=1}^r \hat{\lambda}_i^{1/2} \hat{\mathbf{u}}_i \hat{\mathbf{v}}_i^T$ について、次の結果が得られる．

定理 2 (Theorem 3.2 in Yata and Aoshima [8]). $(A-i), (A-ii)$ のもと、 $d, n \rightarrow \infty$ で以下が成立する．

$$L(\hat{\mathbf{A}}_r|\mathbf{A}) = \frac{r \operatorname{tr}(\boldsymbol{\Sigma}_W)}{n} + o_p(\lambda_{r(A)}).$$

上記の結果より、 $\hat{\mathbf{A}}_r$ が $L(\hat{\mathbf{A}}_r|\mathbf{A}) = o_p(\lambda_{r(A)})$ の意味での一致性を有するには $d, n \rightarrow \infty$ のもと $\operatorname{tr}(\boldsymbol{\Sigma}_W)/(n\lambda_{r(A)}) (= \kappa_r) = o_p(1)$ の条件が必要になる．しかしながら $\operatorname{tr}(\boldsymbol{\Sigma}_W)$ は d に比例して増大するため、 $d/n \rightarrow \infty$ となる高次元小標本設定においては厳しい条件であると言わざるを得ない．

この問題に対し、Yata and Aoshima [8] は特異値 $\lambda_{i(A)}$ の推定において、ノイズ掃き出し法に基づく次の推定量を提案した．

$$\tilde{\lambda}_{i(r)} = \hat{\lambda}_i - \frac{\operatorname{tr}(\mathbf{S}_D) - \sum_{j=1}^r \hat{\lambda}_j}{n - r} \quad (i = 1, \dots, r).$$

上記の推定量 $\tilde{\lambda}_{i(r)}$ に対し、次の結果が得られる．

定理 3 (Corollary 4.1 in Yata and Aoshima [8]). $(A-i), (A-ii)$ のもと、各 $i \leq r$ に対し $d, n \rightarrow \infty$ で以下が成立する．

$$\frac{\tilde{\lambda}_{i(r)}}{\lambda_{i(A)}} = 1 + o_p(1).$$

上記の結果より、 $\tilde{\lambda}_{i(r)}$ の性能は κ_i に依存せず、高次元でも精度の良い推定量となることがわかる．Yata and Aoshima [8] は $\tilde{\lambda}_{i(r)}$ に基づく新たな $\mathbf{A}$ の推定量 $\tilde{\mathbf{A}}_r = \sum_{i=1}^r \tilde{\lambda}_{i(r)}^{1/2} \hat{\mathbf{u}}_i \hat{\mathbf{v}}_i^T$ を提案し、次の結果を得た．

定理 4 (Theorem 4.2 in Yata and Aoshima [8]). $\delta_i = \mathbf{u}_{i(A)}^T \mathbf{W} \mathbf{v}_{i(A)} / (n\lambda_{i(A)})^{1/2}$, $i = 1, \dots, r$ とする． $(A-i), (A-ii)$ のもと、 $d, n \rightarrow \infty$ で以下が成立する．

$$L(\tilde{\mathbf{A}}_r|\mathbf{A}) = 2 \sum_{i=1}^r \lambda_{i(A)} (1 + \delta_i) \left\{ 1 - \frac{1 + \delta_i}{(1 + \kappa_i + 2\delta_i)^{1/2}} \right\} + o_p(\lambda_{r(A)})$$

$$\text{and } \delta_i = o_p\{(\lambda_{r(A)}/\lambda_{i(A)})^{1/2}\} = o_p(1).$$

任意の $\kappa_i > 0$ に対し、 $2\{1 - 1/(1 + \kappa_i)^{1/2}\} < \kappa_i$, $i = 1, \dots, r$ が成り立つことに注意する．したがって漸近的に $L(\tilde{\mathbf{A}}_r|\mathbf{A}) < L(\hat{\mathbf{A}}_r|\mathbf{A})$ となるため、 $\tilde{\mathbf{A}}_r$ は $\hat{\mathbf{A}}_r$ よりも高精度な $\mathbf{A}$ の推定量となることがわかる．一方で、 $\tilde{\mathbf{A}}_r$ には依然として標本左特異ベクトル $\hat{\mathbf{u}}_i$ が用いられているため、そのノルム損失は κ_i に依存し、 $\hat{\mathbf{A}}$ と同様に $L(\tilde{\mathbf{A}}_r|\mathbf{A}) = o_p(\lambda_{r(A)})$ の意味での一致性には $\lim_{d, n \rightarrow \infty} \kappa_r = o(1)$ の条件が必要であった．

3 自動スパース推定による信号行列の再構築

前節より、 $\mathbf{A}$ の推定には左特異ベクトル $\mathbf{u}_{i(A)}$ の推定が重要であることがわかる。そこで、 $\mathbf{u}_{i(A)}$ の推定量として、ノイズ掃き出し法に基づく推定量

$$\tilde{\mathbf{u}}_{i(r)} = \sqrt{\frac{\hat{\lambda}_i}{\tilde{\lambda}_{i(r)}}} \hat{\mathbf{u}}_i \quad (i = 1, \dots, r)$$

を考える。 $\tilde{\mathbf{u}}_{i(r)}$ に対し、次の結果が得られる。

命題 1. $(A-i), (A-ii)$ のもと、各 $i \leq r$ に対し $d, n \rightarrow \infty$ で以下が成立する。

$$\mathbf{u}_{i(A)}^T \tilde{\mathbf{u}}_{i(r)} = 1 + o_p(1), \quad \text{and} \quad \|\tilde{\mathbf{u}}_{i(r)}\|^2 = 1 + \kappa_i \{1 + o_p(1)\} + o_p(1).$$

上記の結果より、

$$\|\tilde{\mathbf{u}}_{i(r)} - \mathbf{u}_{i(A)}\|^2 = \kappa_i \{1 + o_p(1)\} + o_p(1)$$

となることに注意する。したがって $\tilde{\mathbf{u}}_{i(r)}$ のノルム損失は依然として κ_i に依存するため、 $\tilde{\mathbf{u}}_{i(r)}$ を用いて $\mathbf{A}$ を再構築しても損失を改善することはできない。一方で、命題 1 は各 $i \leq r$ に対し

$$\begin{aligned} \tilde{\mathbf{u}}_{i(r)} &= \{1 + o_p(1)\} \mathbf{u}_{i(A)} + \boldsymbol{\xi}_i, \\ \boldsymbol{\xi}_i^T \mathbf{u}_{i(A)} &= 0, \quad \text{and} \quad \|\boldsymbol{\xi}_i\|^2 = \kappa_i \{1 + o_p(1)\} + o_p(1) \end{aligned}$$

となる確率変数ベクトル $\boldsymbol{\xi}_i$ が存在することを意味している。この性質に基づき、自動スパース推定によって $\tilde{\mathbf{u}}_{i(r)}$ を補正した推定量 $\hat{\mathbf{u}}_{i(r)}$ について、次の結果が得られる。

定理 5. $(A-i), (A-ii)$ と適当な正則条件のもと、各 $i \leq r$ に対し $d, n \rightarrow \infty$ で以下が成立する。

$$\|\hat{\mathbf{u}}_{i(r)} - \mathbf{u}_{i(A)}\|^2 = o_p(1).$$

本講演では、この $\hat{\mathbf{u}}_{i(r)}$ に基づく新たな信号行列の再構築手法を提案し、理論的かつ数値的に既存手法と比較する。さらに、信号行列のランク数 r は一般には未知であるため、これを推定するための方法論も与えた上で、実際の高次元データを解析し、性能を検証する。

参考文献

- [1] Aoshima, M. and Yata, K. (2018). Two-sample tests for high-dimension, strongly spiked eigenvalue models. *Statistica Sinica*, 28:43–62.
- [2] Chen, S. X. and Qin, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics*, 38(2):808–835.
- [3] Negahban, S. and Wainwright, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *The Annals of Statistics*, 39(2):1069–1097.

- [4] Rohde, A. and Tsybakov, A. B. (2011). Estimation of high-dimensional low-rank matrices. *The Annals of Statistics*, 39(2):887–930.
- [5] Shabalin, A. A. and Nobel, A. B. (2013). Reconstruction of a low-rank matrix in the presence of Gaussian noise. *Journal of Multivariate Analysis*, 118:67–76.
- [6] Umino, T., Yata, K., and Aoshima, M. (2025). Automatic sparse estimation of the high-dimensional cross-covariance matrix. *Revised in Journal of Multivariate Analysis*.
- [7] Yata, K. and Aoshima, M. (2012). Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations. *Journal of Multivariate Analysis*, 105(1):193–215.
- [8] Yata, K. and Aoshima, M. (2016). Reconstruction of a high-dimensional low-rank matrix. *Electronic Journal of Statistics*, 10(1):895–917.
- [9] Yata, K. and Aoshima, M. (2025). Automatic sparse pca for high-dimensional data. *Statistica Sinica*, 35:1069–1090.