ISSN 0090-5364 (print)
ISSN 2168-8966 (online)

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

Vol. 52, No. 5—October 2024

THE ANNALS OF STATISTICS Vol. 52, No. 5, pp. 1873–2481 October 2024

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

The purpose of the Institute is to foster the development and dissemination of the theory and applications of statistics and probability.

The Annals of Statistics **[ISSN 0090-5364 (print); ISSN 2168-8966 (online)],** Volume 52, Number 5, October 2024. Published bimonthly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, OH 44094, USA. Periodicals postage paid at Cleveland, Ohio, and at additional mailing offices.

POSTMASTER: Send address changes to *The Annals of Statistics*, Institute of Mathematical Statistics, Dues and Subscriptions Office, PO Box 729, Middletown, MD 21769, USA.

The Annals of Statistics
2024, Vol. 52, No. 5, 1873–1898
https://doi.org/10.1214/24-AOS2375

SPECTRAL STATISTICS OF SAMPLE BLOCK CORRELATION MATRICES

BY ZHIGANG BAO[1,a], JIANG HU[2,b], XIAOCONG XU[3,d] AND XIAOZHUO ZHANG[2,c]

[1]*Department of Mathematics, University of Hong Kong,* [a]*zgbao@hku.hk*

[2]*Key Laboratory of Applied Statistics of MOE, School of Mathematics and Statistics, Northeast Normal University,* [b]*huj156@nenu.edu.cn,* [c]*zhangxz722@nenu.edu.cn*

[3]*Department of Mathematics, Hong Kong University of Science and Technology,* [d]*xxuay@connect.ust.hk*

A fundamental concept in multivariate statistics, the sample correlation matrix, is often used to infer the correlation/dependence structure among random variables, when the population mean and covariance are unknown. A natural block extension of it, the *sample block correlation matrix*, is proposed to take on the same role, when random variables are generalized to random subvectors. In this paper, we establish a spectral theory of the sample block correlation matrices and apply it to group independent tests and related problems, under the high-dimensional setting. More specifically, we consider a random vector of dimension p, consisting of k subvectors of dimension p_t's, where p_t's can vary from 1 to order p. Our primary goal is to investigate the dependence of the k subvectors. We construct a random matrix model called sample block correlation matrix based on N samples for this purpose. The spectral statistics of the sample block correlation matrix include the classical Wilks' statistic and Schott's statistic as special cases. It turns out that the spectral statistics do not depend on the unknown population mean and covariance, under the null hypothesis that the subvectors are independent. Further, the limiting behavior of the spectral statistics can be described with the aid of the free probability theory. Specifically, under three different settings of possibly N-dependent k and p_t's, we show that the empirical spectral distribution of the sample block correlation matrix converges to the free Poisson binomial distribution, free Poisson distribution (Marchenko–Pastur law) and free Gaussian distribution (semicircle law), respectively. We then further derive the CLTs for the linear spectral statistics of the block correlation matrix under a general setting. Our results are established under the general distribution assumption on the random vector. It turns out that the CLTs are universal and do not depend on the 4th cumulants of the vector components, due to a self-normalizing effect of the correlation-type matrices. We further derive the CLT under the alternative hypothesis and discuss the power of our statistics. Based on our theory, real data analysis on stock return data and gene data is also conducted.

REFERENCES

[1] AKHIEZER, N. I. (2021). *The Classical Moment Problem and Some Related Questions in Analysis*. *Classics in Applied Mathematics* **82**. SIAM, Philadelphia, PA. MR4191205

[2] BAI, Z. D. and SILVERSTEIN, J. W. (2008). CLT for linear spectral statistics of large-dimensional sample covariance matrices. In *Advances in Statistics* 281–333. World Scientific, Singapore.

[3] BAI, Z. D. and YAO, J. (2005). On the convergence of the spectral empirical process of Wigner matrices. *Bernoulli* **11** 1059–1092. MR2189081 https://doi.org/10.3150/bj/1137421640

[4] BAO, Z., ERDŐS, L. and SCHNELLI, K. (2016). Local stability of the free additive convolution. *J. Funct. Anal*. **271** 672–719. MR3506962 https://doi.org/10.1016/j.jfa.2016.04.006

[5] BAO, Z., ERDŐS, L. and SCHNELLI, K. (2020). Spectral rigidity for addition of random matrices at the regular edge. *J. Funct. Anal*. **279** 108639. MR4102163 https://doi.org/10.1016/j.jfa.2020.108639

MSC2020 subject classifications. Primary 60B20, 62E20; secondary 60E05.

Key words and phrases. Random matrix, spectral statistics, free probability theory, central limit theorem, group independence test.

[6] BAO, Z. and HE, Y. (2023). Quantitative CLT for linear eigenvalue statistics of Wigner matrices. *Ann. Appl. Probab.* **33** 5171–5207. MR4677731 https://doi.org/10.1214/23-aap1945

[7] BAO, Z., HU, J., PAN, G. and ZHOU, W. (2017). Test of independence for high-dimensional random vectors based on freeness in block correlation matrices. *Electron. J. Stat.* **11** 1527–1548. MR3635921 https://doi.org/10.1214/17-EJS1259

[8] BAO, Z., HU, J., XU, X. and ZHANG, X. (2024). Supplement to "Spectral statistics of sample block correlation matrices." https://doi.org/10.1214/24-AOS2375SUPP

[9] BAO, Z., LIN, L.-C., PAN, G. and ZHOU, W. (2015). Spectral statistics of large dimensional Spearman's rank correlation matrix and its application. *Ann. Statist.* **43** 2588–2623. MR3405605 https://doi.org/10.1214/15-AOS1353

[10] BELINSCHI, S. T. and BERCOVICI, H. (2007). A new approach to subordination results in free probability. *J. Anal. Math.* **101** 357–365. MR2346550 https://doi.org/10.1007/s11854-007-0013-1

[11] BI, D., HAN, X., NIE, A. and YANG, Y. Spiked Eigenvalues of High-Dimensional Sample Autocovariance Matrices: CLT and Applications.

[12] BICKEL, P. J. and LEVINA, E. (2008). Regularized estimation of large covariance matrices. *Ann. Statist.* **36** 199–227. MR2387969 https://doi.org/10.1214/009053607000000758

[13] BODNAR, T., DETTE, H. and PAROLYA, N. (2019). Testing for independence of large dimensional vectors. *Ann. Statist.* **47** 2977–3008. MR3988779 https://doi.org/10.1214/18-AOS1771

[14] CAI, T., LIU, W. and LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.* **106** 594–607. MR2847973 https://doi.org/10.1198/jasa.2011.tm10155

[15] CAI, T. T., ZHANG, C.-H. and ZHOU, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* **38** 2118–2144. MR2676885 https://doi.org/10.1214/09-AOS752

[16] CHEN, H. and JIANG, T. (2018). A study of two high-dimensional likelihood ratio tests under alternative hypotheses. *Random Matrices Theory Appl.* **7** 1750016. MR3756424 https://doi.org/10.1142/S2010326317500162

[17] CHISTYAKOV, G. P. and GÖTZE, F. (2011). The arithmetic of distributions in free probability theory. *Cent. Eur. J. Math.* **9** 997–1050. MR2824443 https://doi.org/10.2478/s11533-011-0049-4

[18] CIPOLLONI, G., ERDŐS, L. and SCHRÖDER, D. (2023). Functional central limit theorems for Wigner matrices. *Ann. Appl. Probab.* **33** 447–489. MR4551555 https://doi.org/10.1214/22-aap1820

[19] DETTE, H. and DÖRNEMANN, N. (2020). Likelihood ratio tests for many groups in high dimensions. *J. Multivariate Anal.* **178** 104605. MR4079037 https://doi.org/10.1016/j.jmva.2020.104605

[20] DEVIJVER, E. and GALLOPIN, M. (2018). Block-diagonal covariance selection for high-dimensional Gaussian graphical models. *J. Amer. Statist. Assoc.* **113** 306–314. MR3803466 https://doi.org/10.1080/01621459.2016.1247002

[21] DONOHO, D., GAVISH, M. and JOHNSTONE, I. (2018). Optimal shrinkage of eigenvalues in the spiked covariance model. *Ann. Statist.* **46** 1742–1778. MR3819116 https://doi.org/10.1214/17-AOS1601

[22] DÖRNEMANN, N. (2023). Likelihood ratio tests under model misspecification in high dimensions. *J. Multivariate Anal.* **193** Paper No. 105122. MR4504582 https://doi.org/10.1016/j.jmva.2022.105122

[23] DÖRNEMANN, N. and HEINY, J. (2022). Limiting spectral distribution for large sample correlation matrices. ArXiv preprint. Available at arXiv:2208.14948.

[24] EL KAROUI, N. (2008). Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Ann. Statist.* **36** 2717–2756. MR2485011 https://doi.org/10.1214/07-AOS559

[25] EL KAROUI, N. (2009). Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond. *Ann. Appl. Probab.* **19** 2362–2405. MR2588248 https://doi.org/10.1214/08-AAP548

[26] ERDŐS, L., KNOWLES, A. and YAU, H.-T. (2013). Averaging fluctuations in resolvents of random band matrices. *Ann. Henri Poincaré* **14** 1837–1926. MR3119922 https://doi.org/10.1007/s00023-013-0235-y

[27] FAMA, E. F. and FRENCH, K. R. (1997). Industry costs of equity. *J. Financ. Econ.* **43** 153–193.

[28] FAN, J., LIAO, Y. and MINCHEVA, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 603–680. MR3091653 https://doi.org/10.1111/rssb.12016

[29] FERREIRA, O. and SVAITER, B. (2012). Kantorovich's theorem on Newton's method. ArXiv preprint. Available at arXiv:1209.5704.

[30] FRAZEE, A. C., LANGMEAD, B. and LEEK, J. T. (2011). ReCount: A multi-experiment resource of analysis-ready RNA-seq gene count datasets. *BMC Bioinform.* **12** 449.

[31] GAO, J., HAN, X., PAN, G. and YANG, Y. (2017). High dimensional correlation matrices: The central limit theorem and its applications. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 677–693. MR3641402 https://doi.org/10.1111/rssb.12189

[32] HAN, Q., JIANG, T. and SHEN, Y. (2023). Contiguity under high-dimensional Gaussianity with applications to covariance testing. *Ann. Appl. Probab.* **33** 4272–4321. MR4674051 https://doi.org/10.1214/22-aap1917

[33] HE, Y. and KNOWLES, A. (2020). Mesoscopic eigenvalue density correlations of Wigner matrices. *Probab. Theory Related Fields* **177** 147–216. MR4095015 https://doi.org/10.1007/s00440-019-00946-w

[34] HOTELLING, H. (1936). Relations between two sets of variates. *Biometrika* **28** 321–377.

[35] JIANG, D., BAI, Z. and ZHENG, S. (2013). Testing the independence of sets of large-dimensional variables. *Sci. China Math.* **56** 135–147. MR3016588 https://doi.org/10.1007/s11425-012-4501-0

[36] JIANG, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **14** 865–880. MR2052906 https://doi.org/10.1214/105051604000000143

[37] JIANG, T. (2004). The limiting distributions of eigenvalues of sample correlation matrices. *Sankhyā* **66** 35–48. MR2082906

[38] JIANG, T. and YANG, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Statist.* **41** 2029–2074. MR3127857 https://doi.org/10.1214/13-AOS1134

[39] KHORUNZHY, A. M., KHORUZHENKO, B. A. and PASTUR, L. A. (1996). Asymptotic properties of large random matrices with independent entries. *J. Math. Phys.* **37** 5033–5060. MR1411619 https://doi.org/10.1063/1.531589

[40] LAM, C. and FAN, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Statist.* **37** 4254–4278. MR2572459 https://doi.org/10.1214/09-AOS720

[41] LANDON, B. and SOSOE, P. (2022). Almost-optimal bulk regularity conditions in the CLT for Wigner matrices. ArXiv preprint. Available at arXiv:2204.03419.

[42] LEE, J. O. and SCHNELLI, K. (2018). Local law and Tracy–Widom limit for sparse random matrices. *Probab. Theory Related Fields* **171** 543–616. MR3800840 https://doi.org/10.1007/s00440-017-0787-8

[43] LI, W. and YAO, J. (2018). On structure testing for component covariance matrices of a high dimensional mixture. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 293–318. MR3763693 https://doi.org/10.1111/rssb.12248

[44] LI, Z., HAN, F. and YAO, J. (2020). Asymptotic joint distribution of extreme eigenvalues and trace of large sample covariance matrix in a generalized spiked population model. *Ann. Statist.* **48** 3138–3160. MR4185803 https://doi.org/10.1214/19-AOS1882

[45] LI, Z., WANG, C. and WANG, Q. (2023). On eigenvalues of a high-dimensional Kendall's rank correlation matrix with dependence. *Sci. China Math.* **66** 2615–2640. MR4658666 https://doi.org/10.1007/s11425-022-2031-2

[46] LI, Z., WANG, Q. and LI, R. (2021). Central limit theorem for linear spectral statistics of large dimensional Kendall's rank correlation matrices and its applications. *Ann. Statist.* **49** 1569–1593. MR4298873 https://doi.org/10.1214/20-aos2013

[47] LYTOVA, A. and PASTUR, L. (2009). Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Ann. Probab.* **37** 1778–1840. MR2561434 https://doi.org/10.1214/09-AOP452

[48] MINGO, J. A., ŚNIADY, P. and SPEICHER, R. (2007). Second order freeness and fluctuations of random matrices. II. Unitary random matrices. *Adv. Math.* **209** 212–240. MR2294222 https://doi.org/10.1016/j.aim.2006.05.003

[49] MINGO, J. A., ŚNIADY, P. and SPEICHER, R. (2007). Second order freeness and fluctuations of random matrices. II. Unitary random matrices. *Adv. Math.* **209** 212–240. MR2294222 https://doi.org/10.1016/j.aim.2006.05.003

[50] MINGO, J. A. and SPEICHER, R. (2006). Second order freeness and fluctuations of random matrices. I. Gaussian and Wishart matrices and cyclic Fock spaces. *J. Funct. Anal.* **235** 226–270. MR2216446 https://doi.org/10.1016/j.jfa.2005.10.007

[51] ORTEGA, J. M. (1968). The Newton–Kantorovich theorem. *Amer. Math. Monthly* **75** 658–660. MR0231218 https://doi.org/10.2307/2313800

[52] PAROLYA, N., HEINY, J. and KUROWICKA, D. (2024). Logarithmic law of large random correlation matrices. *Bernoulli* **30** 346–370. MR4665581 https://doi.org/10.3150/23-bej1600

[53] PERROT-DOCKÈS, M., LÉVY-LEDUC, C. and RAJJOU, L. (2022). Estimation of large block structured covariance matrices: Application to 'multi-omic' approaches to study seed quality. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **71** 119–147. MR4376848 https://doi.org/10.1111/rssc.12524

[54] QIU, Y. and CHEN, S. X. (2012). Test for bandedness of high-dimensional covariance matrices and bandwidth estimation. *Ann. Statist.* **40** 1285–1314. MR3015026 https://doi.org/10.1214/12-AOS1002

[55] SHCHERBINA, M. (2011). Central limit theorem for linear eigenvalue statistics of the Wigner and sample covariance random matrices. *J. Math. Phys. Anal. Geom.* **7** 176–192. MR2829615

[56] VOICULESCU, D. (1985). Symmetries of some reduced free product C^*-algebras. In *Operator Algebras and Their Connections with Topology and Ergodic Theory* (*Buşteni*, 1983). *Lecture Notes in Math*. **1132** 556–588. Springer, Berlin. MR0799593 https://doi.org/10.1007/BFb0074909

[57] VOICULESCU, D. (1986). Addition of certain noncommuting random variables. *J. Funct. Anal*. **66** 323–346. MR0839105 https://doi.org/10.1016/0022-1236(86)90062-5

[58] VOICULESCU, D. (1991). Limit laws for random matrices and free products. *Invent. Math*. **104** 201–220. MR1094052 https://doi.org/10.1007/BF01245072

[59] VOICULESCU, D. (1993). The analogues of entropy and of Fisher's information measure in free probability theory. I. *Comm. Math. Phys*. **155** 71–92. MR1228526

[60] WILKS, S. (1935). On the independence of k sets of normally distributed statistical variables. *Econometrica* 309–326.

[61] WU, W. B. and POURAHMADI, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika* **90** 831–844. MR2024760 https://doi.org/10.1093/biomet/90.4.831

[62] YAMADA, Y., HYODO, M. and NISHIYAMA, T. (2017). Testing block-diagonal covariance structure for high-dimensional data under non-normality. *J. Multivariate Anal*. **155** 305–316. MR3607897 https://doi.org/10.1016/j.jmva.2016.12.009

[63] YIN, Y., LI, C., TIAN, G.-L. and ZHENG, S. (2022). Spectral properties of rescaled sample correlation matrix. *Statist. Sinica* **32** 2007–2022. MR4478187

[64] ZHENG, S., BAI, Z. and YAO, J. (2017). CLT for eigenvalue statistics of large-dimensional general Fisher matrices with applications. *Bernoulli* **23** 1130–1178. MR3606762 https://doi.org/10.3150/15-BEJ772

[65] ZHENG, S., CHEN, Z., CUI, H. and LI, R. (2019). Hypothesis testing on linear structures of high-dimensional covariance matrix. *Ann. Statist*. **47** 3300–3334. MR4025743 https://doi.org/10.1214/18-AOS1779

The Annals of Statistics
2024, Vol. 52, No. 5, 1899–1926
https://doi.org/10.1214/24-AOS2385

JOINT SEQUENTIAL DETECTION AND ISOLATION FOR DEPENDENT DATA STREAMS

BY ANAMITRA CHAUDHURI[a] AND GEORGIOS FELLOURIS[b]

Department of Statistics, University of Illinois, Urbana-Champaign, [a]*ac34@illinois.edu,* [b]*fellouri@illinois.edu*

The problem of joint sequential detection and isolation is considered in the context of multiple, not necessarily independent, data streams. A multiple testing framework is proposed, where each hypothesis corresponds to a different subset of data streams, the sample size is a stopping time of the observations, and the probabilities of four kinds of error are controlled below distinct, user-specified levels. Two of these errors reflect the detection component of the formulation, whereas the other two the isolation component. The optimal expected sample size is characterized to a first-order asymptotic approximation as the error probabilities go to 0. Different asymptotic regimes, expressing different prioritizations of the detection and isolation tasks, are considered. A novel, versatile family of testing procedures is proposed, in which two distinct, in general, statistics are computed for each hypothesis, one addressing the detection task and the other the isolation task. Tests in this family, of various computational complexities, are shown to be asymptotically optimal under different setups. The general theory is applied to the detection and isolation of anomalous, not necessarily independent, data streams, as well as to the detection and isolation of an unknown dependence structure.

REFERENCES

[1] ANANDKUMAR, A., TONG, L. and SWAMI, A. (2009). Detection of Gauss–Markov random fields with nearest-neighbor dependency. *IEEE Trans. Inf. Theory* **55** 816–827. MR2597269 https://doi.org/10.1109/TIT.2008.2009855

[2] ARIAS-CASTRO, E., BUBECK, S. and LUGOSI, G. (2012). Detection of correlations. *Ann. Statist.* **40** 412–435. MR3014312 https://doi.org/10.1214/11-AOS964

[3] BALAKIRSKY, V. B., GHAZARYAN, A. R. and VINCK, A. J. H. (2007). Testing the independence of two non-stationary random processes with applications to biometric authentication. In 2007 *IEEE International Symposium on Information Theory* 2671–2675. https://doi.org/10.1109/ISIT.2007.4557622

[4] BARTROFF, J. (2018). Multiple hypothesis tests controlling generalized error rates for sequential data. *Statist. Sinica* **28** 363–398. MR3752265

[5] BARTROFF, J. and LAI, T. L. (2010). Multistage tests of multiple hypotheses. *Comm. Statist. Theory Methods* **39** 1597–1607.

[6] BARTROFF, J., LAI, T. L. and SHIH, M.-C. (2013). *Sequential Experimentation in Clinical Trials: Design and Analysis. Springer Series in Statistics*. Springer, New York. MR2987767 https://doi.org/10.1007/978-1-4614-6114-2

[7] BARTROFF, J. and SONG, J. (2014). Sequential tests of multiple hypotheses controlling type I and II familywise error rates. *J. Statist. Plann. Inference* **153** 100–114. MR3229025 https://doi.org/10.1016/j.jspi.2014.05.010

[8] BOLOGNANI, S., BOF, N., MICHELOTTI, D., MURARO, R. and SCHENATO, L. (2013). Identification of power distribution network topology via voltage correlation analysis. In 52*nd IEEE Conference on Decision and Control* 1659–1664. https://doi.org/10.1109/CDC.2013.6760120

[9] CAI, T. T. and JIANG, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39** 1496–1525. MR2850210 https://doi.org/10.1214/11-AOS879

[10] CAI, T. T. and MA, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** 2359–2388. MR3160557 https://doi.org/10.3150/12-BEJ455

MSC2020 subject classifications. Primary 62L10, 62L05; secondary 62J15.
Key words and phrases. Sequential multiple testing, detection and isolation, anomaly detection, asymptotic optimality, dependence structure.

[11] CHAUDHURI, A. and FELLOURIS, G. (2020). Sequential detection and isolation of a correlated pair. In 2020 *IEEE International Symposium on Information Theory* 1141–1146. https://doi.org/10.1109/ISIT44484.2020.9174318

[12] CHAUDHURI, A. and FELLOURIS, G. (2024). Supplement to "Joint sequential detection and isolation for dependent data streams." https://doi.org/10.1214/24-AOS2385SUPP

[13] CHOI, S. C. (1971). Sequential test for correlation coefficients. *J. Amer. Statist. Assoc.* **66** 575–576. https://doi.org/10.1080/01621459.1971.10482308

[14] COHEN, K. and ZHAO, Q. (2015). Active hypothesis testing for anomaly detection. *IEEE Trans. Inf. Theory* **61** 1432–1450. MR3318002 https://doi.org/10.1109/TIT.2014.2387857

[15] DE, S. K. and BARON, M. (2012). Step-up and step-down methods for testing multiple hypotheses in sequential experiments. *J. Statist. Plann. Inference* **142** 2059–2070. MR2903412 https://doi.org/10.1016/j.jspi.2012.02.005

[16] DE, S. K. and BARON, M. (2012). Sequential Bonferroni methods for multiple hypothesis testing with strong control of family-wise error rates I and II. *Sequential Anal.* **31** 238–262. MR2911288 https://doi.org/10.1080/07474946.2012.665730

[17] DE, S. K. and BARON, M. (2015). Sequential tests controlling generalized familywise error rates. *Stat. Methodol.* **23** 88–102. MR3278804 https://doi.org/10.1016/j.stamet.2014.10.001

[18] DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. MR2065195 https://doi.org/10.1214/009053604000000265

[19] FELLOURIS, G. and TARTAKOVSKY, A. G. (2017). Multichannel sequential detection—Part I: Non-i.i.d. data. *IEEE Trans. Inf. Theory* **63** 4551–4571. MR3666976 https://doi.org/10.1109/TIT.2017.2689785

[20] FORESTI, G. L., REGAZZONI, C. S. and VARSHNEY, P. K. (2003). *Multisensor Surveillance Systems: The Fusion Perspective*. Springer, Berlin.

[21] HE, X. and BARTROFF, J. (2021). Asymptotically optimal sequential FDR and pFDR control with (or without) prior information on the number of signals. *J. Statist. Plann. Inference* **210** 87–99. MR4102184 https://doi.org/10.1016/j.jspi.2020.05.002

[22] HEYDARI, J. and TAJER, A. (2017). Quickest search for local structures in random graphs. *IEEE Trans. Signal Inf. Process. Netw.* **3** 526–538. MR3687956 https://doi.org/10.1109/TSIPN.2017.2731125

[23] HEYDARI, J. and TAJER, A. (2018). Quickest localization of anomalies in power grids: A stochastic graphical framewor. *IEEE Trans. Smart Grid* **9** 4679–4688.

[24] HEYDARI, J., TAJER, A. and POOR, H. V. (2016). Quickest detection of Markov networks. In 2016 *IEEE International Symposium on Information Theory* 1341–1345. https://doi.org/10.1109/ISIT.2016.7541517

[25] HEYDARI, J., TAJER, A. and POOR, H. V. (2016). Quickest linear search over correlated sequences. *IEEE Trans. Inf. Theory* **62** 5786–5808. MR3552424 https://doi.org/10.1109/TIT.2016.2593772

[26] HOLM, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6** 65–70. MR0538597

[27] INGSTER, Y. I. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods Statist.* **6** 47–69. MR1456646

[28] KOWALSKI, C. J. (1971). The OC and ASN functions of some SPRT's for the correlation coefficient. *Technometrics* **13** 833–841.

[29] KU, C.-J. and FINE, T. L. (2005). Testing for stochastic independence: Application to blind source separation. *IEEE Trans. Signal Process.* **53** 1815–1826. MR2143108 https://doi.org/10.1109/TSP.2005.845458

[30] LI, J. and CHEN, S. X. (2012). Two sample tests for high-dimensional covariance matrices. *Ann. Statist.* **40** 908–940. MR2985938 https://doi.org/10.1214/12-AOS993

[31] LI, Y., NITINAWARAT, S. and VEERAVALLI, V. V. (2014). Universal outlier hypothesis testing. *IEEE Trans. Inf. Theory* **60** 4066–4082. MR3225950 https://doi.org/10.1109/TIT.2014.2317691

[32] MALLOY, M. L. and NOWAK, R. D. (2014). Sequential testing for sparse recovery. *IEEE Trans. Inf. Theory* **60** 7862–7873. MR3285750 https://doi.org/10.1109/TIT.2014.2363846

[33] PÉREZ-ORTIZ, M. F., LARDY, T. DE HEIDE, R. and GRÜNWALD, P. (2024). E-statistics, group invariance and anytime valid testing. *Ann. Statist.* To appear. Available at arXiv:2208.07610.

[34] PRADHAN, M. and SATHE, Y. S. (1975). An unbiased estimator and a sequential test for the correlation coefficient. *J. Amer. Statist. Assoc.* **70** 160–161. https://doi.org/10.1080/01621459.1975.10480278

[35] RAPPAPORT, T. S. et al. (1996). *Wireless Communications: Principles and Practice*, 2nd ed. Prentice Hall, New York.

[36] SCHIPPER, M. and MEELIS, E. (2003). Making sequential analysis of environmental monitoring data feasible by simplifying the covariance matrix structure. *J. Agric. Biol. Environ. Stat.* **8** 122–137.

[37] SOLO, V. and PASHA, A. (2010). Testing for independence between a point process and an analog signal. In 2010 *IEEE International Conference on Acoustics, Speech and Signal Processing* 3762–3765. https://doi.org/10.1109/ICASSP.2010.5495860

[38] SONG, Y. and FELLOURIS, G. (2017). Asymptotically optimal, sequential, multiple testing procedures with prior information on the number of signals. *Electron. J. Stat.* **11** 338–363. MR3606774 https://doi.org/10.1214/17-EJS1223

[39] SONG, Y. and FELLOURIS, G. (2019). Sequential multiple testing with generalized error control: An asymptotic optimality theory. *Ann. Statist.* **47** 1776–1803. MR3911130 https://doi.org/10.1214/18-AOS1737

[40] TARTAKOVSKY, A., NIKIFOROV, I. and BASSEVILLE, M. (2015). *Sequential Analysis: Hypothesis Testing and Changepoint Detection. Monographs on Statistics and Applied Probability* **136**. CRC Press, Boca Raton, FL. MR3241619

[41] TARTAKOVSKY, A. G. An efficient adaptive sequential procedure for detecting targets. In *Proceedings, IEEE Aerospace Conference* **4** 4–4. https://doi.org/10.1109/aero.2002.1036875

[42] TARTAKOVSKY, A. G., LI, X. R. and YARALOV, G. (2003). Sequential detection of targets in multichannel systems. *IEEE Trans. Inf. Theory* **49** 425–445. MR1966790 https://doi.org/10.1109/TIT.2002.807288

[43] TSOPELAKOS, A. and FELLOURIS, G. (2023). Sequential anomaly detection under sampling constraints. *IEEE Trans. Inf. Theory* **69** 8126–8146. MR4692659 https://doi.org/10.1109/tit.2022.3177142

[44] VEERAVALLI, V. V., BASAR, T. and POOR, H. V. (1993). Decentralized sequential detection with a fusion center performing the sequential test. *IEEE Trans. Inf. Theory* **39** 433–442. https://doi.org/10.1109/18.212274

[45] VEERAVALLI, V. V., BASAR, T. and POOR, H. V. (1994). Minimax robust decentralized detection. *IEEE Trans. Inf. Theory* **40** 35–40. https://doi.org/10.1109/18.272453

[46] WALD, A. (1945). Sequential tests of statistical hypotheses. *Ann. Math. Stat.* **16** 117–186. MR0013275 https://doi.org/10.1214/aoms/1177731118

[47] WALD, A. (1947). *Sequential Analysis*. Wiley, New York. MR0020764

[48] WOLDE-TSADIK, G. (1976). A generalization of an SPRT for the correlation coefficient. *J. Amer. Statist. Assoc.* **71** 709–710. MR0428640

[49] WOODROOFE, M. (1979). Repeated likelihood ratio tests. *Biometrika* **66** 453–463. MR0556732 https://doi.org/10.1093/biomet/66.3.453

The Annals of Statistics
2024, Vol. 52, No. 5, 1927–1952
https://doi.org/10.1214/24-AOS2399
This research was funded, in whole or in part, by NSF Grant DMS-1953686.

OPTIMAL POLICY EVALUATION USING KERNEL-BASED TEMPORAL DIFFERENCE METHODS

BY YAQI DUAN[1,a], MENGDI WANG[2,b] AND MARTIN J. WAINWRIGHT[3,c]

[1]*Leonard N. Stern School of Business, New York University,* [a]*yaqi.duan@stern.nyu.edu*
[2]*Department of ECE, Princeton University,* [b]*mengdiw@princeton.edu*
[3]*Departments of EECS and Mathematics, Massachusetts Institute of Technology,* [c]*wainwrigwork@gmail.com*

We study nonparametric methods for estimating the value function of an infinite-horizon discounted Markov reward process (MRP). We analyze the kernel-based least-squares temporal difference (LSTD) estimate, which can be understood either as a nonparametric instrumental variables method, or as a projected approximation to the Bellman fixed point equation. Our analysis imposes no assumptions on the transition operator of the Markov chain, but rather only conditions on the reward function and population-level kernel LSTD solutions. Using empirical process theory and concentration inequalities, we establish a nonasymptotic upper bound on the error with explicit dependence on the effective horizon $H = (1 - \gamma)^{-1}$ of the Markov reward process, the eigenvalues of the associated kernel operator, as well as the instance-dependent variance of the Bellman residual error. In addition, we prove minimax lower bounds over subclasses of MRPs, which shows that our guarantees are optimal in terms of the sample size n and the effective horizon H. Whereas existing worst-case theory predicts cubic scaling (H^3) in the effective horizon, our theory reveals a much wider range of scalings, depending on the kernel, the stationary distribution, and the variance of the Bellman residual error. Notably, it is only parametric and near-parametric problems that can ever achieve the worst-case cubic scaling.

REFERENCES

[1] BAGNELL, J. A. and SCHNEIDER, J. (2003). Policy search in kernel Hilbert space. Technical Report, Carnegie Mellon Univ., Pittsburgh, PA.

[2] BARRETO, A. M. S., PRECUP, D. and PINEAU, J. (2016). Practical kernel-based reinforcement learning. *J. Mach. Learn. Res.* **17** 1–70. MR3517090

[3] BELLMAN, R. E. and DREYFUS, S. E. (1962). *Applied Dynamic Programming*. Princeton Univ. Press, Princeton, NJ. MR0140369

[4] BERLINET, A. and THOMAS-AGNAN, C. (2004). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic, Boston, MA. With a preface by Persi Diaconis. MR2239907 https://doi.org/10.1007/978-1-4419-9096-9

[5] BERTSEKAS, D. P. (1996). *Neuro-Dynamic Programming*, 1st ed. *Athena Scientific Optimization and Computation Series*. Athena Scientific, Belmont, MA. MR3444832

[6] BERTSEKAS, D. P. (1995). *Dynamic Programming and Optimal Control*. *Vol. I*. *Approximate Dynamic Programming*, 2th ed. Athena Scientific, Belmont, MA.

[7] BERTSEKAS, D. P. (2011). *Dynamic Programming and Optimal Control*. *Vol. II*. *Approximate Dynamic Programming*, 3th ed. Athena Scientific, Belmont, MA. MR3642732

[8] BLACKWELL, D. (1962). Discrete dynamic programming. *Ann. Math. Stat.* **33** 719–726. MR0149965 https://doi.org/10.1214/aoms/1177704593

[9] BLACKWELL, D. (1965). Discounted dynamic programming. *Ann. Math. Stat.* **36** 226–235. MR0173536 https://doi.org/10.1214/aoms/1177700285

MSC2020 subject classifications. Primary 62G05; secondary 62M05.
Key words and phrases. Sequential decision-making, dynamic programming, reinforcement learning, Markov reward process, nonparametric estimation, policy evaluation, temporal difference learning, reproducing kernel Hilbert space.

[10] BOUCHERIE, R. J. and VAN DIJK, N. M., eds. (2017). *Markov Decision Processes in Practice*. *International Series in Operations Research & Management Science* **248**. Springer, Cham. MR3618832 https://doi.org/10.1007/978-3-319-47766-4

[11] BRADTKE, S. J. and BARTO, A. G. (1996). Linear least-squares algorithms for temporal difference learning. *Mach. Learn.* **22** 33–57.

[12] CHEN, X. and REISS, M. (2011). On rate optimality for ill-posed inverse problems in econometrics. *Econometric Theory* **27** 497–521. MR2806258 https://doi.org/10.1017/S0266466610000381

[13] DAI, B., HE, N., PAN, Y., BOOTS, B. and SONG, L. (2017). Learning from conditional distributions via dual embeddings. In *Artificial Intelligence and Statistics* 1458–1467. PMLR.

[14] DUAN, Y. and WAINWRIGHT, M. J. (2022). Policy evaluation from a single path: Multi-step methods mixing and mis-specification. Technical Report, MIT. Available at arXiv:2211.03899.

[15] DUAN, Y., WANG, M. and WAINWRIGHT, M. J. (2024). Supplement to "Optimal policy evaluation using kernel-based temporal difference methods." https://doi.org/10.1214/24-AOS2399SUPP

[16] FAN, J., WANG, Z., XIE, Y. and YANG, Z. (2020). A theoretical analysis of deep Q-learning. In *Learning for Dynamics and Control* 486–489. PMLR.

[17] FARAHMAND, A., GHAVAMZADEH, M., SZEPESVÁRI, C. and MANNOR, S. (2016). Regularized policy iteration with nonparametric function spaces. *J. Mach. Learn. Res.* **17** 4809–4874. MR3555030

[18] FENG, Y., LI, L. and LIU, Q. (2019). A kernel loss for solving the Bellman equation. *Adv. Neural Inf. Process. Syst.* 15456–15467.

[19] FENG, Y., REN, T., TANG, Z. and LIU, Q. (2020). Accountable off-policy evaluation with kernel Bellman statistics. In *International Conference on Machine Learning* 3102–3111. PMLR.

[20] GHESHLAGHI AZAR, M., MUNOS, R. and KAPPEN, H. J. (2013). Minimax PAC bounds on the sample complexity of reinforcement learning with a generative model. *Mach. Learn.* **91** 325–349. MR3064431 https://doi.org/10.1007/s10994-013-5368-1

[21] GRUNEWALDER, S., LEVER, G., BALDASSARRE, L., PONTIL, M. and GRETTON, A. (2012). Modelling transition dynamics in MDPs with RKHS embeddings. Technical Report, UCL.

[22] GU, C. (2002). *Smoothing Spline ANOVA Models*. *Springer Series in Statistics*. Springer, New York. MR1876599 https://doi.org/10.1007/978-1-4757-3683-0

[23] HASMINSKII, R. Z. (1978). A lower bound for risks of nonparametric density estimates in the uniform metric. *Theory Probab. Appl.* **23** 794–798. MR0516279

[24] JIANG, N. and LI, L. (2016). Doubly robust off-policy value evaluation for reinforcement learning. In *International Conference on Machine Learning* 652–661. PMLR.

[25] KALLUS, N. and UEHARA, M. (2019). Efficiently breaking the curse of horizon: Double reinforcement learning in infinite-horizon processes. *stat* **1050** 12.

[26] KALLUS, N. and UEHARA, M. (2020). Double reinforcement learning for efficient off-policy evaluation in Markov decision processes. *J. Mach. Learn. Res.* **21** 1–63. MR4209453

[27] KHAMARU, K., PANANJADY, A., RUAN, F., WAINWRIGHT, M. J. and JORDAN, M. I. (2021). Is temporal difference learning optimal? An instance-dependent analysis. *SIAM J. Math. Data Sci.* **3** 1013–1040. MR4320891 https://doi.org/10.1137/20M1331524

[28] KIMELDORF, G. and WAHBA, G. (1971). Some results on Tchebycheffian spline functions. *J. Math. Anal. Appl.* **33** 82–95. MR0290013 https://doi.org/10.1016/0022-247X(71)90184-3

[29] KOPPEL, A., WARNELL, G., STUMP, E., STONE, P. and RIBEIRO, A. (2020). Policy evaluation in continuous MDPs with efficient kernelized gradient temporal difference. *IEEE Trans. Automat. Control* **66** 1856–1863. MR4240216 https://doi.org/10.1109/tac.2020.3029315

[30] LONG, J., HAN, J. and WEINAN, E. (2022). An L^2 analysis of reinforcement learning in high dimensions with kernel and neural network approximation. *CSIAM Trans. Appl. Math.* **3** 191–220. MR4456674 https://doi.org/10.4208/csiam-am.so-2021-0026

[31] MENDELSON, S. (2002). Geometric parameters of kernel machines. In *Computational Learning Theory* (*Sydney*, 2002). *Lecture Notes in Computer Science* **2375** 29–43. Springer, Berlin. MR2040403 https://doi.org/10.1007/3-540-45435-7_3

[32] MOU, W., LI, C. J., WAINWRIGHT, M. J., BARTLETT, P. L. and JORDAN, M. I. (2020). On linear stochastic approximation: Fine-grained Polyak–Ruppert and non-asymptotic concentration. In *Conference on Learning Theory* (*COLT*) **125** 2947–2997.

[33] MOU, W., PANANJADY, A. and WAINWRIGHT, M. J. (2023). Optimal oracle inequalities for projected fixed-point equations, with applications to policy evaluation. *Math. Oper. Res.* **48** 2308–2336. MR4675955

[34] MUNOS, R. and SZEPESVÁRI, C. (2008). Finite-time bounds for fitted value iteration. *J. Mach. Learn. Res.* **1** 815–857. MR2417255

[35] NEWEY, W. K. and POWELL, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica* **71** 1565–1578. MR2000257 https://doi.org/10.1111/1468-0262.00459

[36] NGUYEN-TANG, T., GUPTA, S., TRAN-THE, H., VENKATESH, S. et al. (2021). Sample complexity of offline reinforcement learning with deep ReLU networks. arXiv preprint. Available at arXiv:2103.06671.
[37] ORMONEIT, D. and SEN, Ś. (2002). Kernel-based reinforcement learning. *Mach. Learn.* **49** 161–178.
[38] PANANJADY, A. and WAINWRIGHT, M. J. (2021). Instance-dependent ℓ_∞-bounds for policy evaluation in tabular reinforcement learning. *IEEE Trans. Inf. Theory* **67** 566–585. MR4231973 https://doi.org/10.1109/TIT.2020.3027316
[39] PUTERMAN, M. L. (2005). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley, New York. MR1270015
[40] RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2012). Minimax-optimal rates for sparse additive models over kernel classes via convex programming. *J. Mach. Learn. Res.* **13** 389–427. MR2913704
[41] SHAWE-TAYLOR, J., CRISTIANINI, N. et al. (2004). *Kernel Methods for Pattern Analysis*. Cambridge Univ. Press, Cambridge.
[42] SOBEL, M. J. (1982). The variance of discounted Markov decision processes. *J. Appl. Probab.* **19** 794–802. MR0675143 https://doi.org/10.2307/3213832
[43] STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053. MR0673642
[44] SUTTON, R. S. (1988). Learning to predict via the methods of temporal differences. *Mach. Learn.* **3** 9–44.
[45] SUTTON, R. S. and BARTO, A. G. (2018). *Reinforcement Learning: An Introduction*, 2nd ed. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3889951
[46] TAYLOR, G. and PARR, R. (2009). Kernelized value function approximation for reinforcement learning. In *Proceedings of the* 26*th Annual International Conference on Machine Learning* 1017–1024.
[47] TSITSIKLIS, J. N. and VAN ROY, B. (1997). An analysis of temporal-difference learning with function approximation. *IEEE Trans. Automat. Control* **42** 1075–1081. MR1454208 https://doi.org/10.1109/9.580874
[48] VAN DE GEER, S. (2000). *Empirical Processes in M-Estimation*. Cambridge Univ. Press, Cambridge.
[49] WAINWRIGHT, M. J. (2019). *High-Dimensional Statistics: A Non-asymptotic Viewpoint*. *Cambridge Series in Statistical and Probabilistic Mathematics* **48**. Cambridge Univ. Press, Cambridge. MR3967104 https://doi.org/10.1017/9781108627771
[50] WHITE, H. (1982). Instrumental variables regression with independent observations. *Econometrica* **50** 483–499. MR0662289 https://doi.org/10.2307/1912639
[51] WOOLDRIDGE, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. MIT Press, Cambridge, MA. MR2768559
[52] XIE, T., MA, Y. and WANG, Y.-X. (2019). Towards optimal off-policy evaluation for reinforcement learning with marginalized importance sampling. In *Advances in Neural Information Processing Systems* 9668–9678.
[53] YANG, Y., PILANCI, M. and WAINWRIGHT, M. J. (2017). Randomized sketches for kernels: Fast and optimal nonparametric regression. *Ann. Statist.* **45** 991–1023. MR3662446 https://doi.org/10.1214/16-AOS1472
[54] YIN, M. and WANG, Y.-X. (2020). Asymptotically efficient off-policy evaluation for tabular reinforcement learning. arXiv preprint. Available at arXiv:2001.10742.
[55] YIN, M. and WANG, Y.-X. (2021). Towards instance-optimal offline reinforcement learning with pessimism. *Adv. Neural Inf. Process. Syst.* **34** 4065–4078.
[56] YU, B. (1997). Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam* 423–435. Springer, New York. MR1462963
[57] YU, H. and BERTSEKAS, D. P. (2010). Error bounds for approximations from projected linear equations. *Math. Oper. Res.* **35** 306–329. MR2674722 https://doi.org/10.1287/moor.1100.0441
[58] ZHANG, T. (2005). Learning bounds for kernel regression using effective data dimensionality. *Neural Comput.* **17** 2077–2098. MR2175849 https://doi.org/10.1162/0899766054323008

The Annals of Statistics
2024, Vol. 52, No. 5, 1953–1977
https://doi.org/10.1214/24-AOS2407

IMPROVED COVARIANCE ESTIMATION: OPTIMAL ROBUSTNESS AND SUB-GAUSSIAN GUARANTEES UNDER HEAVY TAILS

BY ROBERTO I. OLIVEIRA[1,a] AND ZORAIDA F. RICO[2,b]

[1]*Instituto de Matemática Pura e Aplicada (IMPA),* [a]*rimfo@impa.br*
[2]*Department of Statistics, Columbia University,* [b]*zoraida.f.rico@columbia.edu*

We present an estimator of the covariance matrix Σ of random d-dimensional vector from an i.i.d. sample of size n. Our sole assumption is that this vector satisfies a bounded $L^p - L^2$ moment assumption over its one-dimensional marginals, for some $p \geq 4$. Given this, we show that Σ can be estimated from the sample with the same high-probability error rates that the sample covariance matrix achieves in the case of Gaussian data. This holds even though we allow for very general distributions that may not have moments of order $> p$. Moreover, our estimator can be made to be optimally robust to adversarial contamination. This result improves the recent contributions by Mendelson and Zhivotovskiy and Catoni and Giulini, and matches parallel work by Abdalla and Zhivotovskiy (the exact relationship with this last work is described in the paper).

REFERENCES

[1] ABDALLA, P. and ZHIVOTOVSKIY, N. (2024). Covariance estimation: Optimal dimension-free guarantees for adversarial corruption and heavy tails. *J. Eur. Math. Soc.* **(online)**. https://doi.org/10.4171/JEMS/1505
[2] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford. MR3185193 https://doi.org/10.1093/acprof:oso/9780199535255.001.0001
[3] BRAILOVSKAYA, T. and VAN HANDEL, R. (2022). Universality and sharp matrix concentration inequalities arXiv preprint. Available at arXiv:2201.05142.
[4] CATONI, O. (2012). Challenging the empirical mean and empirical variance: A deviation study. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 1148–1185. MR3052407 https://doi.org/10.1214/11-AIHP454
[5] CATONI, O. (2016). PAC-Bayesian bounds for the Gram matrix and least squares regression with a random design. arXiv preprint. Available at arXiv:1603.05229.
[6] CATONI, O. and GIULINI, I. (2018). Dimension-free PAC-Bayesian bounds for the estimation of the mean of a random vector. *Statist. Theory*. arXiv.
[7] DEPERSIN, J. and LECUÉ, G. (2022). Optimal robust mean and location estimation via convex programs with respect to any pseudo-norms. *Probab. Theory Related Fields* **183** 997–1025. MR4453320 https://doi.org/10.1007/s00440-022-01127-y
[8] DEPERSIN, J. and LECUÉ, G. (2022). Robust sub-Gaussian estimation of a mean vector in nearly linear time. *Ann. Statist.* **50** 511–536. MR4382026 https://doi.org/10.1214/21-aos2118
[9] DEVROYE, L., LERASLE, M., LUGOSI, G. and OLIVEIRA, R. I. (2016). Sub-Gaussian mean estimators. *Ann. Statist.* **44** 2695–2725. MR3576558 https://doi.org/10.1214/16-AOS1440
[10] DIAKONIKOLAS, I., KAMATH, G., KANE, D., LI, J., MOITRA, A. and STEWART, A. (2019). Robust estimators in high-dimensions without the computational intractability. *SIAM J. Comput.* **48** 742–864. MR3945261 https://doi.org/10.1137/17M1126680
[11] DIAKONIKOLAS, I. and KANE, D. M. (2019). Recent advances in algorithmic high-dimensional robust statistics. arXiv preprint. Available at arXiv:1911.05911.
[12] DONOHO, D. and HUBER, P. J. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann*. *Wadsworth Statist./Probab. Ser.* 157–184. Wadsworth, Belmont, CA. MR0689745
[13] HOPKINS, S. B. (2020). Mean estimation with sub-Gaussian rates in polynomial time. *Ann. Statist.* **48** 1193–1213. MR4102693 https://doi.org/10.1214/19-AOS1843

MSC2020 subject classifications. Primary 62G35; secondary 62H12.
Key words and phrases. Covariance estimation, robust statistics, heavy-tailed data.

[14] HOPKINS, S. B., LI, J. and ZHANG, F. (2020). Robust and heavy-tailed mean estimation made simple, via regret minimization. In NeurIPS.

[15] HUBER, P. J. (1981). *Robust Statistics*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0606374

[16] JOLY, E., LUGOSI, G. and OLIVEIRA, R. I. (2017). On the estimation of the mean of a random vector. *Electron. J. Stat.* **11** 440–451. MR3619312 https://doi.org/10.1214/17-EJS1228

[17] KOLTCHINSKII, V. and LOUNICI, K. (2017). Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli* **23** 110–133. MR3556768 https://doi.org/10.3150/15-BEJ730

[18] LEDOUX, M. (2001). *The Concentration of Measure Phenomenon*. *Mathematical Surveys and Monographs* **89**. Amer. Math. Soc., Providence, RI. MR1849347 https://doi.org/10.1090/surv/089

[19] LOUNICI, K. (2014). High-dimensional covariance matrix estimation with missing observations. *Bernoulli* **20** 1029–1058. MR3217437 https://doi.org/10.3150/12-BEJ487

[20] LUGOSI, G. and MENDELSON, S. (2019). Near-optimal mean estimators with respect to general norms. *Probab. Theory Related Fields* **175** 957–973. MR4026610 https://doi.org/10.1007/s00440-019-00906-4

[21] LUGOSI, G. and MENDELSON, S. (2019). Sub-Gaussian estimators of the mean of a random vector. *Ann. Statist.* **47** 783–794. MR3909950 https://doi.org/10.1214/17-AOS1639

[22] LUGOSI, G. and MENDELSON, S. (2021). Robust multivariate mean estimation: The optimality of trimmed mean. *Ann. Statist.* **49** 393–410. MR4206683 https://doi.org/10.1214/20-AOS1961

[23] MENDELSON, S. (2021). Approximating L_p unit balls via random sampling. *Adv. Math.* **386** Paper No. 107829, 20. MR4270522 https://doi.org/10.1016/j.aim.2021.107829

[24] MENDELSON, S. and PAOURIS, G. (2014). On the singular values of random matrices. *J. Eur. Math. Soc.* (*JEMS*) **16** 823–834. MR3191978 https://doi.org/10.4171/JEMS/448

[25] MENDELSON, S. and ZHIVOTOVSKIY, N. (2020). Robust covariance estimation under L_4-L_2 norm equivalence. *Ann. Statist.* **48** 1648–1664. MR4124338 https://doi.org/10.1214/19-AOS1862

[26] MINASYAN, A. and ZHIVOTOVSKIY, N. (2023). Statistically optimal robust mean and covariance estimation for anisotropic Gaussians. arXiv preprint. Available at arXiv:2301.09024.

[27] MINSKER, S. (2015). Geometric median and robust estimation in Banach spaces. *Bernoulli* **21** 2308–2335. MR3378468 https://doi.org/10.3150/14-BEJ645

[28] MINSKER, S. (2017). On some extensions of Bernstein's inequality for self-adjoint operators. *Statist. Probab. Lett.* **127** 111–119. MR3648301 https://doi.org/10.1016/j.spl.2017.03.020

[29] MINSKER, S. (2018). Sub-Gaussian estimators of the mean of a random matrix with heavy-tailed entries. *Ann. Statist.* **46** 2871–2903. MR3851758 https://doi.org/10.1214/17-AOS1642

[30] MINSKER, S. and WEI, X. (2020). Robust modifications of U-statistics and applications to covariance estimation problems. *Bernoulli* **26** 694–727. MR4036049 https://doi.org/10.3150/19-BEJ1149

[31] OLIVEIRA, R. I. (2016). The lower tail of random quadratic forms with applications to ordinary least squares. *Probab. Theory Related Fields* **166** 1175–1194. MR3568047 https://doi.org/10.1007/s00440-016-0738-9

[32] OLIVEIRA, R. I. and RESENDE, L. (2023). Trimmed sample means for robust uniform mean estimation and regression. arXiv preprint. Available at arXiv:2302.06710.

[33] OLIVEIRA, R. I. and RICO, Z. F. (2024). Supplement to "Improved covariance estimation: optimal robustness and sub-Gaussian guarantees under heavy tails." https://doi.org/10.1214/24-AOS2407SUPP

[34] OSTROVSKII, D. M. and RUDI, A. (2019). Affine invariant covariance estimation for heavy-tailed distributions. In *Proceedings of the Thirty-Second Conference on Learning Theory* (A. Beygelzimer and D. Hsu, eds.). *Proceedings of Machine Learning Research* **99** 2531–2550. PMLR.

[35] RICO, Z. F. (2022). Optimal statistical estimation: Sub-Gaussian properties, heavy-tailed data, and robustness Ph.D. thesis, Instituto de Matemática Pura e Aplicada (IMPA).

[36] TIKHOMIROV, K. (2018). Sample covariance matrices of heavy-tailed distributions. *Int. Math. Res. Not. IMRN* 20 6254–6289. MR3872323 https://doi.org/10.1093/imrn/rnx067

[37] TROPP, J. A. (2015). An introduction to matrix concentration inequalities. *Found. Trends® Mach. Learn.* **8** 1–230. https://doi.org/10.1561/2200000048

[38] VERSHYNIN, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge Univ. Press, Cambridge. MR2963170

[39] ZHIVOTOVSKIY, N. (2024). Dimension-free bounds for sums of independent matrices and simple tensors via the variational principle. *Electron. J. Probab.* **29** Paper No. 13, 28. MR4693860 https://doi.org/10.1214/23-ejp1021

The Annals of Statistics
2024, Vol. 52, No. 5, 1978–2003
https://doi.org/10.1214/24-AOS2409
This research was funded, in whole or in part, by UKRI Engineering and Physical Sciences Research Council, EP/R013381/1.

DEBIASED INVERSE PROPENSITY SCORE WEIGHTING FOR ESTIMATION OF AVERAGE TREATMENT EFFECTS WITH HIGH-DIMENSIONAL CONFOUNDERS

BY YUHAO WANG[1,a] AND RAJEN D. SHAH[2,b]

[1]*Institute for Interdisciplinary Information Sciences, Tsinghua University,* [a]*yuhaow@tsinghua.edu.cn*
[2]*Statistical Laboratory, University of Cambridge,* [b]*r.shah@statslab.cam.ac.uk*

We consider estimation of average treatment effects given observational data with high-dimensional pretreatment variables. Existing methods for this problem typically assume some form of sparsity for the regression functions. In this work, we introduce a debiased inverse propensity score weighting (DIPW) scheme for average treatment effect estimation that delivers $\sqrt{n}$-consistent estimates when the propensity score follows a sparse logistic regression model; the outcome regression functions are permitted to be arbitrarily complex. We further demonstrate how confidence intervals centred on our estimates may be constructed. Our theoretical results quantify the price to pay for permitting the regression functions to be unestimable, which shows up as an inflation of the variance of the estimator compared to the semiparametric efficient variance by a constant factor, under mild conditions. We also show that when outcome regressions can be estimated consistently, our estimator achieves semiparametric efficiency. As our results accommodate arbitrary outcome regression functions, averages of transformed responses under each treatment may also be estimated at the $\sqrt{n}$ rate. Thus, for example, the variances of the potential outcomes may be estimated. We discuss extensions to estimating linear projections of the heterogeneous treatment effect function and explain how propensity score models with more general link functions may be handled within our framework. An R package `dipw` implementing our methodology is available on CRAN.

REFERENCES

APS, M. (2019). Rmosek: the R to MOSEK Optimization Interface R package version 9.0.96.

ATHEY, S., IMBENS, G. W. and WAGER, S. (2018). Approximate residual balancing: Debiased inference of average treatment effects in high dimensions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 597–623. MR3849336 https://doi.org/10.1111/rssb.12268

BELLONI, A., CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. and HANSEN, C. (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica* **85** 233–298. MR3611771 https://doi.org/10.3982/ECTA12723

BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Rev. Econ. Stud.* **81** 608–650. MR3207983 https://doi.org/10.1093/restud/rdt044

BRADIC, J., CHERNOZHUKOV, V., NEWEY, W. K. and ZHU, Y. (2019). Minimax semiparametric learning with approximate sparsity. ArXiv preprint. Available at arXiv:1912.12213.

BRADIC, J., WAGER, S. and ZHU, Y. (2019). Sparsity double robust inference of average treatment effects. arXiv preprint. Available at arXiv:1905.00744.

BREIMAN, L. (2001). Random forests. *Mach. Learn.* **45** 5–32.

BREIMAN, L. (2017). Classification and regression trees. Routledge.

CHERNOZHUKOV, V., CHETVERIKOV, D., DEMIRER, M., DUFLO, E., HANSEN, C., NEWEY, W. and ROBINS, J. (2018a). Double/debiased machine learning for treatment and structural parameters. *Econom. J.* **21** C1–C68. MR3769544 https://doi.org/10.1111/ectj.12097

MSC2020 subject classifications. Primary 62D20, 62E17, 62F12; secondary 62J12.
Key words and phrases. Causal inference, high dimensional statistics, inverse propensity score weighting, debiased inference, model robustness.

CHERNOZHUKOV, V., DEMIRER, M., DUFLO, E. and FERNANDEZ-VAL, I. (2018b). Generic machine learning inference on heterogenous treatment effects in randomized experiments. Technical Report, National Bureau of Economic Research.

CHERNOZHUKOV, V., NEWEY, W. K., ROBINS, J. and SINGH, R. (2018). Double/de-biased machine learning of global and local parameters using regularized Riesz representers. ArXiv preprint. Available at arXiv:1802.08667.

DUKES, O. and VANSTEELANDT, S. (2021). Inference for treatment effect parameters in potentially misspecified high-dimensional models. *Biometrika* **108** 321–334. MR4259134 https://doi.org/10.1093/biomet/asaa071

FARRELL, M. H. (2015). Robust inference on average treatment effects with possibly more covariates than observations. *J. Econometrics* **189** 1–23. MR3397349 https://doi.org/10.1016/j.jeconom.2015.06.017

FRIEDMAN, J., HASTIE, T. J. and TIBSHIRANI, R. J. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33** 1.

GUO, F. R. and SHAH, R. D. (2023). Rank-transformed subsampling: Inference for multiple data splitting and exchangeable p-values. arXiv preprint. Available at arXiv:2301.02739.

HIRSHBERG, D. A. and WAGER, S. (2021). Augmented minimax linear estimation. *Ann. Statist.* **49** 3206–3227. MR4352528 https://doi.org/10.1214/21-aos2080

IMBENS, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Rev. Econ. Stat.* **86** 4–29.

JANKOVÁ, J., SHAH, R. D., BÜHLMANN, P. and SAMWORTH, R. J. (2020). Goodness-of-fit testing in high dimensional generalized linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 773–795. MR4112784 https://doi.org/10.1111/rssb.12371

JAVANMARD, A. and MONTANARI, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909. MR3277152

LONSDALE, J., THOMAS, J., SALVATORE, M., PHILLIPS, R., LO, E., SHAD, S., HASZ, R., WALTERS, G., GARCIA, F. et al. (2013). The genotype-tissue expression (GTEx) project. *Nat. Genet.* **45** 580–585.

LUNCEFORD, J. K. and DAVIDIAN, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Stat. Med.* **23** 2937–2960. https://doi.org/10.1002/sim.1903

NEYMAN, J. (1923). Sur les applications de la théorie des probabilités aux experiences agricoles: Essai des principes. *Rocz. Nauk Rol.* **10** 1–51.

NING, Y., SIDA, P. and IMAI, K. (2020). Robust estimation of causal effects via a high-dimensional covariate balancing propensity score. *Biometrika* **107** 533–554. MR4138975 https://doi.org/10.1093/biomet/asaa020

ROBINS, J. M. and ROTNITZKY, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *J. Amer. Statist. Assoc.* **90** 122–129. MR1325119

ROBINS, J. M., ROTNITZKY, A. and ZHAO, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89** 846–866. MR1294730

RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66** 688.

SCHARFSTEIN, D. O., ROTNITZKY, A. and ROBINS, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J. Amer. Statist. Assoc.* **94** 1096–1146. With comments and a rejoinder by the authors. MR1731478 https://doi.org/10.2307/2669923

SHAH, R. D., FROT, B., THANEI, G.-A. and MEINSHAUSEN, N. (2020). Right singular vector projection graphs: Fast high dimensional covariance matrix estimation under latent confounding. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 361–389. MR4084168 https://doi.org/10.1111/rssb.12359

SMUCLER, E., ROTNITZKY, A. and ROBINS, J. M. (2019). A unifying approach for doubly-robust ℓ_1 regularized estimation of causal contrasts. ArXiv preprint. Available at arXiv:1904.03737.

TAN, Z. (2020). Model-assisted inference for treatment effects using regularized calibrated estimation with high-dimensional data. *Ann. Statist.* **48** 811–837. MR4102677 https://doi.org/10.1214/19-AOS1824

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242

VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. and DEZEURE, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** 1166–1202. MR3224285 https://doi.org/10.1214/14-AOS1221

VAN DE GEER, S. A. et al. (2008). High-dimensional generalized linear models and the lasso. *Ann. Statist.* **36** 614–645. MR2396809 https://doi.org/10.1214/009053607000000929

VAN DER LAAN, M. J. and RUBIN, D. (2006). Targeted maximum likelihood learning. *Int. J. Biostat.* **2** Art. 11, 40. MR2306500 https://doi.org/10.2202/1557-4679.1043

WANG, Y. and SHAH, R. D. (2024). Supplement to "Debiased inverse propensity score weighting for estimation of average treatment effects with high-dimensional confounders." https://doi.org/10.1214/24-AOS2409SUPPA, https://doi.org/10.1214/24-AOS2409SUPPB

WRIGHT, M. N. and ZIEGLER, A. (2015). ranger: a fast implementation of random forests for high dimensional data in C++ and R. arXiv preprint. Available at arXiv:1508.04409.

ZHANG, C.-H. and ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 217–242. MR3153940 https://doi.org/10.1111/rssb.12026

The Annals of Statistics
2024, Vol. 52, No. 5, 2004–2033
https://doi.org/10.1214/24-AOS2418

LEAVE-ONE-OUT SINGULAR SUBSPACE PERTURBATION ANALYSIS FOR SPECTRAL CLUSTERING

BY ANDERSON Y. ZHANG[1,a] AND HARRISON Y. ZHOU[2,b]

[1]*Department of Statistics and Data Science, University of Pennsylvania,* [a]*ayz@wharton.upenn.edu*
[2]*Department of Statistics and Data Science, Yale University,* [b]*huibin.zhou@yale.edu*

The singular subspaces perturbation theory is of fundamental importance in probability and statistics. It has various applications across different fields. We consider two arbitrary matrices where one is a leave-one-column-out submatrix of the other one and establish a novel perturbation upper bound for the distance between the two corresponding singular subspaces. It is well suited for mixture models and results in a sharper and finer statistical analysis than classical perturbation bounds such as Wedin's theorem. Empowered by this leave-one-out perturbation theory, we provide a deterministic entrywise analysis for the performance of spectral clustering under mixture models. Our analysis leads to an explicit exponential error rate for spectral clustering of sub-Gaussian mixture models. For the mixture of isotropic Gaussians, the rate is optimal under a weaker signal-to-noise condition than that of Löffler et al. (2021).

REFERENCES

[1] ABBE, E., FAN, J. and WANG, K. (2022). An ℓ_p theory of PCA and spectral clustering. *Ann. Statist.* **50** 2359–2385. MR4474494 https://doi.org/10.1214/22-aos2196

[2] ABBE, E., FAN, J., WANG, K. and ZHONG, Y. (2020). Entrywise eigenvector analysis of random matrices with low expected rank. *Ann. Statist.* **48** 1452–1474. MR4124330 https://doi.org/10.1214/19-AOS1854

[3] AGTERBERG, J., LUBBERTS, Z. and PRIEBE, C. E. (2022). Entrywise estimation of singular vectors of low-rank matrices with heteroskedasticity and dependence. *IEEE Trans. Inf. Theory* **68** 4618–4650. MR4449064 https://doi.org/10.1109/tit.2022.3159085

[4] BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2567175 https://doi.org/10.1007/978-1-4419-0661-8

[5] BELABBAS, M.-A. and WOLFE, P. J. (2009). Spectral methods in machine learning and new strategies for very large datasets. *Proc. Natl. Acad. Sci. USA* **106** 369–374.

[6] BLUM, A., COJA-OGHLAN, A., FRIEZE, A. and ZHOU, S. (2007). Separating populations with wide data: A spectral analysis. In *Algorithms and Computation. Lecture Notes in Computer Science* **4835** 439–451. Springer, Berlin. MR2472630 https://doi.org/10.1007/978-3-540-77120-3_39

[7] CAI, C., LI, G., CHI, Y., POOR, H. V. and CHEN, Y. (2021). Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees. *Ann. Statist.* **49** 944–967. MR4255114 https://doi.org/10.1214/20-aos1986

[8] CAI, T., LI, H. and MA, R. (2021). Optimal structured principal subspace estimation: Metric entropy and minimax rates. *J. Mach. Learn. Res.* **22** Paper No. 46, 45. MR4253739

[9] CAI, T. T. and ZHANG, A. (2018). Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *Ann. Statist.* **46** 60–89. MR3766946 https://doi.org/10.1214/17-AOS1541

[10] CAPE, J., TANG, M. and PRIEBE, C. E. (2019). The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *Ann. Statist.* **47** 2405–2439. MR3988761 https://doi.org/10.1214/18-AOS1752

[11] CHEN, Y., CHI, Y., FAN, J. and MA, C. (2021). Spectral methods for data science: A statistical perspective. *Found. Trends Mach. Learn.* **14** 566–806.

MSC2020 subject classifications. 62H30.
Key words and phrases. Mixture model, spectral clustering, singular subspace, spectral perturbation, leave-one-out analysis.

[12] DAVIS, C. and KAHAN, W. M. (1970). The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.* **7** 1–46. MR0264450 https://doi.org/10.1137/0707001
[13] DAVIS, D., DIAZ, M. and WANG, K. (2021). Clustering a mixture of Gaussians with unknown covariance. Preprint. Available at arXiv:2110.01602.
[14] DING, L. and CHEN, Y. (2020). Leave-one-out approach for matrix completion: Primal and dual analysis. *IEEE Trans. Inf. Theory* **66** 7274–7301. MR4173640 https://doi.org/10.1109/TIT.2020.2992769
[15] FAN, J., WANG, W. and ZHONG, Y. (2017). An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. *J. Mach. Learn. Res.* **18** Paper No. 207, 42. MR3827095
[16] GIRAUD, C. and VERZELEN, N. (2018). Partial recovery bounds for clustering with the relaxed K-means. *Math. Stat. Learn.* **1** 317–374. MR4059724
[17] HAN, R., LUO, Y., WANG, M. and ZHANG, A. R. (2022). Exact clustering in tensor block model: Statistical optimality and computational limit. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 1666–1698. MR4515554
[18] JIN, J. (2015). Fast community detection by SCORE. *Ann. Statist.* **43** 57–89. MR3285600 https://doi.org/10.1214/14-AOS1265
[19] KANNAN, R. and VEMPALA, S. (2008). Spectral algorithms. *Found. Trends Theor. Comput. Sci.* **4** 157–288. MR2558901 https://doi.org/10.1561/0400000025
[20] KISELEV, V. Y., ANDREWS, T. S. and HEMBERG, M. (2019). Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat. Rev. Genet.* **20** 273–282. https://doi.org/10.1038/s41576-018-0088-9
[21] KOLTCHINSKII, V. and LOUNICI, K. (2016). Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 1976–2013. MR3573302 https://doi.org/10.1214/15-AIHP705
[22] KOLTCHINSKII, V. and XIA, D. (2016). Perturbation of linear forms of singular vectors under Gaussian noise. In *High Dimensional Probability VII. Progress in Probability* **71** 397–423. Springer, Cham. MR3565274 https://doi.org/10.1007/978-3-319-40519-3_18
[23] LEI, J. and LIN, K. Z. (2023). Bias-adjusted spectral clustering in multi-layer stochastic block models. *J. Amer. Statist. Assoc.* **118** 2433–2445. MR4681594 https://doi.org/10.1080/01621459.2022.2054817
[24] LEI, J. and RINALDO, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann. Statist.* **43** 215–237. MR3285605 https://doi.org/10.1214/14-AOS1274
[25] LEI, L. (2019). Unified $\ell_{2\to\infty}$ eigenspace perturbation theory for symmetric random matrices. Preprint. Available at arXiv:1909.04798.
[26] LÖFFLER, M., ZHANG, A. Y. and ZHOU, H. H. (2021). Optimality of spectral clustering in the Gaussian mixture model. *Ann. Statist.* **49** 2506–2530. MR4338373 https://doi.org/10.1214/20-aos2044
[27] LU, Y. and ZHOU, H. H. (2016). Statistical and computational guarantees of lloyd's algorithm and its variants. Preprint. Available at arXiv:1612.02099.
[28] MA, C., WANG, K., CHI, Y. and CHEN, Y. (2018). Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion, and blind deconvolution. In *International Conference on Machine Learning* 3345–3354. PMLR.
[29] MONTI, S., TAMAYO, P., MESIROV, J. and GOLUB, T. (2003). Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Mach. Learn.* **52** 91–118.
[30] NDAOUD, M. (2022). Sharp optimal recovery in the two component Gaussian mixture model. *Ann. Statist.* **50** 2096–2126. MR4474484 https://doi.org/10.1214/22-aos2178
[31] NDAOUD, M., SIGALLA, S. and TSYBAKOV, A. B. (2022). Improved clustering algorithms for the bipartite stochastic block model. *IEEE Trans. Inf. Theory* **68** 1960–1975. MR4395508 https://doi.org/10.1109/tit.2021.3130683
[32] NEWMAN, M. E. (2013). Spectral methods for community detection and graph partitioning. *Phys. Rev. B* **88** 042822.
[33] QIN, T. and ROHE, K. (2013). Regularized spectral clustering under the degree-corrected stochastic blockmodel. *Adv. Neural Inf. Process. Syst.* 3120–3128.
[34] ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. MR2893856 https://doi.org/10.1214/11-AOS887
[35] SCHIEBINGER, G., WAINWRIGHT, M. J. and YU, B. (2015). The geometry of kernelized spectral clustering. *Ann. Statist.* **43** 819–846. MR3325711 https://doi.org/10.1214/14-AOS1283
[36] SRIVASTAVA, P. R., SARKAR, P. and HANASUSANTO, G. A. (2023). A robust spectral clustering algorithm for sub-Gaussian mixture models with outliers. *Oper. Res.* **71** 224–244. MR4560196 https://doi.org/10.1287/opre.2022.2317
[37] STEWART, G. W. (1990). Matrix perturbation theory.
[38] TIBSHIRANI, R., WALTHER, G. and HASTIE, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 411–423. MR1841503 https://doi.org/10.1111/1467-9868.00293

[39] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

[40] VON LUXBURG, U. (2007). A tutorial on spectral clustering. *Stat*. *Comput*. **17** 395–416. MR2409803 https://doi.org/10.1007/s11222-007-9033-z

[41] VON LUXBURG, U., BELKIN, M. and BOUSQUET, O. (2008). Consistency of spectral clustering. *Ann*. *Statist*. **36** 555–586. MR2396807 https://doi.org/10.1214/009053607000000640

[42] WANG, J. (2010). Consistent selection of the number of clusters via crossvalidation. *Biometrika* **97** 893–904. MR2746159 https://doi.org/10.1093/biomet/asq061

[43] WANG, K., YAN, Y. and DIAZ, M. (2020). Efficient clustering for stretched mixtures: Landscape and optimality. *Adv. Neural Inf. Process*. *Syst*. **33** 21309–21320.

[44] WEDIN, P. (1972). Perturbation bounds in connection with singular value decomposition. *BIT* **12** 99–111. MR0309968 https://doi.org/10.1007/bf01932678

[45] YU, Y., WANG, T. and SAMWORTH, R. J. (2015). A useful variant of the Davis-Kahan theorem for statisticians. *Biometrika* **102** 315–323. MR3371006 https://doi.org/10.1093/biomet/asv008

[46] ZHANG, A. Y. and ZHOU, H. Y. (2024). Supplement to "Leave-one-out singular subspace perturbation analysis for spectral clustering." https://doi.org/10.1214/24-AOS2418SUPP

[47] ZHOU, Z. and AMINI, A. A. (2019). Analysis of spectral clustering algorithms for community detection: The general bipartite setting. *J. Mach*. *Learn*. *Res*. **20** Paper No. 47, 47. MR3948087

The Annals of Statistics
2024, Vol. 52, No. 5, 2034–2058
https://doi.org/10.1214/24-AOS2420

TESTING HIGH-DIMENSIONAL REGRESSION COEFFICIENTS IN LINEAR MODELS

BY ALEX ZHAO[1,a], CHANGCHENG LI[2,d], RUNZE LI[1,b] AND ZHE ZHANG[1,c]

[1] *Department of Statistics, Pennsylvania State University at University Park,* [a] *alexzhao@alumni.psu.edu,* [b] *rzli@psu.edu,* [c] *zxz288@psu.edu*

[2] *School of Mathematical Sciences, Dalian University of Technology,* [d] *lichangcheng@dlut.edu.cn*

This paper is concerned with statistical inference for regression coefficients in high-dimensional linear regression models. We propose a new method for testing the coefficient vector of the high-dimensional linear models, and establish the asymptotic normality of our proposed test statistic with the aid of the martingale central limit theorem. We derive the asymptotical relative efficiency (ARE) of the proposed test with respect to the test proposed in Zhong and Chen (*J. Amer. Statist. Assoc.* **106** (2011) 260–274), and show that the ARE is always greater or equal to one under the local alternative studied in this paper. Our numerical studies imply that the proposed test with critical values derived from its asymptotical normal distribution may retain Type I error rate very well. Our numerical comparison demonstrates the proposed test performs better than existing ones in terms of powers. We further illustrate our proposed method with a real data example.

REFERENCES

ARIAS-CASTRO, E., CANDÈS, E. J. and PLAN, Y. (2011a). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *Ann. Statist.* **39** 2533–2556. MR2906877 https://doi.org/10.1214/11-AOS910

ARIAS-CASTRO, E., CANDÈS, E. J. and PLAN, Y. (2011b). Supplement to "Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism.". https://doi.org/10.1214/11-AOS910SUPP

BAI, Z. and SARANADASA, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6** 311–329. MR1399305

CHEN, L. S., PAUL, D., PRENTICE, R. L. and WANG, P. (2011). A regularized Hotelling's T^2 test for pathway analysis in proteomic studies. *J. Amer. Statist. Assoc.* **106** 1345–1360. MR2896840 https://doi.org/10.1198/jasa.2011.ap10599

CHEN, S. X. and QIN, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. MR2604697 https://doi.org/10.1214/09-AOS716

CHUDIK, A., KAPETANIOS, G. and PESARAN, M. H. (2018). A one covariate at a time, multiple testing approach to variable selection in high-dimensional linear regression models. *Econometrica* **86** 1479–1512. MR3843496 https://doi.org/10.3982/ECTA14176

CUI, H., GUO, W. and ZHONG, W. (2018). Test for high-dimensional regression coefficients using refitted cross-validation variance estimation. *Ann. Statist.* **46** 958–988. MR3797993 https://doi.org/10.1214/17-AOS1573

CUI, X., LI, R., YANG, G. and ZHOU, W. (2020). Empirical likelihood test for a large-dimensional mean vector. *Biometrika* **107** 591–607. MR4138978 https://doi.org/10.1093/biomet/asaa005

GOEMAN, J. J., VAN DE GEER, S. A. and VAN HOUWELINGEN, H. C. (2006). Testing against a high dimensional alternative. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 477–493. MR2278336 https://doi.org/10.1111/j.1467-9868.2006.00551.x

HUANG, Y., LI, C., LI, R. and YANG, S. (2022). An overview of tests on high-dimensional means. *J. Multivariate Anal.* **188** Paper No. 104813, 14 pp. MR4353841 https://doi.org/10.1016/j.jmva.2021.104813

LAN, W., WANG, H. and TSAI, C.-L. (2014). Testing covariates in high-dimensional regression. *Ann. Inst. Statist. Math.* **66** 279–301. MR3171406 https://doi.org/10.1007/s10463-013-0414-0

LIU, W. and LUO, S. (2014). Hypothesis testing for high-dimensional regression models. Technical report. Available at https://math.sjtu.edu.cn/faculty/weidongl/Publication/regression.pdf.

MSC2020 subject classifications. Primary 62J05, 62F03; secondary 62J20.
Key words and phrases. Hotelling T^2-test, power enhancement, test of high-dimensional means.

PRAT, A., BIANCHINI, G., THOMAS, M., BELOUSOV, A., CHEANG, M. C., KOEHLER, A., GÓMEZ, P., SEMIGLAZOV, V., EIERMANN, W. et al. (2014). Based PAM50 subtype predictor identifies higher responses and improved survival outcomes in HER2-positive breast cancer in the NOAH study. *Clin. Cancer Res.* **20** 511–521.

VERSHYNIN, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science. Cambridge Series in Statistical and Probabilistic Mathematics* **47**. Cambridge Univ. Press, Cambridge. MR3837109 https://doi.org/10.1017/9781108231596

XIA, Y., CAI, T. and CAI, T. T. (2018). Two-sample tests for high-dimensional linear regression with an application to detecting interactions. *Statist. Sinica* **28** 63–92. MR3752252

ZHAO, A., LI, C., LI, R. and ZHANG, Z. (2024). Supplement to "Testing high-dimensional regression coefficients in linear models." https://doi.org/10.1214/24-AOS2420SUPP

ZHONG, P.-S. and CHEN, S. X. (2011). Tests for high-dimensional regression coefficients with factorial designs. *J. Amer. Statist. Assoc.* **106** 260–274. MR2816719 https://doi.org/10.1198/jasa.2011.tm10284

The Annals of Statistics
2024, Vol. 52, No. 5, 2059–2080
https://doi.org/10.1214/24-AOS2421

A CONFORMAL TEST OF LINEAR MODELS VIA PERMUTATION-AUGMENTED REGRESSIONS

BY LEYING GUAN[a]

Department of Biostatistics, Yale University, [a]*leying.guan@yale.edu*

Permutation tests are widely recognized as robust alternatives to tests based on normal theory. Random permutation tests have been frequently employed to assess the significance of variables in linear models. Despite their widespread use, existing random permutation tests lack finite-sample and assumption-free guarantees for controlling type I error in partial correlation tests. To address this ongoing challenge, we have developed a conformal test through permutation-augmented regressions, which we refer to as PALMRT. PALMRT not only achieves power competitive with conventional methods but also provides reliable control of type I errors at no more than 2α, given any targeted level α, for arbitrary fixed designs and error distributions. We have confirmed this through extensive simulations.

Compared to the cyclic permutation test (CPT) and residual permutation test (RPT), which also offer theoretical guarantees, PALMRT does not compromise as much on power or set stringent requirements on the sample size, making it suitable for diverse biomedical applications. We further illustrate the differences in a long-Covid study where PALMRT validated key findings previously identified using the t-test after multiple corrections, while both CPT and RPT suffered from a drastic loss of power and failed to identify any discoveries. We endorse PALMRT as a robust and practical hypothesis test in scientific research for its superior error control, power preservation, and simplicity.

REFERENCES

[1] ANDERSON, M. J. and ROBINSON, J. (2001). Permutation tests for linear models. *Aust*. *N*. *Z*. *J*. *Stat*. **43** 75–88.

[2] BARBER, R. F. and CANDÈS, E. J. (2015). Controlling the false discovery rate via knockoffs. *Ann*. *Statist*. **43**.

[3] BARBER, R. F., CANDES, E. J., RAMDAS, A. and TIBSHIRANI, R. J. (2021). Predictive inference with the jackknife+. *Ann*. *Statist*. **49**.

[4] DAVISON, A. C. and HINKLEY, D. V. (1997). *Bootstrap Methods and Their Application* **1**. Cambridge University Press, Cambridge.

[5] DICICCIO, T. J. and EFRON, B. (1996). Bootstrap confidence intervals. *Statist*. *Sci*. **11** 189–228.

[6] DICICCIO, T. J. and ROMANO, J. P. (1988). A review of bootstrap confidence intervals. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **50** 338–354.

[7] DRAPER, N. R. and STONEMAN, D. M. (1966). Testing for the inclusion of variables in einear regression by a randomisation technique. *Technometrics* **8** 695–699.

[8] EDGINGTON, E. and ONGHENA, P. (2007). *Randomization Tests*. CRC press, Boca Raton.

[9] EFRON, B. (1987). Better bootstrap confidence intervals. *J*. *Amer*. *Statist*. *Assoc*. **82** 171–185.

[10] EFRON, B. and NARASIMHAN, B. (2020). The automatic construction of bootstrap confidence intervals. *J*. *Comput*. *Graph*. *Statist*. **29** 608–619. https://doi.org/10.1080/10618600.2020.1714633

[11] EFRON, B. and TIBSHIRANI, R. J. (1994). *An Introduction to the Bootstrap*. CRC press, Boca Raton.

[12] FISHER, R. A. (1922). The goodness of fit of regression formulae, and the distribution of regression coefficients. *J*. *R*. *Stat*. *Soc*. **85** 597–612.

[13] FISHER, R. A. (1970). Statistical methods for research workers. In *Breakthroughs in Statistics*: *Methodology and Distribution* 66–70. Springer, Berlin.

MSC2020 subject classifications. 62J20, 62H15, 62F35.
Key words and phrases. Partial correlation, random permutation, assumption-free, conformal test.

[14] FISHER, R. A. et al. (1924). 036: On a distribution yielding the error functions of several well known statistics.

[15] FREEDMAN, D. and LANE, D. (1983). A nonstochastic interpretation of reported significance levels. *J. Bus. Econom. Statist.* **1** 292–298.

[16] GARTHWAITE, P. H. (1996). Confidence intervals from randomization tests. *Biometrics* 1387–1393.

[17] GUAN, L. (2024). Supplement to "A conformal test of linear models via permutation-augmented regressions." https://doi.org/10.1214/24-AOS2421SUPP

[18] GUPTA, C., KUCHIBHOTLA, A. K. and RAMDAS, A. (2022). Nested conformal prediction and quantile out-of-bag ensemble methods. *Pattern Recognit.* **127** 108496.

[19] HALL, P. (1988). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist.* 927–953.

[20] HALL, P. and WILSON, S. R. (1991). Two guidelines for bootstrap hypothesis testing. *Biometrics* 757–762.

[21] HAN, Y., XU, M. and GUAN, L. (2023). Conformalized semi-supervised random forest for classification and abnormality detection. ArXiv preprint arXiv:2302.02237.

[22] KENNEDY, F. E. (1995). Randomization tests in econometrics. *J. Bus. Econom. Statist.* **13** 85–94.

[23] KIM, B., XU, C. and BARBER, R. (2020). Predictive inference is free with the jackknife+-after-bootstrap. *Adv. Neural Inf. Process. Syst.* **33** 4138–4149.

[24] KLEIN, J., WOOD, J., JAYCOX, J., DHODAPKAR, R. M., LU, P., GEHLHAUSEN, J. R., TABACHNIKOVA, A., GREENE, K., TABACOF, L. et al. (2023). Distinguishing features of long COVID identified through immune profiling. *Nature* 1–3.

[25] LEI, L. and BICKEL, P. J. (2021). An assumption-free exact test for fixed-design linear models with exchangeable errors. *Biometrika* **108** 397–412.

[26] MANLY, B. F. (2006). *Randomization, Bootstrap and Monte Carlo Methods in Biology* **70**. CRC press, Boca Raton.

[27] MEINSHAUSEN, N. (2015). Group bound: Confidence intervals for groups of variables in sparse high dimensional regression without assumptions on the design. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 923–945.

[28] PITMAN, E. J. G. (1937). Significance tests which may be applied to samples from any populations. II. The correlation coefficient test. *Suppl. J. R. Stat. Soc.* **4** 225–232.

[29] REN, Z. and BARBER, R. F. (2022). Derandomized knockoffs: Leveraging e-values for false discovery rate control. ArXiv preprint arXiv:2205.15461.

[30] TER BRAAK, C. J. (1992). Permutation versus bootstrap significance tests in multiple regression and ANOVA. In *Bootstrapping and Related Techniques: Proceedings of an International Conference, Held in Trier, FRG, June* 4–8, 1990 79–85. Springer, Berlin.

[31] VOVK, V., NOURETDINOV, I., MANOKHIN, V. and GAMMERMAN, A. (2018). Cross-conformal predictive distributions. In *Conformal and Probabilistic Prediction and Applications* 37–51. PMLR.

[32] VOVK, V. and WANG, R. (2021). E-values: Calibration, combination and applications. *Ann. Statist.* **49** 1736–1754.

[33] WANG, R. and RAMDAS, A. (2022). False discovery rate control with e-values. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 822–852.

[34] WEN, K., WANG, T. and WANG, Y. (2022). Residual permutation test for high-dimensional regression coefficient testing. ArXiv preprint arXiv:2211.16182.

[35] WESTFALL, P. H. and YOUNG, S. S. (1993). *Resampling-Based Multiple Testing: Examples and Methods for P-Value Adjustment* **279**. Wiley, New York.

[36] WINKLER, A. M., RIDGWAY, G. R., WEBSTER, M. A., SMITH, S. M. and NICHOLS, T. E. (2014). Permutation inference for the general linear model. *NeuroImage* **92** 381–397. https://doi.org/10.1016/j.neuroimage.2014.01.060

The Annals of Statistics
2024, Vol. 52, No. 5, 2081–2111
https://doi.org/10.1214/24-AOS2423

ESTIMATING A DENSITY NEAR AN UNKNOWN MANIFOLD: A BAYESIAN NONPARAMETRIC APPROACH

BY CLÉMENT BERENFELD[1,a], PAUL ROSA[2,b] AND JUDITH ROUSSEAU[2,c]

[1]*Institut für Mathematik, Universität Potsdam,* [a]*berenfeld@uni-potsdam.de*
[2]*Department of Statistics, University of Oxford,* [b]*paul.rosa@jesus.ox.ac.uk,* [c]*judith.rousseau@stats.ox.ac.uk*

We study the Bayesian density estimation of data living in the offset of an unknown submanifold of the Euclidean space. In this perspective, we introduce a new notion of anisotropic Hölder for the underlying density and obtain posterior rates that are minimax optimal and adaptive to the regularity of the density, to the intrinsic dimension of the manifold, and to the size of the offset, provided that the latter is not too small—while still allowed to go to zero. Our Bayesian procedure, based on location-scale mixtures of Gaussians, appears to be convenient to implement and yields good practical results, even for quite singular data.

REFERENCES

[1] AAMARI, E. and LEVRARD, C. (2019). Nonasymptotic rates for manifold, tangent space and curvature estimation. *Ann. Statist.* **47** 177–204.

[2] ALEXANDER, S. B. and BISHOP, R. L. (2006). Gauss equation and injectivity radii for subspaces in spaces of curvature bounded above. *Geom. Dedicata* **117** 65–84.

[3] ARJOVSKY, M. and BOTTOU, L. (2017). Towards principled methods for training generative adversarial networks. https://doi.org/10.48550/ARXIV.1701.04862

[4] ARJOVSKY, M., CHINTALA, S. and BOTTOU, L. (2017). Wasserstein GAN. https://doi.org/10.48550/ARXIV.1701.07875

[5] BELKIN, M. and NIYOGI, P. (2001). Laplacian eigenmaps and spectral techniques for embedding and clustering. *Adv. Neural Inf. Process. Syst.* **14**.

[6] BELKIN, M. and NIYOGI, P. (2001). Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in Neural Information Processing Systems* (T. Dietterich, S. Becker and Z. Ghahramani, eds.) **14**. MIT Press.

[7] BERENFELD, C. and HOFFMANN, M. (2021). Density estimation on an unknown submanifold. *Electron. J. Stat.* **15** 2179–2223.

[8] BERENFELD, C., ROSA, P. and ROUSSEAU, J. (2024). Supplement to "Estimating a density near an unknown manifold: a Bayesian nonparametric approach." https://doi.org/10.1214/24-AOS2423SUPP

[9] BINGHAM, E., CHEN, J. P., JANKOWIAK, M., OBERMEYER, F., PRADHAN, N., KARALETSOS, T., SINGH, R., SZERLIP, P., HORSFALL, P. and GOODMAN, N. D. (2019). Pyro: Deep universal probabilistic programming. *J. Mach. Learn. Res.* **20** 973–978.

[10] BOISSONNAT, J.-D., LIEUTIER, A. and WINTRAECKEN, M. (2019). The reach, metric distortion, geodesic convexity and the variation of tangent spaces. *J. Appl. Comput. Topol.* **3** 29–58. https://doi.org/10.1007/s41468-019-00029-8

[11] CAMERLENGHI, F., DOLERA, E., FAVARO, S. and MAININI, E. (2022). Wasserstein posterior contraction rates in non-dominated Bayesian nonparametric models. Preprint. Available at arXiv:2201.12225.

[12] CANALE, A. and DE BLASI, P. (2017). Posterior asymptotics of nonparametric location-scale mixtures for multivariate density estimation. *Bernoulli* **23** 379–404.

[13] CAPITAO-MINICONI, J. and GASSIAT, É. (2023). Deconvolution of spherical data corrupted with unknown noise. *Electron. J. Stat.* **17** 607–649.

[14] CHAE, M., KIM, D., KIM, Y. and LIN, L. (2021). A likelihood approach to nonparametric estimation of a singular distribution using deep generative models. Preprint. Available at arXiv:2105.04046.

[15] CHEN, Y.-C., GENOVESE, C. R. and WASSERMAN, L. (2015). Asymptotic theory for density ridges. *Ann. Statist.* **43** 1896–1928.

MSC2020 subject classifications. Primary 62G07, 62G20; secondary 53A07.
Key words and phrases. Density estimation, Bayesian nonparametrics, minimax adaptive estimation, posterior concentration rates, manifold learning.

[16] CLEANTHOUS, G., GEORGIADIS, A. G. and PORCU, E. (2019). Minimax density estimation on Sobolev spaces with dominating mixed smoothness. Preprint. Available at arXiv:1906.06835.

[17] COMTE, F. and LACOUR, C. (2013). Anisotropic adaptive kernel deconvolution. In *Annales de l'IHP Probabilités et statistiques* **49** 569–609.

[18] DIVOL, V. (2020). Minimax adaptive estimation in manifold inference. https://doi.org/10.48550/ARXIV.2001.04896

[19] DIVOL, V. (2021). Reconstructing measures on manifolds: An optimal transport approach. Preprint. Available at arXiv:2102.07595.

[20] DUNSON, D., CHEN, M., SILVA, J., PAISLEY, J., WANG, C. and CARIN, L. (2010). Compressive sensing on manifolds using a nonparametric mixture of factor analyzers: Algorithm and performance bounds. *IEEE Trans. Signal Process.* **58** 6140–6155.

[21] DUNSON, D. B. and WU, N. (2021). Inferring manifolds from noisy data using Gaussian processes. Preprint. Available at arXiv:2110.07478.

[22] FEDERER, H. (1959). Curvature measures. *Trans. Amer. Math. Soc.* **93** 418–491.

[23] FEFFERMAN, C., MITTER, S. and NARAYANAN, H. (2016). Testing the manifold hypothesis. *J. Amer. Math. Soc.* **29** 983–1049.

[24] GENOVESE, C. R., PERONE-PACIFICO, M., VERDINELLI, I. and WASSERMAN, L. (2012). Manifold estimation and singular deconvolution under Hausdorff loss. *Ann. Statist.* **40**. https://doi.org/10.1214/12-aos994

[25] GENOVESE, C. R., PERONE-PACIFICO, M., VERDINELLI, I. and WASSERMAN, L. (2014). Nonparametric ridge estimation. *Ann. Statist.* **42** 1511–1545.

[26] GHAHRAMANI, Z., HINTON, G. E. et al. (1996). The EM algorithm for mixtures of factor analyzers Technical Report, Technical Report CRG-TR-96-1, University of Toronto.

[27] GHOSAL, S., GHOSH, J. K. and VAN DER VAART, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531.

[28] GHOSAL, S. and VAN DER VAART, A. (2007). Posterior convergence rates of Dirichlet mixtures at smooth densities. *Ann. Statist.* 697–723.

[29] GHOSAL, S. and VAN DER VAART, A. W. (2001). Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.* **29** 1233–1263.

[30] GHOSAL, S. and VAN DER VAART, A. (2017). *Fundamentals of Nonparametric Bayesian Inference. Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

[31] GOLDENSHLUGER, A. and LEPSKI, O. (2011). Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39** 1608–1632.

[32] GOLDENSHLUGER, A. and LEPSKI, O. (2014). On adaptive minimax density estimation on Rd. *Probab. Theory Related Fields* **159** 479–543.

[33] GOODFELLOW, I., BENGIO, Y. and COURVILLE, A. (2016). *Deep Learning. Adapt. Comput. Mach. Learn.* MIT Press, Cambridge, MA.

[34] GOODFELLOW, I. J., POUGET-ABADIE, J., MIRZA, M., XU, B., WARDE-FARLEY, D., OZAIR, S., COURVILLE, A. and BENGIO, Y. (2014). Generative adversarial networks. https://doi.org/0.48550/ARXIV.1406.2661

[35] HOFFMAN, M. and LEPSKI, O. (2002). Random rates in anisotropic regression (with a discussion and a rejoinder by the authors). *Ann. Statist.* **30** 325–396.

[36] HORVAT, C. and PFISTER, J.-P. (2021). Density estimation on low-dimensional manifolds: An inflation-deflation approach. Preprint. Available at arXiv:2105.12152.

[37] KERKYACHARIAN, G., LEPSKI, O. and PICARD, D. (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields* **121** 137–170.

[38] KIM, J., SHIN, J., RINALDO, A. and WASSERMAN, L. (2019). Uniform convergence rate of the kernel density estimator adaptive to intrinsic volume dimension. 3398–3407.

[39] KINGMA, D. P. and WELLING, M. (2013). Auto-encoding variational Bayes. https://doi.org/10.48550/ARXIV.1312.6114

[40] KLEIN, W., PLOMP, R. and POLS, L. C. (1970). Vowel spectra, vowel spaces, and vowel identification. *J. Acoust. Soc. Am.* **48** 999–1009.

[41] KRUIJER, W., ROUSSEAU, J. and VAN DER VAART, A. (2010). Adaptive Bayesian density estimation with location-scale mixtures. *Electron. J. Stat.* **4** 1225–1257.

[42] LEE, J. A. and VERLEYSEN, M. (2007). *Nonlinear Dimensionality Reduction* **1**. Springer, Berlin.

[43] LEE, J. M. (2006). *Riemannian Manifolds: An Introduction to Curvature* **176**. Springer, Berlin.

[44] MA, Y. and FU, Y. (2012). *Manifold Learning Theory and Applications* **434**. CRC Press, Boca Raton.

[45] MARKWICK, D. and ROSS, G. J. (2020). dirichletprocess: Build Dirichlet process objects for Bayesian modelling. https://cran.r-project.org/package=dirichletprocess.

[46] MAUGIS-RABUSSEAU, C. and MICHEL, B. (2013). Adaptive density estimation for clustering with Gaussian mixtures. *ESAIM Probab. Stat*. **17** 698–724.

[47] MEGUELATI, K., FONTEZ, B., HILGERT, N. and MASSEGLIA, F. (2019). Dirichlet process mixture models made scalable and effective by means of massive distribution. In *Proceedings of the* 34*th ACM/SIGAPP Symposium on Applied Computing* 502–509.

[48] MENGERSEN, K., SANTOS-FERNANDEZ, E., DENTI, F., MIRA, A. and VARGHESE, A. (2022). On the intrinsic dimensionality of Covid-19 data: A global perspective. https://doi.org/10.48550/ARXIV.2203.04165

[49] MUKHOPADHYAY, M., LI, D., DUNSON, D. B. et al. (2020). Estimating densities with non-linear support by using Fisher–Gaussian kernels. *J. Roy. Statist. Soc. Ser. B* **82** 1249–1271.

[50] NAULET, Z. and ROUSSEAU, J. (2017). Posterior concentration rates for mixtures of normals in random design regression. *Electron. J. Stat*. **11** 4065–4102.

[51] NEAL, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *J. Comput. Graph. Statist*. **9** 249–265.

[52] NIKOL'SKII, S. M. (2012). *Approximation of Functions of Several Variables and Imbedding Theorems* **205**. Springer, Berlin.

[53] NIYOGI, P., SMALE, S. and WEINBERGER, S. (2008). Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom*. **39** 419–441.

[54] OZAKIN, A. and GRAY, A. (2009). Submanifold density estimation. *Adv. Neural Inf. Process. Syst*. **22**.

[55] PUCHKIN, N. and SPOKOINY, V. G. (2022). Structure-adaptive manifold estimation. *J. Mach. Learn. Res*. **23** 1–62.

[56] ROČKOVÁ, V. and ROUSSEAU, J. (2023). Ideal Bayesian spatial adaptation. *J. Amer. Statist. Assoc*. **0** 1–14. https://doi.org/10.1080/01621459.2023.2241705

[57] ROUSSEAU, J. and SCRICCIOLO, C. (2023). Wasserstein convergence in Bayesian and frequentist deconvolution models. Preprint. Available at arXiv:2309.15300.

[58] ROWEIS, S. T. and SAUL, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science* **290** 2323–2326. https://doi.org/10.1126/science.290.5500.2323

[59] SCHÖLKOPF, B., SMOLA, A. and MÜLLER, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput*. **10** 1299–1319.

[60] SHEN, W., TOKDAR, S. T. and GHOSAL, S. (2013). Adaptive Bayesian multivariate density estimation with Dirichlet mixtures. *Biometrika* **100** 623–640.

[61] TANG, R. and YANG, Y. (2022). Minimax rate of distribution estimation on unknown submanifold under adversarial losses. Preprint. Available at arXiv:2202.09030.

[62] TENENBAUM, J. B., DE SILVA, V. and LANGFORD, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science* **290** 2319–2323.

[63] VINCENT, P. and BENGIO, Y. (2002). Manifold parzen windows. *Adv. Neural Inf. Process. Syst*. **15**.

[64] WEINBERGER, K. Q. and SAUL, L. K. (2006). Unsupervised learning of image manifolds by semidefinite programming. *Int. J. Comput. Vis*. **70** 77–90.

The Annals of Statistics
2024, Vol. 52, No. 5, 2112–2138
https://doi.org/10.1214/24-AOS2424

EXACT MINIMAX OPTIMALITY OF SPECTRAL METHODS IN PHASE SYNCHRONIZATION AND ORTHOGONAL GROUP SYNCHRONIZATION

BY ANDERSON YE ZHANG[a]

Department of Statistics and Data Science, University of Pennsylvania, [a]*ayz@wharton.upenn.edu*

We study the performance of the spectral method for the phase synchronization problem with additive Gaussian noises and incomplete data. The spectral method utilizes the leading eigenvector of the data matrix followed by a normalization step. We prove that it achieves the minimax lower bound of the problem with a matching leading constant under a squared ℓ_2 loss. This shows that the spectral method has the same performance as more sophisticated procedures including maximum likelihood estimation, generalized power method, and semidefinite programming, as long as consistent parameter estimation is possible. To establish our result, we first have a novel choice of the population eigenvector, which enables us to establish the exact recovery of the spectral method when there is no additive noise. We then develop a new perturbation analysis toolkit for the leading eigenvector and show it can be well-approximated by its first-order approximation with a small ℓ_2 error. We further extend our analysis to establish the exact minimax optimality of the spectral method for the orthogonal group synchronization.

REFERENCES

[1] ABBE, E., FAN, J., WANG, K. and ZHONG, Y. (2020). Entrywise eigenvector analysis of random matrices with low expected rank. *Ann. Statist.* **48** 1452–1474. MR4124330 https://doi.org/10.1214/19-AOS1854

[2] ABBE, E., MASSOULIÉ, L., MONTANARI, A., SLY, A. and SRIVASTAVA, N. (2018). Group synchronization on grids. *Math. Stat. Learn.* **1** 227–256. MR4059722 https://doi.org/10.4171/msl/6

[3] AGTERBERG, J., LUBBERTS, Z. and PRIEBE, C. E. (2022). Entrywise estimation of singular vectors of low-rank matrices with heteroskedasticity and dependence. *IEEE Trans. Inf. Theory* **68** 4618–4650. MR4449064 https://doi.org/10.1109/tit.2022.3159085

[4] ARIE-NACHIMSON, M., KOVALSKY, S. Z., KEMELMACHER-SHLIZERMAN, I., SINGER, A. and BASRI, R. (2012). Global motion estimation from point matches. In 2012 *Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission* 81–88. IEEE, New York.

[5] BANDEIRA, A. S., BOUMAL, N. and SINGER, A. (2017). Tightness of the maximum likelihood semidefinite relaxation for angular synchronization. *Math. Program.* **163** 145–167. MR3632977 https://doi.org/10.1007/s10107-016-1059-6

[6] BOUCHERON, S., LUGOSI, G. and BOUSQUET, O. (2003). Concentration inequalities. In *Summer School on Machine Learning* 208–240. Springer, Berlin.

[7] BOUMAL, N. (2016). Nonconvex phase synchronization. *SIAM J. Optim.* **26** 2355–2377. MR3566919 https://doi.org/10.1137/16M105808X

[8] BOUMAL, N., SINGER, A. and ABSIL, P.-A. (2013). Robust estimation of rotations from relative measurements by maximum likelihood. In 52*nd IEEE Conference on Decision and Control* 1156–1161. IEEE, New York.

[9] CAI, C., LI, G., CHI, Y., POOR, H. V. and CHEN, Y. (2021). Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees. *Ann. Statist.* **49** 944–967. MR4255114 https://doi.org/10.1214/20-aos1986

[10] CAPE, J., TANG, M. and PRIEBE, C. E. (2019). The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *Ann. Statist.* **47** 2405–2439. MR3988761 https://doi.org/10.1214/18-AOS1752

[11] CHEN, Y., CHI, Y., FAN, J., MA, C. et al. (2021). Spectral methods for data science: A statistical perspective. *Found. Trends Mach. Learn.* **14** 566–806.

MSC2020 subject classifications. 62C20.
Key words and phrases. Spectral method, spectral perturbation, synchronization, minimax risk.

[12] CUCURINGU, M. (2016). Sync-rank: Robust ranking, constrained ranking and rank aggregation via eigenvector and SDP synchronization. *IEEE Trans. Netw. Sci. Eng.* **3** 58–79. MR3477931 https://doi.org/10.1109/TNSE.2016.2523761

[13] DAVIS, C. and KAHAN, W. M. (1970). The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.* **7** 1–46. MR0264450 https://doi.org/10.1137/0707001

[14] FAN, J., WANG, W. and ZHONG, Y. (2017). An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. *J. Mach. Learn. Res.* **18** Paper No. 207, 42 pp. MR3827095

[15] FAN, Y., KHOO, Y. and ZHAO, Z. (2022). Joint community detection and rotational synchronization via semidefinite programming. *SIAM J. Math. Data Sci.* **4** 1052–1081. MR4467534 https://doi.org/10.1137/21M1419702

[16] FILBIR, F., KRAHMER, F. and MELNYK, O. (2021). On recovery guarantees for angular synchronization. *J. Fourier Anal. Appl.* **27** Paper No. 31, 26 pp. MR4240787 https://doi.org/10.1007/s00041-021-09834-1

[17] GAO, C. and ZHANG, A. Y. (2021). Exact minimax estimation for phase synchronization. *IEEE Trans. Inf. Theory* **67** 8236–8247. MR4346085 https://doi.org/10.1109/TIT.2021.3112712

[18] GAO, C. and ZHANG, A. Y. (2022). SDP achieves exact minimax optimality in phase synchronization. *IEEE Trans. Inf. Theory* **68** 5374–5390. MR4476402 https://doi.org/10.1109/tit.2022.3167603

[19] GAO, C. and ZHANG, A. Y. (2023). Optimal orthogonal group synchronization and rotation group synchronization. *Inf. Inference* **12** 591–632. MR4565748 https://doi.org/10.1093/imaiai/iaac022

[20] IWEN, M. A., PRESKITT, B., SAAB, R. and VISWANATHAN, A. (2020). Phase retrieval from local measurements: Improved robustness via eigenvector-based angular synchronization. *Appl. Comput. Harmon. Anal.* **48** 415–444. MR4016998 https://doi.org/10.1016/j.acha.2018.06.004

[21] JAVANMARD, A., MONTANARI, A. and RICCI-TERSENGHI, F. (2016). Phase transitions in semidefinite relaxations. *Proc. Natl. Acad. Sci. USA* **113** E2218–E2223. MR3494080 https://doi.org/10.1073/pnas.1523097113

[22] LEI, L. (2019). Unified $\ell_{2\to\infty}$ eigenspace perturbation theory for symmetric random matrices. Preprint. Available at arXiv:1909.04798.

[23] LELARGE, M. and MIOLANE, L. (2019). Fundamental limits of symmetric low-rank matrix estimation. *Probab. Theory Related Fields* **173** 859–929. MR3936148 https://doi.org/10.1007/s00440-018-0845-x

[24] LERMAN, G. and SHI, Y. (2022). Robust group synchronization via cycle-edge message passing. *Found. Comput. Math.* **22** 1665–1741. MR4509114 https://doi.org/10.1007/s10208-021-09532-w

[25] LING, S. (2022). Near-optimal performance bounds for orthogonal and permutation group synchronization via spectral methods. *Appl. Comput. Harmon. Anal.* **60** 20–52. MR4387245 https://doi.org/10.1016/j.acha.2022.02.003

[26] LING, S. (2022). Improved performance guarantees for orthogonal group synchronization via generalized power method. *SIAM J. Optim.* **32** 1018–1048. MR4425911 https://doi.org/10.1137/20M1389571

[27] LING, S. (2023). Solving orthogonal group synchronization via convex and low-rank optimization: Tightness and landscape analysis. *Math. Program.* **200** 589–628. MR4590243 https://doi.org/10.1007/s10107-022-01896-3

[28] LIU, H., YUE, M.-C. and SO, A. M.-C. (2017). On the estimation performance and convergence rate of the generalized power method for phase synchronization. *SIAM J. Optim.* **27** 2426–2446. MR3735300 https://doi.org/10.1137/16M110109X

[29] PERRY, A., WEIN, A. S., BANDEIRA, A. S. and MOITRA, A. (2016). Optimality and sub-optimality of PCA for spiked random matrices and synchronization. Preprint. Available at arXiv:1609.05573.

[30] PERRY, A., WEIN, A. S., BANDEIRA, A. S. and MOITRA, A. (2018). Message-passing algorithms for synchronization problems over compact groups. *Comm. Pure Appl. Math.* **71** 2275–2322. MR3862091 https://doi.org/10.1002/cpa.21750

[31] PRESKITT, B. P. (2018). Phase retrieval from locally supported measurements. University of California, San Diego. MR3908057

[32] ROMANOV, E. and GAVISH, M. (2020). The noise-sensitivity phase transition in spectral group synchronization over compact groups. *Appl. Comput. Harmon. Anal.* **49** 935–970. MR4135427 https://doi.org/10.1016/j.acha.2019.05.002

[33] SHEN, Y., HUANG, Q., SREBRO, N. and SANGHAVI, S. (2016). Normalized spectral map synchronization. *Adv. Neural Inf. Process. Syst.* **29**.

[34] SHI, Y. and LERMAN, G. (2020). Message passing least squares framework and its application to rotation synchronization. In *International Conference on Machine Learning* 8796–8806. PMLR.

[35] SINGER, A. (2011). Angular synchronization by eigenvectors and semidefinite programming. *Appl. Comput. Harmon. Anal.* **30** 20–36. MR2737931 https://doi.org/10.1016/j.acha.2010.02.001

[36] SINGER, A. and SHKOLNISKY, Y. (2011). Three-dimensional structure determination from common lines in cryo-EM by eigenvectors and semidefinite programming. *SIAM J. Imaging Sci.* **4** 543–572. MR2810897 https://doi.org/10.1137/090767777

[37] WANG, L. and SINGER, A. (2013). Exact and stable recovery of rotations for robust synchronization. *Inf. Inference* **2** 145–193. MR3311446 https://doi.org/10.1093/imaiai/iat005
[38] ZHANG, A. Y. (2024). Supplement to "Exact minimax optimality of spectral methods in phase synchronization and orthogonal group synchronization." https://doi.org/10.1214/24-AOS2424SUPP
[39] ZHONG, Y. and BOUMAL, N. (2018). Near-optimal bounds for phase synchronization. *SIAM J. Optim.* **28** 989–1016. MR3782406 https://doi.org/10.1137/17M1122025
[40] ZHU, L., WANG, J. and SO, A. M.-C. (2021). Orthogonal group synchronization with incomplete measurements: Error bounds and linear convergence of the generalized power method. Preprint. Available at arXiv:2112.06556.

The Annals of Statistics
2024, Vol. 52, No. 5, 2139–2166
https://doi.org/10.1214/24-AOS2425

EFFICIENCY IN LOCAL DIFFERENTIAL PRIVACY

BY LUKAS STEINBERGER[a]

Department of Statistics and OR, University of Vienna, [a]*lukas.steinberger@univie.ac.at*

We develop a theory of asymptotic efficiency in regular parametric models when data confidentiality is ensured by local differential privacy (LDP). Even though efficient parameter estimation is a classical and well-studied problem in mathematical statistics, it leads to several nontrivial obstacles that need to be tackled when dealing with the LDP case. Starting from a regular parametric model $\mathcal{P} = (P_\theta)_{\theta\in\Theta}$, $\Theta \subseteq \mathbb{R}^p$, for the i.i.d. unobserved sensitive data $X_1, \dots, X_n$, we establish local asymptotic mixed normality (along subsequences) of the model

$$Q^{(n)}\mathcal{P} = (Q^{(n)} P_\theta^n)_{\theta\in\Theta}$$

generating the sanitized observations $Z_1, \dots, Z_n$, where $Q^{(n)}$ is an arbitrary sequence of sequentially interactive privacy mechanisms. This result readily implies convolution and local asymptotic minimax theorems. In case $p = 1$, the optimal asymptotic variance is found to be the inverse of the supremal Fisher information $\sup_{Q\in\mathcal{Q}_\alpha} I_\theta(Q\mathcal{P}) \in \mathbb{R}$, where the supremum runs over all α-differentially private (marginal) Markov kernels. We present an algorithm for finding a (nearly) optimal privacy mechanism $\hat{Q}$ and an estimator $\hat{\theta}_n(Z_1, \dots, Z_n)$ based on the corresponding sanitized data that achieves this asymptotically optimal variance.

REFERENCES

ACHARYA, J., CANONNE, C. L., TYAGI, H. and SUN, Z. (2022). The role of interactivity in structured estimation. In *Conference on Learning Theory* 1328–1355. PMLR.

BARNES, L. P., CHEN, W.-N. and ÖZGÜR, A. (2020). Fisher information under local differential privacy. *IEEE J. Sel. Areas Inf. Theory* **1** 645–659.

BASAWA, I. V. and SCOTT, D. J. (1983). *Asymptotic Optimal Inference for Nonergodic Models*. *Lecture Notes in Statistics* **17**. Springer, New York-Berlin. MR0688650

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. *Johns Hopkins Series in the Mathematical Sciences*. Johns Hopkins Univ. Press, Baltimore, MD. MR1245941

BUTUCEA, C., DUBOIS, A., KROLL, M. and SAUMARD, A. (2020). Local differential privacy: Elbow effect in optimal density estimation and adaptation over Besov ellipsoids. *Bernoulli* **26** 1727–1764. MR4091090 https://doi.org/10.3150/19-BEJ1165

BUTUCEA, C., ROHDE, A. and STEINBERGER, L. (2023). Interactive versus noninteractive locally differentially private estimation: Two elbows for the quadratic functional. *Ann. Statist.* **51** 464–486. MR4600989 https://doi.org/10.1214/22-aos2254

CAI, T. T., WANG, Y. and ZHANG, L. (2021). The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy. *Ann. Statist.* **49** 2825–2850. MR4338894 https://doi.org/10.1214/21-aos2058

DINUR, I. and NISSIM, K. (2003). Revealing information while preserving privacy. In *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* 202–210. ACM, New York.

DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2018). Minimax optimal procedures for locally private estimation. *J. Amer. Statist. Assoc.* **113** 182–201. MR3803452 https://doi.org/10.1080/01621459.2017.1389735

DUCHI, J. C. and RUAN, F. (2024). The right complexity measure in locally private estimation: It is not the Fisher information. *Ann. Statist.* **52** 1–51. MR4718406 https://doi.org/10.1214/22-aos2227

MSC2020 subject classifications. Primary 62F12, 62F30; secondary 62B15, 62L05.
Key words and phrases. Local differential privacy, efficiency, Fisher information, local asymptotic mixed normality.

DUNSCHE, M., KUTTA, T. and DETTE, H. (2022). Multivariate mean comparison under differential privacy. In *Privacy in Statistical Databases*: *International Conference*, *PSD* 2022 31–45. Springer, Berlin.
DWORK, C. (2008). Differential privacy: A survey of results. In *Theory and Applications of Models of Computation*. *Lecture Notes in Computer Science* **4978** 1–19. Springer, Berlin. MR2472670 https://doi.org/10.1007/978-3-540-79228-4_1
DWORK, C., MCSHERRY, F., NISSIM, K. and SMITH, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography* (S. Halevi and T. Rabin, eds.). *Lecture Notes in Computer Science* **3876** 265–284. Springer, Berlin. MR2241676 https://doi.org/10.1007/11681878_14
DWORK, C. and NISSIM, K. (2004). Privacy-preserving datamining on vertically partitioned databases. In *Advances in Cryptology—CRYPTO* 2004. *Lecture Notes in Computer Science* **3152** 528–544. Springer, Berlin. MR2147523 https://doi.org/10.1007/978-3-540-28628-8_32
EVFIMIEVSKI, A., GEHRKE, J. and SRIKANT, R. (2003). Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* 211–222. ACM, New York.
HÁJEK, J. (1969/70). A characterization of limiting distributions of regular estimates. *Z. Wahrsch. Verw. Gebiete* **14** 323–330. MR0283911 https://doi.org/10.1007/BF00533669
HALL, P. (1977). Martingale invariance principles. *Ann. Probab.* **5** 875–887. MR0517471 https://doi.org/10.1214/aop/1176995657
HÖPFNER, R. (2014). *Asymptotic Statistics*: *With a View to Stochastic Processes*. *De Gruyter Graduate*. de Gruyter, Berlin. MR3185373 https://doi.org/10.1515/9783110250282
HUCKE, U. (2019). Local differential privacy and estimation in the binomial model. Master's thesis Univ. Freiburg.
JEGANATHAN, P. (1980). Asymptotic theory of estimation when the limit of the log-likelihood ratios is mixed normal. PhD thesis, Indian Statistical Institute-Kolkata.
JEGANATHAN, P. (1981). On a decomposition of the limit distribution of a sequence of estimators. *Sankhyā Ser. A* **43** 26–36. MR0656266
JEGANATHAN, P. (1982). On the asymptotic theory of estimation when the limit of the log-likelihood ratios is mixed normal. *Sankhyā Ser. A* **44** 173–212. MR0688800
KAIROUZ, P., OH, S. and VISWANATH, P. (2016). Extremal mechanisms for local differential privacy. *J. Mach. Learn. Res.* **17** Paper No. 17. MR3491111
KALININ, N. and STEINBERGER, L. (2024). Efficient estimation of a Gaussian mean with Local Differential Privacy. Available at arXiv:2402.04840.
LALANNE, C., GARIVIER, A. and GRIBONVAL, R. (2022). On the Statistical Complexity of Estimation and Testing under Privacy Constraints. ArXiv preprint. Available at arXiv:2210.02215.
LE CAM, L. (1960). Locally asymptotically normal families of distributions. Certain approximations to families of distributions and their use in the theory of estimation and testing hypotheses. *Univ. California Publ. Statist.* **3** 37–98. MR0126903
ROHDE, A. and STEINBERGER, L. (2020). Geometrizing rates of convergence under local differential privacy constraints. *Ann. Statist.* **48** 2646–2670. MR4152116 https://doi.org/10.1214/19-AOS1901
SMITH, A. (2008). Efficient, differentially private point estimators. ArXiv preprint. Available at arXiv:0809.4794.
SMITH, A. (2011). Privacy-preserving statistical estimation with optimal convergence rates [extended abstract]. In *STOC'11—Proceedings of the* 43*rd ACM Symposium on Theory of Computing* 813–821. ACM, New York. MR2932032 https://doi.org/10.1145/1993636.1993743
STEINBERGER, L. (2024). Supplement to "Efficiency in local differential privacy." https://doi.org/10.1214/24-AOS2425SUPP
VAN DER VAART, A. W. (2007). *Asymptotic Statistics*, 8th ed. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, New York.
WARNER, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc.* **60** 63–69.
WASSERMAN, L. and ZHOU, S. (2010). A statistical framework for differential privacy. *J. Amer. Statist. Assoc.* **105** 375–389. MR2656057 https://doi.org/10.1198/jasa.2009.tm08651

The Annals of Statistics
2024, Vol. 52, No. 5, 2167–2193
https://doi.org/10.1214/24-AOS2430

WASSERSTEIN GENERATIVE ADVERSARIAL NETWORKS ARE MINIMAX OPTIMAL DISTRIBUTION ESTIMATORS

BY ARTHUR STÉPHANOVITCH[1,a], EDDIE AAMARI[1,b] AND CLÉMENT LEVRARD[2,c]

[1]*Département de Mathématiques et Applications, École Normale Supérieure, Université PSL, CNRS,* [a]*stephanovitch@dma.ens.fr,* [b]*eddie.aamari@ens.fr*

[2]*Institut de recherche mathématique de Rennes, Université de Rennes, CNRS,* [c]*clement.levrard@univ-rennes1.fr*

We provide nonasymptotic rates of convergence of the Wasserstein Generative Adversarial networks (WGAN) estimator. We build neural networks classes representing the generators and discriminators which yield a GAN that achieves the minimax optimal rate for estimating a certain probability measure μ with support in $\mathbb{R}^p$. The probability μ is considered to be the push forward of the Lebesgue measure on the d-dimensional torus $\mathbb{T}^d$ by a map $g^\star : \mathbb{T}^d \to \mathbb{R}^p$ of smoothness $\beta + 1$. Measuring the error with the γ-Hölder Integral Probability Metric (IPM), we obtain up to logarithmic factors, the minimax optimal rate $O(n^{-\frac{\beta+\gamma}{2\beta+d}} \vee n^{-\frac{1}{2}})$ where n is the sample size, β determines the smoothness of the target measure μ, γ is the smoothness of the IPM ($\gamma = 1$ is the Wasserstein case) and $d \leq p$ is the intrinsic dimension of μ. In the process, we derive a sharp interpolation inequality between Hölder IPMs. This novel result of theory of functions spaces generalizes classical interpolation inequalities to the case where the measures involved have densities on different manifolds.

REFERENCES

AAMARI, E. and LEVRARD, C. (2019). Nonasymptotic rates for manifold, tangent space and curvature estimation. *Ann. Statist.* **47** 177–204. MR3909931 https://doi.org/10.1214/18-AOS1685

ARJOVSKY, M., CHINTALA, S. and BOTTOU, L. (2017). Wasserstein generative adversarial networks. In *Proceedings of the* 34*th International Conference on Machine Learning* (D. Precup and Y. W. Teh, eds.) **70** 214–223. PMLR.

BELOMESTNY, D., NAUMOV, A., PUCHKIN, N. and SAMSONOV, S. (2023). Simultaneous approximation of a smooth function and its derivatives by deep neural networks with piecewise-polynomial activations. *Neural Netw.* **161** 242–253. https://doi.org/10.1016/j.neunet.2023.01.035

CHAE, M. (2022). Rates of convergence for nonparametric estimation of singular distributions using generative adversarial networks. Preprint. Available at arXiv:2202.02890.

CHEN, M., LIAO, W., ZHA, H. and ZHAO, T. (2020). Distribution approximation and statistical estimation guarantees of generative adversarial networks. Preprint. Available at arXiv:2002.03938.

DAUBECHIES, I. (1988). Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.* **41** 909–996. MR0951745 https://doi.org/10.1002/cpa.3160410705

DAUBECHIES, I., DEVORE, R., DYM, N., FAIGENBAUM-GOLOVIN, S., KOVALSKY, S. Z., LIN, K.-C., PARK, J., PETROVA, G. and SOBER, B. (2023). Neural network approximation of refinable functions. *IEEE Trans. Inf. Theory* **69** 482–495. MR4544969 https://doi.org/10.1109/tit.2022.3199601

DAUBECHIES, I. and LAGARIAS, J. C. (1991). Two-scale difference equations. I. Existence and global regularity of solutions. *SIAM J. Math. Anal.* **22** 1388–1410. MR1112515 https://doi.org/10.1137/0522089

DE RYCK, T., LANTHALER, S. and MISHRA, S. (2021). On the approximation of functions by tanh neural networks. *Neural Netw.* **143** 732–750. https://doi.org/10.1016/j.neunet.2021.08.015

DIVOL, V. (2022). Measure estimation on manifolds: An optimal transport approach. *Probab. Theory Related Fields* **183** 581–647. MR4421180 https://doi.org/10.1007/s00440-022-01118-z

FEDERER, H. (1959). Curvature measures. *Trans. Amer. Math. Soc.* **93** 418–491. MR0110078 https://doi.org/10.2307/1993504

MSC2020 subject classifications. Primary 62G05; secondary 62E17.

Key words and phrases. Minimax rate, generative model, distribution estimation, manifold, interpolation inequality.

GENOVESE, C. R., PERONE-PACIFICO, M., VERDINELLI, I. and WASSERMAN, L. (2012). Manifold estimation and singular deconvolution under Hausdorff loss. *Ann. Statist.* **40** 941–963. MR2985939 https://doi.org/10.1214/12-AOS994

GINÉ, E. and NICKL, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models. Cambridge Series in Statistical and Probabilistic Mathematics* **40**. Cambridge Univ. Press, New York. MR3588285 https://doi.org/10.1017/CBO9781107337862

GONZÁLEZ-PRIETO, Á., MOZO, A., TALAVERA, E. and GÓMEZ-CANAVAL, S. (2021). Dynamics of Fourier modes in torus generative adversarial networks. *Mathematics* **9** 325.

GOODFELLOW, I. J., POUGET-ABADIE, J., MIRZA, M., XU, B., WARDE-FARLEY, D., OZAIR, S., COURVILLE, A. and BENGIO, Y. (2014). Generative adversarial nets. In *Advances in Neural Information Processing Systems* (Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence and K. Q. Weinberger, eds.) **27** 2672–2680. Curran Associates, Red Hook, NY.

GULRAJANI, I., AHMED, F., ARJOVSKY, M., DUMOULIN, V. and COURVILLE, A. C. (2017). Improved training of Wasserstein GANs. In *Advances in Neural Information Processing Systems* (I. Guyon, U. von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, eds.) **30** 5767–5777. Curran Associates, Red Hook, NY.

HAROSKE, D. D. (2007). *Envelopes and Sharp Embeddings of Function Spaces. Chapman & Hall/CRC Research Notes in Mathematics* **437**. CRC Press/CRC, Boca Raton, FL. MR2262450

KARRAS, T., AITTALA, M., LAINE, S., HÄRKÖNEN, E., HELLSTEN, J., LEHTINEN, J. and AILA, T. (2021). Alias-free generative adversarial networks. Preprint. Available at arXiv:2106.12423.

LI, C.-L., CHANG, W.-C., CHENG, Y., YANG, Y. and PÓCZOS, B. (2017). Mmd gan: Towards deeper understanding of moment matching network.

LIANG, T. (2018). How well generative adversarial networks learn distributions: A nonparametric view.

LIANG, T. (2021). How well generative adversarial networks learn distributions. *J. Mach. Learn. Res.* **22** Paper No. 228, 41 pp. MR4329807

LIU, Y., ZHOU, Y., LIU, X., DONG, F., WANG, C. and WANG, Z. (2019). Wasserstein gan-based small-sample augmentation for new-generation artificial intelligence: A case study of cancer-staging data in biology. *Engineering* **5** 156–163.

LUO, Y. and LU, B.-L. (2018). Eeg data augmentation for emotion recognition using a conditional Wasserstein gan. In 2018 40*th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (*EMBC*) 2535–2538. IEEE.

MROUEH, Y., LI, C.-L., SERCU, T., RAJ, A. and CHENG, Y. (2018). Sobolev GAN. In *International Conference on Learning Representations*.

MÜLLER, A. (1997). Integral probability metrics and their generating classes of functions. *Adv. in Appl. Probab.* **29** 429–443. MR1450938 https://doi.org/10.2307/1428011

PUCHKIN, N., SAMSONOV, S., BELOMESTNY, D., MOULINES, E. and NAUMOV, A. (2024). Rates of convergence for density estimation with generative adversarial networks. *J. Mach. Learn. Res.* **25** Paper No. [29], 47 pp. MR4723879

SCHREUDER, N., BRUNE, V.-E. and DALALYAN, A. S. (2021). Statistical guarantees for generative models without domination. In *Algorithmic Learning Theory. Proc. Mach. Learn. Res.* (*PMLR*) **132** 21. MR4227353

SINGH, S., UPPAL, A., LI, B., LI, C.-L., ZAHEER, M. and POCZOS, B. (2018). Nonparametric density estimation under adversarial losses. In *Advances in Neural Information Processing Systems* (S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi and R. Garnett, eds.) **31**. Curran Associates, Red Hook, NY.

STANCZUK, J., ETMANN, C., KREUSSER, L. M. and SCHÖNLIEB, C.-B. (2021). Wasserstein gans work because they fail (to approximate the Wasserstein distance). Preprint. Available at arXiv:2103.01678.

STÉPHANOVITCH, A., AAMARI, E. and LEVRARD, C. (2024). Supplement to "Wasserstein generative adversarial networks are minimax optimal distribution estimators." https://doi.org/10.1214/24-AOS2430SUPP

TANG, R. and YANG, Y. (2023). Minimax rate of distribution estimation on unknown submanifolds under adversarial losses. *Ann. Statist.* **51** 1282–1308. MR4630949 https://doi.org/10.1214/23-aos2291

TSYBAKOV, A. B. (2004). *Introduction to Nonparametric Estimation. Springer Series in Statistics* Springer, New York. https://doi.org/10.1007/b13794

VILLANI, C. (2009). *Optimal Transport: Old and New. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Springer, Berlin. MR2459454 https://doi.org/10.1007/978-3-540-71050-9

VONDRICK, C., PIRSIAVASH, H. and TORRALBA, A. (2016). Generating videos with scene dynamics. In *Advances in Neural Information Processing Systems* (D. Lee, M. Sugiyama, U. von Luxburg, I. Guyon and R. Garnett, eds.) **29** 613–621. Curran Associates, Red Hook, NY.

YU, L., ZHANG, W., WANG, J. and YU, Y. (2017). SeqGAN: Sequence generative adversarial nets with policy gradient. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence* 2852–2858. AAAI Press, Menlo Park.

The Annals of Statistics
2024, Vol. 52, No. 5, 2194–2216
https://doi.org/10.1214/24-AOS2432

QUANTILE PROCESSES AND THEIR APPLICATIONS IN FINITE POPULATIONS

BY ANURAG DEY[a] AND PROBAL CHAUDHURI[b]

Statistics and Mathematics Unit, Indian Statistical Institute, Kolkata, [a]*deyanuragsaltlake64@gmail.com,* [b]*probalchaudhuri@gmail.com*

The weak convergence of the quantile processes, which are constructed based on different estimators of the finite population quantiles, is shown under various well-known sampling designs based on a superpopulation model. The results related to the weak convergence of these quantile processes are applied to find asymptotic distributions of the smooth L-estimators and the estimators of smooth functions of finite population quantiles. Based on these asymptotic distributions, confidence intervals are constructed for several finite population parameters like the median, the α-trimmed means, the interquartile range and the quantile based measure of skewness. Comparisons of various estimators are carried out based on their asymptotic distributions. We show that the use of the auxiliary information in the construction of the estimators sometimes has an adverse effect on the performances of the smooth L-estimators and the estimators of smooth functions of finite population quantiles under several sampling designs. Further, the performance of each of the above-mentioned estimators sometimes becomes worse under sampling designs, which use the auxiliary information, than their performances under simple random sampling without replacement (SRSWOR).

REFERENCES

[1] BERGER, Y. G. (1998). Rate of convergence to normal distribution for the Horvitz–Thompson estimator. *J. Statist. Plann. Inference* **67** 209–226. MR1624693 https://doi.org/10.1016/S0378-3758(97)00107-9

[2] BILLINGSLEY, P. (2013). *Convergence of Probability Measures*, 2nd ed. Wiley, New York. MR0233396

[3] BOISTARD, H., LOPUHAÄ, H. P. and RUIZ-GAZEN, A. (2017). Functional central limit theorems for single-stage sampling designs. *Ann. Statist.* **45** 1728–1758. MR3670194 https://doi.org/10.1214/16-AOS1507

[4] BONDESSON, L., TRAAT, I. and LUNDQVIST, A. (2006). Pareto sampling versus Sampford and conditional Poisson sampling. *Scand. J. Stat.* **33** 699–720. MR2300911 https://doi.org/10.1111/j.1467-9469.2006.00497.x

[5] BRESLOW, N. E. and WELLNER, J. A. (2007). Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression. *Scand. J. Stat.* **34** 86–102. MR2325244 https://doi.org/10.1111/j.1467-9469.2006.00523.x

[6] CARDOT, H., GOGA, C. and LARDIN, P. (2014). Variance estimation and asymptotic confidence bands for the mean estimator of sampled functional data with high entropy unequal probability sampling designs. *Scand. J. Stat.* **41** 516–534. MR3207184 https://doi.org/10.1111/sjos.12048

[7] CHATTERJEE, A. (2011). Asymptotic properties of sample quantiles from a finite population. *Ann. Inst. Statist. Math.* **63** 157–179. MR2748939 https://doi.org/10.1007/s10463-008-0210-4

[8] CHAUDHURI, A., DIHIDAR, K. and BOSE, M. (2006). On the feasibility of basing Horvitz and Thompson's estimator on a sample by Rao, Hartley, and Cochran's scheme. *Comm. Statist. Theory Methods* **35** 2239–2244. MR2338928 https://doi.org/10.1080/03610920600853464

[9] COCHRAN, W. G. (1977). *Sampling Techniques*, 3rd ed. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0474575

[10] CONTI, P. L. (2014). On the estimation of the distribution function of a finite population under high entropy sampling designs, with applications. *Sankhya B* **76** 234–259. MR3302272 https://doi.org/10.1007/s13571-014-0083-x

MSC2020 subject classifications. Primary 62D05, 60F05; secondary 60B05, 60B10.
Key words and phrases. Auxiliary information, difference estimator, Hadamard differentiability, high entropy sampling design, ratio estimator, regression estimator, RHC sampling design, stratified multistage cluster sampling design, Skorohod metric, sup norm metric.

[11] CONTI, P. L. and MARELLA, D. (2015). Inference for quantiles of a finite population: Asymptotic versus resampling results. *Scand. J. Stat.* **42** 545–561. MR3345121 https://doi.org/10.1111/sjos.12122

[12] DEL MAR RUEDA, M., ARCOS, A. and MARTÍNEZ, M. D. (2003). Difference estimators of quantiles in finite populations. *TEST* **12** 481–496. MR2044321 https://doi.org/10.1007/BF02595726

[13] DEY, A. and CHAUDHURI, P. (2024). On estimators of the mean of infinite dimensional data in finite populations. *Bernoulli* **30** 797–824. MR4665598 https://doi.org/10.3150/23-bej1617

[14] DEY, A. and CHAUDHURI, P. (2024). A comparison of estimators of mean and its functions in finite populations. *Statist. Sinica*. To appear. Available at http://www3.stat.sinica.edu.tw/preprint/SS-2022-0181_Preprint.pdf.

[15] DEY, A. and CHAUDHURI, P. (2024). Supplement to "A comparison of estimators of mean and its functions in finite populations." *Statist. Sinica*. To appear. Available at http://www3.stat.sinica.edu.tw/preprint/supp/2022-0181_supp.pdf.

[16] DEY, A. and CHAUDHURI, P. (2024). Supplement to "Quantile processes and their applications in finite populations." https://doi.org/10.1214/24-AOS2432SUPP

[17] FRANCISCO, C. A. and FULLER, W. A. (1991). Quantile estimation with a complex survey design. *Ann. Statist.* **19** 454–469. MR1091862 https://doi.org/10.1214/aos/1176347993

[18] HÁJEK, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Ann. Math. Stat.* **35** 1491–1523. MR0178555 https://doi.org/10.1214/aoms/1177700375

[19] HAN, Q. and WELLNER, J. A. (2021). Complex sampling designs: Uniform limit theorems and applications. *Ann. Statist.* **49** 459–485. MR4206686 https://doi.org/10.1214/20-AOS1964

[20] HORVITZ, D. G. and THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47** 663–685. MR0053460

[21] KUK, A. Y. C. (1988). Estimation of distribution functions and medians under sampling with unequal probabilities. *Biometrika* **75** 97–103. MR0932822 https://doi.org/10.1093/biomet/75.1.97

[22] KUK, A. Y. C. and MAK, T. K. (1989). Median estimation in the presence of auxiliary information. *J. Roy. Statist. Soc. Ser. B* **51** 261–269. MR1007458

[23] LAHIRI, D. B. (1951). A method of sample selection providing unbiased ratio estimates. *Bull. Int. Stat. Inst.* **33** 133–140.

[24] MIDZUNO, H. (1952). On the sampling system with probability proportionate to sum of sizes. *Ann. Inst. Statist. Math., Tokyo* **3** 99–107. MR0050840 https://doi.org/10.1007/bf02949779

[25] OHLSSON, E. (1989). Asymptotic normality for two-stage sampling from a finite population. *Probab. Theory Related Fields* **81** 341–352. MR0983089 https://doi.org/10.1007/BF00340058

[26] RAO, J. N. K., HARTLEY, H. O. and COCHRAN, W. G. (1962). On a simple procedure of unequal probability sampling without replacement. *J. Roy. Statist. Soc. Ser. B* **24** 482–491. MR0148196

[27] RAO, J. N. K., KOVAR, J. G. and MANTEL, H. J. (1990). On estimating distribution functions and quantiles from survey data using auxiliary information. *Biometrika* **77** 365–375. MR1064809 https://doi.org/10.1093/biomet/77.2.365

[28] RUEDA, M. M., ARCOS, A., MARTÍNEZ-MIRANDA, M. D. and ROMÁN, Y. (2004). Some improved estimators of finite population quantile using auxiliary information in sample surveys. *Comput. Statist. Data Anal.* **45** 825–848. MR2054889 https://doi.org/10.1016/S0167-9473(03)00097-5

[29] SEN, A. R. (1953). On the estimate of the variance in sampling with varying probabilities. *J. Indian Soc. Agricultural Statist.* **5** 119–127. MR0068179

[30] SERFLING, R. J. (2009). *Approximation Theorems of Mathematical Statistics. Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0595165

[31] SHAO, J. (1994). L-statistics in complex survey problems. *Ann. Statist.* **22** 946–967. MR1292550 https://doi.org/10.1214/aos/1176325505

[32] SHI, X. Q., WU, C.-F. J. and CHEN, J. H. (1990). Weak and strong representations for quantile processes from finite populations with application to simulation size in resampling inference. *Canad. J. Statist.* **18** 141–148. MR1067165 https://doi.org/10.2307/3315562

[33] SHORACK, G. R. and WELLNER, J. A. (2009). *Empirical Processes with Applications to Statistics. Classics in Applied Mathematics* **59**. SIAM, Philadelphia, PA. MR3396731 https://doi.org/10.1137/1.9780898719017.ch1

[34] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics. Springer Series in Statistics*. Springer, New York. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2

[35] WANG, J. C. and OPSOMER, J. D. (2011). On asymptotic normality and variance estimation for nondifferentiable survey estimators. *Biometrika* **98** 91–106. MR2804212 https://doi.org/10.1093/biomet/asq077

The Annals of Statistics
2024, Vol. 52, No. 5, 2217–2240
https://doi.org/10.1214/24-AOS2433

A NEW TEST FOR HIGH-DIMENSIONAL TWO-SAMPLE MEAN PROBLEMS WITH CONSIDERATION OF CORRELATION STRUCTURE

BY SONGSHAN YANG[1,a], SHURONG ZHENG[2,b] AND RUNZE LI[3,c]

[1]*The Center for Applied Statistics and Institute of Statistics and Big Data, Renmin University of China,* [a]*yangss@ruc.edu.cn*
[2]*School of Mathematics and Statistics & KLAS, Northeast Normal University,* [b]*zhengsr@nenu.edu.cn*
[3]*Department of Statistics, The Pennsylvania State University,* [c]*rzli@psu.edu*

This paper is concerned with high-dimensional two-sample mean problems, which receive considerable attention in recent literature. To utilize the correlation information among variables for enhancing the power of two-sample mean tests, we consider the setting in which the precision matrix of high-dimensional data possesses a linear structure. Thus, we first propose a new precision matrix estimation procedure with considering its linear structure, and further develop regularization methods to select the true basis matrices and remove irrelevant basis matrices. With the aid of an estimated precision matrix, we propose a new test statistic for the two-sample mean problems by replacing the inverse of sample covariance matrix in the Hotelling test by the estimated precision matrix. The proposed test is applicable for both the low-dimensional setting and high-dimensional setting even if the dimension of the data exceeds the sample size. The limiting null distributions of the proposed test statistic under both null and alternative hypotheses are derived. We further derive the asymptotical power function of the proposed test and compare its asymptotic power with some existing tests. We find the estimation error of the precision matrix does not have impact on the asymptotical power function. Moreover, asymptotic relative efficiency of the proposed test to the classical Hotelling test tends to infinity when the ratio of the dimension of data to the sample size tends to 1. We conduct a Monte Carlo simulation study to assess the finite sample performance of the proposed precision matrix estimation procedure and the proposed high-dimensional two-sample mean test. Our numerical results imply that the proposed regularization method is able to effectively remove irrelevant basis matrices. The proposed test performs well compared with the existing methods especially when the elements of the vector have unequal variances. We also illustrate the proposed methodology by an empirical analysis of a real-world data set.

REFERENCES

ANDERSON, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Statist.* **1** 135–141. MR0331612

BAI, Z. and SARANADASA, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6** 311–329. MR1399305

CAI, T. T., LIU, W. and XIA, Y. (2014). Two-sample test of high dimensional means under dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 349–372. MR3164870 https://doi.org/10.1111/rssb.12034

CHANG, J., ZHENG, C., ZHOU, W.-X. and ZHOU, W. (2017). Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity. *Biometrics* **73** 1300–1310. MR3744543 https://doi.org/10.1111/biom.12695

CHEN, L. S., PAUL, D., PRENTICE, R. L. and WANG, P. (2011). A regularized Hotelling's T^2 test for pathway analysis in proteomic studies. *J. Amer. Statist. Assoc.* **106** 1345–1360. MR2896840 https://doi.org/10.1198/jasa.2011.ap10599

CHEN, S. X., LI, J. and ZHONG, P.-S. (2019). Two-sample and ANOVA tests for high dimensional means. *Ann. Statist.* **47** 1443–1474. MR3911118 https://doi.org/10.1214/18-AOS1720

MSC2020 subject classifications. Primary 62H15; secondary 62J07.
Key words and phrases. Power enhancement, precision matrix estimation, random matrix theory.

CHEN, S. X. and QIN, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. MR2604697 https://doi.org/10.1214/09-AOS716

DEMPSTER, A. P. (1958). A high dimensional two sample significance test. *Ann. Math. Stat.* **29** 995–1010. MR0112207 https://doi.org/10.1214/aoms/1177706437

FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581 https://doi.org/10.1198/016214501753382273

FAN, J., LI, R., ZHANG, C.-H. and ZOU, H. (2020). *Statistical Foundations of Data Science*. CRC Press/CRC, Boca Raton.

FENG, L., JIANG, T., LI, X. and LIU, B. (2024). Asymptotic independence of the sum and maximum of dependent random variables with applications to high-dimensional tests. *Statist. Sinica* **34** 1745–1763. MR4764695

FENG, L., ZOU, C., WANG, Z. and ZHU, L. (2015). Two-sample Behrens-Fisher problem for high-dimensional data. *Statist. Sinica* **25** 1297–1312. MR3409068

HUANG, Y., LI, C., LI, R. and YANG, S. (2022). An overview of tests on high-dimensional means. *J. Multivariate Anal.* **188** Paper No. 104813, 14. MR4353841 https://doi.org/10.1016/j.jmva.2021.104813

LI, C. and LI, R. (2022). Linear hypothesis testing in linear models with high-dimensional responses. *J. Amer. Statist. Assoc.* **117** 1738–1750. MR4528467 https://doi.org/10.1080/01621459.2021.1884561

LI, H., AUE, A., PAUL, D., PENG, J. and WANG, P. (2020). An adaptable generalization of Hotelling's T^2 test in high dimension. *Ann. Statist.* **48** 1815–1847. MR4124345 https://doi.org/10.1214/19-AOS1869

LIU, W., YU, X. and LI, R. (2022). Multiple-splitting projection test for high-dimensional mean vectors. *J. Mach. Learn. Res.* **23** Paper No. [71], 27. MR4576656

LOPES, M. E., JACOB, L. J. and WAINWRIGHT, M. J. (2011). A more powerful two-sample test in high dimensions using random projection. ArXiv e-prints.

MUIRHEAD, R. J. (2009). *Aspects of Multivariate Statistical Theory. Wiley Series in Probability and Mathematical Statistics*. Wiley, New York.

PAN, G. M. and ZHOU, W. (2011). Central limit theorem for Hotelling's T^2 statistic under large dimension. *Ann. Appl. Probab.* **21** 1860–1910. MR2884053 https://doi.org/10.1214/10-AAP742

QU, A., LINDSAY, B. G. and LI, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika* **87** 823–836. MR1813977 https://doi.org/10.1093/biomet/87.4.823

SRIVASTAVA, M. S. (2009). A test for the mean vector with fewer observations than the dimension under non-normality. *J. Multivariate Anal.* **100** 518–532. MR2483435 https://doi.org/10.1016/j.jmva.2008.06.006

SRIVASTAVA, M. S. and DU, M. (2008). A test for the mean vector with fewer observations than the dimension. *J. Multivariate Anal.* **99** 386–402. MR2396970 https://doi.org/10.1016/j.jmva.2006.11.002

SRIVASTAVA, R., LI, P. and RUPPERT, D. (2016). RAPTT: An exact two-sample test in high dimensions using random projections. *J. Comput. Graph. Statist.* **25** 954–970. MR3533647 https://doi.org/10.1080/10618600.2015.1062771

WANG, L., KIM, Y. and LI, R. (2013). Calibrating nonconvex penalized regression in ultra-high dimension. *Ann. Statist.* **41** 2505–2536. MR3127873 https://doi.org/10.1214/13-AOS1159

WANG, L., PENG, B. and LI, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *J. Amer. Statist. Assoc.* **110** 1658–1669. MR3449062 https://doi.org/10.1080/01621459.2014.988215

WANG, R. and XU, W. (2022). An approximate randomization test for the high-dimensional two-sample Behrens-Fisher problem under arbitrary covariances. *Biometrika* **109** 1117–1132. MR4519119 https://doi.org/10.1093/biomet/asac014

XU, G., LIN, L., WEI, P. and PAN, W. (2016). An adaptive two-sample test for high-dimensional means. *Biometrika* **103** 609–624. MR3551787 https://doi.org/10.1093/biomet/asw029

YANG, S., ZHENG, S. and LI, R. (2024). Supplement to "A new test for high-dimensional two-sample mean problems with consideration of correlation structure." https://doi.org/10.1214/24-AOS2433SUPP

ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. MR2604701 https://doi.org/10.1214/09-AOS729

ZHENG, S., CHEN, Z., CUI, H. and LI, R. (2019). Hypothesis testing on linear structures of high-dimensional covariance matrix. *Ann. Statist.* **47** 3300–3334. MR4025743 https://doi.org/10.1214/18-AOS1779

The Annals of Statistics
2024, Vol. 52, No. 5, 2241–2267
https://doi.org/10.1214/24-AOS2434

ON THE EXISTENCE OF POWERFUL P-VALUES AND E-VALUES FOR COMPOSITE HYPOTHESES

BY ZHENYUAN ZHANG[1,a], AADITYA RAMDAS[2,b] AND RUODU WANG[3,c]

[1]*Department of Mathematics, Stanford University,* [a]*zzy@stanford.edu*
[2]*Department of Statistics & Data Science, Carnegie Mellon University,* [b]*aramdas@cmu.edu*
[3]*Department of Statistics and Actuarial Science, University of Waterloo,* [c]*wang@uwaterloo.ca*

Given a composite null $\mathcal{P}$ and composite alternative $\mathcal{Q}$, when and how can we construct a p-value whose distribution is exactly uniform under the null, and stochastically smaller than uniform under the alternative? Similarly, when and how can we construct an e-value whose expectation exactly equals one under the null, but its expected logarithm under the alternative is positive? We answer these basic questions, and other related ones, when $\mathcal{P}$ and $\mathcal{Q}$ are convex polytopes (in the space of probability measures). We prove that such constructions are possible if and only if $\mathcal{Q}$ does not intersect the span of $\mathcal{P}$. If the p-value is allowed to be stochastically larger than uniform under $P \in \mathcal{P}$, and the e-value can have expectation at most one under $P \in \mathcal{P}$, then it is achievable whenever $\mathcal{P}$ and $\mathcal{Q}$ are disjoint. More generally, even when $\mathcal{P}$ and $\mathcal{Q}$ are not polytopes, we characterize the existence of a bounded nontrivial e-variable whose expectation exactly equals one under any $P \in \mathcal{P}$. The proofs utilize recently developed techniques in simultaneous optimal transport. A key role is played by coarsening the filtration: sometimes, no such p-value or e-value exists in the richest data filtration, but it does exist in some reduced filtration, and our work provides the first general characterization of this phenomenon. We also provide an iterative construction that explicitly constructs such processes, and under certain conditions it finds the one that grows fastest under a specific alternative Q. We discuss implications for the construction of composite nonnegative (super)martingales, and end with some conjectures and open problems.

REFERENCES

BELL, R. and COVER, T. M. (1988). Game-theoretic optimal portfolios. *Manage. Sci.* **34** 724–733. MR0943277 https://doi.org/10.1287/mnsc.34.6.724

BERGER, A. (1951). On uniformly consistent tests. *Ann. Math. Stat.* **22** 289–293. MR0042653 https://doi.org/10.1214/aoms/1177729649

BERTANHA, M. and MOREIRA, M. J. (2020). Impossible inference in econometrics: Theory and applications. *J. Econometrics* **218** 247–270. MR4149226 https://doi.org/10.1016/j.jeconom.2020.04.016

BREIMAN, L. (1960). Optimal gambling systems for favorable games. In *Proc. 4th Berkeley Sympos. Math. Statist. And Prob., Vol. I* 65–78. Univ. California Press, Berkeley. MR0135630

DARLING, D. A. and ROBBINS, H. (1967). Confidence sequences for mean, variance, and median. *Proc. Natl. Acad. Sci. USA* **58** 66–68. MR0215406 https://doi.org/10.1073/pnas.58.1.66

GANGRADE, A., RINALDO, A. and RAMDAS, A. (2023). A sequential test for log-concavity. arXiv preprint. Available at arXiv:2301.03542.

GRÜNWALD, P. D., DE HEIDE, R. and KOOLEN, W. (2024). Safe testing. *J. Roy. Statist. Soc. Ser. B.*

HOEFFDING, W. and WOLFOWITZ, J. (1958). Distinguishability of sets of distributions. (The case of independent and identically distributed chance variables). *Ann. Math. Stat.* **29** 700–718. MR0095555 https://doi.org/10.1214/aoms/1177706531

HOWARD, S. R., RAMDAS, A., MCAULIFFE, J. and SEKHON, J. (2021). Time-uniform, nonparametric, nonasymptotic confidence sequences. *Ann. Statist.* **49** 1055–1080. MR4255119 https://doi.org/10.1214/20-aos1991

MSC2020 subject classifications. 62B15, 62G10, 49Q22, 60G42.
Key words and phrases. p-values, e-values, composite hypothesis testing, convex order, simultaneous optimal transport, nonnegative martingale.

KELLY, J. L. JR. (1956). A new interpretation of information rate. *Bell Syst. Tech. J.* **35** 917–926. MR0090494 https://doi.org/10.1002/j.1538-7305.1956.tb03809.x

KRAFT, C. (1955). Some conditions for consistency and uniform consistency of statistical procedures. *Univ. California Publ. Statist.* **2** 125–141. MR0073896

KÜHNEL, W. (2015). *Differential Geometry: Curves—Surfaces—Manifolds*, 3rd ed. *Student Mathematical Library* **77**. Am. Math. Soc., Providence, RI. Translated from the 2013 German edition by Bruce Hunt, with corrections and additions by the author. MR3443721 https://doi.org/10.1090/stml/077

LARDY, T., GRÜNWALD, P. and HARREMOËS, P., (2023). Universal reverse information projections and optimal e-statistics. To appear in *IEEE Transactions on Information Theory*. Available at arXiv:2306.16646.

LARSSON, M., RAMDAS, A. and RUF, J. (2024). The numeraire e-variable and reverse information projection. arXiv preprint. Available at arXiv:2402.18810.

RAMDAS, A., GRÜNWALD, P., VOVK, V. and SHAFER, G. (2023). Game-theoretic statistics and safe anytime-valid inference. *Statist. Sci.* **38** 576–601. MR4665027 https://doi.org/10.1214/23-sts894

RAMDAS, A., RUF, J., LARSSON, M. and KOOLEN, W. (2020). Admissible anytime-valid sequential inference must rely on nonnegative martingales. Available at arXiv:2009.03167.

RAMDAS, A., RUF, J., LARSSON, M. and KOOLEN, W. M. (2022). Testing exchangeability: Fork-convexity, supermartingales and e-processes. *Internat. J. Approx. Reason.* **141** 83–109. MR4364897 https://doi.org/10.1016/j.ijar.2021.06.017

ROBBINS, H. and SIEGMUND, D. (1974). The expected sample size of some tests of power one. *Ann. Statist.* **2** 415–436. MR0448750

SANTAMBROGIO, F. (2015). *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling. Progress in Nonlinear Differential Equations and Their Applications* **87**. Birkhäuser, Cham. MR3409718 https://doi.org/10.1007/978-3-319-20828-2

SHAFER, G. (2021). Testing by betting: A strategy for statistical and scientific communication. *J. Roy. Statist. Soc. Ser. A* **184** 407–478. MR4255905 https://doi.org/10.1111/rssa.12647

SHAFER, G., SHEN, A., VERESHCHAGIN, N. and VOVK, V. (2011). Test martingales, Bayes factors and p-values. *Statist. Sci.* **26** 84–101. MR2849911 https://doi.org/10.1214/10-STS347

SHAFER, G. and VOVK, V. (2001). *Probability and Finance: It's Only a Game!. Wiley Series in Probability and Statistics. Financial Engineering Section*. Wiley-Interscience, New York. MR1852450 https://doi.org/10.1002/0471249696

SHAFER, G. and VOVK, V. (2019). *Game-Theoretic Foundations for Probability and Finance*. Wiley, Hoboken, NJ.

SHAKED, M. and SHANTHIKUMAR, J. G. (2007). *Stochastic Orders. Springer Series in Statistics*. Springer, New York. MR2265633 https://doi.org/10.1007/978-0-387-34675-5

SHEN, J., SHEN, Y., WANG, B. and WANG, R. (2019). Distributional compatibility for change of measures. *Finance Stoch.* **23** 761–794. MR3968284 https://doi.org/10.1007/s00780-019-00393-4

SIMON, B. (2011). *Convexity: An Analytic Viewpoint. Cambridge Tracts in Mathematics* **187**. Cambridge Univ. Press, Cambridge. MR2814377 https://doi.org/10.1017/CBO9780511910135

SIMONS, G. (1970). A martingale decomposition theorem. *Ann. Math. Stat.* **41** 1102–1104. MR0261678 https://doi.org/10.1214/aoms/1177696991

STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Stat.* **36** 423–439. MR0177430 https://doi.org/10.1214/aoms/1177700153

TESCHL, G. (2012). *Ordinary Differential Equations and Dynamical Systems. Graduate Studies in Mathematics* **140**. Am. Math. Soc., Providence, RI. MR2961944 https://doi.org/10.1090/gsm/140

VILLANI, C. (2009). *Optimal Transport: Old and New. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Springer, Berlin. MR2459454 https://doi.org/10.1007/978-3-540-71050-9

VOVK, V. (2021). Testing randomness online. *Statist. Sci.* **36** 595–611. MR4323055 https://doi.org/10.1214/20-sts817

VOVK, V. and WANG, R. (2021). E-values: Calibration, combination and applications. *Ann. Statist.* **49** 1736–1754. MR4298879 https://doi.org/10.1214/20-aos2020

VOVK, V. and WANG, R. (2024). Nonparametric e-tests of symmetry. *N. Engl. J. Statist. Data Sci.* 1–10.

WANG, R. and RAMDAS, A. (2022). False discovery rate control with e-values. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 822–852. MR4460577

WANG, R. and ZHANG, Z. (2023). Simultaneous optimal transport. arXiv preprint. Available at arXiv:2201.03483v2.

WASSERMAN, L., RAMDAS, A. and BALAKRISHNAN, S. (2020). Universal inference. *Proc. Natl. Acad. Sci. USA* **117** 16880–16890. MR4242731 https://doi.org/10.1073/pnas.1922664117

WAUDBY-SMITH, I. and RAMDAS, A. (2024). Estimating means of bounded random variables by betting. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **86** 1–27. MR4716192 https://doi.org/10.1093/jrsssb/qkad009

WILKS, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Stat.* **9** 60–62.

ZHANG, Z., RAMDAS, A. and WANG, R. (2024). Supplement to "On the existence of powerful p-values and e-values for composite hypotheses." https://doi.org/10.1214/24-AOS2434SUPP

The Annals of Statistics
2024, Vol. 52, No. 5, 2268–2292
https://doi.org/10.1214/24-AOS2435

ENVIRONMENT INVARIANT LINEAR LEAST SQUARES

BY JIANQING FAN[1,a], CONG FANG[2,c], YIHONG GU[1,b] AND TONG ZHANG[3,d]

[1]*Department of Operations Research and Financial Engineering, Princeton University,* [a]*jqfan@princeton.edu,* [b]*yihongg@princeton.edu*

[2]*National Key Lab of General Artificial Intelligence, School of Intelligence Science and Technology, Peking University,* [c]*congfang@pku.edu.cn*

[3]*Siebel School of Computing and Data Science, University of Illinois Urbana-Champaign,* [d]*tongzhang@tongzhang-ml.org*

This paper considers a multi-environment linear regression model in which data from multiple experimental settings are collected. The joint distribution of the response variable and covariates may vary across different environments, yet the conditional expectations of the response variable, given the unknown set of important variables, are invariant. Such a statistical model is related to the problem of endogeneity, causal inference, and transfer learning. The motivation behind it is illustrated by how the goals of prediction and attribution are inherent in estimating the true parameter and the important variable set. We construct a novel *environment invariant linear least squares (EILLS)* objective function, a multi-environment version of linear least squares regression that leverages the above conditional expectation invariance structure and heterogeneity among different environments to determine the true parameter. Our proposed method is applicable without any additional structural knowledge and can identify the true parameter under a near-minimal identification condition related to the heterogeneity of the environments. We establish nonasymptotic ℓ_2 error bounds on the estimation error for the EILLS estimator in the presence of spurious variables. Moreover, we further show that the ℓ_0 penalized EILLS estimator can achieve variable selection consistency in high-dimensional regimes. These nonasymptotic results demonstrate the sample efficiency of the EILLS estimator and its capability to circumvent the curse of endogeneity in an algorithmic manner without any additional prior structural knowledge. To the best of our knowledge, this paper is the first to realize statistically efficient invariance learning in the general linear model.

REFERENCES

ALDRICH, J. (1989). Autonomy *Oxf. Econ. Pap.* **41** 15–34.

ARJOVSKY, M., BOTTOU, L., GULRAJANI, I. and LOPEZ-PAZ, D. (2019). Invariant risk minimization. arXiv preprint. Available at arXiv:1907.02893.

BERTSIMAS, D., KING, A. and MAZUMDER, R. (2016). Best subset selection via a modern optimization lens. *Ann. Statist.* **44** 813–852. MR3476618 https://doi.org/10.1214/15-AOS1388

BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. MR2533469 https://doi.org/10.1214/08-AOS620

BOLLEN, K. A. (1989). *Structural Equations with Latent Variables. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Wiley, New York. A Wiley-Interscience Publication. MR0996025 https://doi.org/10.1002/9781118619179

BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer Series in Statistics*. Springer, Heidelberg. MR2807761 https://doi.org/10.1007/978-3-642-20192-9

CANDES, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *Ann. Statist.* **35** 2313–2351. MR2382644 https://doi.org/10.1214/009053606000001523

CHEN, Y. and BÜHLMANN, P. (2021). Domain adaptation under structural causal models. *J. Mach. Learn. Res.* **22** 11856–11935. MR4353040 https://doi.org/10.1007/s11081-020-09512-z

MSC2020 subject classifications. Primary 62J05; secondary 62D20.

Key words and phrases. Least squares, endogeneity, multiple environments, invariance, heterogeneity, structural causal model, invariant risk minimization.

ČUKLINA, J., LEE, C. H., WILLIAMS, E. G., SAJIC, T., COLLINS, B. C., RODRÍGUEZ MARTÍNEZ, M., SHARMA, V. S., WENDT, F., GOETZE, S. et al. (2021). Diagnostics and correction of batch effects in large-scale proteomic studies: A tutorial. *Mol. Syst. Biol.* **17** e10240.

DAWID, A. P. and DIDELEZ, V. (2010). Identifying the consequences of dynamic treatment strategies: A decision-theoretic overview. *Stat. Surv.* **4** 184–231. MR2740837 https://doi.org/10.1214/10-SS081

DIDELEZ, V., DAWID, P. and GENELETTI, S. (2012). Direct and indirect effects of sequential treatments. arXiv preprint. Available at arXiv:1206.6840.

EFRON, B. (2020). Prediction, estimation, and attribution. *Int. Stat. Rev.* **88** S28–S59.

ENGLE, R. F., HENDRY, D. F. and RICHARD, J.-F. (1983). Exogeneity. *Econometrica* **51** 277–304. MR0688727 https://doi.org/10.2307/1911990

FAN, J., FANG, C., GU, Y. and ZHANG, T. (2024). Supplement to "Environment Invariant Linear Least Squares." https://doi.org/10.1214/24-AOS2435SUPPA, https://doi.org/10.1214/24-AOS2435SUPPB

FAN, J., HAN, F. and LIU, H. (2014). Challenges of big data analysis. *Nat. Sci. Rev.* **1** 293–314.

FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581 https://doi.org/10.1198/016214501753382273

FAN, J., LI, R., ZHANG, C.-H. and ZOU, H. (2020). *Statistical Foundations of Data Science*. CRC Press, Boca Raton.

FAN, J. and LIAO, Y. (2014). Endogeneity in high dimensions. *Ann. Statist.* **42** 872–917. MR3210990 https://doi.org/10.1214/13-AOS1202

FAN, J. and LV, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 849–911. MR2530322 https://doi.org/10.1111/j.1467-9868.2008.00674.x

GAUSS, C. F. (2011). *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium. Cambridge Library Collection*. Cambridge Univ. Press, Cambridge. Reprint of the 1809 original. MR2858122 https://doi.org/10.1017/CBO9780511841705.010

GEIRHOS, R., JACOBSEN, J.-H., MICHAELIS, C., ZEMEL, R., BRENDEL, W., BETHGE, M. and WICHMANN, F. A. (2020). Shortcut learning in deep neural networks. *Nat. Mach. Intell.* **2** 665–673.

GHASSAMI, A., SALEHKALEYBAR, S., KIYAVASH, N. and ZHANG, K. (2017). Learning causal structures using regression invariance. *Adv. Neural Inf. Process. Syst.* **30**.

GU, Y., FANG, C., BÜHLMANN, P. and FAN, J. (2024). Causality pursuit from heterogeneous environments via neural adversarial invariance learning. arXiv preprint. Available at arXiv:2405.04715.

HAAVELMO, T. (1944). The probability approach in econometrics. *Econometrica* **12** **S** iii–115. MR0010953 https://doi.org/10.2307/1906935

HE, Y.-B. and GENG, Z. (2008). Active learning of causal networks with intervention experiments and optimal designs. *J. Mach. Learn. Res.* **9** 2523–2547. MR2460892

HEINZE-DEML, C. and MEINSHAUSEN, N. (2021). Conditional variance penalties and domain shift robustness. *Mach. Learn.* **110** 303–348. MR4207502 https://doi.org/10.1007/s10994-020-05924-1

HEINZE-DEML, C., PETERS, J. and MEINSHAUSEN, N. (2018). Invariant causal prediction for nonlinear models. *J. Causal Inference* **6** Art. No. 20170016, 35. MR4335430 https://doi.org/10.1515/jci-2017-0016

JANG, E., GU, S. and POOLE, B. (2016). Categorical reparameterization with Gumbel-softmax. ArXiv preprint. Available at arXiv:1611.01144.

KAMATH, P., TANGELLA, A., SUTHERLAND, D. and SREBRO, N. (2021). Does invariant risk minimization capture invariance? In *International Conference on Artificial Intelligence and Statistics* 4069–4077.

KRUEGER, D., CABALLERO, E., JACOBSEN, J.-H., ZHANG, A., BINAS, J., ZHANG, D., LE PRIOL, R. and COURVILLE, A. (2021). Out-of-distribution generalization via risk extrapolation (rex). In *International Conference on Machine Learning* 5815–5826.

LEGENDRE, A.-M. (1805). *Nouvelles méthodes pour la détermination des orbites des comètes* [*New Methods for the Determination of the Orbits of Comets*]. F. Didot, Paris. (in French).

LU, C., WU, Y., HERNÁNDEZ-LOBATO, J. M. and SCHÖLKOPF, B. (2021). Nonlinear invariant risk minimization: a causal approach. ArXiv preprint. Available at arXiv:2102.12353.

MEINSHAUSEN, N., HAUSER, A., MOOIJ, J. M., PETERS, J., VERSTEEG, P. and BÜHLMANN, P. (2016). Methods for causal inference from gene perturbation experiments and validation. *Proc. Natl. Acad. Sci. USA* **113** 7361–7368.

GONG, M., ZHANG, K., LIU, T., TAO, D., GLYMOUR, C. and SCHÖLKOPF, B. (2016). Domain adaptation with conditional transferable components. In *International Conference on Machine Learning* 2839–2848.

MUANDET, K., BALDUZZI, D. and SCHÖLKOPF, B. (2013). Domain generalization via invariant feature representation. In *International Conference on Machine Learning* 10–18.

PEARL, J., GLYMOUR, M. and JEWELL, N. P. (2016). *Causal Inference in Statistics: A Primer*. Wiley, Chichester. MR3497861

PETERS, J., BÜHLMANN, P. and MEINSHAUSEN, N. (2016). Causal inference by using invariant prediction: Identification and confidence intervals. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 947–1012. With comments and a rejoinder. MR3557186 https://doi.org/10.1111/rssb.12167

PFISTER, N., BÜHLMANN, P. and PETERS, J. (2019). Invariant causal prediction for sequential data. *J. Amer. Statist. Assoc.* **114** 1264–1276. MR4011778 https://doi.org/10.1080/01621459.2018.1491403

PFISTER, N., WILLIAMS, E. G., PETERS, J., AEBERSOLD, R. and BÜHLMANN, P. (2021). Stabilizing variable selection and regression. *Ann. Appl. Stat.* **15** 1220–1246. MR4317406 https://doi.org/10.1214/21-aoas1487

ROJAS-CARULLA, M., SCHÖLKOPF, B., TURNER, R. and PETERS, J. (2018). Invariant models for causal transfer learning. *J. Mach. Learn. Res.* **19** 1309–1342. MR3862443

ROSENFELD, E., RAVIKUMAR, P. and RISTESKI, A. (2021). The risks of invariant risk minimization. In *International Conference on Learning Representations*.

ROTHENHÄUSLER, D., BÜHLMANN, P. and MEINSHAUSEN, N. (2019). Causal Dantzig: Fast inference in linear structural equation models with hidden variables under additive interventions. *Ann. Statist.* **47** 1688–1722. MR3911127 https://doi.org/10.1214/18-AOS1732

ROTHENHÄUSLER, D., MEINSHAUSEN, N., BÜHLMANN, P. and PETERS, J. (2021). Anchor regression: Heterogeneous data meet causality. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 215–246. MR4250274 https://doi.org/10.1111/rssb.12398

SAGAWA, S., KOH, P. W., HASHIMOTO, T. B. and LIANG, P. (2020). Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization. In *International Conference on Learning Representations*.

SCHÖLKOPF, B., JANZING, D., PETERS, J., SGOURITSA, E., ZHANG, K. and MOOIJ, J. (2012). On causal and anticausal learning. arXiv preprint. Available at arXiv:1206.6471.

STIGLER, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty Before* 1900. The Belknap Press of Harvard Univ. Press, Cambridge, MA. MR0852410

TIBSHIRANI, R. (1997). The lasso method for variable selection in the Cox model. *Stat. Med.* **16** 385–395. https://doi.org/10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3

TORRALBA, A. and EFROS, A. A. (2011). Unbiased look at dataset bias. In *CVPR* 2011 1521–1528. IEEE Press, New York.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics. Springer Series in Statistics*. Springer, New York. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2

VLADIMIROVA, M., GIRARD, S., NGUYEN, H. and ARBEL, J. (2020). Sub-Weibull distributions: Generalizing sub-Gaussian and sub-exponential properties to heavier tailed distributions. *Stat* **9** e318, 8. MR4193421 https://doi.org/10.1007/s40065-018-0218-4

WAINWRIGHT, M. J. (2019). *High-Dimensional Statistics: A Non-asymptotic Viewpoint. Cambridge Series in Statistical and Probabilistic Mathematics* **48**. Cambridge Univ. Press, Cambridge. MR3967104 https://doi.org/10.1017/9781108627771

WANG, Z. and VEITCH, V. (2023). The Causal Structure of Domain Invariant Supervised Representation Learning. Stat 1050 7.

YIN, M., WANG, Y. and BLEI, D. M. (2024). Optimization-based causal estimation from heterogeneous environments. *J. Mach. Learn. Res.* **25** Paper No. [168], 44. MR4777410

ZHANG, A., LYLE, C., SODHANI, S., FILOS, A., KWIATKOWSKA, M., PINEAU, J., GAL, Y. and PRECUP, D. (2020). Invariant causal prediction for block mdps. In *International Conference on Machine Learning* 11214–11224.

ZHANG, C.-H. and ZHANG, T. (2012). A general theory of concave regularization for high-dimensional sparse estimation problems. *Statist. Sci.* **27** 576–593. MR3025135 https://doi.org/10.1214/12-STS399

ZHANG, T. (2011). Adaptive forward-backward greedy algorithm for learning sparse representations. *IEEE Trans. Inf. Theory* **57** 4689–4708. MR2840485 https://doi.org/10.1109/TIT.2011.2146690

ZHAO, P. and YU, B. (2006). On model selection consistency of Lasso. *J. Mach. Learn. Res.* **7** 2541–2563. MR2274449

The Annals of Statistics
2024, Vol. 52, No. 5, 2293–2317
https://doi.org/10.1214/24-AOS2436

GAUSSIAN APPROXIMATION FOR NONSTATIONARY TIME SERIES WITH OPTIMAL RATE AND EXPLICIT CONSTRUCTION

BY SOHAM BONNERJEE[1,a], SAYAR KARMAKAR[2,c] AND WEI BIAO WU[1,b]

[1]*Department of Statistics, University of Chicago,* [a]*sohambonnerjee@uchicago.edu,* [b]*wbwu@galton.uchicago.edu*
[2]*Department of Statistics, University of Florida,* [c]*sayarkarmakar@ufl.edu*

Statistical inference for time series such as curve estimation for time-varying models or testing for existence of a change point have garnered significant attention. However, these works are generally restricted to the assumption of independence and/or stationarity at its best. The main obstacle is that the existing Gaussian approximation results for nonstationary processes only provide an existential proof, and thus they are difficult to apply. In this paper, we provide two clear paths to construct such a Gaussian approximation for nonstationary series. While the first one is theoretically more natural, the second one is practically implementable. Our Gaussian approximation results are applicable for a very large class of nonstationary time series, obtain optimal rates and yet have good applicability. Building on such approximations, we also show theoretical results for change-point detection and simultaneous inference in presence of nonstationary errors. Finally, we substantiate our theoretical results with simulation studies and real data analysis.

REFERENCES

[1] ADAK, S. (1998). Time-dependent spectral analysis of nonstationary time series. *J. Amer. Statist. Assoc.* **93** 1488–1501. MR1666643 https://doi.org/10.2307/2670062

[2] ANDREOU, E. and GHYSELS, E. (2009). Structural breaks in financial time series. *Handbook of Financial Time Series* 839–870.

[3] ANDREWS, A. P., ANDREWS, E. W. and CASTELLANOS, F. R. (2003). The Northern Maya collapse and its aftermath. *Ancient Mesoamerica* **14** 151–156.

[4] ANDREWS, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* **61** 821–856. MR1231678 https://doi.org/10.2307/2951764

[5] AUE, A. and HORVÁTH, L. (2013). Structural breaks in time series. *J. Time Series Anal.* **34** 1–16. MR3008012 https://doi.org/10.1111/j.1467-9892.2012.00819.x

[6] BAI, J. and PERRON, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66** 47–78. MR1616121 https://doi.org/10.2307/2998540

[7] BENNETT, W. R. (1958). Statistics of regenerative digital transmission. *Bell Syst. Tech. J.* **37** 1501–1542. MR0102138 https://doi.org/10.1002/j.1538-7305.1958.tb01560.x

[8] BERCU, B., GAMBOA, F. and ROUAULT, A. (1997). Large deviations for quadratic forms of stationary Gaussian processes. *Stochastic Process. Appl.* **71** 75–90. MR1480640 https://doi.org/10.1016/S0304-4149(97)00071-9

[9] BERKES, I., LIU, W. and WU, W. B. (2014). Komlós–Major–Tusnády approximation under dependence. *Ann. Probab.* **42** 794–817. MR3178474 https://doi.org/10.1214/13-AOP850

[10] BLOOMFIELD, P., HURD, H. L. and LUND, R. B. (1994). Periodic correlation in stratospheric ozone data. *J. Time Series Anal.* **15** 127–150. MR1263886 https://doi.org/10.1111/j.1467-9892.1994.tb00181.x

[11] BOROVKOV, A. (1973). Notes on inequalities for sums of independent variables. *Theory Probab. Appl.* **17** 556.

[12] BROWN, R. L., DURBIN, J. and EVANS, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *J. Roy. Statist. Soc. Ser. B* **37** 149–192. MR0378310

[13] BÜHLMANN, P. (1998). Sieve bootstrap for smoothing in nonstationary time series. *Ann. Statist.* **26** 48–83. MR1611804 https://doi.org/10.1214/aos/1030563978

MSC2020 subject classifications. 62M10, 62M15, 62E17.
Key words and phrases. Gaussian approximation, nonstationary time series, simultaneous confidence band, change-point testing.

[14] CAI, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *J. Econometrics* **136** 163–188. MR2328589 https://doi.org/10.1016/j.jeconom.2005.08.004

[15] CARLETON, C. (2017). Archaeological and Palaeoenvironmental Time-series Analysis. Theses, Department of Archaeology, Simon Fraser Univ.

[16] CARLSTEIN, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist.* **14** 1171–1179. MR0856813 https://doi.org/10.1214/aos/1176350057

[17] CARLSTEIN, E., DO, K.-A., HALL, P., HESTERBERG, T. and KÜNSCH, H. R. (1998). Matched-block bootstrap for dependent data. *Bernoulli* **4** 305–328. MR1653268 https://doi.org/10.2307/3318719

[18] CHOW, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica* **28** 591–605. MR0141193 https://doi.org/10.2307/1910133

[19] CSÖRGŐ, M. and HORVÁTH, L. (1997). *Limit Theorems in Change-Point Analysis*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR2743035

[20] DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37. MR1429916 https://doi.org/10.1214/aos/1034276620

[21] DAHLHAUS, R. (2000). A likelihood approximation for locally stationary processes. *Ann. Statist.* **28** 1762–1794. MR1835040 https://doi.org/10.1214/aos/1015957480

[22] DAUBECHIES, I. (1992). *Ten Lectures on Wavelets*. *CBMS-NSF Regional Conference Series in Applied Mathematics* **61**. SIAM, Philadelphia, PA. MR1162107 https://doi.org/10.1137/1.9781611970104

[23] DAVIS, R. A., LEE, T. C. M. and RODRIGUEZ-YAM, G. A. (2006). Structural break estimation for nonstationary time series models. *J. Amer. Statist. Assoc.* **101** 223–239. MR2268041 https://doi.org/10.1198/016214505000000745

[24] DEMAREST, A. (2004). *Ancient Maya*: *The Rise and Fall of a Rainforest Civilization*. *Case Studies in Early Societies*. Cambridge Univ. Press, Cambridge.

[25] DIAZ, H. and TROUET, V. (2014). Some perspectives on societal impacts of past climatic changes. *History Compass* **12** 160–177.

[26] DONOHO, D. L. (1995). De-noising by soft-thresholding. *IEEE Trans. Inf. Theory* **41** 613–627. MR1331258 https://doi.org/10.1109/18.382009

[27] DOOB, J. L. (1949). Heuristic approach to the Kolmogorov–Smirnov theorems. *Ann. Math. Stat.* **20** 393–403. MR0030732 https://doi.org/10.1214/aoms/1177729991

[28] ELDAN, R., MIKULINCER, D. and ZHAI, A. (2020). The CLT in high dimensions: Quantitative bounds via martingale embedding. *Ann. Probab.* **48** 2494–2524. MR4152649 https://doi.org/10.1214/20-AOP1429

[29] ERDÖS, P. and KAC, M. (1946). On certain limit theorems of the theory of probability. *Bull. Amer. Math. Soc.* **52** 292–302. MR0015705 https://doi.org/10.1090/S0002-9904-1946-08560-2

[30] EUBANK, R. L. and SPECKMAN, P. L. (1993). Confidence bands in nonparametric regression. *J. Amer. Statist. Assoc.* **88** 1287–1301. MR1245362

[31] FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. *Monographs on Statistics and Applied Probability* **66**. CRC Press, London. MR1383587

[32] FAN, J. and ZHANG, W. (1999). Statistical estimation in varying coefficient models. *Ann. Statist.* **27** 1491–1518. MR1742497 https://doi.org/10.1214/aos/1017939139

[33] FAN, J. and ZHANG, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scand. J. Stat.* **27** 715–731. MR1804172 https://doi.org/10.1111/1467-9469.00218

[34] FAUST, B. B. (2001). Maya environmental successes and failures in the Yucatan Peninsula. *Environ. Sci. Policy* **4** 153–169.

[35] FRANKS, L. E. (1969). *Signal Theory*. *Information Theory Series*. Prentice-Hall, New York.

[36] GARDNER, W. A. et al. (1994). *Cyclostationarity in Communications and Signal Processing* **1**. IEEE Press, New York.

[37] GILL, R. B. (2000). *The Great Maya Droughts*: *Water*, *Life*, *and Death*. Univ. New Mexico Press.

[38] GILL, R. B., MAYEWSKI, P. A., NYBERG, J., HAUG, G. H. and PETERSON, L. C. (2007). Drought and the Maya Collapse. *Ancient Mesoamerica* **18** 283–302.

[39] GOLDEN, C. W. and BORGSTEDE, G. (2004). *Continuities and Changes in Maya Archaeology*: *Perspectives at the Millennium*. Routledge, London.

[40] GUNN, J. D., MATHENY, R. T. and FOLAN, W. J. (2002). Climate-change studies in the Maya area: A diachronic analysis. *Ancient Mesoamerica* **13** 79–84.

[41] HAAR, A. (1910). Zur Theorie der orthogonalen Funktionensysteme. *Math. Ann.* **69** 331–371. MR1511592 https://doi.org/10.1007/BF01456326

[42] HALL, P. (1985). Resampling a coverage pattern. *Stochastic Process. Appl.* **20** 231–246. MR0808159 https://doi.org/10.1016/0304-4149(85)90212-1

[43] HANSON, D. L. and WRIGHT, F. T. (1971). A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Stat.* **42** 1079–1083. MR0279864 https://doi.org/10.1214/aoms/1177693335

[44] HÄRDLE, W. (1986). A note on jackknifing kernel regression function estimators. *IEEE Trans. Inf. Theory* **32** 298–300. MR0838421 https://doi.org/10.1109/TIT.1986.1057142

[45] HODELL, D. A., BRENNER, M. and CURTIS, J. H. (2005). Terminal classic drought in the northern Maya lowlands inferred from multiple sediment cores in Lake Chichancanab (Mexico). *Quat. Sci. Rev.* **24** 1413–1427.

[46] HODELL, D. A., BRENNER, M., CURTIS, J. H. and GUILDERSON, T. (2001). Solar forcing of drought frequency in the Maya lowlands. *Science* **292** 1367–1370. https://doi.org/10.1126/science.1057759

[47] HODELL, D. A., CURTIS, J. H. and BRENNER, M. (1995). Possible role of climate in the collapse of Classic Maya civilization. *Nature* **375** 391–394.

[48] HOOVER, D. R., RICE, J. A., WU, C. O. and YANG, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85** 809–822. MR1666699 https://doi.org/10.1093/biomet/85.4.809

[49] HUANG, J. Z., WU, C. O. and ZHOU, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statist. Sinica* **14** 763–788. MR2087972

[50] JANDHYALA, V., FOTOPOULOS, S., MACNEILL, I. and LIU, P. (2013). Inference for single and multiple change-points in time series. *J. Time Series Anal.* **34** 423–446. MR3070866 https://doi.org/10.1111/jtsa.12035

[51] JOHNSTONE, I. M. and SILVERMAN, B. W. (1997). Wavelet threshold estimators for data with correlated noise. *J. Roy. Statist. Soc. Ser. B* **59** 319–351. MR1440585 https://doi.org/10.1111/1467-9868.00071

[52] KAKIZAWA, Y. (2007). Moderate deviations for quadratic forms in Gaussian stationary processes. *J. Multivariate Anal.* **98** 992–1017. MR2325456 https://doi.org/10.1016/j.jmva.2006.07.004

[53] KARMAKAR, S., RICHTER, S. and WU, W. B. (2022). Simultaneous inference for time-varying models. *J. Econometrics* **227** 408–428. MR4384679 https://doi.org/10.1016/j.jeconom.2021.03.002

[54] KARMAKAR, S. and WU, W. B. (2020). Optimal Gaussian approximation for multiple time series. *Statist. Sinica* **30** 1399–1417. MR4257539 https://doi.org/10.5705/ss.202017.0303

[55] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131. MR0375412 https://doi.org/10.1007/BF00533093

[56] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrsch. Verw. Gebiete* **34** 33–58. MR0402883 https://doi.org/10.1007/BF00532688

[57] KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217–1241. MR1015147 https://doi.org/10.1214/aos/1176347265

[58] LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data. Springer Series in Statistics*. Springer, New York. MR2001447 https://doi.org/10.1007/978-1-4757-3803-2

[59] LEYBOURNE, S. J. and MCCABE, B. P. M. (1989). On the distribution of some test statistics for coefficient constancy. *Biometrika* **76** 169–177. MR0991435 https://doi.org/10.1093/biomet/76.1.169

[60] LIN, C.-F. J. and TERÄSVIRTA, T. (1999). Testing parameter constancy in linear models against stochastic stationary parameters. *J. Econometrics* **90** 193–213. MR1703341 https://doi.org/10.1016/S0304-4076(98)00041-4

[61] LIN, D. Y. and YING, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *J. Amer. Statist. Assoc.* **96** 103–126. MR1952726 https://doi.org/10.1198/016214501750333018

[62] LIU, R. Y. and SINGH, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap* (*East Lansing*, *MI*, 1990). *Wiley Ser. Probab. Math. Statist. Probab. Math. Statist.* 225–248. Wiley, New York. MR1197787

[63] LIU, W. and LIN, Z. (2009). Strong approximation for a class of stationary processes. *Stochastic Process. Appl.* **119** 249–280. MR2485027 https://doi.org/10.1016/j.spa.2008.01.012

[64] LIU, W. and WU, W. B. (2010). Asymptotics of spectral density estimates. *Econometric Theory* **26** 1218–1245. MR2660298 https://doi.org/10.1017/S026646660999051X

[65] LU, Q., LUND, R. and LEE, T. C. M. (2010). An MDL approach to the climate segmentation problem. *Ann. Appl. Stat.* **4** 299–319. MR2758173 https://doi.org/10.1214/09-AOAS289

[66] LUCERO, L. J., GUNN, J. D. and SCARBOROUGH, V. L. (2011). Climate change and classic Maya water management. *Water* **3** 479–494.

[67] MCGONIGLE, E. T., KILLICK, R. and NUNES, M. A. (2022). Modelling time-varying first and second-order structure of time series via wavelets and differencing. *Electron. J. Stat.* **16** 4398–4448. MR4474578 https://doi.org/10.1214/22-ejs2044

[68] MIES, F. and STELAND, A. (2023). Sequential Gaussian approximation for nonstationary time series in high dimensions. *Bernoulli* **29** 3114–3140. MR4632133 https://doi.org/10.3150/22-bej1577

[69] NABEYA, S. and TANAKA, K. (1988). Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative. *Ann. Statist.* **16** 218–235. MR0924867 https://doi.org/10.1214/aos/1176350701

[70] NAGAEV, S. V. (1979). Large deviations of sums of independent random variables. *Ann. Probab.* **7** 745–789. MR0542129

[71] NAPOLITANO, A. (2016). Cyclostationarity: New trends and applications. *Signal Process.* **120** 385–408.

[72] NEWEY, W. K. and WEST, K. D. (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55** 703–708. MR0890864 https://doi.org/10.2307/1913610

[73] NYBLOM, J. (1989). Testing for the constancy of parameters over time. *J. Amer. Statist. Assoc.* **84** 223–230. MR0999682

[74] PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–115. MR0088850 https://doi.org/10.1093/biomet/41.1-2.100

[75] PAGE, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika* **42** 523–527. MR0072412 https://doi.org/10.1093/biomet/42.3-4.523

[76] PARZEN, E. and PAGANO, M. (1979). An approach to modeling seasonally stationary time series. *J. Econometrics* **9** 137–153.

[77] PERCIVAL, D. B. and MONDAL, D. (2012). 22—a wavelet variance primer. In *Time Series Analysis: Methods and Applications* (T. Subba Rao, S. Subba Rao and C. R. Rao, eds.). *Handbook of Statistics* **30** 623–657. Elsevier.

[78] PERRON, P. et al. (2006). Dealing with structural breaks. *Palgrave Handbook of Econometrics* **1** 278–352.

[79] PLOBERGER, W., KRÄMER, W. and KONTRUS, K. (1989). A new test for structural stability in the linear regression model. *J. Econometrics* **40** 307–318. MR0994952 https://doi.org/10.1016/0304-4076(89)90087-0

[80] POLITIS, D. N. and ROMANO, J. P. (1994). The stationary bootstrap. *J. Amer. Statist. Assoc.* **89** 1303–1313. MR1310224

[81] POLITIS, D. N. and ROMANO, J. P. (1995). Bias-corrected nonparametric spectral estimation. *J. Time Series Anal.* **16** 67–103. MR1323618 https://doi.org/10.1111/j.1467-9892.1995.tb00223.x

[82] POLITIS, D. N. and WHITE, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Rev.* **23** 53–70. MR2041534 https://doi.org/10.1081/ETC-120028836

[83] PRIESTLEY, M. B. (1981). *Spectral Analysis and Time Series. Probability and Mathematical Statistics: A Series of Monographs and Textbooks v.* 1. Academic Press.

[84] RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2168993

[85] REEVES, J., CHEN, J., WANG, X. L., LUND, R. and LU, Q. Q. (2007). A review and comparison of changepoint detection techniques for climate data. *J. Appl. Meteorol. Climatol.* **46** 900–915.

[86] RICHTER, S. and DAHLHAUS, R. (2019). Cross validation for locally stationary processes. *Ann. Statist.* **47** 2145–2173. MR3953447 https://doi.org/10.1214/18-AOS1743

[87] ROBBINS, M. W., LUND, R. B., GALLAGHER, C. M. and LU, Q. (2011). Changepoints in the North Atlantic tropical cyclone record. *J. Amer. Statist. Assoc.* **106** 89–99. MR2816704 https://doi.org/10.1198/jasa.2011.ap10023

[88] RUDELSON, M. and VERSHYNIN, R. (2013). Hanson–Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.* **18** no. 82. MR3125258 https://doi.org/10.1214/ECP.v18-2865

[89] SAKHANENKO, A. I. (2006). Estimates in the invariance principle in terms of truncated power moments. *Sibirsk. Mat. Zh.* **47** 1355–1371. MR2302850 https://doi.org/10.1007/s11202-006-0119-1

[90] STOUMBOS, Z. G., REYNOLDS JR., M. R., RYAN, T. P. and WOODALL, W. H. (2000). The state of statistical process control as we proceed into the 21st century. *J. Amer. Statist. Assoc.* **95** 992–998.

[91] TURNER, B. L. and SABLOFF, J. A. (2012). Classic period collapse of the Central Maya Lowlands: Insights about human–environment relationships for sustainability. *Proc. Natl. Acad. Sci. USA* **109** 13908–13914.

[92] VON SACHS, R. and MACGIBBON, B. (2000). Non-parametric curve estimation by wavelet thresholding with locally stationary errors. *Scand. J. Stat.* **27** 475–499. MR1795776 https://doi.org/10.1111/1467-9469.00202

[93] WRIGHT, F. T. (1973). A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric. *Ann. Probab.* **1** 1068–1070. MR0353419 https://doi.org/10.1214/aop/1176996815

[94] WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. MR2172215 https://doi.org/10.1073/pnas.0506715102

[95] WU, W. B. and ZHAO, Z. (2007). Inference of trends in time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 391–410. MR2323759 https://doi.org/10.1111/j.1467-9868.2007.00594.x

[96] WU, W. B. and ZHOU, Z. (2011). Gaussian approximations for non-stationary multiple time series. *Statist. Sinica* **21** 1397–1413. MR2827528 https://doi.org/10.5705/ss.2008.223

[97] XIAO, H. and WU, W. B. (2012). Covariance matrix estimation for stationary time series. *Ann. Statist.* **40** 466–493. MR3014314 https://doi.org/10.1214/11-AOS967

[98] YOFFEE, N. and COWGILL, G. L. (1991). *The Collapse of Ancient States and Civilizations. Book Collections on Project MUSE*. Univ. Arizona Press.

[99] ZAITSEV, A. Y. (2000). Multidimensional version of a result of Sakhanenko in the invariance principle for vectors with finite exponential moments. I. *Teor. Veroyatn. Primen.* **45** 718–738. MR1968723 https://doi.org/10.1137/S0040585X97978555

[100] ZAITSEV, A. Y. (2001). Multidimensional version of a result of Sakhanenko in the invariance principle for vectors with finite exponential moments. II. *Teor. Veroyatn. Primen.* **46** 535–561. MR1978667 https://doi.org/10.1137/S0040585X97979123

[101] ZAITSEV, A. Y. (2001). Multidimensional version of a result of Sakhanenko in the invariance principle for vectors with finite exponential moments. III. *Teor. Veroyatn. Primen.* **46** 744–769. MR1971831 https://doi.org/10.1137/S0040585X97979305

[102] ZHANG, D. and WU, W. B. (2021). Convergence of covariance and spectral density estimates for high-dimensional locally stationary processes. *Ann. Statist.* **49** 233–254. MR4206676 https://doi.org/10.1214/20-AOS1954

[103] ZHANG, T. and WU, W. B. (2012). Inference of time-varying regression models. *Ann. Statist.* **40** 1376–1402. MR3015029 https://doi.org/10.1214/12-AOS1010

[104] ZHANG, T. and WU, W. B. (2015). Time-varying nonlinear regression models: Nonparametric estimation and model selection. *Ann. Statist.* **43** 741–768. MR3319142 https://doi.org/10.1214/14-AOS1299

[105] ZHANG, W., LEE, S.-Y. and SONG, X. (2002). Local polynomial fitting in semivarying coefficient model. *J. Multivariate Anal.* **82** 166–188. MR1918619 https://doi.org/10.1006/jmva.2001.2012

[106] ZHAO, Z. and LI, X. (2013). Inference for modulated stationary processes. *Bernoulli* **19** 205–227. MR3019492 https://doi.org/10.3150/11-BEJ399

[107] ZHOU, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *J. Amer. Statist. Assoc.* **108** 726–740. MR3174655 https://doi.org/10.1080/01621459.2013.787184

[108] ZHOU, Z. and WU, W. B. (2010). Simultaneous inference of linear models with time varying coefficients. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 513–531. MR2758526 https://doi.org/10.1111/j.1467-9868.2010.00743.x

[109] BONNERJEE, S., KARMAKAR, S. and WU, W. B. (2024). Supplement to "Gaussian approximation for nonstationary time series with optimal rate and explicit construction." https://doi.org/10.1214/24-AOS2436SUPP

The Annals of Statistics
2024, Vol. 52, No. 5, 2318–2348
https://doi.org/10.1214/24-AOS2437

COMPUTATIONAL LOWER BOUNDS FOR GRAPHON ESTIMATION VIA LOW-DEGREE POLYNOMIALS

BY YUETIAN LUO[1,a] AND CHAO GAO[2,b]

[1]*Data Science Institute, University of Chicago,* [a]*yuetian@uchicago.edu*
[2]*Department of Statistics, University of Chicago,* [b]*chaogao@uchicago.edu*

Graphon estimation has been one of the most fundamental problems in network analysis and has received considerable attention in the past decade. From the statistical perspective, the minimax error rate of graphon estimation has been established by (*Ann. Statist.* **43** (2015) 2624–2652) for both stochastic block model (SBM) and nonparametric graphon estimation. The statistical optimal estimators are based on constrained least squares and have computational complexity exponential in the dimension. From the computational perspective, the best-known, polynomial-time estimator is based on universal singular value thresholding (USVT), but it can only achieve a much slower estimation error rate than the minimax one. It is natural to wonder if such a gap is essential. The computational optimality of the USVT or the existence of a computational barrier in graphon estimation has been a long-standing open problem. In this work, we take the first step toward it and provide rigorous evidence for the computational barrier in graphon estimation via low-degree polynomials. Specifically, in SBM graphon estimation, we show that for low-degree polynomial estimators, their estimation error rates cannot be significantly better than that of the USVT under a wide range of parameter regimes and in nonparametric graphon estimation, we show low-degree polynomial estimators achieve estimation error rates strictly slower than the minimax rate. Our results are proved based on the recent development of low-degree polynomials by (*Ann. Statist.* **50** (2022) 1833–1858), while we overcome a few key challenges in applying it to the general graphon estimation problem. By leveraging our main results, we also provide a computational lower bound on the clustering error for community detection in SBM with a growing number of communities and this yields a new piece of evidence for the conjectured Kesten–Stigum threshold for efficient community recovery. Finally, we extend our computational lower bounds to sparse graphon estimation and biclustering with additive Gaussian noise, and provide discussion on the optimality of our results.

REFERENCES

ABBE, E. (2017). Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** 6446–6531. MR3827065

ABBE, E., BANDEIRA, A. S. and HALL, G. (2016). Exact recovery in the stochastic block model. *IEEE Trans. Inf. Theory* **62** 471–487. MR3447993 https://doi.org/10.1109/TIT.2015.2490670

ABBE, E. and SANDON, C. (2018). Proof of the achievability conjectures for the general stochastic block model. *Comm. Pure Appl. Math.* **71** 1334–1406. MR3812075 https://doi.org/10.1002/cpa.21719

AIROLDI, E. M., COSTA, T. B. and CHAN, S. H. (2013). Stochastic blockmodel approximation of a graphon: Theory and consistent estimation. *Adv. Neural Inf. Process. Syst.* **26**.

ALDOUS, D. J. (1981). Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* **11** 581–598. MR0637937 https://doi.org/10.1016/0047-259X(81)90099-3

ARPINO, G. and VENKATARAMANAN, R. (2023). Statistical–computational tradeoffs in mixed sparse linear regression. In *Proceedings of Thirty Sixth Conference on Learning Theory* **195** 921–986.

MSC2020 subject classifications. Primary 68Q17, 05C80, 62H30; secondary 62C20, 62G05.
Key words and phrases. Graphon estimation, computational lower bound, low-degree polynomials, community detection, Kesten–Stigum threshold, statistical–computational tradeoffs.

AUDDY, A. and YUAN, M. (2023). Large Dimensional Independent Component Analysis: Statistical Optimality and Computational Tractability. ArXiv preprint. Available at arXiv:2303.18156. MR4639578

BALAKRISHNAN, S., KOLAR, M., RINALDO, A., SINGH, A. and WASSERMAN, L. (2011). Statistical and computational tradeoffs in biclustering. In *NIPS* 2011 *Workshop on Computational Trade-Offs in Statistical Learning*, *Vol*. 4.

BANDEIRA, A. S., BANKS, J., KUNISKY, D., MOORE, C. and WEIN, A. (2021). Spectral planting and the hardness of refuting cuts, colorability, and communities in random graphs. In *Conference on Learning Theory* 410–473. PMLR.

BANDEIRA, A. S., EL ALAOUI, A., HOPKINS, S., SCHRAMM, T., WEIN, A. S. and ZADIK, I. (2022). The Franz-Parisi criterion and computational trade-offs in high dimensional statistics. *Adv. Neural Inf. Process. Syst.* **35** 33831–33844.

BANDEIRA, A. S., KUNISKY, D. and WEIN, A. S. (2020). Computational hardness of certifying bounds on constrained PCA problems. In 11*th Innovations in Theoretical Computer Science Conference*. *LIPIcs*. *Leibniz Int. Proc. Inform.* **151** Art. No. 78, 29. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR4048181

BANKS, J., MOHANTY, S. and RAGHAVENDRA, P. (2021). Local statistics, semidefinite programming, and community detection. In *Proceedings of the* 2021 *ACM-SIAM Symposium on Discrete Algorithms* (*SODA*) 1298–1316. SIAM, Philadelphia, PA. MR4262512 https://doi.org/10.1137/1.9781611976465.79

BANKS, J., MOORE, C., NEEMAN, J. and NETRAPALLI, P. (2016). Information-theoretic thresholds for community detection in sparse networks. In *Conference on Learning Theory* 383–416. PMLR.

BANKS, J., MOORE, C., VERSHYNIN, R., VERZELEN, N. and XU, J. (2018). Information-theoretic bounds and phase transitions in clustering, sparse PCA, and submatrix localization. *IEEE Trans. Inf. Theory* **64** 4872–4994. MR3819345 https://doi.org/10.1109/tit.2018.2810020

BARAK, B., HOPKINS, S., KELNER, J., KOTHARI, P. K., MOITRA, A. and POTECHIN, A. (2019). A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM J. Comput.* **48** 687–735. MR3945259 https://doi.org/10.1137/17M1138236

BERTHET, Q. and RIGOLLET, P. (2013). Complexity theoretic lower bounds for sparse principal component detection. In *Conference on Learning Theory* 1046–1066.

BICKEL, P. J. and CHEN, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA* **106** 21068–21073.

BICKEL, P. J., CHEN, A. and LEVINA, E. (2011). The method of moments and degree distributions for network models. *Ann. Statist.* **39** 2280–2301. MR2906868 https://doi.org/10.1214/11-AOS904

BORGS, C., CHAYES, J. and SMITH, A. (2015). Private graphon estimation for sparse graphs. *Adv. Neural Inf. Process. Syst.* **28**.

BORGS, C., CHAYES, J. T., COHN, H. and GANGULY, S. (2021). Consistent nonparametric estimation for heavy-tailed sparse graphs. *Ann. Statist.* **49** 1904–1930. MR4319235 https://doi.org/10.1214/20-aos1985

BORGS, C., CHAYES, J. T., COHN, H. and ZHAO, Y. (2018). An L^p theory of sparse graph convergence II: LD convergence, quotients and right convergence. *Ann. Probab.* **46** 337–396. MR3758733 https://doi.org/10.1214/17-AOP1187

BORGS, C., CHAYES, J. T., COHN, H. and ZHAO, Y. (2019). An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions. *Trans. Amer. Math. Soc.* **372** 3019–3062. MR3988601 https://doi.org/10.1090/tran/7543

BORGS, C., CHAYES, J. T., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2008). Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.* **219** 1801–1851. MR2455626 https://doi.org/10.1016/j.aim.2008.07.008

BORGS, C., CHAYES, J. T., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2012). Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Ann. of Math.* (2) **176** 151–219. MR2925382 https://doi.org/10.4007/annals.2012.176.1.2

BRENNAN, M. and BRESLER, G. (2020). Reducibility and statistical–computational gaps from secret leakage. In *Conference on Learning Theory* 648–847. PMLR.

BRENNAN, M., BRESLER, G. and HULEIHEL, W. (2018). Reducibility and computational lower bounds for problems with planted sparse structure. In *Conference on Learning Theory* 48–166. PMLR.

BRENNAN, M. S., BRESLER, G., HOPKINS, S., LI, J. and SCHRAMM, T. (2021). Statistical query algorithms and low degree tests are almost equivalent. In *Conference on Learning Theory* 774–774. PMLR.

BRESLER, G. and HUANG, B. (2022). The algorithmic phase transition of random k-SAT for low degree polynomials. In 2021 *IEEE* 62*nd Annual Symposium on Foundations of Computer Science—FOCS* 2021 298–309. IEEE Comput. Soc., Los Alamitos, CA. MR4399691

BUTUCEA, C. and INGSTER, Y. I. (2013). Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli* **19** 2652–2688. MR3160567 https://doi.org/10.3150/12-BEJ470

BUTUCEA, C., INGSTER, Y. I. and SUSLINA, I. A. (2015). Sharp variable selection of a sparse submatrix in a high-dimensional noisy matrix. *ESAIM Probab. Stat.* **19** 115–134. MR3374872 https://doi.org/10.1051/ps/2014017

CAI, T. T., LIANG, T. and RAKHLIN, A. (2017). Computational and statistical boundaries for submatrix localization in a large noisy matrix. *Ann. Statist.* **45** 1403–1430. MR3670183 https://doi.org/10.1214/16-AOS1488

CHAN, S. and AIROLDI, E. (2014). A consistent histogram estimator for exchangeable graph models. In *International Conference on Machine Learning* 208–216. PMLR.

CHANDRASEKARAN, V. and JORDAN, M. I. (2013). Computational and statistical tradeoffs via convex relaxation. *Proc. Natl. Acad. Sci. USA* **110** E1181–E1190. MR3047651 https://doi.org/10.1073/pnas.1302293110

CHATTERJEE, S. (2015). Matrix estimation by universal singular value thresholding. *Ann. Statist.* **43** 177–214. MR3285604 https://doi.org/10.1214/14-AOS1272

CHEN, Y. and XU, J. (2016). Statistical–computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *J. Mach. Learn. Res.* **17** 882–938. MR3491121

CHI, E. C., ALLEN, G. I. and BARANIUK, R. G. (2017). Convex biclustering. *Biometrics* **73** 10–19. MR3632347 https://doi.org/10.1111/biom.12540

CHIN, P., RAO, A. and VU, V. (2015). Stochastic block model and community detection in sparse graphs: A spectral algorithm with optimal rate of recovery. In *Conference on Learning Theory* 391–423. PMLR.

CHOI, D. (2017). Co-clustering of nonsmooth graphons. *Ann. Statist.* **45** 1488–1515. MR3670186 https://doi.org/10.1214/16-AOS1497

CHOI, D. and WOLFE, P. J. (2014). Co-clustering separately exchangeable network data. *Ann. Statist.* **42** 29–63. MR3161460 https://doi.org/10.1214/13-AOS1173

DADON, M., HULEIHEL, W. and BENDORY, T. (2024). Detection and recovery of hidden submatrices. *IEEE Trans. Signal Inf. Process. Netw.* **10** 69–82. MR4692776 https://doi.org/10.1109/tsipn.2024.3352264

DECELLE, A., KRZAKALA, F., MOORE, C. and ZDEBOROVÁ, L. (2011). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* **84** 066106.

DIACONIS, P. and JANSON, S. (2008). Graph limits and exchangeable random graphs. *Rend. Mat. Appl.* (7) **28** 33–61. MR2463439

DIAKONIKOLAS, I. and KANE, D. (2022). Non-Gaussian component analysis via lattice basis reduction. In *Conference on Learning Theory* 4535–4547. PMLR.

DIAKONIKOLAS, I., KANE, D. M. and STEWART, A. (2017). Statistical query lower bounds for robust estimation of high-dimensional Gaussians and Gaussian mixtures (extended abstract). In 58*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 73–84. IEEE Comput. Soc., Los Alamitos, CA. MR3734219 https://doi.org/10.1109/FOCS.2017.16

DIAKONIKOLAS, I., KONG, W. and STEWART, A. (2019). Efficient algorithms and lower bounds for robust linear regression. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms* 2745–2754. SIAM, Philadelphia, PA. MR3909639 https://doi.org/10.1137/1.9781611975482.170

DING, Y., KUNISKY, D., WEIN, A. S. and BANDEIRA, A. S. (2024). Subexponential-time algorithms for sparse PCA. *Found. Comput. Math.* **24** 865–914. MR4760356 https://doi.org/10.1007/s10208-023-09603-0

DONIER-MEROZ, E., DALALYAN, A. S., KRAMARZ, F., CHONÉ, P. and D'HAULTFOEUILLE, X. (2023). Graphon Estimation in bipartite graphs with observable edge labels and unobservable node labels. ArXiv preprint. Available at arXiv:2304.03590.

DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2009). Message-passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci.* **106** 18914–18919.

FEI, Y. and CHEN, Y. (2020). Achieving the Bayes error rate in synchronization and block models by SDP, robustly. *IEEE Trans. Inf. Theory* **66** 3929–3953. MR4115142 https://doi.org/10.1109/TIT.2020.2966438

FELDMAN, V., GRIGORESCU, E., REYZIN, L., VEMPALA, S. S. and XIAO, Y. (2017). Statistical algorithms and a lower bound for detecting planted cliques. *J. ACM* **64** Art. 8, 37. MR3664576 https://doi.org/10.1145/3046674

FELDMAN, V., PERKINS, W. and VEMPALA, S. (2018). On the complexity of random satisfiability problems with planted solutions. *SIAM J. Comput.* **47** 1294–1338. MR3827195 https://doi.org/10.1137/16M1078471

GAMARNIK, D., JAGANNATH, A. and WEIN, A. S. (2020). Low-degree hardness of random optimization problems. In 2020 *IEEE* 61*st Annual Symposium on Foundations of Computer Science* 131–140. IEEE Comput. Soc., Los Alamitos, CA. MR4232029 https://doi.org/10.1109/FOCS46700.2020.00021

GAMARNIK, D. and SUDAN, M. (2014). Limits of local algorithms over sparse random graphs [extended abstract]. In *ITCS*'14—*Proceedings of the* 2014 *Conference on Innovations in Theoretical Computer Science* 369–375. ACM, New York. MR3359490

GAO, C., LU, Y., MA, Z. and ZHOU, H. H. (2016). Optimal estimation and completion of matrices with biclustering structures. *J. Mach. Learn. Res.* **17** 5602–5630. MR3569248

GAO, C., LU, Y. and ZHOU, H. H. (2015). Rate-optimal graphon estimation. *Ann. Statist.* **43** 2624–2652. MR3405606 https://doi.org/10.1214/15-AOS1354

GAO, C. and MA, Z. (2021). Minimax rates in network analysis: Graphon estimation, community detection and hypothesis testing. *Statist. Sci.* **36** 16–33. MR4194201 https://doi.org/10.1214/19-STS736

GAO, C., MA, Z. and ZHOU, H. H. (2017). Sparse CCA: Adaptive estimation and computational barriers. *Ann. Statist.* **45** 2074–2101. MR3718162 https://doi.org/10.1214/16-AOS1519

GAUCHER, S. and KLOPP, O. (2021). Optimality of variational inference for stochasticblock model with missing links. *Adv. Neural Inf. Process. Syst.* **34** 19947–19959.

GIRVAN, M. and NEWMAN, M. E. J. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **99** 7821–7826. MR1908073 https://doi.org/10.1073/pnas.122653799

GOLDENBERG, A., ZHENG, A. X., FIENBERG, S. E., AIROLDI, E. M. et al. (2010). A survey of statistical network models. *Found. Trends Mach. Learn.* **2** 129–233.

GUÉDON, O. and VERSHYNIN, R. (2016). Community detection in sparse networks via Grothendieck's inequality. *Probab. Theory Related Fields* **165** 1025–1049. MR3520025 https://doi.org/10.1007/s00440-015-0659-z

HAJEK, B., WU, Y. and XU, J. (2015). Computational lower bounds for community detection on random graphs. In *Conference on Learning Theory* 899–928.

HARTIGAN, J. A. (1972). Direct clustering of a data matrix. *J. Amer. Statist. Assoc.* **67** 123–129.

HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. MR0718088 https://doi.org/10.1016/0378-8733(83)90021-7

HOLMGREN, J. and WEIN, A. S. (2021). Counterexamples to the low-degree conjecture. In 12*th Innovations in Theoretical Computer Science Conference*. *LIPIcs. Leibniz Int. Proc. Inform.* **185** Art. No. 75, 9. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR4214319

HOOVER, D. N. (1979). Relations on probability spaces and arrays of random variables. Preprint, Institute for Advanced Study, Princeton, NJ 2 275.

HOPKINS, S. B., KOTHARI, P. K., POTECHIN, A., RAGHAVENDRA, P., SCHRAMM, T. and STEURER, D. (2017). The power of sum-of-squares for detecting hidden structures. In 58*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 720–731. IEEE Comput. Soc., Los Alamitos, CA. MR3734275 https://doi.org/10.1109/FOCS.2017.72

HOPKINS, S. B. and STEURER, D. (2017). Efficient Bayesian estimation from few samples: Community detection and related problems. In 58*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 379–390. IEEE Comput. Soc., Los Alamitos, CA. MR3734245 https://doi.org/10.1109/FOCS.2017.42

HOPKINS, S. B. K. (2018). Statistical Inference and the Sum of Squares Method. Ph.D. Thesis. MR3864930

JIN, J. (2015). Fast community detection by SCORE. *Ann. Statist.* **43** 57–89. MR3285600 https://doi.org/10.1214/14-AOS1265

JIN, J., KE, T., TURNER, P. and ZHANG, A. (2022). Phase transition for detecting a small community in a large network. In *The* 11*th International Conference on Learning Representations*.

KESTEN, H. and STIGUM, B. P. (1966). Additional limit theorems for indecomposable multidimensional Galton–Watson processes. *Ann. Math. Stat.* **37** 1463–1481. MR0200979 https://doi.org/10.1214/aoms/1177699139

KLOPP, O., TSYBAKOV, A. B. and VERZELEN, N. (2017). Oracle inequalities for network models and sparse graphon estimation. *Ann. Statist.* **45** 316–354. MR3611494 https://doi.org/10.1214/16-AOS1454

KLOPP, O. and VERZELEN, N. (2019). Optimal graphon estimation in cut distance. *Probab. Theory Related Fields* **174** 1033–1090. MR3980311 https://doi.org/10.1007/s00440-018-0878-1

KOEHLER, F. and MOSSEL, E. (2022). Reconstruction on trees and low-degree polynomials. *Adv. Neural Inf. Process. Syst.* **35** 18942–18954.

KOLAR, M., BALAKRISHNAN, S., RINALDO, A. and SINGH, A. (2011). Minimax localization of structural information in large noisy matrices. In *Advances in Neural Information Processing Systems* 909–917.

KUNISKY, D., WEIN, A. S. and BANDEIRA, A. S. (2022). Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *Mathematical Analysis, Its Applications and Computation*. *Springer Proc. Math. Stat.* **385** 1–50. Springer, Cham. MR4461037 https://doi.org/10.1007/978-3-030-97127-4_1

LEI, J. and RINALDO, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann. Statist.* **43** 215–237. MR3285605 https://doi.org/10.1214/14-AOS1274

LI, X., CHEN, Y. and XU, J. (2021). Convex relaxation methods for community detection. *Statist. Sci.* **36** 2–15. MR4194200 https://doi.org/10.1214/19-STS715

LI, Y., SHAH, D., SONG, D. and YU, C. L. (2020). Nearest neighbors for matrix estimation interpreted as blind regression for latent variable model. *IEEE Trans. Inf. Theory* **66** 1760–1784. MR4077516 https://doi.org/10.1109/tit.2019.2950299

LÖFFLER, M., WEIN, A. S. and BANDEIRA, A. S. (2022). Computationally efficient sparse clustering. *Inf. Inference* **11** 1255–1286. MR4526323 https://doi.org/10.1093/imaiai/iaac019

LOVÁSZ, L. (2012). *Large Networks and Graph Limits*. *American Mathematical Society Colloquium Publications* **60**. Amer. Math. Soc., Providence, RI. MR3012035 https://doi.org/10.1090/coll/060

LOVÁSZ, L. and SZEGEDY, B. (2006). Limits of dense graph sequences. *J. Combin. Theory Ser. B* **96** 933–957. MR2274085 https://doi.org/10.1016/j.jctb.2006.05.002

LUO, Y. and GAO, C. (2024). Supplement to "Computational Lower Bounds for Graphon Estimation via Low-degree Polynomials." https://doi.org/10.1214/24-AOS2437SUPP

LUO, Y. and ZHANG, A. R. (2022a). Tensor clustering with planted structures: Statistical optimality and computational limits. *Ann. Statist.* **50** 584–613. MR4382029 https://doi.org/10.1214/21-aos2123

LUO, Y. and ZHANG, A. R. (2022b). Tensor-on-tensor regression: Riemannian optimization, over-parameterization, statistical–computational gap, and their interplay. arXiv preprint. Available at arXiv:2206.08756.

LYU, Z. and XIA, D. (2023). Optimal estimation and computational limit of low-rank Gaussian mixtures. *Ann. Statist.* **51** 646–667. MR4600996 https://doi.org/10.1214/23-aos2264

MA, T. and WIGDERSON, A. (2015). Sum-of-squares lower bounds for sparse PCA. In *Advances in Neural Information Processing Systems* 1612–1620.

MA, Z. and WU, Y. (2015). Computational barriers in minimax submatrix detection. *Ann. Statist.* **43** 1089–1116. MR3346698 https://doi.org/10.1214/14-AOS1300

MANKAD, S. and MICHAILIDIS, G. (2014). Biclustering three-dimensional data arrays with plaid models. *J. Comput. Graph. Statist.* **23** 943–965. MR3270705 https://doi.org/10.1080/10618600.2013.851608

MAO, C. and WEIN, A. S. (2021). Optimal spectral recovery of a planted vector in a subspace. ArXiv preprint. Available at arXiv:2105.15081.

MAO, C., WEIN, A. S. and ZHANG, S. (2023). Detection-recovery gap for planted dense cycles. In *The Thirty Sixth Annual Conference on Learning Theory* 2440–2481. PMLR.

MASSOULIÉ, L. (2014). Community detection thresholds and the weak Ramanujan property. In *STOC'*14*—Proceedings of the* 2014 *ACM Symposium on Theory of Computing* 694–703. ACM, New York. MR3238997

MILLER, K. T., JORDAN, M. and GRIFFITHS, T. (2009). Nonparametric latent feature models for link prediction. In *Advances in Neural Information Processing Systems*, *Vol.* 22.

MONTANARI, A. and WEIN, A. S. (2022). Equivalence of approximate message passing and low-degree polynomials in rank-one matrix estimation. ArXiv preprint. Available at arXiv:2212.06996.

MOORE, C. (2017). The computer science and physics of community detection: Landscapes, phase transitions, and hardness. *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS* **121** 26–61. MR3699594

MOSSEL, E., NEEMAN, J. and SLY, A. (2015a). Consistency thresholds for the planted bisection model [extended abstract]. In *STOC'*15*—Proceedings of the* 2015 *ACM Symposium on Theory of Computing* 69–75. ACM, New York. MR3388184

MOSSEL, E., NEEMAN, J. and SLY, A. (2015b). Reconstruction and estimation in the planted partition model. *Probab. Theory Related Fields* **162** 431–461. MR3383334 https://doi.org/10.1007/s00440-014-0576-6

MOSSEL, E., NEEMAN, J. and SLY, A. (2018). A proof of the block model threshold conjecture. *Combinatorica* **38** 665–708. MR3876880 https://doi.org/10.1007/s00493-016-3238-8

OLHEDE, S. C. and WOLFE, P. J. (2014). Network histograms and universality of blockmodel approximation. *Proc. Natl. Acad. Sci. USA* **111** 14722–14727.

PANANJADY, A. and SAMWORTH, R. J. (2022). Isotonic regression with unknown permutations: Statistics, computation and adaptation. *Ann. Statist.* **50** 324–350.

PENSKY, M. (2019). Dynamic network models and graphon estimation. *Ann. Statist.* **47**.

ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915.

ROHE, K., QIN, T. and YU, B. (2016). Co-clustering directed graphs to discover asymmetries and directional communities. *Proc. Natl. Acad. Sci. USA* **113** 12679–12684.

ROSSMAN, B. (2008). On the constant-depth complexity of k-clique. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing* 721–730. ACM, New York.

SCHRAMM, T. and WEIN, A. S. (2022). Computational barriers to estimation from low-degree polynomials. *Ann. Statist.* **50** 1833–1858.

SHAH, D. and LEE, C. (2018). Reducing crowdsourcing to graphon estimation, statistically. In *International Conference on Artificial Intelligence and Statistics* 1741–1750. PMLR.

SHAH, N., BALAKRISHNAN, S., GUNTUBOYINA, A. and WAINWRIGHT, M. (2016). Stochastically transitive models for pairwise comparisons: Statistical and computational issues. In *International Conference on Machine Learning* 11–20. PMLR.

SISCHKA, B. and KAUERMANN, G. (2022). Stochastic block smooth graphon model. arXiv preprint. Available at arXiv:2203.13304.

WANG, T., BERTHET, Q. and SAMWORTH, R. J. (2016). Statistical and computational trade-offs in estimation of sparse principal components. *Ann. Statist.* **44** 1896–1930.

WASSERMAN, S. and FAUST, K. (1994). Social Network Analysis: Methods and Applications.

WEIN, A. S. (2022). Optimal low-degree hardness of maximum independent set. *Math. Stat. Learn.* **4** 221–251.

WEIN, A. S. (2023). Average-case complexity of tensor decomposition for low-degree polynomials. In *Proceedings of the* 55*th Annual ACM Symposium on Theory of Computing* 1685–1698.

WOLFE, P. J. and OLHEDE, S. C. (2013). Nonparametric graphon estimation. ArXiv preprint. Available at arXiv: 1309.5936.

WU, Y. and XU, J. (2021). Statistical problems with planted structures: Information-theoretical and computational limits. In *Information-Theoretic Methods in Data Science* 383. 13.

XU, J. (2018). Rates of convergence of spectral methods for graphon estimation. In *International Conference on Machine Learning* 5433–5442. PMLR.

YANG, J., HAN, C. and AIROLDI, E. (2014). Nonparametric estimation and testing of exchangeable graph models. In *Artificial Intelligence and Statistics* 1060–1067. PMLR.

ZADIK, I., SONG, M. J., WEIN, A. S. and BRUNA, J. (2022). Lattice-based methods surpass sum-of-squares in clustering. In *Conference on Learning Theory* 1247–1248. PMLR.

ZHANG, A. and XIA, D. (2018). Tensor svd: Statistical and computational limits. *IEEE Trans. Inf. Theory* **64** 7311–7338.

ZHANG, A. Y. and ZHOU, H. H. (2016). Minimax rates of community detection in stochastic block models. *Ann. Statist.* 2252–2280.

ZHANG, Y., LEVINA, E. and ZHU, J. (2017). Estimating network edge probabilities by neighbourhood smoothing. *Biometrika* **104** 771–783.

The Annals of Statistics
2024, Vol. 52, No. 5, 2349–2374
https://doi.org/10.1214/24-AOS2438

A NONPARAMETRIC TEST FOR ELLIPTICAL DISTRIBUTION BASED ON KERNEL EMBEDDING OF PROBABILITIES

BY YIN TANG[a] AND BING LI[b]

Department of Statistics, Pennsylvania State University, [a]*yqt5219@psu.edu,* [b]*bxl9@psu.edu*

Elliptical distribution is a basic assumption underlying many multivariate statistical methods. For example, in sufficient dimension reduction and statistical graphical models, this assumption is routinely imposed to simplify the data dependence structure. Before applying such methods, we need to decide whether the data are elliptically distributed. Currently existing tests either focus exclusively on spherical distributions, or rely on bootstrap to determine the null distribution, or require specific forms of the alternative distribution. In this paper, we introduce a general nonparametric test for elliptical distribution based on kernel embedding of the probability measure that embodies the two properties that characterize an elliptical distribution: namely, after centering and rescaling, (1) the direction and length of the random vector are independent, and (2) the directional vector is uniformly distributed on the unit sphere. We derive the asymptotic distributions of the test statistic via von Mises expansion, develop the sample-level procedure to determine the rejection region, and establish the consistency and validity of the proposed test. We also develop the concentration bounds of the test statistic, allowing the dimension to grow with the sample size, and further establish the consistency in this high-dimension setting. We compare our method with several existing methods via simulation studies, and apply our test to a SENIC dataset with and without a transformation aimed to achieve ellipticity.

REFERENCES

ALBISETTI, I., BALABDAOUI, F. and HOLZMANN, H. (2020). Testing for spherical and elliptical symmetry. *J. Multivariate Anal.* **180** 104667. MR4143534 https://doi.org/10.1016/j.jmva.2020.104667

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley-Interscience, Hoboken, NJ. MR1990662

BABIĆ, S., GELBGRAS, L., HALLIN, M. and LEY, C. (2021). Optimal tests for elliptical symmetry: Specified and unspecified location. *Bernoulli* **27** 2189–2216. MR4303880 https://doi.org/10.3150/20-BEJ1305

BABIĆ, S., LEY, C. and PALANGETIĆ, M. (2021). The R journal: Elliptical symmetry tests in R. **13** 661–672. https://doi.org/10.32614/RJ-2021-078

BARINGHAUS, L. (1991). Testing for spherical symmetry of a multivariate distribution. *Ann. Statist.* **19** 899–917. MR1105851 https://doi.org/10.1214/aos/1176348127

CAMBANIS, S., HUANG, S. and SIMONS, G. (1981). On the theory of elliptically contoured distributions. *J. Multivariate Anal.* **11** 368–385. MR0629795 https://doi.org/10.1016/0047-259X(81)90082-8

CASSART, D., HALLIN, M. and PAINDAVEINE, D. (2008). Optimal detection of Fechner-asymmetry. *J. Statist. Plann. Inference* **138** 2499–2525. MR2414260 https://doi.org/10.1016/j.jspi.2007.10.011

COOK, R. D. and LI, B. (2002). Dimension reduction for conditional mean in regression. *Ann. Statist.* **30** 455–474. MR1902895 https://doi.org/10.1214/aos/1021379861

DUCHESNE, P. and LAFAYE DE MICHEAUX, P. (2010). Computing the distribution of quadratic forms: Further comparisons between the Liu–Tang–Zhang approximation and exact methods. *Comput. Statist. Data Anal.* **54** 858–862. MR2580921 https://doi.org/10.1016/j.csda.2009.11.025

EATON, M. L. (1986). A characterization of spherical distributions. *J. Multivariate Anal.* **20** 272–276. MR0866075 https://doi.org/10.1016/0047-259X(86)90083-7

FERNHOLZ, L. T. (1983). *von Mises Calculus for Statistical Functionals*. *Lecture Notes in Statistics* **19**. Springer, New York. MR0713611 https://doi.org/10.1007/978-1-4612-5604-5

MSC2020 subject classifications. Primary 62G10, 62G20; secondary 62H10.

Key words and phrases. Elliptical distribution, reproducing kernel Hilbert space, kernel embedding of probability, von Mises expansion.

FUKUMIZU, K., GRETTON, A., LANCKRIET, G., SCHÖLKOPF, B. and SRIPERUMBUDUR, B. K. (2009). Kernel choice and classifiability for RKHS embeddings of probability distributions. In *Advances in Neural Information Processing Systems* (Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams and A. Culotta, eds.) **22**. Curran Associates, Red Hook.

GRETTON, A., BORGWARDT, K. M., RASCH, M., SCHÖLKOPF, B. and SMOLA, A. J. (2007). A kernel method for the two-sample-problem. In *Advances in Neural Information Processing Systems* 19: *Proceedings of the* 2006 *Conference* MIT Press, Cambridge. https://doi.org/10.7551/mitpress/7503.003.0069

GRETTON, A., BORGWARDT, K. M., RASCH, M. J., SCHÖLKOPF, B. and SMOLA, A. (2012). A kernel two-sample test. *J. Mach. Learn. Res.* **13** 723–773. MR2913716

GRETTON, A., BOUSQUET, O., SMOLA, A. and SCHÖLKOPF, B. (2005). Measuring statistical dependence with Hilbert-Schmidt norms. In *Algorithmic Learning Theory*. *Lecture Notes in Computer Science* **3734** 63–77. Springer, Berlin. MR2255909 https://doi.org/10.1007/11564089_7

GRETTON, A., FUKUMIZU, K., HARCHAOUI, Z. and SRIPERUMBUDUR, B. K. (2009). A fast, consistent kernel two-sample test. In *Advances in Neural Information Processing Systems* (Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams and A. Culotta, eds.) **22**. Curran Associates, Red Hook.

GRETTON, A., FUKUMIZU, K., TEO, C., SONG, L., SCHÖLKOPF, B. and SMOLA, A. (2008). A kernel statistical test of independence. In *Advances in Neural Information Processing Systems* (J. Platt, D. Koller, Y. Singer and S. Roweis, eds.) **20**. Curran Associates, Red Hook.

GUELLA, J. C. (2022). On Gaussian kernels on Hilbert spaces and kernels on hyperbolic spaces. *J. Approx. Theory* **279** 105765. MR4414825 https://doi.org/10.1016/j.jat.2022.105765

HALEY, R. W., QUADE, D., FREEMAN, H. E. and BENNETT, J. V. (1980). The SENIC project. Study on the efficacy of nosocomial infection control (SENIC project). Summary of study design. *Amer. J. Epidemiol.* **111** 472–485. https://doi.org/10.1093/oxfordjournals.aje.a112928

HENZE, N., HLÁVKA, Z. and MEINTANIS, S. G. (2014). Testing for spherical symmetry via the empirical characteristic function. *Statistics* **48** 1282–1296. MR3269735 https://doi.org/10.1080/02331888.2013.832764

HUFFER, F. W. and PARK, C. (2007). A test for elliptical symmetry. *J. Multivariate Anal.* **98** 256–281. MR2301752 https://doi.org/10.1016/j.jmva.2005.09.011

KARIYA, T. and EATON, M. L. (1977). Robust tests for spherical symmetry. *Ann. Statist.* **5** 206–215. MR0428590

KOKOSZKA, P. and REIMHERR, M. (2017). *Introduction to Functional Data Analysis*. *Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. MR3793167

KOLTCHINSKII, V. I. and LI, L. (1998). Testing for spherical symmetry of a multivariate distribution. *J. Multivariate Anal.* **65** 228–244. MR1625889 https://doi.org/10.1006/jmva.1998.1743

KUTNER, M., NACHTSHEIM, C., NETER, J. and LI, W. (2005). *Applied Linear Statistical Models*. *McGrwa-Hill International Edition*. McGraw-Hill, New York.

LI, B. (2018). *Sufficient Dimension Reduction*: *Methods and Applications with R*. *Chapman & Hall/CRC Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton.

LI, B. and DONG, Y. (2009). Dimension reduction for nonelliptically distributed predictors. *Ann. Statist.* **37** 1272–1298. MR2509074 https://doi.org/10.1214/08-AOS598

LI, B. and SOLEA, E. (2018). A nonparametric graphical model for functional data with application to brain networks based on fMRI. *J. Amer. Statist. Assoc.* **113** 1637–1655. MR3902235 https://doi.org/10.1080/01621459.2017.1356726

LI, K.-C. (1991). Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.* **86** 316–342. MR1137117

LI, K.-C. and DUAN, N. (1989). Regression analysis under link violation. *Ann. Statist.* **17** 1009–1052. MR1015136 https://doi.org/10.1214/aos/1176347254

LIANG, J., FANG, K.-T. and HICKERNELL, F. J. (2008). Some necessary uniform tests for spherical symmetry. *Ann. Inst. Statist. Math.* **60** 679–696. MR2434417 https://doi.org/10.1007/s10463-007-0121-9

LIU, H., HAN, F. and ZHANG, C.-H. (2012). Transelliptical graphical models. In *Advances in Neural Information Processing Systems* (F. Pereira, C. J. Burges, L. Bottou and K. Q. Weinberger, eds.) **25**. Curran Associates, Red Hook.

MANZOTTI, A., PÉREZ, F. J. and QUIROZ, A. J. (2002). A statistic for testing the null hypothesis of elliptical symmetry. *J. Multivariate Anal.* **81** 274–285. MR1906381 https://doi.org/10.1006/jmva.2001.2007

NARASIMHAN, B., JOHNSON, S. G., HAHN, T., BOUVIER, A. and KIÊU, K. (2023). Cubature: Adaptive multivariate integration over hypercubes. R package version 2.0.4.6.

PAINDAVEINE, D. (2012). Elliptical symmetry. *Encycl. EnvironMetrics* 802–807.

QIU, W. and JOE, H. (2020). ClusterGeneration: Random cluster generation (with specified degree of separation). R package version 1.3.7.

SCHMIDT, R. (2002). Tail dependence for elliptically contoured distributions. *Math. Methods Oper. Res.* **55** 301–327. https://doi.org/10.1007/s001860200191

SCHÖLKOPF, B., HERBRICH, R. and SMOLA, A. J. (2001). A generalized representer theorem. In *Computational Learning Theory* (D. Helmbold and B. Williamson, eds.) 416–426. Springer, Berlin, Heidelberg.

SCHOTT, J. R. (2002). Testing for elliptical symmetry in covariance-matrix-based analyses. *Statist. Probab. Lett.* **60** 395–404. MR1947179 https://doi.org/10.1016/S0167-7152(02)00306-1

SEJDINOVIC, D., SRIPERUMBUDUR, B., GRETTON, A. and FUKUMIZU, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.* **41** 2263–2291. MR3127866 https://doi.org/10.1214/13-AOS1140

SRIPERUMBUDUR, B., FUKUMIZU, K. and LANCKRIET, G. (2010). On the relation between universality, characteristic kernels and RKHS embedding of measures. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (Y. W. Teh and M. Titterington, eds.). *Proceedings of Machine Learning Research* **9** 773–780. PMLR, Chia Laguna Resort, Sardinia, Italy.

SRIPERUMBUDUR, B. K., FUKUMIZU, K. and LANCKRIET, G. R. G. (2011c). Universality, characteristic kernels and RKHS embedding of measures. *J. Mach. Learn. Res.* **12** 2389–2410. MR2825431

SRIPERUMBUDUR, B. K., GRETTON, A., FUKUMIZU, K., SCHÖLKOPF, B. and LANCKRIET, G. R. G. (2010). Hilbert space embeddings and metrics on probability measures. *J. Mach. Learn. Res.* **11** 1517–1561. MR2645460

SZABÓ, Z. and SRIPERUMBUDUR, B. K. (2017). Characteristic and universal tensor product kernels. *J. Mach. Learn. Res.* **18** 233. MR3845532

SZÉKELY, G. J. and RIZZO, M. L. (2009). Brownian distance covariance. *Ann. Appl. Stat.* **3** 1236–1265. MR2752127 https://doi.org/10.1214/09-AOAS312

SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. MR2382665 https://doi.org/10.1214/009053607000000505

TANG, Y. and LI, B. (2024). Supplement to "A nonparametric test for elliptical distribution based on kernel embedding of probabilities." https://doi.org/10.1214/24-AOS2438SUPP

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

VOGEL, D. and FRIED, R. (2011). Elliptical graphical modelling. *Biometrika* **98** 935–951. MR2860334 https://doi.org/10.1093/biomet/asr037

YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 49–67. MR2212574 https://doi.org/10.1111/j.1467-9868.2005.00532.x

ZHOU, D.-X. (2008). Derivative reproducing properties for kernel methods in learning theory. *J. Comput. Appl. Math.* **220** 456–463. MR2444183 https://doi.org/10.1016/j.cam.2007.08.023

The Annals of Statistics
2024, Vol. 52, No. 5, 2375–2399
https://doi.org/10.1214/24-AOS2439

SIMULTANEOUS STATISTICAL INFERENCE FOR SECOND ORDER PARAMETERS OF TIME SERIES UNDER WEAK CONDITIONS

BY YUNYI ZHANG[1,a], EFSTATHIOS PAPARODITIS[2,b] AND DIMITRIS N. POLITIS[3,c]

[1]*School of Data Science, The Chinese University of Hong Kong, Shenzhen,* [a]*zhangyunyi@cuhk.edu.cn*
[2]*Cyprus Academy of Sciences, Letters and Arts,* [b]*stathisp@ucy.ac.cy*
[3]*Department of Mathematics and Halicioglu Data Science Institute, University of California, San Diego,* [c]*dpolitis@ucsd.edu*

Strict stationarity is an assumption commonly used in time-series analysis in order to derive asymptotic distributional results for second-order statistics, like sample autocovariances and sample autocorrelations. Focusing on weak stationarity, this paper derives the asymptotic distribution of the maximum of sample autocovariances and sample autocorrelations under weak conditions by using Gaussian approximation techniques. The asymptotic theory for parameter estimators obtained by fitting a (linear) autoregressive model to a general weakly stationary time series is revisited and a Gaussian approximation theorem for the maximum of the estimators of the autoregressive coefficients is derived. To perform statistical inference for the aforementioned second-order parameters of interest, a bootstrap algorithm, the so-called second-order wild bootstrap is applied. Consistency of the bootstrap procedure is proven without imposing strict stationary conditions or structural process assumptions, like linearity. The good finite sample performance of the second-order wild bootstrap is demonstrated by means of simulations.

REFERENCES

AGRESTI, A. and KATERI, M. (2021). *Foundations of Statistics for Data Scientists*. CRC Press, Boca Raton.

BLOOMFIELD, P. (2000). *Fourier Analysis of Time Series*: *An Introduction*, 2dn ed. *Wiley Series in Probability and Statistics*: *Applied Probability and Statistics*. Wiley, New York. MRMR1884963 https://doi.org/10.1002/0471722235

BRAUMANN, A., KREISS, J.-P. and MEYER, M. (2021). Simultaneous inference for autocovariances based on autoregressive sieve bootstrap. *J. Time Series Anal*. **42** 534–553. MRMR4325664 https://doi.org/10.1111/jtsa.12604

BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series*: *Theory and Methods*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MRMR1093459 https://doi.org/10.1007/978-1-4419-0320-4

CHANG, J., CHEN, X. and WU, M. (2023). Central limit theorems for high dimensional dependent data. Available at arXiv:2104.12929.

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist*. **41** 2786–2819. MRMR3161448 https://doi.org/10.1214/13-AOS1161

DAHLHAUS, R., RICHTER, S. and WU, W. B. (2019). Towards a general theory for nonlinear locally stationary processes. *Bernoulli* **25** 1013–1044. MRMR3920364 https://doi.org/10.3150/17-bej1011

DAS, S. and POLITIS, D. N. (2021). Predictive inference for locally stationary time series with an application to climate data. *J. Amer. Statist. Assoc*. **116** 919–934. MRMR4270034 https://doi.org/10.1080/01621459.2019.1708368

DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory*: *An Introduction*. *Springer Series in Operations Research and Financial Engineering*. Springer, New York. MRMR2234156 https://doi.org/10.1007/0-387-34471-3

DETTE, H., WU, W. and ZHOU, Z. (2019). Change point analysis of correlation in non-stationary time series. *Statist. Sinica* **29** 611–643. MRMR3931381

MSC2020 subject classifications. 62M10, 62G09, 62G20.
Key words and phrases. Time series, weak stationarity, autoregressive model, second-order statistics, wild bootstrap.

DURBIN, J. and KOOPMAN, S. J. (2012). *Time Series Analysis by State Space Methods*. Oxford Univ. Press, London. https://doi.org/10.1093/acprof:oso/9780199641178.001.0001

FOLLAND, G. B. (1999). *Real Analysis*: *Modern Techniques and Their Applications*, 2nd ed. *Pure and Applied Mathematics* (*New York*). Wiley, New York. A Wiley-Interscience Publication. MRMR1681462

FRAGKESKOU, M. and PAPARODITIS, E. (2016). Inference for the fourth-order innovation cumulant in linear time series. *J. Time Series Anal.* **37** 240–266. MRMR3511584 https://doi.org/10.1111/jtsa.12160

FRAGKESKOU, M. and PAPARODITIS, E. (2018). Extending the range of validity of the autoregressive (sieve) bootstrap. *J. Time Series Anal.* **39** 356–379. MRMR3796524 https://doi.org/10.1111/jtsa.12275

GIACOMINI, R., POLITIS, D. N. and WHITE, H. (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory* **29** 567–589. MRMR3064050 https://doi.org/10.1017/S0266466612000655

GONÇALVES, S. and KILIAN, L. (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *J. Econometrics* **123** 89–120. MRMR2125439 https://doi.org/10.1016/j.jeconom.2003.10.030

HALL, P. and HUANG, L.-S. (2001). Nonparametric kernel regression subject to monotonicity constraints. *Ann. Statist.* **29** 624–647. MRMR1865334 https://doi.org/10.1214/aos/1009210683

HAYFIELD, T. and RACINE, J. S. (2008). Nonparametric econometrics: The np package. *J. Stat. Softw.* **27** 1–32. https://doi.org/10.18637/jss.v027.i05

KELEJIAN, H. H. and PRUCHA, I. R. (2007). HAC estimation in a spatial framework. *J. Econometrics* **140** 131–154. MRMR2395919 https://doi.org/10.1016/j.jeconom.2006.09.005

KIM, M. S. and SUN, Y. (2011). Spatial heteroskedasticity and autocorrelation consistent estimation of covariance matrix. *J. Econometrics* **160** 349–371. MRMR2748557 https://doi.org/10.1016/j.jeconom.2010.10.002

KREISS, J.-P. and PAPARODITIS, E. (2003). Autoregressive-aided periodogram bootstrap for time series. *Ann. Statist.* **31** 1923–1955. MRMR2036395 https://doi.org/10.1214/aos/1074290332

KREISS, J.-P. and PAPARODITIS, E. (2025). *Bootstrap for Time Series*: *Theory and Applications*. Springer, New York. To appear.

KREISS, J.-P. and PAPARODITIS, E. (2023). Bootstrapping Whittle estimators. *Biometrika* asac044. https://doi.org/10.1093/biomet/asac044

KREISS, J.-P., PAPARODITIS, E. and POLITIS, D. N. (2011). On the range of validity of the autoregressive sieve bootstrap. *Ann. Statist.* **39** 2103–2130. MRMR2893863 https://doi.org/10.1214/11-AOS900

KURISU, D., KATO, K. and SHAO, X. (2021). Gaussian approximation and spatially dependent wild bootstrap for high-dimensional spatial data. Available at arXiv:2103.10720.

LAI, T. L., LIU, H. and XING, H. (2005). Autoregressive models with piecewise constant volatility and regression parameters. *Statist. Sinica* **15** 279–301. MRMR2190207

LEE, J. and RAO, S. S. (2017). A note on general quadratic forms of nonstationary stochastic processes. *Statistics* **51** 949–968. MRMR3698495 https://doi.org/10.1080/02331888.2017.1318880

LJUNG, G. M. and BOX, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika* **65** 297–303.

MCELROY, T. S. and POLITIS, D. N. (2020). *Time Series*: *A First Course with a Bootstrap Starter*. CRC Press, Boca Raton.

MCMURRY, T. L. and POLITIS, D. N. (2004). Nonparametric regression with infinite order flat-top kernels. *J. Nonparametr. Stat.* **16** 549–562. MRMR2073041 https://doi.org/10.1080/10485250310001622596

MCMURRY, T. L. and POLITIS, D. N. (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *J. Time Series Anal.* **31** 471–482. MRMR2732601 https://doi.org/10.1111/j.1467-9892.2010.00679.x

MEYER, M., PAPARODITIS, E. and KREISS, J.-P. (2020). Extending the validity of frequency domain bootstrap methods to general stationary processes. *Ann. Statist.* **48** 2404–2427. MRMR4134800 https://doi.org/10.1214/19-AOS1892

MIES, F. and STELAND, A. (2022). Sequential Gaussian approximation for nonstationary time series in high dimensions. Available at arXiv:2203.03237.

MOSCONE, F. and TOSETTI, E. (2012). HAC estimation in spatial panels. *Econom. Lett.* **117** 60–65. MRMR2943479 https://doi.org/10.1016/j.econlet.2012.04.006

NARDI, Y. and RINALDO, A. (2011). Autoregressive process modeling via the Lasso procedure. *J. Multivariate Anal.* **102** 528–549. MRMR2755014 https://doi.org/10.1016/j.jmva.2010.10.012

NASON, G. (2013). A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 879–904. MRMR3124795 https://doi.org/10.1111/rssb.12015

PAPARODITIS, E. (2010). Validating stationarity assumptions in time series analysis by rolling local periodograms. *J. Amer. Statist. Assoc.* **105** 839–851. MRMR2724865 https://doi.org/10.1198/jasa.2010.tm08243

POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling*. *Springer Series in Statistics*. Springer, New York. MRMR1707286 https://doi.org/10.1007/978-1-4612-1554-7

POLITIS, D. N. and WHITE, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Rev.* **23** 53–70. MRMR2041534 https://doi.org/10.1081/ETC-120028836

PRIESTLEY, M. B. and SUBBA RAO, T. (1969). A test for non-stationarity of time-series. *J. Roy. Statist. Soc. Ser. B* **31** 140–149. MRMR0269062

PUCHSTEIN, R. and PREUSS, P. (2016). Testing for stationarity in multivariate locally stationary processes. *J. Time Series Anal.* **37** 3–29. MRMR3439530 https://doi.org/10.1111/jtsa.12133

SHAO, X. (2010). The dependent wild bootstrap. *J. Amer. Statist. Assoc.* **105** 218–235. With supplementary material available online. MRMR2656050 https://doi.org/10.1198/jasa.2009.tm08744

SHAO, X. and WU, W. B. (2007). Asymptotic spectral theory for nonlinear time series. *Ann. Statist.* **35** 1773–1801. MRMR2351105 https://doi.org/10.1214/009053606000001479

VOGT, M. (2012). Nonparametric regression for locally stationary time series. *Ann. Statist.* **40** 2601–2633. MRMR3097614 https://doi.org/10.1214/12-AOS1043

WHITE, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48** 817–838. MRMR575027 https://doi.org/10.2307/1912934

WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. MRMR2172215 https://doi.org/10.1073/pnas.0506715102

WU, W. B. and POURAHMADI, M. (2009). Banding sample autocovariance matrices of stationary processes. *Statist. Sinica* **19** 1755–1768. MRMR2589209

WU, W. B. and SHAO, X. (2007). A limit theorem for quadratic forms and its applications. *Econometric Theory* **23** 930–951. MRMR2396738 https://doi.org/10.1017/S0266466607070399

WU, W. B. and ZHOU, Z. (2011). Gaussian approximations for non-stationary multiple time series. *Statist. Sinica* **21** 1397–1413. MRMR2827528 https://doi.org/10.5705/ss.2008.223

XIAO, H. and WU, W. B. (2014). Portmanteau test and simultaneous inference for serial covariances. *Statist. Sinica* **24** 577–599. MRMR3235390

ZHANG, D. and WU, W. B. (2017). Gaussian approximation for high dimensional time series. *Ann. Statist.* **45** 1895–1919. MRMR3718156 https://doi.org/10.1214/16-AOS1512

ZHANG, D. and WU, W. B. (2018). Asymptotic theory for estimators of high-order statistics of stationary processes. *IEEE Trans. Inf. Theory* **64** 4907–4922. MRMR3819347 https://doi.org/10.1109/tit.2017.2764480

ZHANG, D. and WU, W. B. (2021). Convergence of covariance and spectral density estimates for high-dimensional locally stationary processes. *Ann. Statist.* **49** 233–254. MRMR4206676 https://doi.org/10.1214/20-AOS1954

ZHANG, T. and WU, W. B. (2015). Time-varying nonlinear regression models: Nonparametric estimation and model selection. *Ann. Statist.* **43** 741–768. MRMR3319142 https://doi.org/10.1214/14-AOS1299

ZHANG, X. and CHENG, G. (2018). Gaussian approximation for high dimensional vector under physical dependence. *Bernoulli* **24** 2640–2675. MRMR3779697 https://doi.org/10.3150/17-BEJ939

ZHANG, Y., PAPARODITIS, E. and POLITIS, D. N. (2024). Supplement to "Simultaneous Statistical Inference for Second Order Parameters of Time Series under Weak Conditions." https://doi.org/10.1214/24-AOS2439SUPP

ZHANG, Y. and POLITIS, D. N. (2022). Ridge regression revisited: Debiasing, thresholding and bootstrap. *Ann. Statist.* **50** 1401–1422. https://doi.org/10.1214/21-AOS2156

ZHANG, Y. and POLITIS, D. N. (2023). Debiased and thresholded ridge regression for linear models with heteroskedastic and correlated errors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **85** 327–355.

ZHOU, Z. (2014). Inference of weighted V-statistics for nonstationary time series and its applications. *Ann. Statist.* **42** 87–114. MRMR3161462 https://doi.org/10.1214/13-AOS1184

ZHOU, Z. (2015). Inference for non-stationary time series regression with or without inequality constraints. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 349–371. MRMR3310530 https://doi.org/10.1111/rssb.12077

The Annals of Statistics
2024, Vol. 52, No. 5, 2400–2430
https://doi.org/10.1214/24-AOS2440

MULTIVARIATE TREND FILTERING FOR LATTICE DATA

BY VEERANJANEYULU SADHANALA[1,a], YU-XIANG WANG[2,b], ADDISON J. HU[3,c] AND RYAN J. TIBSHIRANI[4,d]

[1]*Google,* [a]*veerus@google.com*
[2]*Department of Computer Science, University of California,* [b]*yuxiangw@ucsd.edu*
[3]*Department of Statistics, Carnegie Mellon University,* [c]*mail@huisaddison.com*
[4]*Department of Statistics, University of California,* [d]*ryantibs@berkeley.edu*

We study a multivariate version of trend filtering, called Kronecker trend filtering or KTF, for the case in which the design points form a lattice in d dimensions. KTF is a natural extension of univariate trend filtering (*Int. J. Comput. Vis.* **70** (2006) 214–255; *SIAM Rev.* **51** (2009) 339–360; *Ann. Statist.* **42** (2014) 285–323), and is defined by minimizing a penalized least squares problem whose penalty term sums the absolute (higher-order) differences of the parameter to be estimated along each of the coordinate directions. The corresponding penalty operator can be written in terms of Kronecker products of univariate trend filtering penalty operators, hence the name Kronecker trend filtering. Equivalently, one can view KTF in terms of an ℓ_1-penalized basis regression problem where the basis functions are tensor products of falling factorial functions, which is a piecewise polynomial (discrete spline) basis that underlies univariate trend filtering.

This paper is a unification and extension of the results in (In *Advances in Neural Information Processing Systems* (2016); in *Advances in Neural Information Processing Systems* (2017)). We develop a complete set of theoretical results that describe the behavior of kth-order Kronecker trend filtering in d dimensions, for every $k \geq 0$ and $d \geq 1$. This reveals a number of interesting phenomena, including the dominance of KTF over linear smoothers in estimating heterogeneously smooth functions, and a phase transition at $d = 2(k+1)$, a boundary past which (on the high dimension-to-smoothness side) linear smoothers fail to be consistent entirely. We also leverage recent results on discrete splines from (Tibshirani (2020)), in particular, discrete spline interpolation results that enable us to extend the KTF estimate to any off-lattice location in constant-time (independent of the size of the lattice n).

REFERENCES

ALMANSA, A., BALLESTER, C., CASELLES, V. and HARO, G. (2008). A TV based restoration model with local constraints. *J. Sci. Comput.* **34** 209–236. MR2377617 https://doi.org/10.1007/s10915-007-9160-x

BARBERO, Á. and SRA, S. (2018). Modular proximal optimization for multidimensional total-variation regularization. *J. Mach. Learn. Res.* **19** 56. MR3899758

BIBAUT, A. F. and VAN DER LAAN, M. J. (2019). Fast rates for empirical risk minimization over càdlàg functions with bounded sectional variation norm. Available at arXiv:1907.09244.

BIRGÉ, L. and MASSART, P. (2001). Gaussian model selection. *J. Eur. Math. Soc.* (*JEMS*) **3** 203–268. MR1848946 https://doi.org/10.1007/s100970100031

BOYD, S., PARIKH, N., CHU, E., PELEATO, B. and ECKSTEIN, J. (2011). Distributed optimization and statistical learning via the alternative direction method of multipliers. *Found. Trends Mach. Learn.* **3** 1–122.

CANDÈS, E. J. and GUO, F. (2002). New multiscale transforms, minimum total variation synthesis: Applications to edge-preserving image reconstruction. *Signal Process.* **82** 1519–1543.

CHAMBOLLE, A. (2004a). An algorithm for total variation minimization and applications. *J. Math. Imaging Vision* **20** 89–97.

MSC2020 subject classifications. 62G08, 62G20.
Key words and phrases. Trend filtering, nonparametric regression, minimax rates, local adaptivity.

CHAMBOLLE, A. (2005). Total variation minimization and a class of binary MRF models. In *Energy Minimization Methods in Computer Vision and Pattern Recognition* 136–152. Springer, Berlin.

CHAMBOLLE, A. and LIONS, P.-L. (1997). Image recovery via total variation minimization and related problems. *Numer. Math.* **76** 167–188. MR1440119 https://doi.org/10.1007/s002110050258

CHAN, T., MARQUINA, A. and MULET, P. (2000). High-order total variation-based image restoration. *SIAM J. Sci. Comput.* **22** 503–516. MR1780611 https://doi.org/10.1137/S1064827598344169

CHAN, T. F. and ESEDOḠLU, S. (2005). Aspects of total variation regularized L^1 function approximation. *SIAM J. Appl. Math.* **65** 1817–1837. MR2177726 https://doi.org/10.1137/040604297

CHATTERJEE, S. and GOSWAMI, S. (2021a). Adaptive estimation of multivariate piecewise polynomials and bounded variation functions by optimal decision trees. *Ann. Statist.* **49** 2531–2551. MR4338374 https://doi.org/10.1214/20-aos2045

CHATTERJEE, S. and GOSWAMI, S. (2021b). New risk bounds for 2D total variation denoising. *IEEE Trans. Inf. Theory* **67** 4060–4091. MR4289366 https://doi.org/10.1109/TIT.2021.3059657

CHUI, C. K. (1992). *An Introduction to Wavelets. Wavelet Analysis and Its Applications* **1**. Academic Press, Boston, MA. MR1150048

CHUI, C. K., STÖCKLER, J. and WARD, J. D. (1992). Compactly supported box-spline wavelets. *Approx. Theory Appl.* **8** 77–100. MR1195176

DALALYAN, A. S., HEBIRI, M. and LEDERER, J. (2017). On the prediction performance of the Lasso. *Bernoulli* **23** 552–581. MR3556784 https://doi.org/10.3150/15-BEJ756

DAUBECHIES, I. (1992). *Ten Lectures on Wavelets. CBMS-NSF Regional Conference Series in Applied Mathematics* **61**. SIAM, Philadelphia, PA. MR1162107 https://doi.org/10.1137/1.9781611970104

DEL ÁLAMO, M., LI, H. and MUNK, A. (2021). Frame-constrained total variation regularization for white noise regression. *Ann. Statist.* **49** 1318–1346. MR4298866 https://doi.org/10.1214/20-aos2001

DEVORE, R. A., KONYAGIN, S. V. and TEMLYAKOV, V. N. (1998). Hyperbolic wavelet approximation. *Constr. Approx.* **14** 1–26. MR1486387 https://doi.org/10.1007/s003659900060

DEVORE, R. A. and LORENTZ, G. G. (1993). *Constructive Approximation. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **303**. Springer, Berlin. MR1261635

DEVORE, R. A. and LUCIER, B. J. (1992). Wavelets. *Acta Numer.* **1** 1–56.

DONG, Y., HINTERMÜLLER, M. and RINCON-CAMACHO, M. M. (2011). Automated regularization parameter selection in multi-scale total variation models for image restoration. *J. Math. Imaging Vision* **40** 82–104. MR2782120 https://doi.org/10.1007/s10851-010-0248-9

DONOHO, D. L. (1997). CART and best-ortho-basis: A connection. *Ann. Statist.* **25** 1870–1911. MR1474073 https://doi.org/10.1214/aos/1069362377

DONOHO, D. L. and JOHNSTONE, I. M. (1998). Minimax estimation via wavelet shrinkage. *Ann. Statist.* **26** 879–921. MR1635414 https://doi.org/10.1214/aos/1024691081

EFRON, B. (1986). How biased is the apparent error rate of a prediction rule? *J. Amer. Statist. Assoc.* **81** 461–470. MR0845884

FANG, B., GUNTUBOYINA, A. and SEN, B. (2021). Multivariate extensions of isotonic regression and total variation denoising via entire monotonicity and Hardy-Krause variation. *Ann. Statist.* **49** 769–792. MR4255107 https://doi.org/10.1214/20-aos1977

GAVISH, M., NADLER, B. and COIFMAN, R. (2010). Multiscale wavelets on trees, graphs and high dimensional data: Theory and applications to semi supervised learning. In *Proceedings of the Annual Conference on Learning Theory*.

GÖBEL, F., BLANCHARD, G. and VON LUXBURG, U. (2018). Construction of tight frames on graphs and application to denoising. In *Handbook of Big Data Analytics. Springer Handb. Comput. Stat.* 503–522. Springer, Cham. MR3932016

GOLDENSHLUGER, A. and LEPSKI, O. (2008). Universal pointwise selection rule in multivariate function estimation. *Bernoulli* **14** 1150–1190. MR2543590 https://doi.org/10.3150/08-BEJ144

GOLDENSHLUGER, A. and LEPSKI, O. (2009). Structural adaptation via $\mathbb{L}_p$-norm oracle inequalities. *Probab. Theory Related Fields* **143** 41–71. MR2449122 https://doi.org/10.1007/s00440-007-0119-5

GOLDENSHLUGER, A. and LEPSKI, O. (2011). Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39** 1608–1632. MR2850214 https://doi.org/10.1214/11-AOS883

GOLDENSHLUGER, A. and LEPSKI, O. (2014). On adaptive minimax density estimation on R^d. *Probab. Theory Related Fields* **159** 479–543. MR3230001 https://doi.org/10.1007/s00440-013-0512-1

GOLDENSHLUGER, A. V. and LEPSKI, O. V. (2013). General selection rule from a family of linear estimators. *Theory Probab. Appl.* **57** 209–226. MR3201652 https://doi.org/10.1137/S0040585X97985923

GUNTUBOYINA, A., LIEU, D., CHATTERJEE, S. and SEN, B. (2020). Adaptive risk bounds in univariate total variation denoising and trend filtering. *Ann. Statist.* **48** 205–229. MR4065159 https://doi.org/10.1214/18-AOS1799

HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models*. *Monographs on Statistics and Applied Probability* **43**. CRC Press, London. MR1082147

HUTTER, J.-C. and RIGOLLET, P. (2016). Optimal rates for total variation denoising. In *Proceedings of the Annual Conference on Learning Theory*.

JOHNSTONE, I. M. (2015). *Gaussian Estimation*: *Sequence and Wavelet Models*. Cambridge Univ. Press, Cambridge. Draft version.

KERKYACHARIAN, G., LEPSKI, O. and PICARD, D. (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields* **121** 137–170. MR1863916 https://doi.org/10.1007/PL00008800

KERKYACHARIAN, G., LEPSKI, O. V. and PICARD, D. (2008). Nonlinear estimation in anisotropic multi-index denoising. Sparse case. *Theory Probab. Appl*. **52** 58–77.

KI, D., FANG, B. and GUNTUBOYINA, A. (2024). MARS via LASSO. *Ann*. *Statist*. **52** 1102–1126. MR4784071 https://doi.org/10.1214/24-aos2384

KIM, S.-J., KOH, K., BOYD, S. and GORINEVSKY, D. (2009). l_1 trend filtering. *SIAM Rev.* **51** 339–360. MR2505584 https://doi.org/10.1137/070690274

KOENKER, R., NG, P. and PORTNOY, S. (1994). Quantile smoothing splines. *Biometrika* **81** 673–680. MR1326417 https://doi.org/10.1093/biomet/81.4.673

KOROSTELEV, A. P. and TSYBAKOV, A. B. (2003). *Minimax Theory of Image Reconstructions*. Springer, Berlin.

LEPSKI, O. (2015). Adaptive estimation over anisotropic functional classes via oracle approach. *Ann*. *Statist*. **43** 1178–1242. MR3346701 https://doi.org/10.1214/14-AOS1306

LEPSKI, O. V., MAMMEN, E. and SPOKOINY, V. G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: An approach based on kernel estimates with variable bandwidth selectors. *Ann*. *Statist*. **25** 929–947. MR1447734 https://doi.org/10.1214/aos/1069362731

LEPSKI, O. V. and SPOKOINY, V. G. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Ann*. *Statist*. **25** 2512–2546. MR1604408 https://doi.org/10.1214/aos/1030741083

LEPSKII, O. V. (1991). On a problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl*. **35** 454–466.

LEPSKII, O. V. (1992). Asymptotically minimax adaptive estimation. I: Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl*. **36** 682–697.

LEPSKII, O. V. (1993). Asymptotically minimax adaptive estimation. II. Schemes without optimal adaptation: Adaptive estimators. *Theory Probab. Appl*. **37** 433–448.

LIN, K., SHARPNACK, J., RINALDO, A. and TIBSHIRANI, R. J. (2017). A sharp error analysis for the fused lasso, with application to approximate changepoint screening. In *Advances in Neural Information Processing Systems*.

LORENTZ, R. A. H. and MADYCH, W. R. (1992). Wavelets and generalized box splines. *Appl*. *Anal*. **44** 51–76. MR1284770 https://doi.org/10.1080/00036819208840068

MADRID PADILLA, O. H., SHARPNACK, J., CHEN, Y. and WITTEN, D. M. (2020). Adaptive nonparametric regression with the K-nearest neighbour fused lasso. *Biometrika* **107** 293–310. MR4108932 https://doi.org/10.1093/biomet/asz071

MADRID PADILLA, O. H., SHARPNACK, J., SCOTT, J. G. and TIBSHIRANI, R. J. (2017). The DFS fused lasso: Linear-time denoising over general graphs. *J*. *Mach*. *Learn*. *Res*. **18** 176. MR3827064

MALLAT, S. (2009). *A Wavelet Tour of Signal Processing*, 3rd ed. Elsevier/Academic Press, Amsterdam. MR2479996

MALLAT, S. G. (1989a). Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbf{R})$. *Trans*. *Amer. Math*. *Soc*. **315** 69–87. MR1008470 https://doi.org/10.2307/2001373

MALLAT, S. G. (1989b). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans*. *Pattern Anal*. *Mach*. *Intell*. **11** 674–693.

MAMMEN, E. and VAN DE GEER, S. (1997). Locally adaptive regression splines. *Ann*. *Statist*. **25** 387–413. MR1429931 https://doi.org/10.1214/aos/1034276635

MEYER, Y. (1987). Principe d'incertitude, bases hilbertiennes et algebres d'operateurs. *Sémin*. *Bourbaki* **145–146** 209–223.

MEYER, Y. (1990). *Ondelettes et Opérateurs*. Hermann, Paris.

MEYER, Y. and ROQUES, S. (1993). *Progress in Wavelet Analysis and Applications*. Atlantica Séguier Frontières.

NEMIROVSKIĬ, A. S., POLYAK, B. T. and TSYBAKOV, A. B. (1984). Signal processing by the nonparametric maximum likelihood method. *Problemy Peredachi Informatsii* **20** 29–46. MR0791733

NEMIROVSKIĬ, A. S., POLYAK, B. T. and TSYBAKOV, A. B. (1985). The rate of convergence of nonparametric estimates of maximum likelihood type. *Problemy Peredachi Informatsii* **21** 17–33. MR0820705

NEUMANN, M. H. (2000). Multivariate wavelet thresholding in anisotropic function spaces. *Statist*. *Sinica* **10** 399–431. MR1769750

NEUMANN, M. H. and VON SACHS, R. (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *Ann. Statist.* **25** 38–76. MR1429917 https://doi.org/10.1214/aos/1034276621

ORTELLI, F. and VAN DE GEER, S. (2021a). Prediction bounds for higher order total variation regularized least squares. *Ann. Statist.* **49** 2755–2773. MR4338382 https://doi.org/10.1214/21-aos2054

ORTELLI, F. and VAN DE GEER, S. (2021b). Tensor denoising with trend filtering. *Math. Stat. Learn.* **4** 87–142. MR4383732 https://doi.org/10.4171/msl/26

RAMDAS, A. and TIBSHIRANI, R. J. (2016). Fast and flexible ADMM algorithms for trend filtering. *J. Comput. Graph. Statist.* **25** 839–858. MR3533641 https://doi.org/10.1080/10618600.2015.1054033

RIEMENSCHNEIDER, S. D. and SHEN, Z. (1992). Wavelets and pre-wavelets in low dimensions. *J. Approx. Theory* **71** 18–38. MR1180872 https://doi.org/10.1016/0021-9045(92)90129-C

RUDIN, L. I. and OSHER, S. (1994). Total variation based image restoration with free local constraints. In *Proceedings of the International Conference on Image Processing* 31–35.

RUDIN, L. I., OSHER, S. and FATERNI, E. (1992). Nonlinear total variation based noise removal algorithms. *Phys. D, Nonlinear Phenom.* **60** 259–268.

SADHANALA, V. (2019). Nonparametric methods with total variation type regularization Ph.D. thesis, Machine Learning Department, Carnegie Mellon Univ.

SADHANALA, V., BASSETT, R., SHARPNACK, J. and MCDONALD, D. J. (2024). Exponential family trend filtering on lattices. *Electron. J. Stat.* **18** 1749–1814. MR4735240 https://doi.org/10.1214/24-ejs2241

SADHANALA, V. and TIBSHIRANI, R. J. (2019). Additive models with trend filtering. *Ann. Statist.* **47** 3032–3068. MR4025734 https://doi.org/10.1214/19-AOS1833

SADHANALA, V., WANG, Y.-X., HU, A. J. and TIBSHIRANI, R. J. (2024). Supplement to “Multivariate Trend Filtering for Lattice Data.” https://doi.org/10.1214/24-AOS2440SUPP

SADHANALA, V., WANG, Y.-X., SHARPNACK, J. and TIBSHIRANI, R. J. (2017). Higher-total variation classes on grids: Minimax theory and trend filtering methods. In *Advances in Neural Information Processing Systems*.

SADHANALA, V., WANG, Y.-X. and TIBSHIRANI, R. J. (2016). Total variation classes beyond 1d: Minimax rates, and the limitations of linear smoothers. In *Advances in Neural Information Processing Systems*.

SHARPNACK, J., SINGH, A. and KRISHNAMURTHY, A. (2013). Detecting activations over graphs using spanning tree wavelet bases. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*.

STEIDL, G., DIDAS, S. and NEUMANN, J. (2006). Splines in higher order TV regularization. *Int. J. Comput. Vis.* **70** 214–255.

STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. MR0630098

TANSEY, W. and SCOTT, J. (2015). A fast and flexible algorithm for the graph-fused lasso. Available at arXiv:1505.06475.

TIBSHIRANI, R. J. (2014). Adaptive piecewise polynomial estimation via trend filtering. *Ann. Statist.* **42** 285–323. MR3189487 https://doi.org/10.1214/13-AOS1189

TIBSHIRANI, R. J. (2020). Divided differences, falling factorials, and discrete splines: Another look at trend filtering and related problems. Available at arXiv:2003.03886.

TIBSHIRANI, R. J. and TAYLOR, J. (2011). The solution path of the generalized lasso. *Ann. Statist.* **39** 1335–1371. MR2850205 https://doi.org/10.1214/11-AOS878

TIBSHIRANI, R. J. and TAYLOR, J. (2012). Degrees of freedom in lasso problems. *Ann. Statist.* **40** 1198–1232. MR2985948 https://doi.org/10.1214/12-AOS1003

TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, Berlin.

VOGEL, C. R. and OMAN, M. E. (1996). Iterative methods for total variation denoising. *SIAM J. Sci. Comput.* **17** 227–238.

WANG, Y.-X., SHARPNACK, J., SMOLA, A. J. and TIBSHIRANI, R. J. (2016). Trend filtering on graphs. *J. Mach. Learn. Res.* **17** 105. MR3543511

WANG, Y.-X., SMOLA, A. and TIBSHIRANI, R. J. (2014). The falling factorial basis and its statistical applications. In *Proceedings of the International Conference on Machine Learning*.

YE, S. S. and MADRID PADILLA, O. H. (2021). Non-parametric quantile regression via the K-NN fused lasso. *J. Mach. Learn. Res.* **22** 111. MR4279762

The Annals of Statistics
2024, Vol. 52, No. 5, 2431–2455
https://doi.org/10.1214/24-AOS2441

COMPUTATIONAL AND STATISTICAL THRESHOLDS IN MULTI-LAYER STOCHASTIC BLOCK MODELS

BY JING LEI[1,a], ANRU R. ZHANG[2,b] AND ZIHAN ZHU[3,c]

[1]*Department of Statistics & Data Science, Carnegie Mellon University,* [a]*jinglei@andrew.cmu.edu*
[2]*Department of Biostatistics & Bioinformatics, Duke University,* [b]*anru.zhang@duke.edu*
[3]*Department of Statistics and Data Science, The Wharton School, University of Pennsylvania,* [c]*zhzhu1@wharton.upenn.edu*

We study the problem of community recovery and detection in multi-layer stochastic block models, focusing on the critical network density threshold for consistent community structure inference. Using a prototypical two-block model, we reveal a computational barrier for such multilayer stochastic block models that does not exist for its single-layer counterpart: When there are no computational constraints, the density threshold depends linearly on the number of layers. However, when restricted to polynomial-time algorithms, the density threshold scales with the square root of the number of layers, assuming correctness of a low-degree polynomial hardness conjecture. Our results provide a nearly complete picture of the optimal inference in multiple-layer stochastic block models and partially settle the open question in (*J. Amer. Statist. Assoc.* **118** (2023) 2433–2445) regarding the optimality of the bias-adjusted spectral method.

REFERENCES

[1] ABBE, E. (2017). Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** Paper No. 177, 86 pp. MR3827065

[2] ABBE, E., BANDEIRA, A. S. and HALL, G. (2016). Exact recovery in the stochastic block model. *IEEE Trans. Inf. Theory* **62** 471–487. MR3447993 https://doi.org/10.1109/TIT.2015.2490670

[3] ABBE, E. and SANDON, C. (2015). Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In 2015 *IEEE* 56*th Annual Symposium on Foundations of Computer Science—FOCS* 2015 670–688. IEEE Computer Soc., Los Alamitos, CA. MR3473334 https://doi.org/10.1109/FOCS.2015.47

[4] AUDDY, A. and YUAN, M. (2023). Large dimensional independent component analysis: Statistical optimality and computational tractability. Preprint. Available at arXiv:2303.18156.

[5] BANDEIRA, A. S., EL ALAOUI, A., HOPKINS, S., SCHRAMM, T., WEIN, A. S. and ZADIK, I. (2022). The Franz–Parisi criterion and computational trade-offs in high dimensional statistics. *Adv. Neural Inf. Process. Syst.* **35** 33831–33844.

[6] BANDEIRA, A. S., KUNISKY, D. and WEIN, A. S. (2020). Computational hardness of certifying bounds on constrained PCA problems. In 11*th Innovations in Theoretical Computer Science Conference*. *LIPIcs. Leibniz Int. Proc. Inform.* **151** Art. No. 78, 29 pp. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR4048181

[7] BARAK, B., HOPKINS, S., KELNER, J., KOTHARI, P. K., MOITRA, A. and POTECHIN, A. (2019). A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM J. Comput.* **48** 687–735. MR3945259 https://doi.org/10.1137/17M1138236

[8] BERTHET, Q. and RIGOLLET, P. (2013). Optimal detection of sparse principal components in high dimension. *Ann. Statist.* **41** 1780–1815. MR3127849 https://doi.org/10.1214/13-AOS1127

[9] BRENNAN, M. and BRESLER, G. (2019). Optimal average-case reductions to sparse pca: From weak assumptions to strong hardness. In *Conference on Learning Theory* 469–470. PMLR.

[10] BRENNAN, M. S., BRESLER, G., HOPKINS, S., LI, J. and SCHRAMM, T. (2021). Statistical query algorithms and low degree tests are almost equivalent. In *Conference on Learning Theory* 774–774. PMLR.

MSC2020 subject classifications. Primary 62H99; secondary 05C82.
Key words and phrases. Community detection and recovery, computational barrier, low-degree polynomial hardness, multilayer network.

[11] BRESLER, G. and HUANG, B. (2022). The algorithmic phase transition of random k-SAT for low degree polynomials. In 2021 *IEEE* 62*nd Annual Symposium on Foundations of Computer Science—FOCS* 2021 298–309. IEEE Computer Soc., Los Alamitos, CA. MR4399691

[12] CHEN, S., LIU, S. and MA, Z. (2022). Global and individualized community detection in inhomogeneous multilayer networks. *Ann. Statist.* **50** 2664–2693. MR4500621 https://doi.org/10.1214/22-aos2202

[13] DAVIS, D., DÍAZ, M. and WANG, K. (2021). Clustering a mixture of Gaussians with unknown covariance. Preprint. Available at arXiv:2110.01602.

[14] DIAKONIKOLAS, I., KANE, D. M., LUO, Y. and ZHANG, A. (2023). Statistical and computational limits for tensor-on-tensor association detection. In *The Thirty Sixth Annual Conference on Learning Theory* 5260–5310. PMLR.

[15] DING, Y., KUNISKY, D., WEIN, A. S. and BANDEIRA, A. S. (2021). The average-case time complexity of certifying the restricted isometry property. *IEEE Trans. Inf. Theory* **67** 7355–7361. MR4345126 https://doi.org/10.1109/TIT.2021.3112823

[16] DING, Y., KUNISKY, D., WEIN, A. S. and BANDEIRA, A. S. (2024). Subexponential-time algorithms for sparse PCA. *Found. Comput. Math.* **24** 865–914. MR4760356 https://doi.org/10.1007/s10208-023-09603-0

[17] DONG, X., FROSSARD, P., VANDERGHEYNST, P. and NEFEDOV, N. (2012). Clustering with multi-layer graphs: A spectral perspective. *IEEE Trans. Signal Process.* **60** 5820–5831. MR2990287 https://doi.org/10.1109/TSP.2012.2212886

[18] GAMARNIK, D., JAGANNATH, A. and WEIN, A. S. (2020). Low-degree hardness of random optimization problems. In 2020 *IEEE* 61*st Annual Symposium on Foundations of Computer Science* 131–140. IEEE Computer Soc., Los Alamitos, CA. MR4232029 https://doi.org/10.1109/FOCS46700.2020.00021

[19] GAO, C., MA, Z., ZHANG, A. Y. and ZHOU, H. H. (2017). Achieving optimal misclassification proportion in stochastic block models. *J. Mach. Learn. Res.* **18** Paper No. 60, 45 pp. MR3687603

[20] GOLDENBERG, A., ZHENG, A. X., FIENBERG, S. E. and AIROLDI, E. M. (2010). A survey of statistical network models. *Found. Trends Mach. Learn.* **2** 129–233.

[21] HAN, Q., XU, K. and AIROLDI, E. (2015). Consistent estimation of dynamic and multi-layer block models. In *International Conference on Machine Learning* 1511–1520.

[22] HAN, R., LUO, Y., WANG, M. and ZHANG, A. R. (2022). Exact clustering in tensor block model: Statistical optimality and computational limit. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 1666–1698. MR4515554

[23] HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. MR0718088 https://doi.org/10.1016/0378-8733(83)90021-7

[24] HOPKINS, S. (2018). Statistical inference and the sum of squares method. Ph.D. thesis, Cornell Univ. MR3864930

[25] HOPKINS, S. B., KOTHARI, P. K., POTECHIN, A., RAGHAVENDRA, P., SCHRAMM, T. and STEURER, D. (2017). The power of sum-of-squares for detecting hidden structures. In 58*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 720–731. IEEE Computer Soc., Los Alamitos, CA. MR3734275 https://doi.org/10.1109/FOCS.2017.72

[26] HOPKINS, S. B. and STEURER, D. (2017). Efficient Bayesian estimation from few samples: Community detection and related problems. In 58*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 379–390. IEEE Computer Soc., Los Alamitos, CA. MR3734245 https://doi.org/10.1109/FOCS.2017.42

[27] IZENMAN, A. J. (2023). *Network Models for Data Science*. Cambridge Univ. Press, Cambridge.

[28] KIVELÄ, M., ARENAS, A., BARTHELEMY, M., GLEESON, J. P., MORENO, Y. and PORTER, M. A. (2014). Multilayer networks. *J. Complex Netw.* **2** 203–271.

[29] KOLACZYK, E. D. (2009). *Statistical Analysis of Network Data: Methods and Models. Springer Series in Statistics*. Springer, New York. MR2724362 https://doi.org/10.1007/978-0-387-88146-1

[30] KUNISKY, D., WEIN, A. S. and BANDEIRA, A. S. (2022). Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *Mathematical Analysis, Its Applications and Computation. Springer Proc. Math. Stat.* **385** 1–50. Springer, Cham. MR4461037 https://doi.org/10.1007/978-3-030-97127-4_1

[31] LE CAM, L. (2012). *Asymptotic Methods in Statistical Decision Theory. Springer Series in Statistics*. Springer, New York. MR0856411 https://doi.org/10.1007/978-1-4612-4946-7

[32] LEI, J., CHEN, K. and LYNCH, B. (2020). Consistent community detection in multi-layer network data. *Biometrika* **107** 61–73. MR4064140 https://doi.org/10.1093/biomet/asz068

[33] LEI, J. and LIN, K. Z. (2023). Bias-adjusted spectral clustering in multi-layer stochastic block models. *J. Amer. Statist. Assoc.* **118** 2433–2445. MR4681594 https://doi.org/10.1080/01621459.2022.2054817

[34] LEI, J. and RINALDO, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann. Statist.* **43** 215–237. MR3285605 https://doi.org/10.1214/14-AOS1274

[35] LEVIN, K., LODHIA, A. and LEVINA, E. (2022). Recovering shared structure from multiple networks with unknown edge distributions. *J. Mach. Learn. Res.* **23** Paper No. 3, 48 pp. MR4420728
[36] LIU, F., CHOI, D., XIE, L. and ROEDER, K. (2018). Global spectral clustering in dynamic networks. *Proc. Natl. Acad. Sci. USA* **115** 927–932. MR3763702 https://doi.org/10.1073/pnas.1718449115
[37] LÖFFLER, M., WEIN, A. S. and BANDEIRA, A. S. (2022). Computationally efficient sparse clustering. *Inf. Inference* **11** 1255–1286. MR4526323 https://doi.org/10.1093/imaiai/iaac019
[38] LUO, Y. and GAO, C. (2023). Computational lower bounds for graphon estimation via low-degree polynomials. Preprint. Available at arXiv:2308.15728.
[39] LUO, Y. and ZHANG, A. R. (2022). Tensor clustering with planted structures: Statistical optimality and computational limits. *Ann. Statist.* **50** 584–613. MR4382029 https://doi.org/10.1214/21-aos2123
[40] LYU, Z. and XIA, D. (2023). Optimal estimation and computational limit of low-rank Gaussian mixtures. *Ann. Statist.* **51** 646–667. MR4600996 https://doi.org/10.1214/23-aos2264
[41] MA, Z. and WU, Y. (2015). Computational barriers in minimax submatrix detection. *Ann. Statist.* **43** 1089–1116. MR3346698 https://doi.org/10.1214/14-AOS1300
[42] MAO, C. and WEIN, A. S. (2021). Optimal spectral recovery of a planted vector in a subspace. Preprint. Available at arXiv:2105.15081.
[43] MAO, C., WEIN, A. S. and ZHANG, S. (2023). Detection-recovery gap for planted dense cycles. In *The Thirty Sixth Annual Conference on Learning Theory* 2440–2481. PMLR.
[44] MATIAS, C. and MIELE, V. (2017). Statistical clustering of temporal networks through a dynamic stochastic block model. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1119–1141. MR3689311 https://doi.org/10.1111/rssb.12200
[45] MONDELLI, M. and MONTANARI, A. (2019). On the connection between learning two-layer neural networks and tensor decomposition. In *The 22nd International Conference on Artificial Intelligence and Statistics* 1051–1060. PMLR.
[46] MOSSEL, E., NEEMAN, J. and SLY, A. (2012). Stochastic block models and reconstruction. Preprint. Available at arXiv:1202.1499.
[47] NEWMAN, M. E. J. (2010). *Networks: An Introduction*. Oxford Univ. Press, Oxford. MR2676073 https://doi.org/10.1093/acprof:oso/9780199206650.001.0001
[48] PAUL, S. and CHEN, Y. (2020). A random effects stochastic block model for joint community detection in multiple networks with applications to neuroimaging. *Ann. Appl. Stat.* **14** 993–1029. MR4117838 https://doi.org/10.1214/20-AOAS1339
[49] PAUL, S. and CHEN, Y. (2020). Spectral and matrix factorization methods for consistent community detection in multi-layer networks. *Ann. Statist.* **48** 230–250. MR4065160 https://doi.org/10.1214/18-AOS1800
[50] PENSKY, M. (2019). Dynamic network models and graphon estimation. *Ann. Statist.* **47** 2378–2403. MR3953455 https://doi.org/10.1214/18-AOS1751
[51] RAGHAVENDRA, P., SCHRAMM, T. and STEURER, D. (2018). High dimensional estimation via sum-of-squares proofs. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro* 2018. *Vol. IV. Invited Lectures* 3389–3423. World Sci. Publ., Hackensack, NJ. MR3966537
[52] SCHRAMM, T. and WEIN, A. S. (2022). Computational barriers to estimation from low-degree polynomials. *Ann. Statist.* **50** 1833–1858. MR4441142 https://doi.org/10.1214/22-aos2179
[53] SKALA, M. (2013). Hypergeometric tail inequalities: Ending the insanity. Preprint. Available at arXiv:1311.5939.
[54] TAN, N. and VENKATARAMANAN, R. (2023). Mixed regression via approximate message passing. *J. Mach. Learn. Res.* **24** Paper No. [317], 44 pp. MR4664754
[55] TANG, W., LU, Z. and DHILLON, I. S. (2009). Clustering with multiple graphs. In *International Conference on Data Mining* (*ICDM*) 1016–1021. IEEE, New York.
[56] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. MR2724359 https://doi.org/10.1007/b13794
[57] XU, K. S. and HERO, A. O. (2014). Dynamic stochastic blockmodels for time-evolving social networks. *IEEE J. Sel. Top. Signal Process.* **8** 552–562.
[58] ZHANG, A. and XIA, D. (2018). Tensor SVD: Statistical and computational limits. *IEEE Trans. Inf. Theory* **64** 7311–7338. MR3876445 https://doi.org/10.1109/TIT.2018.2841377
[59] ZHANG, A. Y. and ZHOU, H. H. (2016). Minimax rates of community detection in stochastic block models. *Ann. Statist.* **44** 2252–2280. MR3546450 https://doi.org/10.1214/15-AOS1428
[60] ZHANG, J. and CAO, J. (2017). Finding common modules in a time-varying network with application to the *Drosophila melanogaster* gene regulation network. *J. Amer. Statist. Assoc.* **112** 994–1008. MR3735355 https://doi.org/10.1080/01621459.2016.1260465

The Annals of Statistics
2024, Vol. 52, No. 5, 2456–2481
https://doi.org/10.1214/24-AOS2443

A GAUSSIAN PROCESS APPROACH TO MODEL CHECKS

BY JUAN CARLOS ESCANCIANO[a]

Department of Economics, Universidad Carlos III de Madrid, [a]*jescanci@eco.uc3m.es*

This paper proposes a Gaussian process (GP) approach for testing conditional moment restrictions. Tests are based on squared Neyman orthogonal function-parametric processes integrated with respect to a GP distribution. This methodology leads to a general unified framework of kernel-based tests having the following properties: (i) bootstrap tests are easy to implement in the presence of nuisance parameters (they are simple quadratic forms, and there is no need to reestimate the nuisance parameters in each bootstrap replication); and (ii) the new tests are valid under general conditions, including higher-order conditional moments of unknown form, regularized estimators (e.g., Lasso) or parameters at the boundary of the parameter space. Novel applications include distance kernel tests for zero conditional treatment effects. The paper introduces Neyman orthogonal kernels, a new asymptotic theory and a detailed local power analysis. Monte Carlo experiments and a real data application illustrate the sensitivity of tests to the dimension of covariates and to the mean and covariance kernel of the GP.

REFERENCES

[1] BARINGHAUS, L. and FRANZ, C. (2004). On a new multivariate two-sample test. *J. Multivariate Anal.* **88** 190–206. MR2021870 https://doi.org/10.1016/S0047-259X(03)00079-4

[2] BICKEL, P. J., RITOV, Y. and STOKER, T. M. (2006). Tailor-made tests for goodness of fit to semiparametric hypotheses. *Ann. Statist.* **34** 721–741. MR2281882 https://doi.org/10.1214/009053606000000137

[3] BIERENS, H. J. (1982). Consistent model specification tests. *J. Econometrics* **20** 105–134. MR0685673 https://doi.org/10.1016/0304-4076(82)90105-1

[4] BIERENS, H. J. and PLOBERGER, W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica* **65** 1129–1151. MR1475077 https://doi.org/10.2307/2171881

[5] BILLINGSLEY, P. (2013). *Convergence of Probability Measures*. Wiley, New York. MR0233396

[6] CHOI, S., HALL, W. J. and SCHICK, A. (1996). Asymptotically uniformly most powerful tests in parametric and semiparametric models. *Ann. Statist.* **24** 841–861. MR1394992 https://doi.org/10.1214/aos/1032894469

[7] CRUMP, R. K., HOTZ, V. J., IMBENS, G. W. and MITNIK, O. A. (2008). Nonparametric tests for treatment effect heterogeneity. *Rev. Econ. Stat.* **90** 389–405.

[8] DELGADO, M. A., DOMINGUEZ, M. A. and LAVERGNE, P. (2006). Consistent tests of conditional moment restrictions". *Ann. Écon. Stat.* **81** 33–67.

[9] DUDLEY, R. M. (1999). *Uniform Central Limit Theorems*. *Cambridge Studies in Advanced Mathematics* **63**. Cambridge Univ. Press, Cambridge. MR1720712 https://doi.org/10.1017/CBO9780511665622

[10] ESCANCIANO, J. C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory* **22** 1030–1051. MR2328527 https://doi.org/10.1017/S0266466606060506

[11] ESCANCIANO, J. C. (2009). On the lack of power of omnibus specification tests. *Econometric Theory* **25** 162–194. MR2472049 https://doi.org/10.1017/S0266466608090051

[12] ESCANCIANO, J. C. (2009). Simple bootstrap tests for conditional moment restrictions. Mimeo. Available at http://www.ecares.org/ecaresdocuments/seminars0809/escanciano.pdf.

[13] ESCANCIANO, J. C. and GOH, S. C. (2014). Specification analysis of linear quantile models. *J. Econometrics* **178** 495–507. MR3132447 https://doi.org/10.1016/j.jeconom.2013.07.006

[14] ESCANCIANO, J. C., SANT'ANNA, P. H. and SONG, X. (2023). Specification tests for generalized propensity scores using double projections. Preprint. Available at arXiv:2003.13803.

[15] ESCANCIANO, J. C. and VELASCO, C. (2006). Generalized spectral tests for the martingale difference hypothesis. *J. Econometrics* **134** 151–185. MR2328319 https://doi.org/10.1016/j.jeconom.2005.06.019

MSC2020 subject classifications. Primary 62E20, 62G10; secondary 62A01.
Key words and phrases. Conditional moment restrictions, Gaussian processes, kernels, goodness-of-fit, Lasso.

[16] FAN, Y. and LI, Q. (2000). Consistent model specification tests: Kernel-based tests versus Bierens' ICM tests. *Econometric Theory* **16** 1016–1041. MR1803718 https://doi.org/10.1017/S0266466600166083

[17] FERNÁNDEZ, T. and RIVERA, N. (2024). A general framework for the analysis of kernel-based tests. *J. Mach. Learn. Res.* **25** Paper No. [95], 40 pp. MR4749131

[18] FERREIRA, J. C. and MENEGATTO, V. A. (2013). Positive definiteness, reproducing kernel Hilbert spaces and beyond. *Ann. Funct. Anal.* **4** 64–88. MR3004212 https://doi.org/10.15352/afa/1399899838

[19] FINKELSTEIN, A., TAUBMAN, S., WRIGHT, B., BERNSTEIN, M., GRUBER, J., NEWHOUSE, J. P., ALLEN, H., BAICKER, K. and OREGON HEALTH STUDY GROUP (2012). The Oregon health insurance experiment: Evidence from the first year. *Q. J. Econ.* **127** 1057–1106.

[20] GINÉ, E. and ZINN, J. (1990). Bootstrapping general empirical measures. *Ann. Probab.* **18** 851–869. MR1055437

[21] GONZÁLEZ-MANTEIGA, W. and CRUJEIRAS, R. M. (2013). An updated review of goodness-of-fit tests for regression models. *TEST* **22** 361–411. MR3093195 https://doi.org/10.1007/s11749-013-0327-5

[22] GRETTON, A., BORGWARDT, K. M., RASCH, M. J., SCHÖLKOPF, B. and SMOLA, A. (2012). A kernel two-sample test. *J. Mach. Learn. Res.* **13** 723–773. MR2913716

[23] HÄRDLE, W. and MAMMEN, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21** 1926–1947. MR1245774 https://doi.org/10.1214/aos/1176349403

[24] HEISS, F., HETZENECKER, S. and OSTERHAUS, M. (2022). Nonparametric estimation of the random coefficients model: An elastic net approach. *J. Econometrics* **229** 299–321. MR4429567 https://doi.org/10.1016/j.jeconom.2020.11.010

[25] HOFMANN, T., SCHÖLKOPF, B. and SMOLA, A. J. (2008). Kernel methods in machine learning. *Ann. Statist.* **36** 1171–1220. MR2418654 https://doi.org/10.1214/009053607000000677

[26] IMBENS, G. W. and RUBIN, D. B. (2015). *Causal Inference—For Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge Univ. Press, New York. MR3309951 https://doi.org/10.1017/CBO9781139025751

[27] JANSSEN, A. (2000). Global power functions of goodness of fit tests. *Ann. Statist.* **28** 239–253. MR1762910 https://doi.org/10.1214/aos/1016120371

[28] JENNRICH, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Stat.* **40** 633–643. MR0238419 https://doi.org/10.1214/aoms/1177697731

[29] KANAGAWA, M., HENNIG, P., SEJDINOVIC, D. and SRIPERUMBUDUR, B. K. (2018). Gaussian processes and kernel methods: a review on connections and equivalences. Preprint. Available at arXiv:1807.02582.

[30] KHMALADZE, È. V. (1981). A martingale approach in the theory of goodness-of-fit tests. *Teor. Veroyatn. Primen.* **26** 246–265. MR0616619

[31] KIM, I., BALAKRISHNAN, S. and WASSERMAN, L. (2020). Robust multivariate nonparametric tests via projection averaging. *Ann. Statist.* **48** 3417–3441. MR4185814 https://doi.org/10.1214/19-AOS1936

[32] KNIGHT, K. and FU, W. (2000). Asymptotics for lasso-type estimators. *Ann. Statist.* **28** 1356–1378. MR1805787 https://doi.org/10.1214/aos/1015957397

[33] LINDSAY, B. G., MARKATOU, M. and RAY, S. (2014). Kernels, degrees of freedom, and power properties of quadratic distance goodness-of-fit tests. *J. Amer. Statist. Assoc.* **109** 395–410. MR3180572 https://doi.org/10.1080/01621459.2013.836972

[34] LINDSAY, B. G., MARKATOU, M., RAY, S., YANG, K. and CHEN, S.-C. (2008). Quadratic distances on probabilities: A unified foundation. *Ann. Statist.* **36** 983–1006. MR2396822 https://doi.org/10.1214/009053607000000956

[35] MAMMEN, E. (1993). Bootstrap and wild bootstrap for high-dimensional linear models. *Ann. Statist.* **21** 255–285. MR1212176 https://doi.org/10.1214/aos/1176349025

[36] MUANDET, K., FUKUMIZU, K., SRIPERUMBUDUR, B. and SCHÖLKOPF, B. (2017). Kernel mean embedding of distributions: A review and beyond. *Found. Trends Mach. Learn.* **10** 1–141.

[37] MUANDET, K., JITKRITTUM, W. and KÜBLER, J. (2020). Kernel conditional moment test via maximum moment restriction. In *Conference on Uncertainty in Artificial Intelligence* 41–50.

[38] NEYMAN, J. (1959). Optimal asymptotic tests of composite statistical hypotheses. In *Probability and Statistics: The Harald Cramér Volume* (U. Grenander, ed.) 213–234. Almqvist & Wiksell, Stockholm. MR0112201

[39] OKUTMUSTUR, B. and GHEONDEA, A. (2010). *Reproducing Kernel Hilbert Spaces. The Basics Bergman Spaces and Interpolation Problems*. LAP LAMBERT Academic Publishing.

[40] ROSENBAUM, P. R. and RUBIN, D. B. (1983). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **45** 212–218.

[41] SANCETTA, A. (2021). Estimation in reproducing Kernel Hilbert spaces with dependent data. *IEEE Trans. Inf. Theory* **67** 1782–1795. MR4282326 https://doi.org/10.1109/tit.2020.3045290

[42] SANCETTA, A. (2022). Testing subspace restrictions in the presence of high dimensional nuisance parameters. *Electron. J. Stat.* **16** 5277–5320. MR4492990 https://doi.org/10.1214/22-ejs2058

[43] SEJDINOVIC, D., SRIPERUMBUDUR, B., GRETTON, A. and FUKUMIZU, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.* **41** 2263–2291. MR3127866 https://doi.org/10.1214/13-AOS1140

[44] SHAO, X. and ZHANG, J. (2014). Martingale difference correlation and its use in high-dimensional variable screening. *J. Amer. Statist. Assoc.* **109** 1302–1318. MR3265698 https://doi.org/10.1080/01621459.2014.887012

[45] SRIPERUMBUDUR, B. K., FUKUMIZU, K. and LANCKRIET, G. R. G. (2011). Universality, characteristic kernels and RKHS embedding of measures. *J. Mach. Learn. Res.* **12** 2389–2410. MR2825431

[46] STEINWART, I. (2019). Convergence types and rates in generic Karhunen-Loève expansions with applications to sample path properties. *Potential Anal.* **51** 361–395. MR4023468 https://doi.org/10.1007/s11118-018-9715-5

[47] STINCHCOMBE, M. B. and WHITE, H. (1998). Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* **14** 295–325. MR1628586 https://doi.org/10.1017/S0266466698143013

[48] STRASSER, H. (1990). Global extrapolation of local efficiency. *Statist. Decisions* **8** 11–26. MR1050041

[49] STUTE, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25** 613–641. MR1439316 https://doi.org/10.1214/aos/1031833666

[50] STUTE, W., GONZÁLEZ MANTEIGA, W. and PRESEDO QUINDIMIL, M. (1998). Bootstrap approximations in model checks for regression. *J. Amer. Statist. Assoc.* **93** 141–149. MR1614600 https://doi.org/10.2307/2669611

[51] SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. MR2382665 https://doi.org/10.1214/009053607000000505

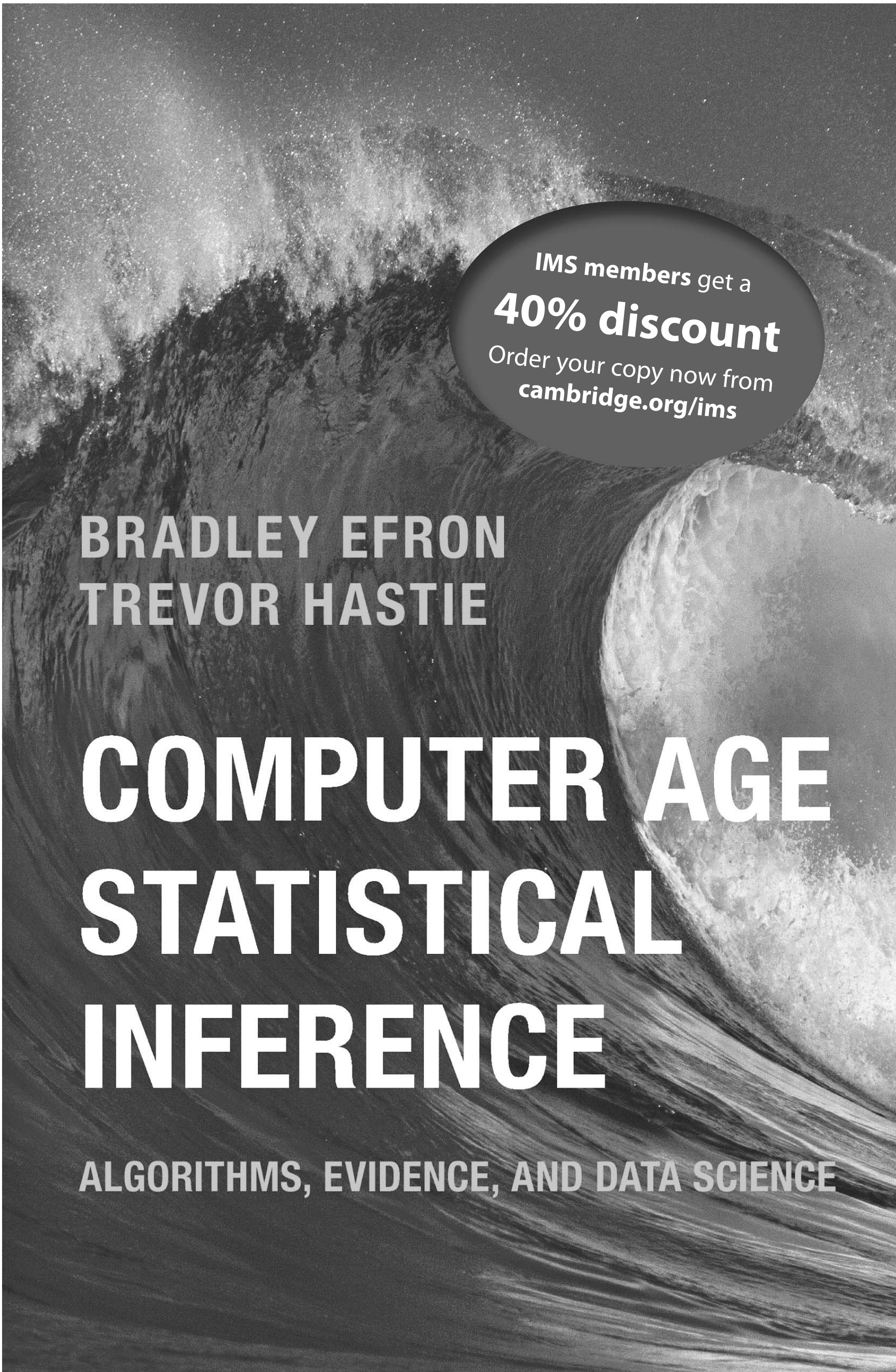
[52] TAN, F. and ZHU, L. (2022). Integrated conditional moment test and beyond: When the number of covariates is divergent. *Biometrika* **109** 103–122. MR4374643 https://doi.org/10.1093/biomet/asab009

[53] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242

[54] VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

[55] VAN DER VAART, A. W. and WELLNER, J. A. (2023). *Weak Convergence and Empirical Processes—with Applications to Statistics. Springer Series in Statistics*. Springer, Cham. MR4628026 https://doi.org/10.1007/978-3-031-29040-4

[56] WU, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* **14** 1261–1350. With discussion and a rejoinder by the author. MR0868303 https://doi.org/10.1214/aos/1176350142


IMS members get a
40% discount
Order your copy now from
cambridge.org/ims
BRADLEY EFRON
TREVOR HASTIE
COMPUTER AGE STATISTICAL INFERENCE
ALGORITHMS, EVIDENCE, AND DATA SCIENCE