International Symposium on New Developments of Theories, Methodologies and Applications for Large Complex Data

October 21-23, 2026

Venue:

iias Hall (located in iias Tsukuba `https://tsukuba.iias.jp/iiashall`)

5-19 Kenkyugakuen, Tsukuba, Ibaraki 305-0817, Japan

Organizers:

Makoto Aoshima (University of Tsukuba)

Kazuyoshi Yata (University of Tsukuba)

Kento Egashira (Tokyo University of Science)

Supported by

Grant-in-Aid for Scientific Research (A) 25H01107 (Project Period: 2025-2030)

"Advancement and Expansion of Theories and Methodologies for Large-Scale Complex Data"

(Principal Investigator: Makoto Aoshima)

Grant-in-Aid for Challenging Research (Exploratory) 26K23798 (Project Period: 2026-2029)

"Building an Innovative Platform for Real-Time Streaming Data Based on High-Dimensional, Low-Sample-Size Theory"

(Principal Investigator: Makoto Aoshima)

Program (UTC+9)

October 21 (Wednesday)

14:00∼14:10 Opening

14:10∼14:50 Kento Egashira[*,a], Kazuyoshi Yata[b], and Makoto Aoshima[b]

[a](Department of Information Science and Technology, Tokyo University of Science)

[b](Institute of Mathematics, University of Tsukuba)

Enhanced Gaussian kernel k-means for high-dimensional, low-sample-size data

15:00∼15:40 Yoshikazu Terada

(Graduate School of Engineering Science, The University of Osaka / RIKEN AIP)

A theory of nonparametric covariance function estimation for discretely observed data

15:50∼16:30 Tsutomu T. Takeuchi

(Division of Particle and Astrophysical Science, Nagoya University)

Applications of the high-dimensional statistics to astrophysics and related fields

16:50∼17:30 Shao-Hsuan Wang

(Graduate Institute of Statistics, National Central University)

A statistical framework for contrastive principal component analysis

17:40∼18:20 Yuan-Tsung Chang[*,a] and Nobuo Shinozaki[b]

[a](The Institute of Statistical Mathematics / Mejiro University)

[b](Faculty of Science and Technology, Keio University)

A note on estimation of restricted (ordered, upper (lower)-bounded) scale parameter of Gamma distribution

October 22 (Thursday)

10:40∼11:50 **Young Researchers Session I**

1. Tetsuya Umino[*,a], Kazuyoshi Yata[b], and Makoto Aoshima[b]

   [a](Graduate School of Science and Technology, University of Tsukuba)

   [b](Institute of Mathematics, University of Tsukuba)

   Data transformation for high-dimension, strongly spiked eigenvalue models using automatic sparse estimation

2. Giheon Seong[*,a], Seungki Hong[b], and Sungkyu Jung[c]

   [a](Department of Statistics, Seoul National University)

   [b](Meritz Fire & Marine Insurance Co., Ltd.)

   [c](Department of Statistics and Institute for Data Innovation in Science, Seoul National University)

   Augmented James-Stein estimation for leading eigenvectors and eigenspaces in high dimensions

3. Atsutomo Yara[*,a] and Yoshikazu Terada[b]

   [a](Graduate School of Engineering Science, The University of Osaka)

   [b](Graduate School of Engineering Science, The University of Osaka / RIKEN AIP)

   Nonparametric maximum likelihood estimation of intensities for covariate-driven Poisson processes

12:00∼12:30 Bilol Banerjee[*,a] and Anil K. Ghosh[b]

[a](Department of Statistics and Data Science, National University of Singapore)

[b](Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, Kolkata)

A nonparametric test of spherical symmetry applicable to high dimensional data

12:30∼14:00 Lunch

14:00∼18:15 **Special Invited and Keynote Sessions**

19:30∼ Dinner

October 23 (Friday)

10:40∼11:50 **Young Researchers Session II**

1. Masaki Hoshino[*,a], Kento Egashira[b], and Hikaru Yamaguchi[b]

   [a](Graduate School of Science and Technology, Tokyo University of Science)

   [b](Department of Information Science and Technology, Tokyo University of Science)

   Group sequential one-sample location testing for high dimensions

2. Kyoowon Kim[*,a] and Sungkyu Jung[b]

   [a](Department of Statistics, Seoul National University)

   [b](Department of Statistics and Institute for Data Innovation in Science, Seoul National University)

   Testing and segmentation of joint and individual components in integrative multi-source factor models

3. Shuntaro Suzuki[*,a] and Yoshikazu Terada[b]

   [a](Graduate School of Engineering Science, The University of Osaka)

   [b](Graduate School of Engineering Science, The University of Osaka / RIKEN AIP)

   Operator learning for functional autoregressive models

12:00∼12:40 Shogo Nakakita

(Graduate School of Engineering Science, The University of Osaka)

Generalization error analysis for linear binary classification under Lipschitz loss functions

12:40∼14:10 Lunch

14:10∼14:50 Koji Tsukuda[*,a] and Kou Fujimori[b]

[a](Faculty of Mathematics, Kyushu University)

[b](Faculty of Economics and Law, Shinshu University)

Asymptotically normal estimators in high-dimensional linear regression

15:00∼15:40 Takahiro Nishiyama[*,a], Masashi Hyodo[b], Shoma Yamashita[c], and Nobumichi Shutoh[d]

[a](Faculty of Science, Tokyo University of Science)

[b](Faculty of Economics, Kanagawa University)

[c](Graduate School of Science, Kobe University)

[d](Laboratory of Medical Data Sciences, Kobe Pharmaceutical University / Graduate School of Science, Kobe University)

Testing homogeneity of high-dimensional covariance matrices under a low-dimensional factor structure

15:40∼15:45 Closing

(∗ Speaker)

October 22 (Thursday)

Special Invited Session

14:00∼14:50 **High-dimensional precision matrix quadratic forms: Estimation framework for $p > n$**

Speaker: Weiming Li
(School of Statistics and Data Science, Shanghai University of Finance and Economics)

Discussion Leader: Kento Egashira (Department of Information Science and Technology, Tokyo University of Science)

15:05∼15:55 **Joint estimation of high-dimensional spiked covariance matrices via a partially shared subspace**

Speaker: Sungkyu Jung
(Department of Statistics and Institute for Data Innovation in Science, Seoul National University)

Joint work with: Changwon Yoon[a], Minwoo Kim[b], and Jeongyoun Ahn[a]
[a](Department of Industrial and Systems Engineering, KAIST)
[b](Department of Statistics, Seoul National University)

Discussion Leader: Kazuyoshi Yata (Institute of Mathematics, University of Tsukuba)

16:10∼17:00 **From high-dimensional classification to clustering: High-impact feature selection via ggCGA and UL-MLR**

Speaker: Shih-Feng Huang
(Graduate Institute of Statistics, National Central University)

Discussion Leader: Yuan-Tsung Chang (The Institute of Statistical Mathematics / Mejiro University)

Keynote Session

17:15∼18:15 **Approaching the frontier in large risky-asset universes: Direct portfolio estimation via CORE**

Speaker: Xinghua Zheng
(Department of ISOM, Hong Kong University of Science and Technology)

Discussion Leader: Shogo Nakakita (Graduate School of Engineering Science, The University of Osaka)

Enhanced Gaussian kernel k-means for high-dimensional, low-sample-size data

Kento Egashira[a], **Kazuyoshi Yata**[b], **Makoto Aoshima**[b]

[a]**Department of Information Science and Technology, Tokyo University of Science**

[b]**Institute of Mathematics, University of Tsukuba**

Clustering high-dimensional data plays an important role in a wide range of applications, including gene expression analysis, where the dimension can be substantially larger than the sample size. In such high-dimensional, low-sample-size (HDLSS) settings, clustering methods are routinely used to identify latent groups of observations. Among them, k-means is one of the most widely used clustering methods because of its simple formulation and computational efficiency. However, high dimensionality can substantially affect the behavior of clustering procedures, and a theoretical understanding of clustering methods under HDLSS asymptotics is therefore important.

Recent research has extensively investigated the asymptotic properties of clustering methods for HDLSS data. The theoretical analysis of k-means clustering in high-dimensional settings has been conducted predominantly under sparsity assumptions. Arias-Castro and Pu [1] developed a sparse k-means method based on a different formulation. Chang et al. [3] introduced a regularization scheme based on an ℓ_∞/ℓ_0-type penalty. Chakraborty and Das [2] considered a lasso-weighted k-means algorithm for high-dimensional data. Yang and Benjamin [6] proposed a sparse extension of the possibilistic c-means algorithm originally introduced by Krishnapuram and Keller [5], adapting it to high-dimensional settings. Despite these advances, the asymptotic properties of the standard k-means algorithm in HDLSS settings remain less well understood, particularly without imposing sparsity assumptions.

In our previous study, Egashira et al. [4], we theoretically investigated the asymptotic properties of k-means under HDLSS settings without imposing sparsity assumptions. We showed that, even when appropriate initial observations are selected, the clustering criterion can contain a bias induced by differences in the within-cluster covariance structures. This bias can affect the clustering result and prevent the recovery of the underlying population structure. To overcome this problem, in Egashira et al. [4], we proposed a bias-corrected k-means procedure that removes the bias arising from the within-cluster covariance structures. We also considered k-means with multiple initial observations and showed that using multiple observations for initialization can relax the conditions required for consistent clustering. These results indicate that both the initialization strategy and the bias inherent in the clustering criterion play important roles in the asymptotic behavior of k-means in HDLSS settings.

Kernel methods provide a natural extension of k-means for capturing nonlinear relationships through a feature-space representation. However, our previous theoretical analysis showed that, under HDLSS asymptotics, Gaussian kernel k-means does not necessarily provide an asymptotic advantage over conventional k-means. More specifically, the consistency conditions of Gaussian kernel k-means and conventional k-means are asymptotically equivalent under the HDLSS setting. This result indicates that, although kernelization provides a nonlinear feature-space representation, it does not by itself eliminate the fundamental difficulty associated with the high-dimensional bias.

In this talk, we investigated incorporating two ideas into Gaussian kernel k-means: multiple initial observations and bias correction. We developed enhanced Gaussian kernel k-means procedures based on these two ideas and derived their asymptotic properties under HDLSS settings. In particular, we clarified how multiple initial observations and bias correction affect the consistency conditions of Gaussian kernel k-means and whether these modifications can improve its asymptotic performance.

Acknowledgments
The research of the first author was partially supported by Grant-in-Aid for Early-Career Scientists, JSPS, under Contract Number 24K20748. The research of the third author was partially supported by Grants-in-Aid for Scientific Research (A) and Challenging Research (Exploratory), JSPS, under Contract Numbers 25H01107 and 26K23798.

References

[1] Arias-Castro, E., Pu, X. (2017) A simple approach to sparse clustering. *Computational Statistics & Data Analysis*, 105, 217–228.

[2] Chakraborty, S., Das, S. (2022). Detecting meaningful clusters from high-dimensional data: a strongly consistent sparse center-based clustering approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44, 2894–2908.

[3] Chang, X., Wang, Y., Li, R., Xu, Z. (2018). Sparse k-means with l_∞/l_0 penalty for high-dimensional data clustering. *Statistica Sinica*, 28, 1265–1284.

[4] Egashira, K., Yata, K., Aoshima, M. (2026). Asymptotic properties of k-means and its bias correction for high-dimension, low-sample-size data. *Journal of Multivariate Analysis*, 216, 105675.

[5] Krishnapuram, R., Keller, J. M. (1993). A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems*, 1, 98–110.

[6] Yang, M.-S., Benjamin, J.B. (2023). Sparse possibilistic c-means clustering with lasso. *Pattern Recognition*, 138, 109348.

A Theory of Nonparametric Covariance Function Estimation for Discretely Observed Data

Yoshikazu Terada[1,2] and Atsutomo Yara[1]

[1] Graduate School of Engineering Science, The University of Osaka, Osaka, Japan
[2] Center for Advanced Integrated Intelligence Research, RIKEN
terada.yoshikazu.es@osaka-u.ac.jp, a-yara@sigmath.es.osaka-u.ac.jp

Recent advances in measurement technology have led to the increasing availability of data in which each subject is observed over a domain, either continuously or at a finite set of locations. In functional data analysis (FDA), such data are viewed as realizations of random elements in a function space, or equivalently as sample paths of stochastic processes [4]. Estimation of the mean and covariance functions is a fundamental problem in FDA. It plays a central role in many subsequent methods, most notably functional principal component analysis and functional linear regression, and is also essential for reconstructing latent trajectories from discretely observed data [6].

When functional data are observed continuously over the domain, covariance estimation is straightforward, as the empirical covariance function is directly available. By contrast, when measurements are available only at discrete locations, covariance function estimation becomes a challenging nonparametric estimation problem.

Consider a centered, square-integrable stochastic process $X(\cdot)$ on a multidimensional domain $\mathcal{T} \subset \mathbb{R}^d$ with the covariance function $K^\circ(s,t) = \mathbb{E}[X(s)X(t)]$ $(s,t \in \mathcal{T})$. We observe i.i.d. replicates $X_1, \ldots, X_n$, but not the trajectories/surfaces themselves. Instead, for the i-th subject, we observe noisy measurements

$$Y_{ij} = X_i(T_{ij}) + \varepsilon_{ij}, \qquad j = 1, \ldots, m_i,$$

where the observed points $T_{ij} \in \mathcal{T}$ are random, and ε_{ij} are mean-zero errors. For simplicity, the number of measurements is the same among the subjects (i.e., $m = m_1 = \cdots = m_n$).

Existing estimators are typically based on smoothing splines [1] or local linear smoothing [7], and their statistical properties have been investigated in considerable detail. Most of the existing literature focuses on functions defined on one-dimensional domains. In many modern applications, however, functions are defined on multidimensional domains, and the covariance function is therefore defined on the corresponding product domain, whose dimension is doubled, thereby exacerbating the curse of dimensionality.

Recent results in nonparametric regression suggest that adaptive deep learning estimators can, for suitably structured function classes, mitigate the curse of dimensionality (e.g., [3]). This fact motivates the use of learning-based estimators for covariance function estimation. For instance, [5] established theoretical

guarantees for deep learning estimators of the mean function, and [2] proposed a deep learning approach to covariance function estimation.

We study a general nonparametric least-squares (LS) approach to estimating K°. More precisely, we consider a learning-based estimator defined as a minimizer of the least-squares loss

$$\frac{1}{n\,m(m-1)}\sum_{i=1}^{n}\sum_{1\le j\neq k\le m}\left\{Y_{ij}Y_{ik}-K(T_{ij},T_{ik})\right\}^2.$$

This loss function is widely used (e.g., [6, 1, 7]).

In this talk, we develop a framework for sharp localization analysis of covariance function estimation. We derive a sharp oracle inequality that treats sparse and dense regimes in a unified manner. As an application, we derive convergence rates for deep learning estimators over several classes of covariance functions.

Our results suggest that, under highly anisotropic smoothness, deep learning estimators can mitigate the curse of dimensionality in covariance function estimation, as also demonstrated in other estimation problems [3, 5]. In the sparse regime, the convergence rate of deep learning estimators matches, up to polylogarithmic factors, the minimax rate for nonparametric regression over $2d$-dimensional domains, as established in [3].

Interestingly, in the dense regime in the one-dimensional setting, for the class of smooth functions considered by [7], deep learning estimators are suboptimal, whereas local linear smoothing attains a faster rate. In contrast, for the structured function classes studied by [1], deep learning estimators attain the same minimax rate as smoothing spline estimators up to polylogarithmic factors, while local linear smoothing becomes suboptimal. Thus, these results reflect a distinctive adaptivity–variance trade-off in covariance function estimation.

References

1. Cai, T.T., Yuan, M.: Nonparametric covariance function estimation for functional and longitudinal data. Technical report, University of Pennsylvania, Philadelphia, PA (2010)
2. Sarkar, S., Panaretos, V.M.: Covnet: Covariance networks for functional data on multidimensional domains. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **84**(5), 1785–1820 (2022)
3. Suzuki, T., Nitanda, A.: Deep learning is adaptive to intrinsic dimensionality of model smoothness in anisotropic besov space. Advances in Neural Information Processing Systems **34**, 3609–3621 (2021)
4. Wang, J.L., Chiou, J.M., Müller, H.G.: Functional data analysis. Annual Review of Statistics and Its Application **3**, 257–295 (2016)
5. Yan, S., Yao, F., Zhou, H.: Deep regression for repeated measurements. Journal of the American Statistical Association **120**(552), 2461–2472 (2025)
6. Yao, F., Müller, H.G., Wang, J.L.: Functional data analysis for sparse longitudinal data. Journal of the American Statistical Association **100**(470), 577–590 (2005)
7. Zhang, X., Wang, J.L.: From sparse to dense functional data and beyond. The Annals of Statistics **44**(5), 2281–2321 (2016)

Applications of the High-Dimensional Statistics to Astrophysics and Related Fields

Tsutomu T. Takeuchi, Yu Ogane, Kazuyoshi Yata, Kento Egashira,
Makoto Aoshima, Kohji Yoshikawa, Nanase Harada, Koichiro Nakanishi

Abstract

High-dimensional statistical analysis extracts intensity and velocity structures from molecular spectra of NGC 253 and identifies contaminated frequency ranges in the larger ARCHEMI cube. Recent extensions to H I absorption forests distinguish physical mean evolution from residual covariance and separate total absorption from persistent spectral morphology using optimal transport.

1 High-dimensional spectral analysis

Spectral maps connect molecular composition, excitation, and gas motions with the spatial organization of star formation. Their frequency coverage often grows faster than the number of spatial samples. With n spectra and d channels, the condition $n \ll d$ defines high-dimensional, low-sample-size (HDLSS) data. Neighboring pixels can also be correlated, so the nominal sample count need not equal the number of independent observations.

For principal component analysis (PCA), the centered data matrix $\widetilde{X} \in \mathbb{R}^{d\times n}$ gives the sample covariance and its dual,

$$\widehat{S} = \frac{\widetilde{X}\widetilde{X}^{\mathsf{T}}}{n-1}, \qquad \widehat{S}^{\mathrm{D}} = \frac{\widetilde{X}^{\mathsf{T}}\widetilde{X}}{n-1}. \tag{1}$$

They share the same nonzero eigenvalues, but the dual problem has dimension n rather than d. Noise-reduction PCA corrects the high-dimensional contribution to estimated eigenvalues under specified population conditions. Automatic sparse PCA identifies the channels carrying the important loadings. Eigenspectra indicate which features vary together; spatial score maps show where that variation occurs. Their joint interpretation is essential for distinguishing physical structure from observational artifacts.

2 NGC 253: molecular gas and data quality

The ALMA Band 7 map of the nuclear starburst in NGC 253 contains $n = 231$ spatial spectra and initially $d = 2248$ channels, with a beam corresponding to approximately 8×5 pc. Frequency-resolved analysis retains information in line wings and profile distortions that could be hidden by replacing each transition with its integrated intensity.

In the original cube, PC1 approximately traces total line intensity, while PC2 follows systemic rotation. Its spatial gradient agrees with the HCN(4–3) velocity field. Sparse loadings identify channels associated with HCN(4–3) and HNC(4–3), linking the statistical directions to recognizable molecular features without prescribing a list of integrated-line observables.

We estimated the velocity field from HCN(4–3), HNC(4–3), and CS(7–6) and shifted the spectra to a common kinematic frame. The corrected coverage contains $d = 1971$ channels. The rotation-induced butterfly pattern in the score diagram disappears, and higher components expose smaller-scale velocity structure. Corrected PC2 emphasizes line wings and a localized expansion on roughly 10 pc scales. PC3 identifies blueshifted profile distortions near the root of the known southwest molecular streamer (Fig. 1). The interpretation follows from both the profiles and their spatial correspondence, not from the arbitrary sign of a principal component. The result localizes an outflow signature; its mass or mass-loss rate requires further physical modeling.

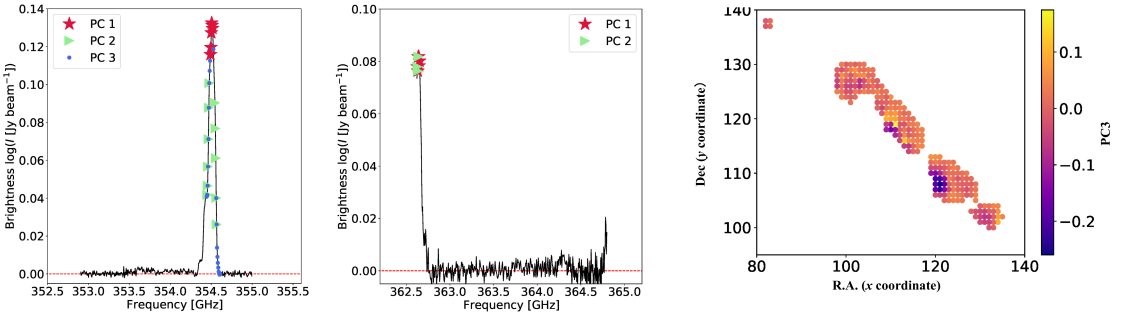


Figure 1: Doppler-corrected NGC 253 analysis. Left: sparse-PCA channels around HCN and HNC. Right: PC3 score map, whose localized signal is associated with the molecular outflow. Panels adapted from Takeuchi et al. (2024).

The ARCHEMI working cube extends the analysis to 85–163 GHz, with $d = 50702$ and $n \lesssim 2000$. Initial component maps contained stripes and gradients inconsistent with the source. The contributing frequency ranges were identified as affected by strong atmospheric absorption. Removing these ranges recovered coherent central structures. This is a quality-control result; physical interpretation of the cleaned cube remains in progress.

3 HI forests: mean evolution and transport shape

HI 21-cm forests probe neutral gas through absorption against background radio sources. Our absorption studies distinguish degeneracies of the observation operator from information lost through compression. For positive total absorption, $g(v) = 1 - e^{-\tau(v)}$ defines $M_g = \int g(v)\,\mathrm{d}v$ and $p_g = g/M_g$. With Q_g the quantile function of p_g, a mass–shape distance is

$$\mathsf{d}_\beta^2(g,h) = \beta\left(\log\frac{M_g}{M_h}\right)^2 + \frac{1}{v_*^2}\int_0^1 [Q_g(q) - Q_h(q)]^2\,\mathrm{d}q, \tag{2}$$

where $\beta > 0$ and $v_* > 0$ are fixed. This measures absorption, not gas mass. In regular profile families, its pullback identifies locally unobservable physical directions. For example, an isolated 21 cm component constrains a column-density-to-spin-temperature ratio rather than both quantities separately.

A TNG50-derived mock contains 12 matched sightlines at three epochs near $z = 10, 9, 8$, with 1024 bins per epoch. The leading eigenvalue of the between-epoch scatter accounts for 95.54% of its trace, yet that trace is only 7.05% of the total pixel scatter. Mean evolution is therefore nearly rank one but low trace. The full nonzero within-epoch covariance span captures just 2.02% of the squared linear mean-contrast norm. A regularized mean-aware direction recovers epoch ordering when all epochs of each sightline are held out together, with performance comparable to integrated absorption.

After mapping each velocity interval to $[0, 1]$ for shape comparison, redshift accounts for 98.50% of log-absorption-mass variance, whereas sightline identity accounts for 90.30% of normalized-quantile variation. Centroid-removed shape correctly matches all sightlines across epochs. These results concern one periodic volume and 12 matched groups, not 36 independent spectra or a unique three-dimensional reconstruction.

Thus, physical information need not occupy the largest covariance mode. Spectral loadings, score maps, mean evolution, and absorption mass–shape coordinates provide complementary routes to its interpretation.

A Statistical Framework for Contrastive Principal Component Analysis

Shao-Hsuan Wang

Graduate Institute of Statistics, National Central University

Principal component analysis (PCA) is one of the most widely used techniques for exploratory data analysis. Contrastive principal component analysis (cPCA), introduced by Abid et al. (2018), generalizes PCA by extracting features that are enriched in a target dataset relative to a background dataset while preserving as much information as possible from the target data. However, in high-dimensional settings, cPCA becomes computationally demanding due to the need to construct the contrastive covariance matrix and perform its eigenvalue decomposition to obtain the leading contrastive principal components.

Despite its growing popularity, a rigorous statistical framework for cPCA has been lacking. In this work, we establish a statistical framework for cPCA and develop its asymptotic theory under the high-dimensional spiked covariance model. Building on these theoretical results, we further propose a data-driven procedure for selecting the contrast parameter, a fundamental issue that has remained largely unresolved.

In this talk, we will present the proposed statistical framework, discuss the main theoretical results, and illustrate their performance through numerical studies, demonstrating the validity of the proposed theory.

A note on estimation of restricted (ordered, upper (lower)-bounded) scale parameter of Gamma distribution

Yuan-Tsung Chang (ISM,Mejiro University) and Shinozaki Nobuo (Keio University)

Summary. First, we introduce the estimation of a linear function of two ordered scale parameters of two Gamma distributions. A necessary and sufficient condition for the ratio of two coefficients, which guarantees that the maximum likelihood estimator (MLE) and the modified MLE dominate the crude unbiased estimator (UE) and the admissible estimator based solely on one sample under quadratic loss are derived, respectively. (Chang and Shinozaki (2002,2008)). Next, an admissible estimator of the upper-bounded scale parameter of a Gamma distribution is derived and the simultaneous estimation of p upper-bounded scale parameters of Gamma distributions are considered. Moreover, simultaneous estimation of two ordered scale parameters of two Gamma distributions is investigated in terms of the mean square error (MSE).

The Gamma distribution plays an important role in reliability theory, as it serves often as a model for lifetime times and waiting times, leading to the necessity of estimating the parameters of a Gamma distribution. Usually, the estimation problem is solved assuming that there are no restrictions with respect to the values of the parameter. However, in reality the parameters are always restricted and very often prior information about the restriction is available. In this case, the usual methods of estimation turn out to be inappropriate. Therefore, we consider for the scale parameter some types of constraints, for example, bounded above or order restrictions between two scale parameters, and derive estimators, which take into account the restrictions. The suggested estimators perform better than the usual estimators.

Problems concerning statistical inference for restricted parameters are of considerable interest. The earlier works related to statistical inference under restrictions are reviewed by Barlow et al (1972), Robertson and Wright (1988). The more recent literature about restricted statistical inference after 1988 is reviewed in two monographs, Silvapulle and Sen (2004), Eeden (2006). Many authors have considered inference on parameters in case of a is simple order restriction or tree order restriction on parameters and have suggested estimators, which improve the performance of unbiased estimators, especially for the normal distribution.

Katz (1961) has shown the admissibility or minimaxity of estimators when an upper or lower bound restriction is given in the case of the expectation parameter in exponential family under squared error loss. Casella and Strawderman (1981) proposed a minimax estimator for a normal mean that is restricted to a bounded interval. Berry (1993) considered admissible minimax estimators of the location parameter for the exponential distribution, when upper bound, lower bound or two-sided restrictions are given. Eeden (1995) suggested an admissible estimator for the scale parameter of a Gamma distribution with given lower bound under scale invariant squared error loss. When there is an order restriction between two scale parameters of two Gamma distributions, Chang and Shinozaki (2002,2008) have considered the estimation of a linear function of two ordered scale parameters of two Gamma distributions. They derived a necessary and sufficient condition for the ratio of two coefficients, which guarantees that the maximum likelihood estimator (MLE) and the modified MLE dominate the crude unbiased estimator (UE) and the admissible estimator based solely on one sample under quadratic loss, respectively. Marchand and Strawderman (2005) improved the minimum risk equivariant estimator of a scale parameter under a lower bound restriction.

Although Katz(1961) has given a proof of the admissibility for an estimator of the expectation parameter in the exponential family, when there is a lower bound(upper bound) restriction, Eeden(1995) has pointed out that there is an error in Katz's proof and showed that Katz's condition is not satisfied for the Gamma scale parameter with lower bound restriction.

On the other hand, Berger(1980) developed a general technique for simultaneous estimation of Gamma scale parameters to improve the standard estimators under several types of loss functions that include the sum of squared error loss and scale invariant squared error loss.

In Section 1, we first introduce the estimation of a linear function of two ordered scale parameters of two Gamma distributions (Chang and Shinozaki (2002, 2008)) in term of MSE and entropy loss. In Section 2, we discuss the admissible estimation of an scale parameter, which is bounded above, of the Gamma distribution under squared error loss. In Section 3, we consider the simultaneous estimation of Gamma scale parameters with upper bound and lower bound restrictions under sum of squared error loss and sum of scale invariant squared error loss. In Section 4, we consider the simultaneous estimation of two ordered scale parameters of two Gamma distributions.

References

1. Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). Statistical Inference under Order Restrictions, Wiley, New York.

2. Berger, J. (1980). Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma scale parameters. *Ann. Statist.*, 8, 545-571.

3. Berry, J.C. (1993). Minimax estimation of a restricted exponential location parameter. *Statist. Decisions*, 11, 307-316.

4. Casella, G. and Strawderman, W.E (1981). Estimating a bounded normal mean. *Ann. Statist.*, 9, 870-878.

5. Chang, Y.-T. (1982). Stein-type estimators for parameters in truncated spaces. *Keio Science and Technology Reports*, 34, 83-95.

6. Chang, Y.-T. and Shinozaki, N. (2002). A comparison of restricted and unrestricted estimators in estimating linear functions of ordered scale parameters of two gamma distributions. *Ann. Inst. Statist. Math.* Vol. 54, No.4, 848-860.

7. Chang, Y.-T. and Shinozaki, N. (2008) Estimation of linear functions of ordered scale parameters of two Gamma distributions under entropy loss. Journal of Japan Statistical Society, **38**, Issue 2, 335-347

8. Katz,M.W. (1961). Admissible and minimax estimates of parameters in truncated spaces. *Ann. Math. Statist.*, 32, 136-142.

9. Marchand, $\grave{E}$. and Strawderman, W.E. (2005) On improving on the minimum risk equivariant estimator of a scale parameter under a lower-bound constraint. *Journal of Statist. planning and inference*, 134. 90-101.

10. Robertson, T., Wright, F. T. and Dykstra, R. L.(1988). Order Restricted Statistical Inference, Wiley, New York.

11. Silvapulle, M. J. and Sen, P. K. (2004). Constrained Statistical Inference, Wiley, New Jersey.

12. Van Eeden,C. (1995). Minimax estimation of a lower-bounded scale parameter of a gamma distribution for scale-invariant squared-error loss. *Canad.J.Statist.*, 23, 245-256.

13. Van Eeden, C (2006). Restricted parameter space estimation problems. Lecture notes in Statistics 188, Springer.

Data Transformation for High-Dimension, Strongly Spiked Eigenvalue Models using Automatic Sparse Estimation

Tetsuya Umino[a], **Kazuyoshi Yata**[b], **and Makoto Aoshima**[b]

[a]Graduate School of Science and Technology, University of Tsukuba
[b]Institute of Mathematics, University of Tsukuba

1 Introduction

A common feature of high-dimensional data is that the data dimension is large while the sample size is relatively small. This is the so-called "HDLSS" or "large p, small n" data situation where $p/n \to \infty$; here p is the data dimension and n is the sample size. Such data situations occur in many areas of modern science such as genetic microarrays, medical imaging, text recognition, finance, and chemometrics.

Suppose we have a sample of n (≥ 4) independent and identically distributed (i.i.d.) p-variate random vectors, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. Here, we consider situations where the data dimension p is very high compared to the sample size n. We also assume that $\boldsymbol{x}_j$ has a mean vector $\mathbf{0}$ and an unknown covariance matrix $\boldsymbol{\Sigma}$ ($\geq \boldsymbol{O}$), and the eigen-decomposition of $\boldsymbol{\Sigma}$ is given by $\sum_{s=1}^{p} \lambda_s \boldsymbol{h}_s \boldsymbol{h}_s^T$, where $\lambda_1 \geq \cdots \geq \lambda_p (\geq 0)$ are the eigenvalues of $\boldsymbol{\Sigma}$ and $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_p$ are the corresponding eigenvectors. Recently, Aoshima and Yata [1] established two disjoint models for high-dimensional data based on the covariance structure of $\boldsymbol{\Sigma}$: the strongly spiked eigenvalue (SSE) model and the non-SSE (NSSE) model. The SSE model is defined by

$$\liminf_{p\to\infty} \frac{\lambda_{\max}(\boldsymbol{\Sigma})}{\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^2)}} > 0, \tag{1}$$

where $\lambda_{\max}(\boldsymbol{\Sigma})$ is the largest eigenvalue of $\boldsymbol{\Sigma}$. On the other hand, the NSSE model is defined by

$$\frac{\lambda_{\max}(\boldsymbol{\Sigma})}{\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^2)}} = 0, \quad \text{as } p \to \infty. \tag{2}$$

High-dimensional statistical methods are often developed under the NSSE model. However, in practice, the NSSE model may not hold due to high correlations between variables or data mixing, meaning that these methods may not provide valid analytical results. In this talk, we propose a data transformation method that transforms high-dimensional data $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]$ generated from the SSE model into a new dataset $\widetilde{\boldsymbol{X}} = [\widetilde{\boldsymbol{x}}_1, \ldots, \widetilde{\boldsymbol{x}}_n]$ that approximately follows the NSSE model.

2 Data Transformation Method using Automatic Sparse PCA

Let $\boldsymbol{S}$ be the sample covariance matrix, and consider its eigen-decomposition as $\boldsymbol{S} = \sum_{s=1}^{p} \hat{\lambda}_s \hat{\boldsymbol{h}}_s \hat{\boldsymbol{h}}_s^T$, where $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p (\geq 0)$. For $l = 1, \ldots, p$, let $\Psi_{(l)} = \sum_{s=l}^{p} \lambda_s^2$. Under the SSE model (1), we assume the following condition:

(A-i) There exists a positive fixed integer k such that $\lambda_1, \ldots, \lambda_k$ are distinct in the sense that $\liminf_{p\to\infty}(\lambda_s/\lambda_t - 1) > 0$ when $1 \leq s < t \leq k$, and λ_k and λ_{k+1} satisfy

$$\liminf_{p\to\infty} \frac{\lambda_k^2}{\Psi_{(k)}} > 0 \quad \text{and} \quad \frac{\lambda_{k+1}^2}{\Psi_{(k+1)}} \to 0 \quad \text{as } p \to \infty.$$

Condition (A-i) was introduced by Aoshima and Yata [1] as a special case of the SSE model. Under (A-i), we define the matrix $\boldsymbol{A}$ by

$$\boldsymbol{A} = \boldsymbol{I}_p - \sum_{s=1}^{k} \boldsymbol{h}_s \boldsymbol{h}_s^T = \sum_{s=k+1}^{p} \boldsymbol{h}_s \boldsymbol{h}_s^T.$$

Then, transforming $\boldsymbol{x}_j$ into $\boldsymbol{A}\boldsymbol{x}_j$ yields data following the NSSE model (2). However, as pointed out by Aoshima and Yata [1], the sample eigenvectors $\hat{\boldsymbol{h}}_s$ cannot ensure an accurate estimation of $\boldsymbol{h}_s$ in high-dimensional settings.

In this talk, focusing on the representation $\boldsymbol{A}\boldsymbol{x}_j = \boldsymbol{x}_j - \sum_{s=1}^{k} (\boldsymbol{x}_j^T \boldsymbol{h}_s) \boldsymbol{h}_s$, we propose a novel method for estimating $\boldsymbol{A}\boldsymbol{x}_j$ by combining the estimators of $(\boldsymbol{x}_j^T \boldsymbol{h}_s)$ and $\boldsymbol{h}_s$. Let $x_{js} = \boldsymbol{x}_j^T \boldsymbol{h}_s$ for $s = 1, \ldots, k$, and let $\tilde{x}_{js}$ be the estimator of x_{js} proposed by Aoshima and Yata [1]. In addition, let $\widetilde{\boldsymbol{h}}_{s*}$, $s = 1, \ldots, k$, be the estimator of $\boldsymbol{h}_s$ obtained by the automatic sparse PCA proposed by Yata and Aoshima [2]. Based on these estimators, for each $j = 1, \ldots, n$, we estimate $\boldsymbol{A}\boldsymbol{x}_j$ as

$$\widetilde{\boldsymbol{x}}_j = \boldsymbol{x}_j - \sum_{s=1}^{k} \tilde{x}_{js} \widetilde{\boldsymbol{h}}_{s*}.$$

Then, we have the following result.

Theorem 1. Let $\widetilde{\boldsymbol{X}} = [\widetilde{\boldsymbol{x}}_1, \ldots, \widetilde{\boldsymbol{x}}_n]$. Under (A-i) and some regularity conditions, it holds that

$$\frac{1}{n} \|\boldsymbol{A}\boldsymbol{X} - \widetilde{\boldsymbol{X}}\|_F^2 = \frac{1}{n} \sum_{j=1}^{n} \|\boldsymbol{A}\boldsymbol{x}_j - \widetilde{\boldsymbol{x}}_j\|^2 = o_P(\lambda_1)$$

as $p, n \to \infty$.

References

[1] Aoshima, M. and Yata, K. (2018). Two-sample tests for high-dimension, strongly spiked eigenvalue models. *Statistica Sinica*, 28:43–62.

[2] Yata, K. and Aoshima, M. (2025). Automatic sparse PCA for high-dimensional data. *Statistica Sinica*, 35:1069–1090.

Augmented James–Stein estimation for leading eigenvectors and eigenspaces in high dimensions

Giheon Seong[1], Seungki Hong[2], Sungkyu Jung[3]
[1]Department of Statistics, Seoul National University, heon1998@snu.ac.kr
[2]Meritz Fire & Marine Insurance Co., Ltd., seungki.hong@meritz.co.kr
[3]Department of Statistics and Institute for Data Innovation in Science, Seoul National University, sungkyu@snu.ac.kr

Abstract

Building on the James–Stein approach to leading eigenvector estimation (Goldberg and Kercheval, 2023), we develop a data-adaptive augmented James–Stein shrinkage framework for estimating leading eigenvectors and eigenspaces under a generalized spiked population model, in the high-dimensional regime where the dimension p and sample size n grow proportionally. For each spiked eigenvector, we construct an augmented target subspace that combines auxiliary information, either from domain knowledge or prior information, with the remaining sample spiked eigenvectors. This augmentation allows information shared across the sample spiked components to be exploited while retaining a fully data-driven shrinkage rule. We show that the resulting eigenvector estimator strictly improves upon standard PCA whenever the target subspace contains nonvanishing information about the population eigenvector, while asymptotically reverting to PCA when the target is uninformative. The individual estimators further yield a nested sequence of estimators for all leading spiked eigenspaces, with analogous dominance properties. For the full spiked eigenspace, the proposed estimator also strictly improves upon the existing HDLSS-motivated shrinkage estimator in the proportional high-dimensional regime, while becoming asymptotically equivalent to it under HDLSS asymptotics. Simulation studies demonstrate substantial finite-sample gains and robustness to misspecification of the number of spikes.

Nonparametric maximum likelihood estimation of intensities for covariate-driven Poisson processes

Atsutomo Yara[1] Yoshikazu Terada[1,2]

[1]The University of Osaka [2]RIKEN AIP

1 Model and motivation

We observe independent and identically distributed pairs (Z_i, X_i), $i = 1, \dots, n$, with $Z_i \in [0,1]^{d_z}$ and $X_i \mid Z_i = z \sim \mathrm{PP}\{S_0(x \mid z)\, d\nu(x)\}$ on a compact set $A \subset \mathbb{R}^{d_x}$. The reference measure ν is known and finite, and S_0 is bounded and nonnegative. For $d_z > 0$, the covariate distribution has a positive density. The sample size n is the number of point patterns, not the total number of events.

For a class $\mathcal{F}_n$ of nonnegative neural network intensity functions, define

$$\widehat{S}_n \in \operatorname*{arg\,min}_{S \in \mathcal{F}_n} \frac{1}{n} \sum_{i=1}^{n} \left\{ - \sum_{j=1}^{N_i} \log S(x_{ij} \mid Z_i) + \int_A S(x \mid Z_i)\, d\nu(x) \right\},$$

where $N_i = X_i(A)$ and x_{ij} are the observed event locations. The excess negative log-likelihood risk can be infinite for vanishing candidate intensities. We instead use the squared Hellinger loss averaged over the covariate distribution,

$$\widetilde{H}^2(S, S_0) = \frac{1}{2} \mathbb{E}_Z \int_A \left\{ \sqrt{S(x \mid Z)} - \sqrt{S_0(x \mid Z)} \right\}^2 d\nu(x).$$

This loss bounds the squared Hellinger distance between the conditional Poisson process laws, averaged over Z (Leskelä, 2024).

2 Main theoretical results

Suppose that $0 \le S \le F < \infty$ for every $S \in \mathcal{F}_n$ and that $S_n^* \in \mathcal{F}_n$ satisfies $S_0 / S_n^* \le c_0^2$, with $0/0 = 0$. A localized likelihood argument based on van de Geer (2000) yields

$$\mathbb{E}[\widetilde{H}^2(\widehat{S}_n, S_0)] \lesssim \delta_n^2 + \widetilde{H}^2(S_n^*, S_0) + n^{-1}, \qquad \sqrt{n}\, \delta_n^2 \gtrsim \Psi_n(\delta_n),$$

where Ψ_n is an upper bound on the local bracketing entropy integral for square roots of midpoint intensities, with $\Psi_n(\delta)/\delta^2$ nonincreasing. Only S_n^* must satisfy the ratio condition; candidates need not be bounded away from zero.

We impose the small value bound condition $(\nu \otimes P_Z)\{S_0 \le t\} \le C_{\mathrm{svb}} t^\alpha$, $t > 0$, with $\alpha \ge 0$, as an intensity analogue of the condition in Bos and Schmidt-Hieber (2022). Put $a = 1 + (\alpha \wedge 1)$. A uniform approximation with error $O(\xi_n)$, truncated to $[\xi_n, F]$, satisfies the ratio condition and has Hellinger approximation error $O\{\xi_n^a (1 + \log(1/\xi_n))\}$.

For $S_0 = g_q \circ \cdots \circ g_0$, each g_i has input dimension $d^{(i)}$, with β_i-Hölder components depending on at most t_i variables. Define $\beta_i^* = \beta_i \prod_{\ell=i+1}^{q} (\beta_\ell \wedge 1)$. With a suitable ReLU network architecture,

$$\mathbb{E}[\widetilde{H}^2(\widehat{S}_n, S_0)] \lesssim \max_{0 \le i \le q} n^{-a\beta_i^* / (a\beta_i^* + t_i)} \log^3 n.$$

For $\alpha = 0$, this rate matches a minimax lower bound up to logarithmic factors on a Euclidean cube with Lebesgue measure, if the covariate density is bounded above and below by positive constants, $t_i \le \min_{0 \le j < i} d^{(j)}$ for $i \ge 1$, and the Hölder radius is sufficiently large.

For $\beta > 0$, a Hölder class defined using a fixed atlas on a known compact smooth manifold without boundary has rate $n^{-a\beta/(a\beta + d_{\mathcal{M}} + d_z)} \log^3 n$. Here $d_{\mathcal{M}}$ is the intrinsic dimension.

3 Numerical examples and application to neural spike trains

Simulations compare DNN–Softplus and DNN–ReLU with kernel estimators on Euclidean domains, the circle, and the sphere, using five sample sizes from 100 to 10000, $d_z \in \{0, 1, 5\}$, and 20 repetitions. For a one-dimensional compositional intensity with $d_z = 1$, the conditional kernel estimator and DNN–Softplus have mean squared Hellinger losses 0.120 and 0.728 at $n = 100$, compared with 0.0329 and 0.00926 at $n = 10000$. On the circle with $d_z = 1$ and $n = 10000$, the kernel estimator using the true geodesic geometry has mean loss 0.00520, compared with 0.00570 for DNN–Softplus. Thus, the relative performance varies with sample size and simulation setting.

The application uses Allen Visual Coding Neuropixels session 715093703 (Siegle et al., 2021). The 3,024 retained Gabor stimulus presentations each last 0.25 seconds and cover 243 combinations of position and orientation. A fixed split, stratified by stimulus condition, gives 2,417 development and 607 test presentations, with all conditions represented in both. Twelve high-activity units in the primary visual cortex are selected using development data. Both DNNs have three ReLU hidden layers of width 64. The covariates are stimulus position and the sine and cosine of twice the orientation angle.

Table 1 reports differences in test negative log-likelihood (NLL), averaged across units with equal weights. Both DNNs have lower average NLL than the conditional kernel estimator. The 11-unit sensitivity analysis excludes unit 950931315, whose test results had been examined in a preliminary analysis. Both analyses are exploratory comparisons within the same session.

Table 1. DNN minus conditional-kernel NLL, in nats per presentation. Negative values favor the DNN. The pointwise 95% intervals are computed by the percentile bootstrap, resampling whole stimulus conditions.

Units	Output activation	Mean difference	95% interval
12	Softplus	−0.02901	[−0.04042, −0.01729]
12	ReLU	−0.02080	[−0.03092, −0.01026]
11	Softplus	−0.02610	[−0.03791, −0.01378]
11	ReLU	−0.01954	[−0.02991, −0.00933]

Each of the 2,000 bootstrap resamples uses the same sampled conditions across units and methods and retains all presentations within each sampled condition. The direct comparison between Softplus and ReLU does not establish a difference between the two output activations.

4 Conclusion

We developed a likelihood-based framework for nonparametric intensity estimation in covariate-dependent Poisson processes using deep neural networks. The proposed oracle inequality leads to convergence rates that adapt to compositional structure and intrinsic dimension. These results demonstrate that deep neural networks can exploit low-dimensional structure in intensity estimation even when the ambient dimension is large. Our numerical experiments further illustrate the practical advantages of the proposed estimators in a range of Euclidean and manifold settings, and the neural data analysis shows their potential for modeling stimulus-dependent spike intensities. Taken together, the theoretical and empirical results provide a foundation for flexible intensity estimation with complex covariate effects.

References

Bos, T. and Schmidt-Hieber, J. (2022). Convergence rates of deep ReLU networks for multiclass classification. *Electronic Journal of Statistics* **16**, 2724–2773.

Leskelä, L. (2024). Information divergences and likelihood ratios of Poisson processes and point patterns. *IEEE Transactions on Information Theory* **70**, 9084–9101.

Siegle, J. H. et al. (2021). Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature* **592**, 86–92.

van de Geer, S. A. (2000). *Empirical Processes in M-estimation*. Cambridge University Press.

A nonparametric test of spherical symmetry applicable to high dimensional data

Bilol Banerjee[1] and Anil K. Ghosh[2]

[1] Department of Statistics and Data Science, National University of Singapore.
[2] Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, Kolkata

Abstract

We develop a test for spherical symmetry of a multivariate distribution P that works well even when the dimension of the data d is larger than the sample size n. We propose a non-negative measure of spherical asymmetry $\zeta(\mathrm{P})$ such that $\zeta(\mathrm{P}) = 0$ if and only if P is spherically symmetric. We construct a consistent estimator of $\zeta(\mathrm{P})$ using the data augmentation method and investigate its large sample properties. The proposed test based on this estimator is calibrated using a novel resampling algorithm. Our test controls the type I error, and it is consistent against general alternatives. We also study its behavior for a sequence of alternatives $(1-\delta_n)F + \delta_n G$, where $\zeta(G) = 0$ but $\zeta(F) > 0$, and $\delta_n \in [0, 1]$. When $\limsup \delta_n < 1$, for any G, the power of our test converges to unity as n increases. However, if $\limsup \delta_n = 1$, the asymptotic power of our test depends on $\lim n(1-\delta_n)^2$. We establish this by proving the minimax rate optimality of our test over a suitable class of alternatives and showing that it is Pitman efficient when $\lim n(1-\delta_n)^2 > 0$. Moreover, our test is provably consistent for high-dimensional data even when d grows with n. When the center of symmetry is unspecified, naive centering may lead to size distortions. We therefore propose a modified procedure based on pairwise differences that retains the desired level and other theoretical properties of the test. Simulation studies illustrate the theoretical findings and demonstrate favorable performance relative to existing methods.

Keywords: Consistency; Contiguous alternatives; Data augmentation; Minimax rate optimality; Pitman efficiency; Spherical symmetry.

High-Dimensional Precision Matrix Quadratic Forms: Estimation Framework for $p > n$

Weiming Li

School of Statistics and Data Science,
Shanghai University of Finance and Economics

We propose a novel estimation framework for quadratic functionals of precision matrices in high-dimensional settings, particularly in regimes where the feature dimension p exceeds the sample size n. Traditional moment-based estimators with bias correction remain consistent when $p < n$, i.e., when p/n tends to a constant less than one. However, they break down entirely once $p > n$, highlighting a fundamental distinction between the two regimes due to rank deficiency and high-dimensional complexity. Our approach resolves these issues by combining a spectral-moment representation with constrained optimization, resulting in consistent estimation under mild moment conditions.

The proposed framework provides a unified approach for inference on a broad class of high-dimensional statistical measures. We illustrate its utility through two representative examples: the optimal Sharpe ratio in portfolio optimization and the multiple correlation coefficient in regression analysis. Simulation studies demonstrate that the proposed estimator effectively overcomes the fundamental $p > n$ barrier where conventional methods fail.

Joint Estimation of High-Dimensional Spiked Covariance Matrices via a Partially Shared Subspace

Changwon Yoon[1], Minwoo Kim[2], **Sungkyu Jung**[2], and Jeongyoun Ahn[1]

[1]KAIST [2]Seoul National University

Abstract. High-dimensional covariance estimation is intrinsically difficult when the number of variables is comparable to, or larger than, the sample size. At the same time, many applications provide an auxiliary data set measured on the same variables and related to the target data through only part of its covariance structure. Pooling the two samples can then reduce variance but may induce substantial bias when dataset-specific directions are present. We propose a *partially shared subspace* (PSS) model and estimation framework that transfers only the covariance information supported by a common signal subspace while retaining distinct structure in each data source.

For independent target and background samples with spiked covariance matrices, after noise-level normalization we write

$$\Sigma_X = Q_S \Lambda_S Q_S^\top + Q_D \Lambda_D Q_D^\top + I_p, \qquad \Sigma_Y = U_S \Gamma_S U_S^\top + U_D \Gamma_D U_D^\top + I_p,$$

where $\mathrm{span}(Q_S) = \mathrm{span}(U_S)$ is an unknown shared subspace of rank r_S, while Q_D and U_D contain dataset-specific spikes. The shared spikes may occur at arbitrary spectral positions and need not correspond to the leading eigenvectors. This formulation strictly generalizes standard common-principal-component and shared-subspace models while allowing the two data sets to have different total spiked ranks and unrelated distinct components.

We work in the proportional-growth regime $p/n_X \to c_X$ and $p/n_Y \to c_Y$. Random matrix theory implies that sample spiked eigenvectors are not consistent, but their alignments converge to deterministic limits. In particular, for population eigenvectors q_i, u_j and their sample counterparts $\widehat{q}_i, \widehat{u}_j$,

$$|\langle \widehat{q}_i, \widehat{u}_j \rangle| \xrightarrow{a.s.} \phi_{c_X}(\lambda_i)\phi_{c_Y}(\gamma_j)|\langle q_i, u_j \rangle|, \qquad \phi_c(\ell) = \sqrt{\frac{1 - c/(\ell-1)^2}{1 + c/(\ell-1)}}.$$

This yields a principal-angle characterization of the shared subspace through the singular values of $\widehat{Q}^\top \widehat{U}$. We develop an iterative estimator that alternates between estimating r_S via a random-matrix-theory threshold and identifying the shared eigenvector indices through alignment with a pooled projection. The corresponding oracle rank and index estimators are consistent under a separation condition between the shared and distinct components.

After estimating the shared subspace, we combine the two samples through a weighted projection

$$\widehat{P}_S^{(\alpha)} = \Pi_{\widehat{r}_S}\left\{\alpha \widehat{P}_X + (1-\alpha)\widehat{P}_Y\right\}, \qquad \widehat{P}_X = \widehat{Q}\widehat{Q}^\top, \quad \widehat{P}_Y = \widehat{U}\widehat{U}^\top.$$

An asymptotic Frobenius-loss calculation gives the approximate optimal weight

$$\alpha^* = \frac{1}{2} + \frac{L_Y - L_X}{2(r_S - T)},$$

so the amount of transfer depends on both sample sizes and spike strengths rather than on sample size alone. The target covariance is then reconstructed by projecting its spiked component onto the estimated shared subspace and its orthogonal complement. This construction also produces a principled estimator of the target-specific subspace, providing a high-dimensional solution to contrastive dimension reduction.

Simulation studies show accurate recovery of the shared rank and uniformly favorable target-covariance estimation across leading or interleaved shared spikes, non-orthogonal distinct subspaces, and mild model misspecification. In an S&P 500 application, using pre-pandemic returns as background improves minimum-variance portfolio construction during the early COVID-19 period; with $p = 499$, $n_X = 61$, and $n_Y = 378$, the method estimates nine shared directions and attains the lowest out-of-sample volatility among the considered estimators. In a second application to low-grade glioma RNA-seq data, using normal brain tissue as background yields a distinct subspace that cleanly separates astrocytoma and oligodendroglioma samples, whereas the shared subspace and ordinary PCA do not. These examples demonstrate that partial structural transfer can improve high-dimensional covariance estimation while preserving scientifically meaningful dataset-specific variation.

Keywords: high-dimensional covariance estimation; spiked covariance model; shared subspace; transfer learning; random matrix theory; contrastive dimension reduction.

From High-Dimensional Classification to Clustering: High-Impact Feature Selection via ggCGA and UL-MLR

Shih-Feng Huang

Graduate Institute of Statistics, National Central University

This study develops a unified framework for high-impact feature (HIF) selection in high-dimensional classification and clustering. We first propose ggCGA, a new feature selection procedure for high-dimensional multinomial logistic regression (MLR), and establish its selection consistency under mild regularity conditions. Building on ggCGA, we further develop UL-MLR, an unsupervised learning framework that extends HIF selection to high-dimensional clustering. UL-MLR iteratively alternates between an MLR step and an unsupervised learning (UL) step. In the MLR step, ggCGA identifies HIFs using pseudo-labels generated by the UL step, while in the UL step, the pseudo-labels are updated based on the selected features. For each fixed iteration, the selection consistency of UL-MLR is established when the pseudo-labels satisfy a proposed local imbalance condition. Through this iterative procedure, UL-MLR jointly refines feature selection and cluster assignment. Extensive simulation studies and real-data applications demonstrate that ggCGA effectively identifies HIFs in high-dimensional classification, while UL-MLR achieves accurate HIF selection and reliable clustering performance in high-dimensional settings.

Approaching the Frontier in Large Risky-Asset Universes: Direct Portfolio Estimation via CORE

Xinghua Zheng

Department of ISOM,
Hong Kong University of Science and Technology

Constructing mean-variance efficient portfolios in large asset universes consisting solely of risky assets is a central problem in financial economics and a cornerstone of empirical asset pricing. We propose CORE (COnstrained sparse Regression for Efficient portfolios), a direct portfolio estimation framework built on a novel linearly constrained regression representation of the optimal portfolio, which bypasses the error-prone estimation of high-dimensional mean vectors and covariance matrices. Theoretically, we establish asymptotic mean-variance efficiency in a regime where the cross-sectional dimension and the sample size proportionally approach infinity, while allowing aggregate gross exposure to grow with the size of the asset universe. We further show that CORE serves as a unified framework encompassing the major portfolio paradigms, including global minimum variance estimation (GMV-CORE), quadratic utility maximization (Utility-CORE), and active index enhancement (IndEnh-CORE). In extensive empirical evaluations, CORE portfolios consistently outperform a broad set of benchmarks including those combining nonlinear shrinkage of the covariance matrix with Bayes-Stein mean shrinkage, as well as factor-augmented alternatives, both gross and net of transaction costs.

Based on joint work with Leheng Chen and Yingying Li.

Group sequential one-sample location testing for high dimensions

Masaki Hoshino[*,a], **Kento Egashira**[b], **Hikaru Yamaguchi**[b]

[a]**Graduate School of Science and Technology, Tokyo University of Science**
[b]**Department of Information Science and Technology, Tokyo University of Science**

Recent advances in measurement and information technologies have made high-dimensional data increasingly common in fields such as genomics and medical imaging. When the dimension is comparable to or larger than the sample size, classical statistical procedures can become difficult to apply. A fundamental problem is testing a population mean vector. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent and identically distributed d-dimensional random vectors with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. We consider

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \qquad \text{versus} \qquad H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0,$$

where $\boldsymbol{\mu}_0$ is a prespecified vector. Hotelling's T^2 test is not directly applicable when $d \geq n$, because the sample covariance matrix is singular. Quadratic-form approaches that avoid covariance matrix inversion, such as the two-sample test of Bai and Saranadasa [1], motivate the construction of our statistic in the one-sample setting.

The asymptotic behavior of high-dimensional test statistics depends critically on the eigenvalue structure of $\boldsymbol{\Sigma}$. Let $\lambda_{(1)} \geq \cdots \geq \lambda_{(d)} \geq 0$ denote the eigenvalues of $\boldsymbol{\Sigma}$, and assume that $\lambda_{(1)} > 0$. Aoshima and Yata [2] distinguished the non-strongly spiked eigenvalue (NSSE) model, under which $\lambda_{(1)}^2/\mathrm{tr}(\boldsymbol{\Sigma}^2) \to 0$, from the strongly spiked eigenvalue (SSE) model, under which $\liminf_{d\to\infty} \lambda_{(1)}^2/\mathrm{tr}(\boldsymbol{\Sigma}^2) > 0$. Thus, the leading eigenvalue makes an asymptotically negligible contribution to the sum of squared eigenvalues under NSSE, but a non-negligible contribution under SSE. Our focus is the uni-strongly spiked eigenvalue (uni-SSE) model of Ishii et al. [3], characterized by

$$\frac{\sum_{i=2}^{d} \lambda_{(i)}^2}{\lambda_{(1)}^2} \to 0 \quad \text{as} \quad d \to \infty.$$

Here, the leading eigencomponent dominates the remaining components in terms of squared eigenvalues. This dominance can substantially alter the behavior of mean-vector test statistics, requiring inference that explicitly accounts for the covariance structure.

When observations become available sequentially, group sequential testing allows the hypothesis to be evaluated at several prespecified interim sample sizes. The procedure can terminate early when sufficient evidence against the null hypothesis has accumulated, while controlling the overall type I error rate. In constructing the proposed test, we accounted for the dependence among statistics based on cumulative observations. We used an O'Brien–Fleming-type α-spending function within the framework of Lan and DeMets [4]. Extending this framework to high-dimensional mean testing requires consideration of both the covariance structure and the dependence across interim analyses.

In our presentation, we introduced a one-sample group sequential test for a high-dimensional mean vector under the uni-SSE model. The proposed statistic is standardized using an estimator of the leading eigenvalue obtained by the cross-data-matrix (CDM) method of Yata and Aoshima [5]. We discussed the joint asymptotic behavior of the test statistics under the null hypothesis at multiple analysis stages as the dimension tends to infinity. Under suitable regularity conditions, this analysis provides a basis for constructing rejection boundaries that asymptotically control the overall type I error rate. The procedure therefore incorporates the dominant eigencomponent into both the standardization of the statistic and the calibration of sequential rejection boundaries.

We also presented numerical results on finite-sample performance, comparing the proposed procedure with the one-sample test T_{CDM} developed under the NSSE model by Aoshima and Yata [6] and the one-sample test T_{DT} developed under the SSE model by Ishii et al. [7]. Under the uni-SSE settings considered, the results showed that the type I error rate of T_{CDM} could deviate from the nominal significance level, whereas the proposed procedure provided more accurate control. Furthermore, when the mean difference was aligned with the leading eigenvector, T_{DT} had lower power than the proposed procedure. These findings support the usefulness of incorporating the dominant eigencomponent when constructing group sequential tests for high-dimensional mean vectors.

Acknowledgments
This research of the second author was partially supported by Grant-in-Aid for Early-Career Scientists, JSPS, under Contract Number 24K20748. This research of the third author was partially supported by Grant-in-Aid for Early-Career Scientists, JSPS, under Contract Number 24K16967.

References

[1] Bai, Z., Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica*, 6, 311–329.

[2] Aoshima, M., Yata, K. (2018). Two-sample tests for high-dimension, strongly spiked eigenvalue models. *Statistica Sinica*, 28, 43–62.

[3] Ishii, A., Yata, K., Aoshima, M. (2021). Hypothesis tests for high-dimensional covariance structures. *Annals of the Institute of Statistical Mathematics*, 73, 599–622.

[4] Lan, K. K. G., DeMets, D. L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika*, 70, 659–663.

[5] Yata, K., Aoshima, M. (2010). Effective PCA for high-dimension, low-sample-size data with singular value decomposition of cross data matrix. *Journal of Multivariate Analysis*, 101, 2060–2077.

[6] Aoshima, M., Yata, K. (2015). Asymptotic normality for inference on multisample, high-dimensional mean vectors under mild conditions. *Methodology and Computing in Applied Probability*, 17, 419–439.

[7] Ishii, A., Yata, K., Aoshima, M. (2019). Inference on high-dimensional mean vectors under the strongly spiked eigenvalue model. *Japanese Journal of Statistics and Data Science*, 2, 105–128.

Testing and segmentation of joint and individual components in integrative multi-source factor models

Kyoowon Kim[1], Sungkyu Jung[2]

[1]Department of Statistics, Seoul National University, kwkim1224@snu.ac.kr

[2]Department of Statistics and Institute for Data Innovation in Science, Seoul National University, sungkyu@snu.ac.kr

Abstract

Disentangling shared (joint) structures from source-specific (individual) variations is a fundamental task in multi-source data integration. Existing joint–individual models often rely on computationally intensive optimization or loose spectral bounds, leading to suboptimal separation accuracy and poor scalability. In this paper, we propose the Multi-Source Sequential Alignment Test (MSSAT). MSSAT leverages the geometric observation that true joint components manifest as closely aligned score subspaces across different data sources. By deriving the asymptotic null distribution of our alignment statistic, we develop a rigorous, resampling-free sequential testing procedure to accurately estimate the joint rank. Extensive simulations and real data applications, including a TCGA multi-omics dataset, demonstrate that MSSAT achieves superior separation accuracy and substantially faster computation compared to competing methods.

関数型自己回帰モデルの作用素学習

大阪大学大学院基礎工学研究科 博士後期課程 2 年 鈴木 俊太郎
大阪大学基礎工学研究科, 理研 AIP 寺田吉壱

1 概要

関数空間 $\mathcal{H} := \{f : [0,1]^2 \to \mathbb{R} \mid \int_{[0,1]^2} |f(u)|^2\, du < \infty\}$ 上の時系列モデルとして以下の非線形関数型自己回帰 (NFAR) を導入する：

$$X_{t+1}(u) = \Psi_0(X_t)(u) + \xi_t(u), \quad u \in [0,1]^d, t \in \mathbb{N}. \tag{1.1}$$

ここで，$\Psi_0 : \mathcal{H} \to \mathcal{H}$ は非線形作用素である．上記の Ψ_0 を推定する問題は, 機械学習において注目を浴びている作用素学習の時系列予測などを含み, 応用上重要な問題である．一方で, NFAR 内の Ψ_0 の推定を扱った先行研究は少なく理論的な基盤も整っていない．例えば，代表的な先行研究である Zhu and Politis (2017) [2] は，Ψ_0 に対するカーネル推定量を提案し，NFAR の指数 β-mixing で推定量の漸近正規性を示しているが, NFAR がこの条件を満たすための十分条件については議論されていない. 加えて, 無限次元値の確率過程においてはこの条件は非自明なものになる．そのため，本研究では無限次元においても有限次元の mixing の理論である Hairer and Mattingly (2011) [1] の結果を適用できる特殊な条件を考察し，NFAR が指数 β-mixing を満たすための十分条件を明らかにする．また，この性質の応用例として，NFAR の Ψ_0 の学習に対して作用素学習を用いた場合のオラクル不等式を導出し, 特定の作用素学習モデルに対する汎化誤差を導出する．

2 必要な記号の定義

作用素 $\Phi : \mathcal{H} \to \mathcal{H}$ にたいして $\mathrm{Im}(\Psi)$ によってその値域を表す. また, $\{\xi_t\}_{t\in\mathbb{Z}}$ はそれぞれが平均が 0 の $\mathcal{H}$-値ガウス確率要素である. 任意の $t \in \mathbb{N}$ に対して ξ_t の共分散関数 $k : [0,1]^d \times [0,1]^d \to \mathbb{R}$ を $k(u,v) := \mathbb{E}[\xi_1(u)\xi_1(v)] \quad u, v \in [0,1]^d$ によって定める. 加えて, 共分散作用素 $\mathcal{Q} : \mathcal{H} \to \mathcal{H}$ を $f \in \mathcal{H}$ に対して, $(\mathcal{Q}f)(u) := \int_{[0,1]^d} k(u,v)f(v)dv, u \in [0,1]^d$ によって定める. さらに, FAR列 $\{X_t\}_{t\in\mathbb{N}}$ の β-mixing 係数を, $\beta(j) := \sup_{n\in\mathbb{N}} \mathbb{E}\Big[\sup_{B\in\mathcal{F}_{j+n}^{\infty}} |\mathbb{P}(B) - \mathbb{P}(B \mid X_n, \dots, X_1)|\Big]$ によって定める. ただし, $\mathcal{F}_{j+n}^{\infty} := \sigma(X_{j+n}, X_{j+n+1}, \dots)$ である.

3 主結果

Theorem 3.1. 適当な正則条件と以下の 2 条件を仮定する:

(A1) $\exists m_0 : \mathcal{H} \to \mathcal{H}$ such that $\Psi_0 = Q \circ m_0$.

(A2) $\exists c_1, c_2 > 0$ with $c_1 \|Q\|_{\mathrm{op}} < 1$ such that $\forall x \in \mathcal{H} : \|m_0(x)\|_{\mathcal{H}} \le c_1\|x\|_{\mathcal{H}} + c_2$.

定数 $C_1, C_2 > 0$ が存在して, $\beta(j) \le C_1 \exp(-C_2 j)$ が成り立つ.

以降は, (1.1) の予測において作用素学習を活用した場合のオラクル不等式を与える. 作用素 Ψ_0 が以下のフォームをもつ場合に注目する.

$$\Psi_0(x)(u) := \int_{[0,1]^d} \psi_0(u, v, X(v))dv, x \in \mathcal{H}, u \in [0,1]^d$$

ただし, $\psi_0 : [0,1]^d \times [0,1]^d \times \mathbb{R} \to \mathbb{R}$ である. 本節では, 関数 ψ_0 を推定することで作用素 Ψ_0 を学習する作用素学習モデルを提案する. モデル (1.1) にしたがう観測 $\{X_t\}_{t=1}^{T+1}$ が与えられているとする. 損失関数 $L(\psi)$ を $L(\psi) := \sum_{t=1}^{T-1} \int \{X_{t+1}(u) - \int \psi(u, v, X_t(v))dv\}^2 du$ によって定める. 損失関数 $L(\psi)$ を用いて推定量 $\widehat{\psi}_T$ を $\widehat{\psi}_T := \arg\min_{\psi \in \mathcal{F}} L(\psi)$ によって定める. ここで, $\mathcal{F}$ は ψ の仮説集合である. $\psi \in \mathcal{F}$ に対応する作用素 Ψ_ψ を

$$\Psi_\psi(x)(u) := \int_{[0,1]^d} \psi(u, v, x(v))dv \mathbf{1}_{\{\|x\|_\infty \leq 1\}}, u \in [0,1]^d, x \in \mathcal{H},$$

によって定める. そして, 推定量 $\widehat{\psi}_T$ を用いて Ψ_0 に対する推定量 $\widehat{\Psi}_T$ を $\widehat{\Psi}_T := \Psi_{\widehat{\psi}}$ によって定める. 以降は, ここまで導入した推定量の評価指標を導入する. 作用素 $\Psi : \mathcal{H} \to \mathcal{H}$ の汎化誤差 $R(\Psi, \Psi_0^*)$ を $R(\Psi, \Psi_0^*) := \mathbb{E}\left[\|\Psi(X^*) - \Psi_0^*(X^*)\|_{\mathcal{H}}^2\right]$, によって定める. ただし, $\Psi_0^* : \mathcal{H} \to \mathcal{H}$ は, $\Psi_0^*(x) := \Psi_0(x)\mathbf{1}_{\{\|x\|_\infty \leq 1\}}$ によって定められる作用素で, X^* は $\{X_t\}_{t=1}^T$ と独立で同分布なコピーである. さらに, 作用素学習モデルの複雑さを表す指標として, covering number $\mathcal{N}(\delta, (\Psi_\psi)_{\psi \in \mathcal{F}}, \|\cdot\|_\infty)$ を導入する. ここで, 作用素 Ψ に対してノルム $\|\Psi\|_\infty$ は, $\|\Psi\|_\infty := \sup_{x \in \mathcal{H}} \|\Psi(x)\|_{\mathcal{H}}$ によって定める.

Theorem 3.2. 適当な正則な条件と FAR が指数 mixing であることを仮定する. $(\Psi_\psi)_{\psi \in \mathcal{F}}$ が適当な条件をみたすとき $R(\Psi_{\widehat{\psi}_T}, \Psi_0^*) \lesssim \inf_{\psi \in \mathcal{F}} R(\Psi_0^*, \Psi_\psi) + (\log \mathcal{N}_T)(\log T)T^{-1}$ がなりたつ. ここで, $\mathcal{N}_T := \mathcal{N}(1/T, (\Psi_\psi)_{\psi \in \mathcal{F}}, \|\cdot\|_\infty)$.

さらに, 定理 3.2 を用いて $\mathcal{F}$ とした場合の汎化誤差を導く. 真の関数 ψ_0 が属すクラスを定めるために記号を追加する. $\mathcal{C}^\beta(Q)$ を滑らかさ β, 上界 Q のヘルダークラスであるとし, 合成関数のクラス $\mathcal{G}_{\text{comp}}(q, d, t, \beta, 1)$ を, $\{g_q \circ \cdots \circ g_0 | g_i = (g_{ij})_j,\ g_i : [a_i, b_i]^{d_i} \to [a_{i+1}, b_{i+1}]^{d_{i+1}}, g_{ij} \in \mathcal{C}^{\beta_i}([a_i, b_i]^{t_i}, Q), \text{for some } |a_i|, |b_i| \leq Q\}$ によって定める.

Corollary 3.1. NFAR が指数 mixing であることと適切な正則条件を仮定する. ハイパーパラメタを適切に定め, $\psi_0 \in \mathcal{G}_{\text{comp}}(q, d, t, \beta, 1)$ とすると, $R(\Psi_{\widehat{\psi}_T}, \Psi_0^*) \lesssim \min_{i=0,\ldots,d} T^{-\frac{\beta_i}{t_i + \beta_i}} (\log T)^4$.

References

[1] M. Hairer and J. C. Mattingly. "Yet another look at Harris' ergodic theorem for Markov chains". In: *Seminar on Stochastic Analysis, Random Fields and Applications VI*. Basel: Springer, 2011, pp. 109–117.

[2] T. Zhu and D. Politis. "Kernel estimates of nonparametric functional autoregression models and their bootstrap approximation". In: *Electron. J. Stat.* 11 (2017), pp. 2876–2906.

GENERALIZATION ERROR ANALYSIS FOR LINEAR BINARY CLASSIFICATION UNDER LIPSCHITZ LOSS FUNCTIONS

SHOGO NAKAKITA

Abstract. We examine the concentration of uniform generalization errors around their expectation in binary linear classification problems via an isoperimetric argument. In particular, we establish Poincaré and log-Sobolev inequalities for the joint distribution of the output labels and the label-weighted input vectors, which we apply to derive concentration bounds. The derived results improve upon existing bounds obtained from general unbounded empirical processes, as well as that tailored specifically to logistic regression.

1. Introduction

Binary classification is one of the most fundamental tasks in machine learning and statistics. It concerns estimating a high-performing model for predicting a $\{\pm 1\}$-valued random variable Y (output label) from its corresponding d-dimensional random vector $\boldsymbol{X}$ (input vector) from a sequence of n independent and identically distributed random variables $\{(\boldsymbol{X}_i, Y_i)\}_{i=1}^n$ whose distributions are the same as $(\boldsymbol{X}, Y)$. Binary linear classifiers are fundamental models in this task; we estimate a reasonable $\hat{\boldsymbol{w}} \in \mathbb{R}^d$ by the training data and use $\mathrm{sgn}(\langle \boldsymbol{X}, \hat{\boldsymbol{w}} \rangle)$ as a prediction for the corresponding value of Y. There are multiple approaches to estimating $\hat{\boldsymbol{w}}$ from training data; in this study, we consider an empirical risk minimization problem such that

$$\text{minimize } \mathcal{R}_n(\boldsymbol{w}, b) \left(= \frac{1}{n} \sum_{i=1}^n \ell\left(Y_i \left(\langle \boldsymbol{X}_i, \boldsymbol{w} \rangle + b\right)\right) \right) \text{ such that } \|\boldsymbol{w}\| \le R_{\boldsymbol{w}},\ |b| \le R_b, \tag{1}$$

where $\ell : \mathbb{R} \to [0, +\infty)$ is an L-Lipschitz continuous function (for some $L > 0$), $\|\cdot\|$ is the Euclidean norm, $R_{\boldsymbol{w}} > 0$ is the radius of the space of $\boldsymbol{w}$, and $R_b \ge 0$ is the radius of the space of b. We use the notation $\mathcal{T} := \{\boldsymbol{w} \in \mathbb{R}^d : \|\boldsymbol{w}\| \le R_{\boldsymbol{w}}\} \times [-R_b, R_b] \subset \mathbb{R}^{d+1}$. We refer to $\mathcal{R}_n$ as the empirical risk function and ℓ as the loss function. The logistic loss function $\ell_{\mathrm{L}}(t) = \log(1 + \exp(-t))$ and hinge loss function $\ell_{\mathrm{H}}(t) = \max\{0, 1 - t\}$ serve as prime examples of Lipschitz loss functions. A basic idea of these loss functions is that their subderivatives satisfy $\ell'(t) \le 0$, and thus they penalize the difference in the signs of Y_i and $\langle \boldsymbol{X}_i, \boldsymbol{w} \rangle + b$. Hence, we can interpret the problem (1) as the selection of $(\boldsymbol{w}, b)$ so that the signs of Y_i and $\langle \boldsymbol{X}_i, \boldsymbol{w} \rangle + b$ are the same as possible.

A motivation for the empirical risk minimization (1) is that we expect that the problem should be close to the following population risk minimization problem:

$$\text{minimize } \mathcal{R}(\boldsymbol{w}, b) \left(= \mathbb{E}\left[\ell\left(Y(\langle \boldsymbol{X}, \boldsymbol{w} \rangle + b)\right)\right] \right) \text{ such that } \|\boldsymbol{w}\| \le R_{\boldsymbol{w}},\ |b| \le R_b, \tag{2}$$

where $\mathcal{R}(\boldsymbol{w}, b)$ is referred to as a population risk function. Under a mild condition, we know that $\mathcal{R}_n(\boldsymbol{w}, b) \to \mathcal{R}(\boldsymbol{w}, b)$ almost surely for every $(\boldsymbol{w}, b) \in \mathbb{R}^{d+1}$, which is a *pointwise* law of large numbers. However, we need a *uniform* law on $\mathcal{T}$ to verify that the optimization problem (1) is indeed close to (2). In particular, the following type of concentration inequalities for given $\delta \in (0, 1]$ has been widely used to characterize the discrepancy between them in studies of machine learning and statistics:

$$\mathbb{P}\left(\sup_{(\boldsymbol{w}, b) \in \mathcal{T}} \left(\mathcal{R}(\boldsymbol{w}, b) - \mathcal{R}_n(\boldsymbol{w}, b)\right) \ge (\text{residual dependent on } (\delta, n, R_{\boldsymbol{w}}, R_b)) \right) \le \delta.$$

This work is supported by JSPS KAKENHI Grant Number JP24K02904. The author used ChatGPT and Google Gemini to improve the grammatical accuracy of the manuscript, as a non-native English speaker, which is checked and corrected by the author.

The term $\sup_{(\boldsymbol{w},b)\in\mathcal{T}}(\mathcal{R}(\boldsymbol{w},b)-\mathcal{R}_n(\boldsymbol{w},b))$ is referred to as a *uniform generalization error* or *worst-case generalization error*. If the residual dependent on $(\delta, n, R_{\boldsymbol{w}}, R_b)$ takes small values even for small $\delta > 0$, we conclude that the problem (1) is close to (2).

In this work, we analyse the discrepancy between the problems (1) and (2) by showing the following concentration of a uniform generalization error around its expectation:

$$\mathbb{P}\left(\sup_{(\boldsymbol{w},b)\in\mathcal{T}}(\mathcal{R}(\boldsymbol{w},b)-\mathcal{R}_n(\boldsymbol{w},b)) \geq \mathbb{E}\left[\sup_{(\boldsymbol{w},b)\in\mathcal{T}}(\mathcal{R}(\boldsymbol{w},b)-\mathcal{R}_n(\boldsymbol{w},b))\right] + (\text{residual})_{\delta,n,R_{\boldsymbol{w}},R_b}\right) \leq \delta. \quad (3)$$

Given a large n, we show that uniform generalization errors concentrate around their expectation in many settings, and the residual terms tend to zero. Hence, our strategy for analysing a uniform generalization error is a divide-and-conquer one; we characterize the behaviour of a uniform generalization error in terms of its expectation, which may be small or large. If both the expectation and residual converge to zero, then we conclude that the empirical risk minimization problem (1) is a good alternative to the population risk minimization problem (2). On the other hand, if the residual converges to zero, but the expectation does not, then the empirical risk minimization problem does not serve as a good approximation of the population risk minimization, and we can identify that the cause of the difficulty is not the random behaviour of uniform generalization errors but the essential difficulty represented by expected uniform generalization errors.

Such a strategy is reasonable and even classical since the evaluation of expected uniform generalization errors is a classical problem in statistical learning theory. For example, the Rademacher complexity argument is a very established strategy to give estimates for the expectation, and it is straightforward to derive such a bound in our settings (see Proposition **??**). Therefore, estimates of the residual term are crucial for evaluating the stochastic behaviour of uniform generalization errors precisely.

The boundedness of loss functions is a widely employed assumption for deriving clean estimates of the residuals. Under this assumption, one can directly apply McDiarmid's inequality to obtain an explicit and simple residual term $\|\ell\|_\infty\sqrt{\log(\delta^{-1})/2n}$ (e.g., see Bach, 2024). On the other hand, although the boundedness of ℓ yields a tractable statistical complexity estimate for the problem, this assumption excludes typical loss functions that render the empirical risk minimization problem (1) a convex optimization problem. Intuitively, this assumption on ℓ makes the problem statistically tractable but computationally intractable.

Therefore, we are motivated to examine the concentration of uniform generalization errors under unbounded loss functions that render $\mathcal{R}_n(\boldsymbol{w}, b)$ convex. The logistic and hinge loss functions are typical examples of such loss functions; they correspond to logistic regression and the soft-margin support vector machine, which are central topics in binary linear classification within machine learning. These problems are computationally tractable, but their statistical complexities have been scarcely investigated. Although the concentration for general unbounded loss functions has been studied (Adamczak, 2008; van de Geer and Lederer, 2013; Lederer and van de Geer, 2014), previous bounds centre around the scaled expectations of uniform generalization errors and contain residual terms that may diverge in realistic scenarios.In particular, since the motivating loss functions are Lipschitz continuous, we focus on Lipschitz continuous ℓ to circumvent the difficulties encountered in previous studies on empirical processes with unbounded loss functions.

References

Adamczak, R. (2008). A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.*, 13:1000–1034.

Bach, F. (2024). *Learning Theory from First Principles*. MIT Press.

Lederer, J. and van de Geer, S. (2014). New concentration inequalities for suprema of empirical processes. *Bernoulli*, 20(4):2020–2038.

van de Geer, S. and Lederer, J. (2013). The Bernstein–Orlicz norm and deviation inequalities. *Probab. Theory Related Fields*, 157(1-2):225–250.

Asymptotically normal estimators in high-dimensional linear regression*

Koji Tsukuda† Kou Fujimori‡

Keywords.
high-dimensional inference; linear regression; simultaneous inference; sparsity; weak convergence

Abstract.
We establish asymptotic normality for estimators in high-dimensional linear regression through weak convergence in a separable Hilbert space, enabling direct application of standard asymptotic tools such as the continuous mapping theorem. The framework allows the number of nonzero coefficients to diverge, provided that only a fixed number have moderate magnitude. As applications, we develop testing procedures for both finite and infinite collections of linear hypotheses.

Notation.

- $\mathbb{N} = \{1, 2, \ldots\}$ and $[k] = \{1, \ldots, k\}$ for $k \in \mathbb{N}$.
- For $q \in \mathbb{N} \cup \{\infty\}$, write $\|\boldsymbol{v}\|_q$ for the ℓ^q norm: for $\boldsymbol{v} = (v_1, \ldots, v_p)^\top \in \mathbb{R}^p$ with $p \in \mathbb{N}$, $\|\boldsymbol{v}\|_q = (\sum_{j=1}^p |v_j|^q)^{1/q}$ $(q < \infty)$, $\|\boldsymbol{v}\|_\infty = \max_{1 \le j \le p} |v_j|$. For an infinite sequence $\boldsymbol{v} = (v_j)_{j \ge 1} \in \ell^q$, the same formulae apply.
- For a vector $\boldsymbol{v} = (v_j)_{j \ge 1}$ and an index set $T \subset \mathbb{N}$ with $|T| < \infty$, write $\boldsymbol{v}_T = (v_j)_{j \in T}$, which is viewed as an element of $\mathbb{R}^{|T|}$. When $\boldsymbol{v} \in \mathbb{R}^p$, we restrict $T \subset [p]$; when $\boldsymbol{v} \in \ell^2$, we allow $T \subset \mathbb{N}$.
- For $\boldsymbol{A} = (A_{ij})_{i,j \in [p]} \in \mathbb{R}^{p \times p}$ and index sets $T, T' \subset [p]$, denote $(A_{ij})_{i \in T, j \in T'}$ by $\boldsymbol{A}_{T,T'}$.
- For a random variable X, denote by $\|X\|_{\psi_2}$ the sub-Gaussian norm of X, if it exists, that is, $\|X\|_{\psi_2} = \inf\{C > 0 : \mathbb{E}[\exp(X^2/C^2)] \le 2\}$. For a d-dimensional random vector $\boldsymbol{X}$, the sub-Gaussian norm of $\boldsymbol{X}$ is $\|\boldsymbol{X}\|_{\psi_2} = \sup_{\|\boldsymbol{u}\|_2 = 1} \|\boldsymbol{u}^\top \boldsymbol{X}\|_{\psi_2}$.
- For a random element X and a sequence $\{X_n\}_{n \in \mathbb{N}}$ of random elements, we write $X_n \to^d X$ and $X_n \to^p X$ for convergence in distribution and in probability as $n \to \infty$, respectively.

*This work was supported by JSPS KAKENHI Grant Numbers 25K07133 (KT) and 26K17032 (KF).
†Faculty of Mathematics, Kyushu University
‡Faculty of Economics and Law, Shinshu University

Main result. (Fujimori and Tsukuda, 2026)
For each $n \in \mathbb{N}$, we work with the triangular array $\{(Y_{i,n}, \boldsymbol{Z}_{i,n}^\top)\}_{i\in[n]}$ satisfying

$$Y_{i,n} = \boldsymbol{\theta}_n^\top \boldsymbol{Z}_{i,n} + \epsilon_{i,n}, \quad i \in [n],$$

where $\boldsymbol{\theta}_n \in \mathbb{R}^{p_n}$ and $\boldsymbol{Z}_{i,n} = (Z_{i1,n}, \ldots, Z_{ip_n,n})^\top \in \mathbb{R}^{p_n}$ for $i \in [n]$. Assume that $\{(\boldsymbol{Z}_{i,n}^\top, \epsilon_{i,n})^\top\}_{i=1}^n$ are i.i.d. with $\mathbb{E}(\boldsymbol{Z}_{1,n}) = \boldsymbol{0}$, $\mathbb{E}(\epsilon_{1,n}) = 0$, $\mathrm{Var}(\epsilon_{1,n}) = \sigma^2 > 0$, and $\boldsymbol{Z}_{1,n}$ is independent of $\epsilon_{1,n}$. Let $\boldsymbol{\theta}_{0,n} = (\theta_{01,n}, \ldots, \theta_{0p_n,n})^\top$ denote the true parameter vector. Define the index sets

$$\begin{aligned} T_\gamma &:= \{j \in [p_n] : |\theta_{0j,n}| > \gamma\}, \\ T_{0,n} &:= \{j \in [p_n] : \theta_{0j,n} \neq 0\}, \\ T_{W,n} &:= \{j \in [p_n] : \theta_{\min,n} \leq |\theta_{0j,n}| \leq \eta_n\}, \end{aligned}$$

where $\gamma > 0$ is a fixed constant independent of n and p_n, $\theta_{\min,n} := \min_{j\in T_{0,n}} |\theta_{0j,n}|$, and η_n is a positive sequence depending on n. We write $s_{0,n} := |T_{0,n}|$ for the support size, which may grow with n, and $s_\gamma := |T_\gamma|$ for the number of coefficients larger than γ in magnitude. Unless otherwise stated, throughout this material, all limits are taken as $n \to \infty$, allowing p_n to diverge.

Assumption 1. (i) T_γ is fixed and independent of n and p_n.

(ii) $T_{0,n} \setminus T_\gamma = T_{W,n}$.

(iii) There exists $\theta^* > 0$, independent of n and p_n, such that $\|\boldsymbol{\theta}_{0,n}\|_\infty \leq \theta^*$.

(iv) σ^2 is fixed and independent of n and p_n.

Let $\hat{\boldsymbol{\theta}}_n = (\hat{\theta}_{n1}, \ldots, \hat{\theta}_{np_n})^\top$ be a sparse estimator for $\boldsymbol{\theta}_n$ such that $\Pr(\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{0,n}\|_\infty > r_n) \to 0$ for some rate $r_n \to 0$. Given a threshold sequence $\{\tau_n\}_{n\in\mathbb{N}}$, define the sequence $\{\hat{T}_n\}_{n\in\mathbb{N}}$ of an estimator for T_γ by

$$\hat{T}_n := \{j \in [p_n] : |\hat{\theta}_{nj}| > \tau_n\}.$$

Remark 1. Under suitable conditions, one may take $r_n \asymp n^{-1/2}(s_{0,n} \log p_n)^{1/2}$.

Assumption 2. $r_n + \eta_n < \tau_n < \gamma - r_n$.

Lemma 1. *Let Assumptions 1–2 hold. Then,* $\Pr(\hat{T}_n = T_\gamma) \to 1$.

The post-selection ordinary least squares estimator $\tilde{\boldsymbol{\theta}}_n = (\tilde{\theta}_{n1}, \ldots, \tilde{\theta}_{np_n})^\top$ is defined as the solution to the estimating equation

$$n^{-1} \sum_{i=1}^n \boldsymbol{Z}_{i,n\hat{T}_n}(Y_{i,n} - \boldsymbol{\theta}_{\hat{T}_n}^\top \boldsymbol{Z}_{i,n\hat{T}_n}) = \boldsymbol{0}, \quad \boldsymbol{\theta}_{\hat{T}_n^c} = \boldsymbol{0}.$$

Consider the ℓ^2-valued random sequence $\{\boldsymbol{\mathcal{R}}_n\}_{n\in\mathbb{N}}$ defined by

$$\langle \boldsymbol{e}_j, \boldsymbol{\mathcal{R}}_n \rangle = \begin{cases} n^{1/2}(\tilde{\theta}_{nj} - \theta_{0j,n}) & (j \in [p_n]) \\ 0 & (j \in \mathbb{N} \setminus [p_n]) \end{cases},$$

where $\{\boldsymbol{e}_j\}_{j\in\mathbb{N}}$ is the canonical basis of ℓ^2. Define $\boldsymbol{J}_n := \mathbb{E}[\boldsymbol{Z}_{1,n}\boldsymbol{Z}_{1,n}^\top]$.

Assumption 3. (i) There exists a $s_\gamma \times s_\gamma$ positive definite matrix $\boldsymbol{J}_{T_\gamma,T_\gamma} = \lim(\boldsymbol{J}_n)_{T_\gamma,T_\gamma}$.

(ii) There exists λ_*, independent of n and p_n, such that $0 < \lambda_* \leq \Lambda_{\min}((\boldsymbol{J}_n)_{T_\gamma,T_\gamma})$, where $\Lambda_{\min}((\boldsymbol{J}_n)_{T_\gamma,T_\gamma})$ is the smallest eigenvalue of $(\boldsymbol{J}_n)_{T_\gamma,T_\gamma}$.

(iii) $\sup_{n\in\mathbb{N}} \max_{j\in[p_n]} \|Z_{1j,n}\|_{\psi_2} < \infty$.

(iv) $\sup_{n\in\mathbb{N}} \|\epsilon_{1,n}\|_{\psi_2} < \infty$.

Assumption 4. There exists $\kappa \in [0, 1/2]$, independent of n and p_n, such that $\sup_{n\in\mathbb{N}} s_{0,n}^{-\kappa} \|\boldsymbol{Z}_{1,nT_{0,n}}\|_{\psi_2} < \infty$ and $n s_{0,n}^{1+2\kappa} \eta_n^2 \to 0$.

Theorem 2. *Let Assumptions 1–4 hold. Then,* $\boldsymbol{\mathcal{R}}_n \to^d \boldsymbol{\mathcal{R}}$ *in* ℓ^2*, where* $\boldsymbol{\mathcal{R}}$ *is a centered Gaussian field satisfying* $\langle \boldsymbol{u}, \boldsymbol{\mathcal{R}} \rangle \sim \mathrm{N}(0, \sigma^2 \boldsymbol{u}_{T_\gamma}^\top (\boldsymbol{J}_{T_\gamma,T_\gamma})^{-1} \boldsymbol{u}_{T_\gamma})$ *for every* $\boldsymbol{u} \in \ell^2$.

References

K. Fujimori, K. Tsukuda (2026). Asymptotically normal estimators in high-dimensional linear regression. arXiv:2602.07480

Testing homogeneity of high-dimensional covariance matrices under a low-dimensional factor structure

Takahiro Nishiyama[a], Masashi Hyodo[b], Shoma Yamashita[c], Nobumichi Shutoh[d]

[a] Faculty of Science, Tokyo University of Science
[b] Faculty of Economics, Kanagawa University
[c] Graduate School of Science, Kobe University
[d] Laboratory of Medical Data Sciences, Kobe Pharmaceutical University / Graduate School of Science, Kobe University

Suppose that a p-dimensional random sample of size n_g is represented by $\mathbf{x}_{g1}, \ldots, \mathbf{x}_{gn_g}$, where the observations are independently drawn from population Π_g $(g = 1, \ldots, k)$ with mean vector $\boldsymbol{\mu}_g = \mathrm{E}(\mathbf{x}_{gi})$ and covariance matrix $\boldsymbol{\Sigma}_g = \mathrm{cov}(\mathbf{x}_{gi})$. We consider the k-sample hypothesis

$$\mathcal{H} : \boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_k \quad \text{versus} \quad \mathcal{A} : \neg\mathcal{H}. \tag{1}$$

Note that hypothesis (1) is equivalent to

$$\mathcal{H} : \|\boldsymbol{\Delta}_{gh}\|_F^2 = 0 \quad \text{for all} \quad g < h,$$

where $\boldsymbol{\Delta}_{gh} = \boldsymbol{\Sigma}_g - \boldsymbol{\Sigma}_h$ and $\|\cdot\|_F$ denotes the Frobenius norm.

Also, the data generation model of population Π_g is assumed to be a latent factor model, which is expressed as

$$\mathbf{x}_g = \boldsymbol{\mu}_g + \mathbf{B}\mathbf{f}_g + \boldsymbol{\epsilon}_g.$$

Here, the elements of the $p \times d$ $(d < p)$ non-random matrix $\mathbf{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_p)^\top$ are referred to as factor loadings. $\mathbf{f}_g \in \mathbb{R}^d$ and $\boldsymbol{\epsilon}_g \in \mathbb{R}^p$ are random vectors for common factors and specific factors, respectively. We assume that $\mathbf{f}_g$ and $\boldsymbol{\epsilon}_g$ are independent of each other and that their mean vectors are $\mathrm{E}(\mathbf{f}_g) = \mathbf{0}$ and $\mathrm{E}(\boldsymbol{\epsilon}_g) = \mathbf{0}$, respectively. Also, we assume that $\mathrm{cov}(\mathbf{f}_g) = \boldsymbol{\Phi}_g$ is a positive definite matrix and $\mathrm{cov}(\boldsymbol{\epsilon}_g) = \boldsymbol{\Psi}_g = \mathrm{diag}(\psi_{g1}, \ldots, \psi_{gp})$. Let $\mathbf{z}_g = (\mathrm{z}_{g1}, \ldots, \mathrm{z}_{gd})^\top = \boldsymbol{\Phi}_g^{-1/2}\mathbf{f}_g$ and $\mathbf{e}_g = (\mathrm{e}_{g1}, \ldots, \mathrm{e}_{gp})^\top = \boldsymbol{\Psi}_g^{-1/2}\boldsymbol{\epsilon}_g$. Furthermore, we assume that z_{gi} are iid with $\mathrm{E}(\mathrm{z}_{gi}) = 0$, $\mathrm{E}(\mathrm{z}_{gi}^2) = 1$, and $\mathrm{E}(\mathrm{z}_{gi}^4) = \kappa_{\mathrm{z}} < \infty$, and e_{gj} are iid with $\mathrm{E}(\mathrm{e}_{gj}) = 0$, $\mathrm{E}(\mathrm{e}_{gj}^2) = 1$, and $\mathrm{E}(\mathrm{e}_{gj}^4) = \kappa_{\mathrm{e}} < \infty$ for $g \in \{1, \ldots, k\}$, $i \in \{1, \ldots, d\}$, and $j \in \{1, \ldots, p\}$.

There are several approaches for testing the equality of high-dimensional covariance matrices. Among them, Li and Chen [3] proposed a pioneering two-sample test based on the asymptotic normality of an unbiased estimator of $\|\boldsymbol{\Delta}_{gh}\|_F^2$. Their asymptotic theory relies on the weak dependence condition

$$\mathrm{tr}(\boldsymbol{\Sigma}_g\boldsymbol{\Sigma}_h\boldsymbol{\Sigma}_{g'}\boldsymbol{\Sigma}_{h'}) = o\{\mathrm{tr}(\boldsymbol{\Sigma}_g\boldsymbol{\Sigma}_h)\mathrm{tr}(\boldsymbol{\Sigma}_{g'}\boldsymbol{\Sigma}_{h'})\}, \quad g, h, g', h' \in \{1, \ldots, k\}, \quad p \to \infty,$$

which is also known as the boundary condition on the maximum eigenvalue. To overcome these limitations, for high-dimensional covariance matrix testing, Hyodo et al. [2] developed a two-sample test for the equality of covariance matrices under a low-dimensional latent factor structure. In this study, we extended the two-sample methodology of Hyodo et al. [2] to the general k-sample problem and developed a new test procerure for (1).

To construct a hypothesis test (1), we defined the test statistic as

$$\widehat{T} = \sum_{g<h}^{k} \frac{n_g n_h}{n} \left\{ \frac{\widehat{\|\boldsymbol{\Delta}_{gh}\|_F^2}}{p^2} + \sum_{\ell \in \{g,h\}} \frac{\widehat{\mathrm{tr}(\mathbf{A}_{\ell\ell}^2)} + \{\widehat{\mathrm{tr}(\mathbf{A}_{\ell\ell})}\}^2}{n_\ell} \right\},$$

where, for $g \in \{1, \ldots, k\}$,

$$\widehat{\mathrm{tr}(\mathbf{A}_{gg})} = \frac{\sum_{i=1}^{\hat{d}} \lambda_i(\mathbf{S}_g)}{p} \quad \text{and} \quad \widehat{\mathrm{tr}(\mathbf{A}_{gg}^2)} = \frac{\sum_{i=1}^{\hat{d}} \{\lambda_i(\mathbf{S}_g)\}^2}{p^2}.$$

Here, $\lambda_i(\mathbf{S}_g)$ is the i-th largest eigenvalue of matrix $\mathbf{S}_g = \{1/(n_g - 1)\} \sum_{i=1}^{n_g} (\mathbf{x}_{gi} - \overline{\mathbf{x}}_g)(\mathbf{x}_{gi} - \overline{\mathbf{x}}_g)^\top$, $\overline{\mathbf{x}}_g = (1/n_g) \sum_{i=1}^{n_g} \mathbf{x}_{gi}$ and $\hat{d}$ is a consistent estimator of d based on the ER method proposed by [1]. Also,

$$\widehat{\|\boldsymbol{\Delta}_{gh}\|_F^2} = \sum_{\ell \in \{g,h\}} \widehat{\|\boldsymbol{\Sigma}_\ell\|_F^2} - 2\mathrm{tr}(\mathbf{S}_g \mathbf{S}_h),$$

where

$$\widehat{\|\boldsymbol{\Sigma}_\ell\|_F^2} = \frac{n_\ell - 1}{n_\ell(n_\ell - 2)(n_\ell - 3)} [(n_\ell - 1)(n_\ell - 2)\mathrm{tr}(\mathbf{S}_\ell^2) + \{\mathrm{tr}(\mathbf{S}_\ell)\}^2 - n_\ell K_\ell],$$

$$K_\ell = \frac{1}{n_\ell - 1} \sum_{i=1}^{n_\ell} \|\mathbf{x}_{\ell i} - \overline{\mathbf{x}}_\ell\|^4.$$

In this study, we derived the limiting null distribution of $\widehat{T}$ under some assumptions, and constructed test procedure for testing (1). Also, we discussed the asymptotic power of our proposed test procedure under a local alternative. Besides, we compared, through simulations, the performance of the proposed test and existing procedures suitable for testing the equality of multi-sample high-dimensional covariance matrices in terms of size control and power.

References

[1] Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica* **81**, 1203–1227.

[2] Hyodo, M., Nishiyama, T., Watanabe, H., Nakagawa, T., Tahata, K. (2025). Equality tests of covariance matrices under a low-dimensional factor structure. *J. Multivariate Anal.* **206**, 105397.

[3] Li, J., Chen, S. X. (2012). Two sample tests for high-dimensional covariance matrices. *Ann. Statist.* **40**, 908–940.