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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Articles

continued

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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Articles—*Continued from front cover*

Erratum

Vol. 50, No. 4—August 2022

THE ANNALS OF STATISTICS Vol. 50, No. 4, pp. 1859–2468 August 2022

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.

IMS OFFICERS

President: Peter Bühlmann, Seminar für Statistik, ETH Zürich, 8092 Zürich, Switzerland

President-Elect: Michael Kosorok, Department of Biostatistics and Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, Chapel Hill, NC 27599, USA

Past President: Krzysztof Burdzy, Department of Mathematics, University of Washington, Seattle, WA 98195-4350, USA

Executive Secretary: Edsel Peña, Department of Statistics, University of South Carolina, Columbia, SC 29208-001, USA

Treasurer: Jiashun Jin, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA

Program Secretary: Annie Qu, Department of Statistics, University of California, Irvine, Irvine, CA 92697-3425, USA

IMS EDITORS

The Annals of Statistics. *Editors*: Enno Mammen, Institute for Applied Mathematics, Heidelberg University, 69120 Heidelberg, Germany. Lan Wang, Miami Herbert Business School, University of Miami, Coral Gables, FL 33124, USA

The Annals of Applied Statistics. *Editor-in-Chief*: Ji Zhu, Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA

The Annals of Probability. *Editors*: Alice Guionnet, Unité de Mathématiques Pures et Appliquées, ENS de Lyon, Lyon, France. Christophe Garban, Institut Camille Jordan, Université Claude Bernard Lyon 1, 69622 Villeurbanne, France

The Annals of Applied Probability. *Editors*: Kavita Ramanan, Division of Applied Mathematics, Brown University, Providence, RI 02912, USA. Qi-Man Shao, Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen, Guangdong 518055, P.R. China

Statistical Science. *Editor*: Sonia Petrone, Department of Decision Sciences, Università Bocconi, 20100 Milano MI, Italy

The IMS Bulletin. *Editor*: Tati Howell, bulletin@imstat.org

The Annals of Statistics **[ISSN 0090-5364 (print); ISSN 2168-8966 (online)],** Volume 50, Number 4, August 2022. 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
2022, Vol. 50, No. 4, 1859–1889
https://doi.org/10.1214/21-AOS2120

CONVERGENCE OF DE FINETTI'S MIXING MEASURE IN LATENT STRUCTURE MODELS FOR OBSERVED EXCHANGEABLE SEQUENCES

BY YUN WEI[1,a] AND XUANLONG NGUYEN[2,b]

[1]*Department of Mathematics, University of Michigan,* [a]*cloudwei@umich.edu*
[2]*Department of Statistics, University of Michigan,* [b]*xuanlong@umich.edu*

Mixtures of product distributions are a powerful device for learning about heterogeneity within data populations. In this class of latent structure models, de Finetti's mixing measure plays the central role for describing the uncertainty about the latent parameters representing heterogeneity. In this paper, posterior contraction theorems for de Finetti's mixing measure arising from finite mixtures of product distributions will be established; under the setting the number of exchangeable sequences of observed variables increases while sequence length(s) may be either fixed or varied. The role of both the number of sequences and the sequence lengths will be carefully examined. In order to obtain concrete rates of convergence, a first-order identifiability theory for finite mixture models and a family of sharp inverse bounds for mixtures of product distributions will be developed via a harmonic analysis of such latent structure models. This theory is applicable to broad classes of probability kernels composing the mixture model of product distributions for both continuous and discrete domain $\mathfrak{X}$. Examples of interest include the case the probability kernel is only weakly identifiable in the sense of (*Ann. Statist.* **44** (2016) 2726–2755), the case where the kernel is itself a mixture distribution as in hierarchical models, and the case the kernel may not have a density with respect to a dominating measure on an abstract domain $\mathfrak{X}$, such as Dirichlet processes.

REFERENCES

[1] ALDOUS, D. J. (1985). Exchangeability and related topics. In *École D'été de Probabilités de Saint-Flour, XIII*—1983. *Lecture Notes in Math*. **1117** 1–198. Springer, Berlin. MR0883646 https://doi.org/10.1007/BFb0099421

[2] ALIPRANTIS, C. D. and BORDER, K. C. (2006). *Infinite Dimensional Analysis*: *A Hitchhiker's Guide*, 3rd ed. Springer, Berlin. MR2378491

[3] ALLMAN, E. S., MATIAS, C. and RHODES, J. A. (2009). Identifiability of parameters in latent structure models with many observed variables. *Ann. Statist.* **37** 3099–3132. MR2549554 https://doi.org/10.1214/09-AOS689

[4] ANTONIAK, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* **2** 1152–1174. MR0365969

[5] BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd ed. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR1324786

[6] BONHOMME, S., JOCHMANS, K. and ROBIN, J.-M. (2016). Non-parametric estimation of finite mixtures from repeated measurements. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 211–229. MR3453653 https://doi.org/10.1111/rssb.12110

[7] CAMERLENGHI, F., DUNSON, D. B., LIJOI, A., PRÜNSTER, I. and RODRÍGUEZ, A. (2019). Latent nested nonparametric priors (with discussion). *Bayesian Anal.* **14** 1303–1356. MR4044854 https://doi.org/10.1214/19-BA1169

[8] CAMERLENGHI, F., LIJOI, A., ORBANZ, P. and PRÜNSTER, I. (2019). Distribution theory for hierarchical processes. *Ann. Statist.* **47** 67–92. MR3909927 https://doi.org/10.1214/17-AOS1678

[9] CHEN, J. H. (1995). Optimal rate of convergence for finite mixture models. *Ann. Statist.* **23** 221–233. MR1331665 https://doi.org/10.1214/aos/1176324464

MSC2020 subject classifications. Primary 62F15, 62G05; secondary 62G20.

Key words and phrases. Mixture of product distributions, hierarchical models, inverse bounds, Fourier analysis, mixtures of repeated measurements, mixtures of grouped observations.

[10] Cruz-Medina, I. R., Hettmansperger, T. P. and Thomas, H. (2004). Semiparametric mixture models and repeated measures: The multinomial cut point model. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **53** 463–474. MR2075502 https://doi.org/10.1111/j.1467-9876.2004.05203.x

[11] Elmore, R., Hall, P. and Neeman, A. (2005). An application of classical invariant theory to identifiability in nonparametric mixtures. *Ann. Inst. Fourier* (*Grenoble*) **55** 1–28. MR2141286

[12] Elmore, R. T., Hettmansperger, T. P. and Thomas, H. (2004). Estimating component cumulative distribution functions in finite mixture models. *Comm. Statist. Theory Methods* **33** 2075–2086. MR2103063 https://doi.org/10.1081/STA-200026574

[13] Elmore, R. T. and Wang, S. (2003). Identifiability and estimation in finite mixture models with multinomial coefficients. Technical Report 03-04, Penn State Univ.

[14] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications. Vol. I*, 3rd ed. Wiley, New York. MR0228020

[15] Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230. MR0350949

[16] Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference. Cambridge Series in Statistical and Probabilistic Mathematics* **44**. Cambridge Univ. Press, Cambridge. MR3587782 https://doi.org/10.1017/9781139029834

[17] Guha, A., Ho, N. and Nguyen, X. (2021). On posterior contraction of parameters and interpretability in Bayesian mixture modeling. *Bernoulli* **27** 2159–2188. MR4303879 https://doi.org/10.3150/20-BEJ1275

[18] Hall, P., Neeman, A., Pakyari, R. and Elmore, R. (2005). Nonparametric inference in multivariate mixtures. *Biometrika* **92** 667–678. MR2202653 https://doi.org/10.1093/biomet/92.3.667

[19] Hall, P. and Zhou, X.-H. (2003). Nonparametric estimation of component distributions in a multivariate mixture. *Ann. Statist.* **31** 201–224. MR1962504 https://doi.org/10.1214/aos/1046294462

[20] Heinrich, P. and Kahn, J. (2018). Strong identifiability and optimal minimax rates for finite mixture estimation. *Ann. Statist.* **46** 2844–2870. MR3851757 https://doi.org/10.1214/17-AOS1641

[21] Hettmansperger, T. P. and Thomas, H. (2000). Almost nonparametric inference for repeated measures in mixture models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 811–825. MR1796294 https://doi.org/10.1111/1467-9868.00266

[22] Hjort, N. L., Holmes, C., Müller, P. and Walker, S. G. (2010). *Bayesian Nonparametrics. Cambridge Series in Statistical and Probabilistic Mathematics* **28**. Cambridge University Press, Cambridge.

[23] Ho, N. and Nguyen, X. (2016). Convergence rates of parameter estimation for some weakly identifiable finite mixtures. *Ann. Statist.* **44** 2726–2755. MR3576559 https://doi.org/10.1214/16-AOS1444

[24] Ho, N. and Nguyen, X. (2016). On strong identifiability and convergence rates of parameter estimation in finite mixtures. *Electron. J. Stat.* **10** 271–307. MR3466183 https://doi.org/10.1214/16-EJS1105

[25] Ho, N. and Nguyen, X. (2019). Singularity structures and impacts on parameter estimation in finite mixtures of distributions. *SIAM J. Math. Data Sci.* **1** 730–758. MR4016133 https://doi.org/10.1137/18M122947X

[26] Kallenberg, O. (2005). *Probabilistic Symmetries and Invariance Principles. Probability and Its Applications* (*New York*). Springer, New York. MR2161313

[27] Lindsay, B. G. (1995). Mixture models: Theory, geometry and applications. In *NSF-CBMS Regional Conference Series in Probability and Statistics* i–163. JSTOR.

[28] Nguyen, X. (2011). Wasserstein distances for discrete measures and convergence in nonparametric mixture models. Technical Report, Univ. of Michigan MI Dept. of Statistics.

[29] Nguyen, X. (2013). Convergence of latent mixing measures in finite and infinite mixture models. *Ann. Statist.* **41** 370–400. MR3059422 https://doi.org/10.1214/12-AOS1065

[30] Nguyen, X. (2015). Posterior contraction of the population polytope in finite admixture models. *Bernoulli* **21** 618–646. MR3322333 https://doi.org/10.3150/13-BEJ582

[31] Nguyen, X. (2016). Borrowing strengh in hierarchical Bayes: Posterior concentration of the Dirichlet base measure. *Bernoulli* **22** 1535–1571. MR3474825 https://doi.org/10.3150/15-BEJ703

[32] Ritchie, A., Vandermeulen, R. A. and Scott, C. (2020). Consistent estimation of identifiable nonparametric mixture models from grouped observations. arXiv preprint. Available at arXiv:2006.07459.

[33] Rodríguez, A., Dunson, D. B. and Gelfand, A. E. (2008). The nested Dirichlet process. *J. Amer. Statist. Assoc.* **103** 1131–1144. MR2528831 https://doi.org/10.1198/016214508000000553

[34] Rousseau, J. and Mengersen, K. (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 689–710. MR2867454 https://doi.org/10.1111/j.1467-9868.2011.00781.x

[35] Teh, Y. W., Jordan, M. I., Beal, M. J. and Blei, D. M. (2006). Hierarchical Dirichlet processes. *J. Amer. Statist. Assoc.* **101** 1566–1581. MR2279480 https://doi.org/10.1198/016214506000000302

[36] TEICHER, H. (1967). Identifiability of mixtures of product measures. *Ann. Math. Stat.* **38** 1300–1302. MR0216635 https://doi.org/10.1214/aoms/1177698805

[37] VANDERMEULEN, R. A. and SCOTT, C. D. (2019). An operator theoretic approach to nonparametric mixture models. *Ann. Statist.* **47** 2704–2733. MR3988770 https://doi.org/10.1214/18-AOS1762

[38] WEI, Y. and NGUYEN, X. (2021). Supplement to "Convergence of de Finetti's mixing measure in latent structure models for observed exchangeable sequences." https://doi.org/10.1214/21-AOS2120SUPP

[39] WEI, Y. and NGUYEN, X. (2021). Convergence of de Finetti's mixing measure in latent structure models for observed exchangeable sequences. arXiv preprint. Available at arXiv:2004.05542v3.

[40] WELLNER, J. and VAN DER VAART, A. (1996). *Weak Convergence and Empirical Processes*: *With Applications to Statistics*. *Springer Series in Statistics*. Springer Science & Business Media.

[41] WU, Y. and YANG, P. (2020). Optimal estimation of Gaussian mixtures via denoised method of moments. *Ann. Statist.* **48** 1981–2007. MR4134783 https://doi.org/10.1214/19-AOS1873

The Annals of Statistics
2022, Vol. 50, No. 4, 1890–1909
https://doi.org/10.1214/21-AOS2139

DETECTING MULTIPLE REPLICATING SIGNALS USING ADAPTIVE FILTERING PROCEDURES

BY JINGSHU WANG[1,a], LIN GUI[1,b], WEIJIE J. SU[2,c], CHIARA SABATTI[3,d] AND ART B. OWEN[3,e]

[1]*Department of Statistics, The University of Chicago,* [a]*jingshuw@uchicago.edu,* [b]*glin6@uchicago.edu*
[2]*Department of Statistics and Data Science, University of Pennsylvania,* [c]*suw@wharton.upenn.edu*
[3]*Department of Statistics, Stanford University,* [d]*sabatti@stanford.edu,* [e]*owen@stanford.edu*

Replicability is a fundamental quality of scientific discoveries: we are interested in those signals that are detectable in different laboratories, different populations, across time etc. Unlike meta-analysis which accounts for experimental variability but does not guarantee replicability, testing a partial conjunction (PC) null aims specifically to identify the signals that are discovered in multiple studies. In many contemporary applications, for example, comparing multiple high-throughput genetic experiments, a large number M of PC nulls need to be tested simultaneously, calling for a multiple comparisons correction. However, standard multiple testing adjustments on the M PC p-values can be severely conservative, especially when M is large and the signals are sparse. We introduce AdaFilter, a new multiple testing procedure that increases power by adaptively filtering out unlikely candidates of PC nulls. We prove that AdaFilter can control FWER and FDR as long as data across studies are independent, and has much higher power than other existing methods. We illustrate the application of AdaFilter with three examples: microarray studies of Duchenne muscular dystrophy, single-cell RNA sequencing of T cells in lung cancer tumors and GWAS for metabolomics.

REFERENCES

[1] BAKER, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature* **533** 452–454.
[2] BARBER, R. F. and RAMDAS, A. (2017). The p-filter: Multilayer false discovery rate control for grouped hypotheses. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1247–1268. MR3689317 https://doi.org/10.1111/rssb.12218
[3] BEGLEY, C. G. and ELLIS, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature* **483** 531–533.
[4] BENJAMINI, Y. and HELLER, R. (2008). Screening for partial conjunction hypotheses. *Biometrics* **64** 1215–1222. MR2522270 https://doi.org/10.1111/j.1541-0420.2007.00984.x
[5] BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. MR1325392
[6] BOGDAN, M., VAN DEN BERG, E., SABATTI, C., SU, W. and CANDÈS, E. J. (2015). SLOPE—adaptive variable selection via convex optimization. *Ann. Appl. Stat.* **9** 1103–1140. MR3418717 https://doi.org/10.1214/15-AOAS842
[7] BOGOMOLOV, M. and HELLER, R. (2018). Assessing replicability of findings across two studies of multiple features. *Biometrika* **105** 505–516. MR3842881 https://doi.org/10.1093/biomet/asy029
[8] DJORDJILOVIĆ, V., HEMERIK, J. and THORESEN, M. (2022). On optimal two-stage testing of multiple mediators. *Biom. J.* https://doi.org/10.1002/bimj.202100190
[9] DJORDJILOVIĆ, V., PAGE, C. M., GRAN, J. M., NØST, T. H., SANDANGER, T. M., VEIERØD, M. B. and THORESEN, M. (2019). Global test for high-dimensional mediation: Testing groups of potential mediators. *Stat. Med.* **38** 3346–3360. MR3979813 https://doi.org/10.1002/sim.8199
[10] DUDOIT, S. and VAN DER LAAN, M. J. (2008). *Multiple Testing Procedures with Applications to Genomics. Springer Series in Statistics.* Springer, New York. MR2373771 https://doi.org/10.1007/978-0-387-49317-6

MSC2020 subject classifications. Primary 62J15; secondary 62P10.
Key words and phrases. Simultaneous signals, meta-analysis, high-throughput experiments, composite null, multiple hypotheses testing.

[11] EFRON, B. (2010). *Large-Scale Inference*: *Empirical Bayes Methods for Estimation*, *Testing*, *and Prediction*. *Institute of Mathematical Statistics* (*IMS*) *Monographs* **1**. Cambridge Univ. Press, Cambridge. MR2724758 https://doi.org/10.1017/CBO9780511761362

[12] FERREIRA, J. A. and ZWINDERMAN, A. H. (2006). On the Benjamini–Hochberg method. *Ann. Statist.* **34** 1827–1849. MR2283719 https://doi.org/10.1214/009053606000000425

[13] FLUTRE, T., WEN, X., PRITCHARD, J. and STEPHENS, M. (2013). A statistical framework for joint eQTL analysis in multiple tissues. *PLoS Genet.* **9** e1003486. https://doi.org/10.1371/journal.pgen.1003486

[14] FRISTON, K. J., PENNY, W. D. and GLASER, D. E. (2005). Conjunction revisited. *NeuroImage* **25** 661–667.

[15] GIRI, A., HELLWEGE, J. N., KEATON, J. M., PARK, J., QIU, C., WARREN, H. R., TORSTENSON, E. S., KOVESDY, C. P., SUN, Y. V. et al. (2019). Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nat. Genet.* **51** 51–62.

[16] GOEMAN, J. J. and SOLARI, A. (2010). The sequential rejection principle of familywise error control. *Ann. Statist.* **38** 3782–3810. MR2766868 https://doi.org/10.1214/10-AOS829

[17] GUO, X., ZHANG, Y., ZHENG, L., ZHENG, C., SONG, J., ZHANG, Q., KANG, B., LIU, Z., JIN, L. et al. (2018). Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat. Med.* **24** 978–985.

[18] HASIN, Y., SELDIN, M. and LUSIS, A. (2017). Multi-omics approaches to disease. *Genome Biol.* **18** 83. https://doi.org/10.1186/s13059-017-1215-1

[19] HELLER, R., BOGOMOLOV, M. and BENJAMINI, Y. (2014). Deciding whether follow-up studies have replicated findings in a preliminary large-scale omics study. *Proc. Natl. Acad. Sci. USA* **111** 16262–16267.

[20] HELLER, R., GOLLAND, Y., MALACH, R. and BENJAMINI, Y. (2007). Conjunction group analysis: An alternative to mixed/random effect analysis. *NeuroImage* **37** 1178–1185.

[21] HELLER, R. and YEKUTIELI, D. (2014). Replicability analysis for genome-wide association studies. *Ann. Appl. Stat.* **8** 481–498. MR3191999 https://doi.org/10.1214/13-AOAS697

[22] HUME, D. (2003). *A Treatise of Human Nature*. Courier Corporation.

[23] INTRODUCTION TO REPLICABILITY IN SCIENCE. Available at http://www.replicability.tau.ac.il/index.php/replicability-in-science.html. Accessed: 2018-08-29.

[24] IRIZARRY, R. A., HOBBS, B., COLLIN, F., BEAZER-BARCLAY, Y. D., ANTONELLIS, K. J., SCHERF, U. et al. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4** 249–264.

[25] JALJULI, I., BENJAMINI, Y., SHENHAV, L., PANAGIOTOU, O. and HELLER, R. (2022). Quantifying replicability and consistency in systematic reviews. *Stat. Biopharm. Res.*. https://doi.org/10.1080/19466315.2022.2050291

[26] KARMAKAR, B., SMALL, D. S. and ROSENBAUM, P. R. (2021). Reinforced designs: Multiple instruments plus control groups as evidence factors in an observational study of the effectiveness of Catholic schools. *J. Amer. Statist. Assoc.* **116** 82–92. MR4227676 https://doi.org/10.1080/01621459.2020.1745811

[27] KOTELNIKOVA, E., SHKROB, M. A., PYATNITSKIY, M. A., FERLINI, A. and DARASELIA, N. (2012). Novel approach to meta-analysis of microarray datasets reveals muscle remodeling-related drug targets and biomarkers in Duchenne muscular dystrophy. *PLoS Comput. Biol.* **8** e1002365.

[28] LEHRER, J. (2010). The truth wears off. *New Yorker*.

[29] LEI, L. and FITHIAN, W. (2018). AdaPT: An interactive procedure for multiple testing with side information. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 649–679. MR3849338 https://doi.org/10.1111/rssb.12253

[30] LI, A. and BARBER, R. F. (2019). Multiple testing with the structure-adaptive Benjamini–Hochberg algorithm. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 45–74. MR3904779

[31] LIU, Z., SHEN, J., BARFIELD, R., SCHWARTZ, J., BACCARELLI, A. A. and LIN, X. (2022). Large-scale hypothesis testing for causal mediation effects with applications in genome-wide epigenetic studies. *J. Amer. Statist. Assoc.* **117** 67–81. MR4399068 https://doi.org/10.1080/01621459.2021.1914634

[32] MARIGORTA, U. M. and NAVARRO, A. (2013). High trans-ethnic replicability of GWAS results implies common causal variants. *PLoS Genet.* **9** e1003566. https://doi.org/10.1371/journal.pgen.1003566

[33] MOONESINGHE, R., KHOURY, M. J. and JANSSENS, A. C. J. W. (2007). Most published research findings are false—but a little replication goes a long way. *PLoS Med.* **4** e28.

[34] OWEN, A. B. (2009). Karl Pearson's meta-analysis revisited. *Ann. Statist.* **37** 3867–3892. MR2572446 https://doi.org/10.1214/09-AOS697

[35] PRICE, C. J. and FRISTON, K. J. (1997). Cognitive conjunction: A new approach to brain activation experiments. *NeuroImage* **5** 261–270.

[36] SMYTH, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3** Art. 3, 29. MR2101454 https://doi.org/10.2202/1544-6115.1027

[37] SUN, W., REICH, B. J., CAI, T. T., GUINDANI, M. and SCHWARTZMAN, A. (2015). False discovery control in large-scale spatial multiple testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 59–83. MR3299399 https://doi.org/10.1111/rssb.12064

[38] TUKEY, J. W. (1953). The problem of multiple comparisons. Technical report, Princeton Univ.

[39] URBUT, S. M., WANG, G., CARBONETTO, P. and STEPHENS, M. (2019). Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat. Genet.* **51** 187–195.

[40] WANG, J., GUI, L., SU, W. J., SABATTI, C. and OWEN, A. B. (2022). Supplement to "Detecting multiple replicating signals using adaptive filtering procedures." https://doi.org/10.1214/21-AOS2139SUPP

[41] WANG, J. and OWEN, A. B. (2019). Admissibility in partial conjunction testing. *J. Amer. Statist. Assoc.* **114** 158–168. MR3941245 https://doi.org/10.1080/01621459.2017.1385465

[42] XIANG, D., ZHAO, S. D. and CAI, T. T. (2019). Signal classification for the integrative analysis of multiple sequences of large-scale multiple tests. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 707–734. MR3997098

[43] ZHANG, N. R., SENBABAOGLU, Y. and LI, J. Z. (2010). Joint estimation of DNA copy number from multiple platforms. *Bioinformatics* **26** 153–160.

[44] ZHAO, S. D. and NGUYEN, Y. T. (2020). Nonparametric false discovery rate control for identifying simultaneous signals. *Electron. J. Stat.* **14** 110–142. MR4047996 https://doi.org/10.1214/19-EJS1663

The Annals of Statistics
2022, Vol. 50, No. 4, 1910–1932
https://doi.org/10.1214/21-AOS2150

OPTIMAL SIGNAL DETECTION IN SOME SPIKED RANDOM MATRIX MODELS: LIKELIHOOD RATIO TESTS AND LINEAR SPECTRAL STATISTICS

BY DEBAPRATIM BANERJEE[a] AND ZONGMING MA[b]

Department of Statistics, The Wharton School, University of Pennsylvania, [a]*dban@wharton.upenn.edu,* [b]*zongming@wharton.upenn.edu*

We study signal detection by likelihood ratio tests in a number of spiked random matrix models, including but not limited to Gaussian mixtures and spiked Wishart covariance matrices. We work directly with multi-spiked cases in these models and with flexible priors on signal components that allow dependence across spikes. We derive asymptotic normality for the log-likelihood ratios when the signal-to-noise ratios are below certain bounds. In addition, the log-likelihood ratios can be asymptotically decomposed as weighted sums of a collection of statistics which we call bipartite signed cycles. Based on this decomposition, we show that below the bounds we could always achieve the asymptotically optimal powers of likelihood ratio tests via tests based on linear spectral statistics which have polynomial time complexity.

REFERENCES

[1] ANDERSON, G. W., GUIONNET, A. and ZEITOUNI, O. (2010). *An Introduction to Random Matrices*. *Cambridge Studies in Advanced Mathematics* **118**. Cambridge Univ. Press, Cambridge. MR2760897

[2] 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

[3] BANERJEE, D. (2018). Contiguity and non-reconstruction results for planted partition models: The dense case. *Electron. J. Probab.* **23** 18. MR3771755 https://doi.org/10.1214/17-EJP128

[4] BANERJEE, D. and MA, Z. (2017). Optimal hypothesis testing for stochastic block models with growing degrees. arXiv preprint. Available at arXiv:1705.05305.

[5] BANERJEE, D. and MA, Z. (2022). Supplement to "Optimal signal detection in some spiked random matrix models: Likelihood ratio tests and linear spectral statistics." https://doi.org/10.1214/21-AOS2150SUPP

[6] BANKS, J., MOORE, C., VERSHYNIN, R., VERZELEN, N. and XU, J. (2017). Information-theoretic bounds and phase transitions in clustering, sparse pca, and submatrix localization. In *Information Theory* (*ISIT*), 2017 *IEEE International Symposium on* 1137–1141. IEEE, Los Alamitos.

[7] BENAYCH-GEORGES, F., GUIONNET, A. and MALE, C. (2014). Central limit theorems for linear statistics of heavy tailed random matrices. *Comm. Math. Phys.* **329** 641–686. MR3210147 https://doi.org/10.1007/s00220-014-1975-3

[8] CHEN, W.-K., HANDSCHY, M. and LERMAN, G. (2018). Phase transition in random tensors with multiple spikes. arXiv preprint. Available at arXiv:1809.06790.

[9] EL ALAOUI, A. and JORDAN, M. I. (2018). Detection limits in the high-dimensional spiked rectangular model. In 31th Annual Conference on Learning Theory (COLT), PMLR 75 410–438.

[10] EL ALAOUI, A., KRZAKALA, F. and JORDAN, M. I. (2017). Finite size corrections and likelihood ratio fluctuations in the spiked wigner model. arXiv preprint. Available at arXiv:1710.02903.

[11] GEMAN, S. (1980). A limit theorem for the norm of random matrices. *Ann. Probab.* **8** 252–261. MR0566592

[12] HSU, D., KAKADE, S. M. and ZHANG, T. (2012). A tail inequality for quadratic forms of subgaussian random vectors. *Electron. Commun. Probab.* **17** 52. MR2994877 https://doi.org/10.1214/ECP.v17-2079

[13] JANSON, S. (1995). Random regular graphs: Asymptotic distributions and contiguity. *Combin. Probab. Comput.* **4** 369–405. MR1377557 https://doi.org/10.1017/S0963548300001735

MSC2020 subject classifications. Primary 62C05, 62F05; secondary 60F05.

Key words and phrases. Contiguity, finite rank deformation, principal component analysis, random graphs, signal detection.

[14] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961 https://doi.org/10.1214/aos/1009210544

[15] JOHNSTONE, I. M. and ONATSKI, A. (2015). Testing in high-dimensional spiked models. *Ann. Statist.* arXiv preprint. Available at arXiv:1509.07269.

[16] JUNG, J. H., CHUNG, H. W. and LEE, J. O. (2021). Detection of signal in the spiked rectangular models. arXiv preprint. Available at arXiv:2104.13517.

[17] LALOUX, L., CIZEAU, P., POTTERS, M. and BOUCHAUD, J.-P. (2000). Random matrix theory and financial correlations. *Int. J. Theor. Appl. Finance* **3** 391–397.

[18] LE CAM, L. (2012). *Asymptotic Methods in Statistical Decision Theory*. Springer, Berlin.

[19] LE CAM, L. and YANG, G. L. (2000). *Asymptotics in Statistics*, 2nd ed. *Springer Series in Statistics*. Springer, New York. Some basic concepts. MR1784901 https://doi.org/10.1007/978-1-4612-1166-2

[20] MARČENKO, V. and PASTUR, L. (1967). Distribution of eigenvalues for some sets of random matrices. *Sb. Math.* **1** 457–483.

[21] MOSSEL, E., NEEMAN, J. and SLY, A. (2015). 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

[22] ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2013). Asymptotic power of sphericity tests for high-dimensional data. *Ann. Statist.* **41** 1204–1231. MR3113808 https://doi.org/10.1214/13-AOS1100

[23] ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2014). Signal detection in high dimension: The multispiked case. *Ann. Statist.* **42** 225–254. MR3189485 https://doi.org/10.1214/13-AOS1181

[24] PATTERSON, N., PRICE, A. L. and REICH, D. (2006). Population structure and eigenanalysis. *PLoS Genet.* **2** e190.

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

[26] SINAI, Y. and SOSHNIKOV, A. (1998). Central limit theorem for traces of large random symmetric matrices with independent matrix elements. *Bol. Soc. Bras. Mat.* **29** 1–24. MR1620151 https://doi.org/10.1007/BF01245866

[27] STEWART, G. W. and SUN, J. G. (1990). *Matrix Perturbation Theory*. *Computer Science and Scientific Computing*. Academic Press, Boston, MA. MR1061154

The Annals of Statistics
2022, Vol. 50, No. 4, 1933–1959
https://doi.org/10.1214/21-AOS2151

ON UNIVERSALLY CONSISTENT AND FULLY DISTRIBUTION-FREE RANK TESTS OF VECTOR INDEPENDENCE

BY HONGJIAN SHI[1,a], MARC HALLIN[2,c], MATHIAS DRTON[1,b] AND FANG HAN[3,d]

[1]*Department of Mathematics, Technical University of Munich,* [a]*hongjian.shi@tum.de,* [b]*mathias.drton@tum.de*
[2]*ECARES and Department of Mathematics, Université Libre de Bruxelles,* [c]*mhallin@ulb.ac.be*
[3]*Department of Statistics, University of Washington,* [d]*fanghan@uw.edu*

Rank correlations have found many innovative applications in the last decade. In particular, suitable rank correlations have been used for consistent tests of independence between pairs of random variables. Using ranks is especially appealing for continuous data as tests become distribution-free. However, the traditional concept of ranks relies on ordering data and is, thus, tied to univariate observations. As a result, it has long remained unclear how one may construct distribution-free yet consistent tests of independence between random vectors. This is the problem addressed in this paper, in which we lay out a general framework for designing dependence measures that give tests of multivariate independence that are not only consistent and distribution-free but which we also prove to be statistically efficient. Our framework leverages the recently introduced concept of center-outward ranks and signs, a multivariate generalization of traditional ranks, and adopts a common standard form for dependence measures that encompasses many popular examples. In a unified study, we derive a general asymptotic representation of center-outward rank-based test statistics under independence, extending to the multivariate setting the classical Hájek asymptotic representation results. This representation permits direct calculation of limiting null distributions and facilitates a local power analysis that provides strong support for the center-outward approach by establishing, for the first time, the nontrivial power of center-outward rank-based tests over root-n neighborhoods within the class of quadratic mean differentiable alternatives.

REFERENCES

AGARWAL, P. K. and SHARATHKUMAR, R. (2014). Approximation algorithms for bipartite matching with metric and geometric costs. In *STOC'*14—*Proceedings of the* 2014 *ACM Symposium on Theory of Computing* 555–564. ACM, New York. MR3238983 https://doi.org/10.1145/2591796.2591844

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR1990662

AZADKIA, M. and CHATTERJEE, S. (2021). A simple measure of conditional dependence. *Ann. Statist.* **49** 3070–3102. MR4352523 https://doi.org/10.1214/21-aos2073

BEIRLANT, J., BUITENDAG, S., DEL BARRIO, E., HALLIN, M. and KAMPER, F. (2020). Center-outward quantiles and the measurement of multivariate risk. *Insurance Math. Econom.* **95** 79–100. MR4154446 https://doi.org/10.1016/j.insmatheco.2020.08.005

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

BERGSMA, W. (2006). A new correlation coefficient, its orthogonal decomposition and associated tests of independence. Preprint. Available at arXiv:math/0604627v1.

BERGSMA, W. (2011). Nonparametric testing of conditional independence by means of the partial copula. Preprint. Available at arXiv:1101.4607v1.

MSC2020 subject classifications. 62G10.

Key words and phrases. Multivariate ranks and signs, center-outward ranks and signs, multivariate dependence measure, independence test, Hájek representation, Le Cam's third lemma.

BERGSMA, W. and DASSIOS, A. (2014). A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli* **20** 1006–1028. MR3178526 https://doi.org/10.3150/13-BEJ514

BERRETT, T. B., KONTOYIANNIS, I. and SAMWORTH, R. J. (2021). Optimal rates for independence testing via U-statistic permutation tests. *Ann. Statist.* **49** 2457–2490. MR4338371 https://doi.org/10.1214/20-aos2041

BHATTACHARYA, B. B. (2019). A general asymptotic framework for distribution-free graph-based two-sample tests. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 575–602. MR3961499

BLUM, J. R., KIEFER, J. and ROSENBLATT, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Stat.* **32** 485–498. MR0125690 https://doi.org/10.1214/aoms/1177705055

CAO, S. and BICKEL, P. J. (2020). Correlations with tailored extremal properties. Preprint. Available at arXiv:2008.10177v2.

CHERNOFF, H. and SAVAGE, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *Ann. Math. Stat.* **29** 972–994. MR0100322 https://doi.org/10.1214/aoms/1177706436

CHERNOZHUKOV, V., GALICHON, A., HALLIN, M. and HENRY, M. (2017). Monge–Kantorovich depth, quantiles, ranks and signs. *Ann. Statist.* **45** 223–256. MR3611491 https://doi.org/10.1214/16-AOS1450

DEB, N., BHATTACHARYA, B. B. and SEN, B. (2021). Efficiency lower bounds for distribution-free Hotelling-type two-sample tests based on optimal transport. Preprint. Available at arXiv:2104.01986v1.

DEB, N. and SEN, B. (2022). Multivariate rank-based distribution-free nonparametric testing using measure transportation. *J. Amer. Statist. Assoc.* To appear. https://doi.org/10.1080/01621459.2021.1923508

DEL BARRIO, E., GONZÁLEZ-SANZ, A. and HALLIN, M. (2020). A note on the regularity of optimal-transport-based center-outward distribution and quantile functions. *J. Multivariate Anal.* **180** 104671, 13 pp. MR4147635 https://doi.org/10.1016/j.jmva.2020.104671

DEL BARRIO, E., CUESTA-ALBERTOS, J. A., HALLIN, M. and MATRÁN, C. (2018). Smooth cyclically monotone interpolation and empirical center-outward distribution functions. Preprint. Available at arXiv:1806.01238v1.

DHAR, S. S., DASSIOS, A. and BERGSMA, W. (2016). A study of the power and robustness of a new test for independence against contiguous alternatives. *Electron. J. Stat.* **10** 330–351. MR3466185 https://doi.org/10.1214/16-EJS1107

DRTON, M., HAN, F. and SHI, H. (2020). High-dimensional consistent independence testing with maxima of rank correlations. *Ann. Statist.* **48** 3206–3227. MR4185806 https://doi.org/10.1214/19-AOS1926

FEUERVERGER, A. (1993). A consistent test for bivariate dependence. *Int. Stat. Rev.* **61** 419–433. MR4205649

FIGALLI, A. (2018). On the continuity of center-outward distribution and quantile functions. *Nonlinear Anal.* **177** 413–421. MR3886582 https://doi.org/10.1016/j.na.2018.05.008

FUKUMIZU, K., BACH, F. R. and GRETTON, A. (2007). Statistical consistency of kernel canonical correlation analysis. *J. Mach. Learn. Res.* **8** 361–383. MR2320675

GABOW, H. N. and TARJAN, R. E. (1989). Faster scaling algorithms for network problems. *SIAM J. Comput.* **18** 1013–1036. MR1015271 https://doi.org/10.1137/0218069

GHOSAL, P. and SEN, B. (2022). Multivariate ranks and quantiles using optimal transport: Consistency, rates and nonparametric testing. *Ann. Statist.* **50** 1012–1037. MR4404927 https://doi.org/10.1214/21-aos2136

GIESER, P. W. (1993). A new nonparametric test for independence between two sets of variates. PhD thesis, University of Florida. Available at https://ufdc.ufl.edu/AA00003658/00001 and https://www.proquest.com/docview/304041219. MR2691505

GIESER, P. W. and RANDLES, R. H. (1997). A nonparametric test of independence between two vectors. *J. Amer. Statist. Assoc.* **92** 561–567. MR1467849 https://doi.org/10.2307/2965704

GRETTON, A., SMOLA, A., BOUSQUET, O., HERBRICH, R., BELITSKI, A., AUGATH, M., MURAYAMA, Y., PAULS, J., SCHÖLKOPF, B. et al. (2005a). Kernel constrained covariance for dependence measurement. In *AISTATS*05 (R. G. Cowell and Z. Ghahramani, eds.) 112–119. Society for Artificial Intelligence and Statistics.

GRETTON, A., BOUSQUET, O., SMOLA, A. and SCHÖLKOPF, B. (2005b). 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., HERBRICH, R., SMOLA, A., BOUSQUET, O. and SCHÖLKOPF, B. (2005c). Kernel methods for measuring independence. *J. Mach. Learn. Res.* **6** 2075–2129. MR2249882

HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests*. Academic Press, New York–London; Academia [Publishing House of the Czechoslovak Academy of Sciences], Prague. MR0229351

HALLIN, M. (1994). On the Pitman non-admissibility of correlogram-based methods. *J. Time Series Anal.* **15** 607–611. MR1312324 https://doi.org/10.1111/j.1467-9892.1994.tb00215.x

HALLIN, M. (2017). On distribution and quantile functions, ranks and signs in $\mathbb{R}^d$: A measure transportation approach. Preprint. Available at https://ideas.repec.org/p/eca/wpaper/2013-258262.html.

HALLIN, M. (2022). Measure transportation and statistical decision theory. *Annu. Rev. Stat. Appl.* **9** 401–424. MR4394914 https://doi.org/10.1146/annurev-statistics-040220-105948

HALLIN, M., HLUBINKA, D. and HUDECOVÁ, Š. (2022). Efficient fully distribution-free center-outward rank tests for multiple-output regression and MANOVA. *J. Amer. Statist. Assoc*. To appear. https://doi.org/10.1080/01621459.2021.2021921

HALLIN, M., LA VECCHIA, D. and LIU, H. (2022). Center-outward R-estimation for semiparametric VARMA models. *J. Amer. Statist. Assoc*. To appear. https://doi.org/10.1080/01621459.2020.1832501

HALLIN, M., MORDANT, G. and SEGERS, J. (2021). Multivariate goodness-of-fit tests based on Wasserstein distance. *Electron. J. Stat*. **15** 1328–1371. MR4255302 https://doi.org/10.1214/21-ejs1816

HALLIN, M. and PAINDAVEINE, D. (2002a). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist*. **30** 1103–1133. MR1926170 https://doi.org/10.1214/aos/1031689019

HALLIN, M. and PAINDAVEINE, D. (2002b). Optimal procedures based on interdirections and pseudo-Mahalanobis ranks for testing multivariate elliptic white noise against ARMA dependence. *Bernoulli* **8** 787–815. MR1963662

HALLIN, M. and PAINDAVEINE, D. (2002c). Multivariate signed ranks: Randles' interdirections or Tyler's angles? In *Statistical Data Analysis Based on the L_1-Norm and Related Methods* (*Neuchâtel*, 2002). *Stat. Ind. Technol*. 271–282. Birkhäuser, Basel. MR2001322

HALLIN, M. and PAINDAVEINE, D. (2008). Chernoff–Savage and Hodges–Lehmann results for Wilks' test of multivariate independence. In *Beyond Parametrics in Interdisciplinary Research*: *Festschrift in Honor of Professor Pranab K. Sen. Inst. Math. Stat*. (*IMS*) *Collect*. **1** 184–196. IMS, Beachwood, OH. MR2462206 https://doi.org/10.1214/193940307000000130

HALLIN, M., DEL BARRIO, E., CUESTA-ALBERTOS, J. and MATRÁN, C. (2021). Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach. *Ann. Statist*. **49** 1139–1165. MR4255122 https://doi.org/10.1214/20-aos1996

HAN, F., CHEN, S. and LIU, H. (2017). Distribution-free tests of independence in high dimensions. *Biometrika* **104** 813–828. MR3737306 https://doi.org/10.1093/biomet/asx050

HAN, F. and LIU, H. (2018). ECA: High-dimensional elliptical component analysis in non-Gaussian distributions. *J. Amer. Statist. Assoc*. **113** 252–268. MR3803462 https://doi.org/10.1080/01621459.2016.1246366

HELLER, R., GORFINE, M. and HELLER, Y. (2012). A class of multivariate distribution-free tests of independence based on graphs. *J. Statist. Plann. Inference* **142** 3097–3106. MR2956796 https://doi.org/10.1016/j.jspi.2012.06.003

HELLER, R. and HELLER, Y. (2016). Multivariate tests of association based on univariate tests. In *Advances in Neural Information Processing Systems* 29 (D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon and R. Garnett, eds.) 208–216. Curran Associates, Red Hook.

HELLER, R., HELLER, Y. and GORFINE, M. (2013). A consistent multivariate test of association based on ranks of distances. *Biometrika* **100** 503–510. MR3068450 https://doi.org/10.1093/biomet/ass070

HOEFFDING, W. (1948). A non-parametric test of independence. *Ann. Math. Stat*. **19** 546–557. MR0029139 https://doi.org/10.1214/aoms/1177730150

KENDALL, M. G. (1938). A new measure of rank correlation. *Biometrika* **30** 81–93.

KIM, I., BALAKRISHNAN, S. and WASSERMAN, L. (2020a). Robust multivariate nonparametric tests via projection averaging. *Ann. Statist*. **48** 3417–3441. MR4185814 https://doi.org/10.1214/19-AOS1936

KIM, I., BALAKRISHNAN, S. and WASSERMAN, L. (2020b). Supplement to "Robust multivariate nonparametric tests via projection averaging". *Ann. Statist*. **48**. MR4185814 https://doi.org/10.1214/19-AOS1936SUPP

KONIJN, H. S. (1956). On the power of certain tests for independence in bivariate populations. *Ann. Math. Stat*. **27** 300–323. MR0079384 https://doi.org/10.1214/aoms/1177728260

KOROLJUK, V. S. and BOROVSKICH, Y. V. (1994). *Theory of U-Statistics*. *Mathematics and Its Applications* **273**. Kluwer Academic, Dordrecht. MR1472486 https://doi.org/10.1007/978-94-017-3515-5

LE CAM, L. and YANG, G. L. (2000). *Asymptotics in Statistics*: *Some Basic Concepts*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR1784901 https://doi.org/10.1007/978-1-4612-1166-2

LEEB, H. and PÖTSCHER, B. M. (2008). Can one estimate the unconditional distribution of post-model-selection estimators? *Econometric Theory* **24** 338–376. MR2422862 https://doi.org/10.1017/S0266466608080158

LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. *Springer Texts in Statistics*. Springer, New York. MR2135927

LIN, J. (2017). Copula versions of RKHS-based and distance-based criteria. PhD thesis, Pennsylvania State University. Available at https://etda.libraries.psu.edu/catalog/14485jul268. MR4172371

LIU, R. Y. and SINGH, K. (1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc*. **88** 252–260. MR1212489

LYONS, R. (2013). Distance covariance in metric spaces. *Ann. Probab.* **41** 3284–3305. MR3127883 https://doi.org/10.1214/12-AOP803

OJA, H. (2010). *Multivariate Nonparametric Methods with R*: *An Approach Based on Spatial Signs and Ranks*. *Lecture Notes in Statistics* **199**. Springer, New York. MR2598854 https://doi.org/10.1007/978-1-4419-0468-3

OJA, H., PAINDAVEINE, D. and TASKINEN, S. (2016). Affine-invariant rank tests for multivariate independence in independent component models. *Electron. J. Stat.* **10** 2372–2419. MR3544291 https://doi.org/10.1214/16-EJS1174

PEARSON, K. (1895). Note on regression and inheritance in the case of two parents. *Proc. R. Soc. Lond.* **58** 240–242.

PINELIS, I. and MOLZON, R. (2016). Optimal-order bounds on the rate of convergence to normality in the multivariate delta method. *Electron. J. Stat.* **10** 1001–1063. MR3486424 https://doi.org/10.1214/16-EJS1133

PURI, M. L. and SEN, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. Wiley, New York. MR0298844

PURI, M. L., SEN, P. K. and GOKHALE, D. V. (1970). On a class of rank order tests for independence in multivariate distributions. *Sankhyā Ser. A* **32** 271–298. MR0298847

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

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics. Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0595165

SHAH, R. D. and PETERS, J. (2020). The hardness of conditional independence testing and the generalised covariance measure. *Ann. Statist.* **48** 1514–1538. MR4124333 https://doi.org/10.1214/19-AOS1857

SHARATHKUMAR, R. and AGARWAL, P. K. (2012). Algorithms for the transportation problem in geometric settings. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms* 306–317. ACM, New York. MR3205219

SHI, H., DRTON, M. and HAN, F. (2022a). Distribution-free consistent independence tests via center-outward ranks and signs. *J. Amer. Statist. Assoc.* **117** 395–410. MR4399094 https://doi.org/10.1080/01621459.2020.1782223

SHI, H., DRTON, M. and HAN, F. (2022b). On the power of Chatterjee's rank correlation. *Biometrika*. To appear. https://doi.org/10.1093/biomet/asab028

SHI, H., DRTON, M., HALLIN, M. and HAN, F. (2021). Center-outward sign-and rank-based quadrant, Spearman, and Kendall tests for multivariate independence. Preprint. Available at arXiv:2111.15567v1.

SHI, H., HALLIN, M., DRTON, M. and HAN, F. (2022). Supplement to "On universally consistent and fully distribution-free rank tests of vector independence." https://doi.org/10.1214/21-AOS2151SUPP

SPEARMAN, C. (1904). The proof and measurement of association between two things. *Amer. J. Psychol.* **15** 72–101.

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. and RIZZO, M. L. (2013). The distance correlation t-test of independence in high dimension. *J. Multivariate Anal.* **117** 193–213. MR3053543 https://doi.org/10.1016/j.jmva.2013.02.012

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

TASKINEN, S., KANKAINEN, A. and OJA, H. (2003). Sign test of independence between two random vectors. *Statist. Probab. Lett.* **62** 9–21. MR1965367 https://doi.org/10.1016/S0167-7152(02)00399-1

TASKINEN, S., KANKAINEN, A. and OJA, H. (2004). Rank scores tests of multivariate independence. In *Theory and Applications of Recent Robust Methods. Stat. Ind. Technol.* 329–341. Birkhäuser, Basel. MR2088309

TASKINEN, S., OJA, H. and RANDLES, R. H. (2005). Multivariate nonparametric tests of independence. *J. Amer. Statist. Assoc.* **100** 916–925. MR2201019 https://doi.org/10.1198/016214505000000097

WEIHS, L., DRTON, M. and MEINSHAUSEN, N. (2018). Symmetric rank covariances: A generalized framework for nonparametric measures of dependence. *Biometrika* **105** 547–562. MR3842884 https://doi.org/10.1093/biomet/asy021

WILKS, S. S. (1935). On the independence of k sets of normally distributed statistical variables. *Econometrica* **3** 309–326.

YANAGIMOTO, T. (1970). On measures of association and a related problem. *Ann. Inst. Statist. Math.* **22** 57–63. https://doi.org/10.1007/BF02506323

ZHU, L., XU, K., LI, R. and ZHONG, W. (2017). Projection correlation between two random vectors. *Biometrika* **104** 829–843. MR3737307 https://doi.org/10.1093/biomet/asx043

ZUO, Y. and HE, X. (2006). On the limiting distributions of multivariate depth-based rank sum statistics and related tests. *Ann. Statist.* **34** 2879–2896. MR2329471 https://doi.org/10.1214/009053606000000876

The Annals of Statistics
2022, Vol. 50, No. 4, 1960–1991
https://doi.org/10.1214/21-AOS2165

LIMIT THEOREMS FOR DISTRIBUTIONS INVARIANT UNDER GROUPS OF TRANSFORMATIONS

BY MORGANE AUSTERN[1,a] AND PETER ORBANZ[2,b]

[1]*Department of Statistics, Harvard University,* [a]*maustern@fas.harvard.edu*
[2]*Gatsby Computational Neuroscience Unit, University College London,* [b]*porbanz@gatsby.ucl.ac.uk*

A distributional symmetry is invariance of a distribution under a group of transformations. Exchangeability and stationarity are examples. We explain that a result of ergodic theory implies a law of large numbers for such invariant distributions: If the group satisfies suitable conditions, expectations can be estimated by averaging over subsets of transformations, and these estimators are strongly consistent. We show that, if a mixing condition holds, the averages also satisfy a central limit theorem, a Berry–Esseen bound, and concentration. These are extended further to apply to triangular arrays, to randomly subsampled averages, and to a generalization of U-statistics. As applications, we obtain a general limit theorem for exchangeable random structures, and new results on stationary random fields, network models, and a class of marked point processes. We also establish asymptotic normality of the empirical entropy for a large class of processes. Some known results are recovered as special cases, and can hence be interpreted as an outcome of symmetry. The proofs adapt Stein's method.

REFERENCES

[1] ALFSEN, E. M. (1971). *Compact Convex Sets and Boundary Integrals*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete*, *Band* 57. Springer, New York. MR0445271

[2] AMBROISE, C. and MATIAS, C. (2012). New consistent and asymptotically normal parameter estimates for random-graph mixture models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 3–35. MR2885838 https://doi.org/10.1111/j.1467-9868.2011.01009.x

[3] AUSTERN, M. and ORBANZ, P. (2022). Supplement to "Limit theorems for distributions invariant under groups of transformations." https://doi.org/10.1214/21-AOS2165SUPP

[4] BARYSHNIKOV, YU. and YUKICH, J. E. (2005). Gaussian limits for random measures in geometric probability. *Ann. Appl. Probab.* **15** 213–253. MR2115042 https://doi.org/10.1214/105051604000000594

[5] 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

[6] BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd ed. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR1324786

[7] BOLTHAUSEN, E. (1982). On the central limit theorem for stationary mixing random fields. *Ann. Probab.* **10** 1047–1050. MR0672305

[8] BONDAR, J. V. and MILNES, P. (1981). Amenability: A survey for statistical applications of Hunt–Stein and related conditions on groups. *Z. Wahrsch. Verw. Gebiete* **57** 103–128. MR0623457 https://doi.org/10.1007/BF00533716

[9] BORGS, C., CHAYES, J. T., COHN, H. and HOLDEN, N. (2017). Sparse exchangeable graphs and their limits via graphon processes. *J. Mach. Learn. Res.* **18** Paper No. 210, 71. MR3827098

[10] 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

[11] 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

MSC2020 subject classifications. Primary 62G20; secondary 37A30, 62M99, 60F05, 60G09.

Key words and phrases. Asymptotic normality, central limit theorems, Berry–Esseen theorems, pointwise ergodic theorems, Stein's method, symmetry, exchangeability, ergodicity.

[12] BRADLEY, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** 107–144. MR2178042 https://doi.org/10.1214/154957805100000104

[13] BÜHLMANN, H. (1958). Le problème "limite central" pour les variables aléatoires échangeables. *C. R. Acad. Sci. Paris Sér. I Math.* **246** 534–536. MR0092265

[14] CARON, F. and FOX, E. B. (2017). Sparse graphs using exchangeable random measures. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1295–1366. MR3731666 https://doi.org/10.1111/rssb.12233

[15] CHOI, D. S., WOLFE, P. J. and AIROLDI, E. M. (2012). Stochastic blockmodels with a growing number of classes. *Biometrika* **99** 273–284. MR2931253 https://doi.org/10.1093/biomet/asr053

[16] DAVEZIES, L., D'HAULTFŒUILLE, X. and GUYONVARCH, Y. (2021). Empirical process results for exchangeable arrays. *Ann. Statist.* **49** 845–862. MR4255110 https://doi.org/10.1214/20-aos1981

[17] DIACONIS, P. and JANSON, S. (2008). Graph limits and exchangeable random graphs. *Rend. Mat. Appl.* (7) **28** 33–61. MR2463439

[18] DIACONIS, P. and MALLIARIS, M. (2021). Complexity and randomness in the Heisenberg groups (and beyond). *New Zealand J. Math.* **52** 403–426. MR4378158 https://doi.org/10.53733/134

[19] EAGLESON, G. K. and WEBER, N. C. (1978). Limit theorems for weakly exchangeable arrays. *Math. Proc. Cambridge Philos. Soc.* **84** 123–130. MR0501275 https://doi.org/10.1017/S0305004100054967

[20] EINSIEDLER, M. and WARD, T. (2011). *Ergodic Theory*. Springer, Berlin.

[21] GEORGII, H.-O. (2011). *Gibbs Measures and Phase Transitions*, 2nd ed. *De Gruyter Studies in Mathematics* **9**. de Gruyter, Berlin. MR2807681 https://doi.org/10.1515/9783110250329

[22] HEINRICH, L. and MOLCHANOV, I. S. (1999). Central limit theorem for a class of random measures associated with germ-grain models. *Adv. in Appl. Probab.* **31** 283–314. MR1724553 https://doi.org/10.1239/aap/1029955136

[23] HUG, D., LAST, G. and SCHULTE, M. (2016). Second-order properties and central limit theorems for geometric functionals of Boolean models. *Ann. Appl. Probab.* **26** 73–135. MR3449314 https://doi.org/10.1214/14-AAP1086

[24] JAMES, L. F., LIJOI, A. and PRÜNSTER, I. (2009). Posterior analysis for normalized random measures with independent increments. *Scand. J. Stat.* **36** 76–97. MR2508332 https://doi.org/10.1111/j.1467-9469.2008.00609.x

[25] JANSON, S., ŁUCZAK, T. and RUCINSKI, A. (2000). *Random Graphs*. *Wiley-Interscience Series in Discrete Mathematics and Optimization*. Wiley Interscience, New York. MR1782847 https://doi.org/10.1002/9781118032718

[26] JENSEN, J. L. and KÜNSCH, H. R. (1994). On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes. *Ann. Inst. Statist. Math.* **46** 475–486. MR1309718

[27] KALLENBERG, O. (1999). Multivariate sampling and the estimation problem for exchangeable arrays. *J. Theoret. Probab.* **12** 859–883. MR1702867 https://doi.org/10.1023/A:1021692202530

[28] KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd ed. *Probability and Its Applications* (*New York*). Springer, New York. MR1876169 https://doi.org/10.1007/978-1-4757-4015-8

[29] KALLENBERG, O. (2005). *Probabilistic Symmetries and Invariance Principles*. *Probability and Its Applications* (*New York*). Springer, New York. MR2161313

[30] 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

[31] LINDENSTRAUSS, E. (2001). Pointwise theorems for amenable groups. *Invent. Math.* **146** 259–295. MR1865397 https://doi.org/10.1007/s002220100162

[32] LÖH, C. (2017). *Geometric Group Theory*: *An Introduction*. *Universitext*. Springer, Cham. MR3729310 https://doi.org/10.1007/978-3-319-72254-2

[33] 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

[34] MAITRA, A. (1977). Integral representations of invariant measures. *Trans. Amer. Math. Soc.* **229** 209–225. MR0442197 https://doi.org/10.2307/1998506

[35] ORBANZ, P. and ROY, D. M. (2015). Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Trans. Pattern Anal. Mach. Intell.* **37** 437–461.

[36] PANCHENKO, D. (2013). Spin glass models from the point of view of spin distributions. *Ann. Probab.* **41** 1315–1361. MR3098679 https://doi.org/10.1214/11-AOP696

[37] PARISI, G. (1988). *Statistical Field Theory*. Addison-Wesley, Reading.

[38] PENROSE, M. D. (2007). Gaussian limits for random geometric measures. *Electron. J. Probab.* **12** 989–1035. MR2336596 https://doi.org/10.1214/EJP.v12-429

[39] PITMAN, J. (2006). *Combinatorial Stochastic Processes*. *Lecture Notes in Math.* **1875**. Springer, Berlin. MR2245368

[40] Ross, N. (2011). Fundamentals of Stein's method. *Probab. Surv.* **8** 210–293. MR2861132 https://doi.org/10.1214/11-PS182

[41] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0595165

[42] Shields, P. C. (1996). *The Ergodic Theory of Discrete Sample Paths*. *Graduate Studies in Mathematics* **13**. Amer. Math. Soc., Providence, RI. MR1400225 https://doi.org/10.1090/gsm/013

[43] Stroock, D. W. and Zegarliński, B. (1992). The logarithmic Sobolev inequality for discrete spin systems on a lattice. *Comm. Math. Phys.* **149** 175–193. MR1182416

[44] Varadarajan, V. S. (1963). Groups of automorphisms of Borel spaces. *Trans. Amer. Math. Soc.* **109** 191–220. MR0159923 https://doi.org/10.2307/1993903

[45] Veitch, V. and Roy, D. M. (2019). Sampling and estimation for (sparse) exchangeable graphs. *Ann. Statist.* **47** 3274–3299. MR4025742 https://doi.org/10.1214/18-AOS1778

[46] Weiss, B. (2003). Actions of amenable groups. In *Topics in Dynamics and Ergodic Theory*. *London Mathematical Society Lecture Note Series* **310** 226–262. Cambridge Univ. Press, Cambridge. MR2052281 https://doi.org/10.1017/CBO9780511546716.012

The Annals of Statistics
2022, Vol. 50, No. 4, 1992–2020
https://doi.org/10.1214/21-AOS2167

DISTRIBUTED ADAPTIVE GAUSSIAN MEAN ESTIMATION WITH UNKNOWN VARIANCE: INTERACTIVE PROTOCOL HELPS ADAPTATION

BY T. TONY CAI[a] AND HONGJI WEI[b]

Department of Statistics and Data Science, The Wharton School, University of Pennsylvania, [a]*tcai@wharton.upenn.edu,* [b]*hongjiw@wharton.upenn.edu*

Distributed estimation of a Gaussian mean with unknown variance under communication constraints is studied. Necessary and sufficient communication costs under different types of distributed protocols are derived for any estimator that is adaptively rate-optimal over a range of possible values for the variance. Communication-efficient and statistically optimal procedures are developed.

The analysis reveals an interesting and important distinction among different types of distributed protocols: compared to the independent protocols, interactive protocols such as the sequential and blackboard protocols require less communication costs for rate-optimal adaptive Gaussian mean estimation. The lower bound techniques developed in the present paper are novel and can be of independent interest.

REFERENCES

ACHARYA, J., CANONNE, C. L. and TYAGI, H. (2020). Distributed signal detection under communication constraints. In *Conference on Learning Theory* 41–63. PMLR.

BARNES, L. P., HAN, Y. and ÖZGÜR, A. (2019a). Fisher information for distributed estimation under a blackboard communication protocol. In 2019 *IEEE International Symposium on Information Theory* (*ISIT*) 2704–2708. IEEE, Los Alamitos.

BARNES, L. P., HAN, Y. and ÖZGÜR, A. (2019b). Learning distributions from their samples under communication constraints. CoRR. Available at arXiv:1902.02890.

BATTEY, H., FAN, J., LIU, H., LU, J. and ZHU, Z. (2018). Distributed testing and estimation under sparse high dimensional models. *Ann. Statist.* **46** 1352–1382. MR3798006 https://doi.org/10.1214/17-AOS1587

BICKEL, P. J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Ann. Statist.* **9** 1301–1309. MR0630112

BRAVERMAN, M., GARG, A., MA, T., NGUYEN, H. L. and WOODRUFF, D. P. (2016). Communication lower bounds for statistical estimation problems via a distributed data processing inequality. In *STOC'*16—*Proceedings of the* 48*th Annual ACM SIGACT Symposium on Theory of Computing* 1011–1020. ACM, New York. MR3536632 https://doi.org/10.1145/2897518.2897582

CAI, T. T. and LOW, M. G. (2006). Adaptive confidence balls. *Ann. Statist.* **34** 202–228. MR2275240 https://doi.org/10.1214/009053606000000146

CAI, T. T. and WEI, H. (2020). Distributed Gaussian mean estimation under communication constraints: Optimal rates and communication-efficient algorithms. arXiv preprint. Available at arXiv:2001.08877.

CAI, T. T. and WEI, H. (2022a). Distributed nonparametric function estimation: Optimal rate of convergence and cost of adaptation. *Ann. Statist.* **50** 698–725. MR4404917 https://doi.org/10.1214/21-aos2124

CAI, T. T. and WEI, H. (2022b). Supplement to "Distributed Adaptive Gaussian Mean Estimation with Unknown Variance: Interactive Protocol Helps Adaptation." https://doi.org/10.1214/21-AOS2167SUPP

DEISENROTH, M. and NG, J. W. (2015). Distributed Gaussian processes. In *Proceedings of the* 32*nd International Conference on Machine Learning* (F. Bach and D. Blei, eds.). *Proceedings of Machine Learning Research* **37** 1481–1490.

DIAKONIKOLAS, I., GRIGORESCU, E., LI, J., NATARAJAN, A., ONAK, K. and SCHMIDT, L. (2017). Communication-efficient distributed learning of discrete distributions. In *Advances in Neural Information Processing Systems* 6391–6401.

MSC2020 subject classifications. Primary 62F30; secondary 62B10, 62F12.

Key words and phrases. Adaptive estimation, communication constraints, distributed learning, Gaussian mean, minimax lower bound, optimal rate of convergence.

DONOHO, D. L. and JOHNSTONE, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90** 1200–1224. MR1379464

FAN, J., WANG, D., WANG, K. and ZHU, Z. (2019). Distributed estimation of principal eigenspaces. *Ann. Statist.* **47** 3009–3031. MR4025733 https://doi.org/10.1214/18-AOS1713

GARG, A., MA, T. and NGUYEN, H. (2014). On communication cost of distributed statistical estimation and dimensionality. In *Advances in Neural Information Processing Systems* 2726–2734.

HAN, Y., ÖZGÜR, A. and WEISSMAN, T. (2021). Geometric lower bounds for distributed parameter estimation under communication constraints. *IEEE Trans. Inf. Theory* **67** 8248–8263. MR4346086 https://doi.org/10.1109/TIT.2021.3108952

JOHNSTONE, I. M. (2017). Gaussian Estimation: Sequence and Wavelet Models. Manuscript.

JORDAN, M. I., LEE, J. D. and YANG, Y. (2019). Communication-efficient distributed statistical inference. *J. Amer. Statist. Assoc.* **114** 668–681. MR3963171 https://doi.org/10.1080/01621459.2018.1429274

KIPNIS, A. and DUCHI, J. C. (2017). Mean estimation from adaptive one-bit measurements. In 2017 55*th Annual Allerton Conference on Communication, Control, and Computing* (*Allerton*) 1000–1007.

KLEINER, A., TALWALKAR, A., SARKAR, P. and JORDAN, M. I. (2014). A scalable bootstrap for massive data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 795–816. MR3248677 https://doi.org/10.1111/rssb.12050

KUSHILEVITZ, E. (1997). Communication complexity. In *Advances in Computers* **44** 331–360. Elsevier, Amsterdam.

LEE, J. D., LIU, Q., SUN, Y. and TAYLOR, J. E. (2017). Communication-efficient sparse regression. *J. Mach. Learn. Res.* **18** 5. MR3625709

LIU, J. (2021). A few interactions improve distributed nonparametric estimation. optimally.

POLYAK, B. T. (1990). A new method of stochastic approximation type. *Avtomat. i Telemekh.* **7** 98–107. MR1071220

SZABÓ, B. and VAN ZANTEN, H. (2020a). Adaptive distributed methods under communication constraints. *Ann. Statist.* **48** 2347–2380. MR4134798 https://doi.org/10.1214/19-AOS1890

SZABÓ, B. and VAN ZANTEN, H. (2020b). Distributed function estimation: Adaptation using minimal communication. arXiv preprint. Available at arXiv:2003.12838.

SZABÓ, B., VUURSTEEN, L. and VAN ZANTEN, H. (2020). Optimal distributed testing in high-dimensional gaussian models. CoRR. Available at arXiv:2012.04957.

XIANG, Y. and KIM, Y.-H. (2013). Interactive hypothesis testing against independence. In 2013 *IEEE International Symposium on Information Theory* 2840–2844. IEEE, Los Alamitos.

ZHANG, Y., DUCHI, J., JORDAN, M. I. and WAINWRIGHT, M. J. (2013). Information-theoretic lower bounds for distributed statistical estimation with communication constraints. In *Advances in Neural Information Processing Systems* 2328–2336.

ZHU, Y. and LAFFERTY, J. (2018). Distributed nonparametric regression under communication constraints. arXiv preprint. Available at arXiv:1803.01302.

The Annals of Statistics
2022, Vol. 50, No. 4, 2021–2046
https://doi.org/10.1214/22-AOS2169

ASYMPTOTIC ACCURACY OF THE SADDLEPOINT APPROXIMATION FOR MAXIMUM LIKELIHOOD ESTIMATION

BY JESSE GOODMAN[a]

Department of Statistics, University of Auckland, [a]*jesse.goodman@auckland.ac.nz*

The saddlepoint approximation gives an approximation to the density of a random variable in terms of its moment generating function. When the underlying random variable is itself the sum of n unobserved i.i.d. terms, the basic classical result is that the relative error in the density is of order $1/n$. If instead the approximation is interpreted as a likelihood and maximised as a function of model parameters, the result is an approximation to the maximum likelihood estimate (MLE) that can be much faster to compute than the true MLE. This paper proves the analogous basic result for the approximation error between the saddlepoint MLE and the true MLE: subject to certain explicit identifiability conditions, the error has asymptotic size $O(1/n^2)$ for some parameters and $O(1/n^{3/2})$ or $O(1/n)$ for others. In all three cases, the approximation errors are asymptotically negligible compared to the inferential uncertainty.

The proof is based on a factorisation of the saddlepoint likelihood into an exact and approximate term, along with an analysis of the approximation error in the gradient of the log-likelihood. This factorisation also gives insight into alternatives to the saddlepoint approximation, including a new and simpler saddlepoint approximation, for which we derive analogous error bounds. As a corollary of our results, we also obtain the asymptotic size of the MLE approximation error when the saddlepoint approximation is replaced by the normal approximation.

REFERENCES

[1] BARNDORFF-NIELSEN, O. (2014). *Information and Exponential Families in Statistical Theory*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR3221776 https://doi.org/10.1002/9781118857281

[2] BARNDORFF-NIELSEN, O. E. and KLÜPPELBERG, C. (1999). Tail exactness of multivariate saddlepoint approximations. *Scand. J. Stat.* **26** 253–264. MR1707583 https://doi.org/10.1111/1467-9469.00148

[3] BLEISTEIN, N. and HANDELSMAN, R. A. (1986). *Asymptotic Expansions of Integrals*, 2nd ed. Dover, New York. MR0863284

[4] BUTLER, R. W. (2007). *Saddlepoint Approximations with Applications*. *Cambridge Series in Statistical and Probabilistic Mathematics* **22**. Cambridge Univ. Press, Cambridge. MR2357347 https://doi.org/10.1017/CBO9780511619083

[5] DANIELS, H. E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Stat.* **25** 631–650. MR0066602 https://doi.org/10.1214/aoms/1177728652

[6] DANIELS, H. E. (1982). The saddlepoint approximation for a general birth process. *J. Appl. Probab.* **19** 20–28. MR0644416 https://doi.org/10.2307/3213912

[7] DAVISON, A. C., HAUTPHENNE, S. and KRAUS, A. (2021). Parameter estimation for discretely observed linear birth-and-death processes. *Biometrics* **77** 186–196. MR4229731 https://doi.org/10.1111/biom.13282

[8] DEN HOLLANDER, F. (2000). *Large Deviations*. *Fields Institute Monographs* **14**. Amer. Math. Soc., Providence, RI. MR1739680 https://doi.org/10.1007/s00440-009-0235-5

[9] GOODMAN, J. (2022). Supplement to "Asymptotic accuracy of the saddlepoint approximation for maximum likelihood estimation." https://doi.org/10.1214/22-AOS2169SUPP

[10] JENSEN, J. L. (1988). Uniform saddlepoint approximations. *Adv. in Appl. Probab.* **20** 622–634. MR0955507 https://doi.org/10.2307/1427038

MSC2020 subject classifications. Primary 62F12; secondary 41A60, 62F15.
Key words and phrases. Saddlepoint approximation, maximum likelihood estimation, approximation error.

[11] JENSEN, J. L. (1995). *Saddlepoint Approximations*. *Oxford Statistical Science Series* **16**. The Clarendon Press, Oxford University Press, New York. MR1354837
[12] KOLASSA, J. E. (2006). *Series Approximation Methods in Statistics*, 3rd ed. *Lecture Notes in Statistics* **88**. Springer, New York. MR1295242 https://doi.org/10.1007/978-1-4757-4275-6
[13] LUNDE, B. Å. S., KLEPPE, T. S. and SKAUG, H. J. (2018). Saddlepoint adjusted inversion of characteristic functions. Preprint. Available at arXiv:1811.05678 [stat.CO].
[14] OGDEN, H. E. (2017). On asymptotic validity of naive inference with an approximate likelihood. *Biometrika* **104** 153–164. MR3626485 https://doi.org/10.1093/biomet/asx002
[15] OGDEN, H. E. (2018). On the error in Laplace approximations of high-dimensional integrals. Preprint. Available at arXiv:1808.06341 [math.ST].
[16] PEDELI, X., DAVISON, A. C. and FOKIANOS, K. (2015). Likelihood estimation for the INAR(p) model by saddlepoint approximation. *J. Amer. Statist. Assoc.* **110** 1229–1238. MR3420697 https://doi.org/10.1080/01621459.2014.983230
[17] REID, N. (1988). Saddlepoint methods and statistical inference. *Statist. Sci.* **3** 213–227. MR0968390
[18] RUDIN, W. (1976). *Principles of Mathematical Analysis*, 3rd ed. *International Series in Pure and Applied Mathematics*. McGraw-Hill, New York. MR0385023
[19] WONG, R. (2001). *Asymptotic Approximations of Integrals*. *Classics in Applied Mathematics* **34**. SIAM, Philadelphia, PA. MR1851050 https://doi.org/10.1137/1.9780898719260
[20] WOOD, A. T. A., BOOTH, J. G. and BUTLER, R. W. (1993). Saddlepoint approximations to the CDF of some statistics with nonnormal limit distributions. *J. Amer. Statist. Assoc.* **88** 680–686. MR1224395 https://doi.org/10.2307/2290351
[21] ZHANG, W., BRAVINGTON, M. V. and FEWSTER, R. M. (2019). Fast likelihood-based inference for latent count models using the saddlepoint approximation. *Biometrics* **75** 723–733. MR4012079 https://doi.org/10.1111/biom.13030

The Annals of Statistics
2022, Vol. 50, No. 4, 2047–2071
https://doi.org/10.1214/22-AOS2171

GENERALIZATION ERROR BOUNDS OF DYNAMIC TREATMENT REGIMES IN PENALIZED REGRESSION-BASED LEARNING

BY EUN JEONG OH[a], MIN QIAN[b] AND YING KUEN CHEUNG[c]

Department of Biostatistics, Columbia University, [a]*eo2366@caa.columbia.edu,* [b]*mq2158@cumc.columbia.edu,* [c]*yc632@cumc.columbia.edu*

A dynamic treatment regime (DTR) is a sequence of decision rules, one per stage of intervention, that maps up-to-date patient information to a recommended treatment. Discovering an appropriate DTR for a given disease is a challenging issue especially when a large set of prognostic variables are observed. To address this problem, we propose penalized regression-based learning methods with l_1 penalty to estimate the optimal DTR that would maximize the expected outcome if implemented. We also provide generalization error bounds of the estimated DTR in the setting of finite number of stages with multiple treatment options. We first examine the relationship between value and Q-functions and derive a finite sample upper bound on the difference in values between the optimal and the estimated DTRs. For practical implementation, we develop an algorithm with partial regularization via orthogonality to construct the optimal DTR. The advantages of the proposed methods are demonstrated with extensive simulation studies and data analysis of depression clinical trials.

REFERENCES

BLATT, D., MURPHY, S. A. and ZHU, J. (2004). A-learning for approximate planning.

BUBECK, S., PERCHET, V. and RIGOLLET, P. (2013). Bounded regret in stochastic multiarmed bandits. In *Proceedings of the* 26*th Annual Conference on Learning Theory*.

CHAKRABORTY, B. and MOODIE, E. E. M. (2013). *Statistical Methods for Dynamic Treatment Regimes*: *Reinforcement Learning*, *Causal Inference*, *and Personalized Medicine*. *Statistics for Biology and Health*. Springer, New York. MR3112454 https://doi.org/10.1007/978-1-4614-7428-9

CHEUNG, Y. K., CHAKRABORTY, B. and DAVIDSON, K. W. (2015). Sequential multiple assignment randomized trial (SMART) with adaptive randomization for quality improvement in depression treatment program. *Biometrics* **71** 450–459. MR3366249 https://doi.org/10.1111/biom.12258

DAVIDSON, K. W., RIECKMANN, N., CLEMOW, L., SCHWARTZ, J. E., SHIMBO, D., MEDINA, V., ALBANESE, G., KRONISH, I., HEGEL, M. et al. (2010). Enhanced depression care for patients with acute coronary syndrome and persistent depressive symptoms: Coronary psychosocial evaluation studies randomized controlled trial. *Arch. Intern. Med.* **170** 600–608.

DAVIDSON, K. W., BIGGER, J. T., BURG, M. M., CARNEY, R. M., CHAPLIN, W. F., CZAJKOWSKI, S., DORNELAS, E., DUER-HEFELE, J., FRASURE-SMITH, N. et al. (2013). Centralized, stepped, patient preference-based treatment for patients with post-acute coronary syndrome depression: CODIACS vanguard randomized controlled trial. *J. Am. Med. Assoc. Intern. Med.* **173** 997–1004.

ERTEFAIE, A. and STRAWDERMAN, R. L. (2018). Constructing dynamic treatment regimes over indefinite time horizons. *Biometrika* **105** 963–977. MR3877877 https://doi.org/10.1093/biomet/asy043

ERTEFAIE, A., MCKAY, J. R., OSLIN, D. and STRAWDERMAN, R. L. (2021). Robust Q-learning. *J. Amer. Statist. Assoc.* **116** 368–381. MR4227700 https://doi.org/10.1080/01621459.2020.1753522

FAN, A., LU, W. and SONG, R. (2016). Sequential advantage selection for optimal treatment regime. *Ann. Appl. Stat.* **10** 32–53. MR3480486 https://doi.org/10.1214/15-AOAS849

FOSTER, J. C., TAYLOR, J. M. G. and RUBERG, S. J. (2011). Subgroup identification from randomized clinical trial data. *Stat. Med.* **30** 2867–2880. MR2844689 https://doi.org/10.1002/sim.4322

GEY, S. (2012). Risk bounds for CART classifiers under a margin condition. *Pattern Recognit.* **45** 3523–3534.

MSC2020 subject classifications. Primary 62H99, 62J07; secondary 62P10.

Key words and phrases. Personalized medicine, regression-based learning, treatment decision making, variable selection, backward induction.

HENDERSON, R., ANSELL, P. and ALSHIBANI, D. (2010). Regret-regression for optimal dynamic treatment regimes. *Biometrics* **66** 1192–1201. MR2758507 https://doi.org/10.1111/j.1541-0420.2009.01368.x
JIANG, R., LU, W., SONG, R., HUDGENS, M. G. and NAPRVAVNIK, S. (2017). Doubly robust estimation of optimal treatment regimes for survival data—with application to an HIV/AIDS study. *Ann. Appl. Stat.* **11** 1763–1786. MR3709577 https://doi.org/10.1214/17-AOAS1057
JIANG, B., SONG, R., LI, J. and ZENG, D. (2019). Entropy learning for dynamic treatment regimes. *Statist. Sinica* **29** 1633–1656. MR3970323
LABER, E. B. and ZHAO, Y. Q. (2015). Tree-based methods for individualized treatment regimes. *Biometrika* **102** 501–514. MR3394271 https://doi.org/10.1093/biomet/asv028
LABER, E. B., LIZOTTE, D. J., QIAN, M., PELHAM, W. E. and MURPHY, S. A. (2014). Dynamic treatment regimes: Technical challenges and applications. *Electron. J. Stat.* **8** 1225–1272. MR3263118 https://doi.org/10.1214/14-EJS920
LATTIMORE, T. and MUNOS, R. (2014). Bounded regret for finite-armed structured bandits. In *Advances in Neural Information Processing Systems* **27** 550–558.
LAVORI, P. W., DAWSON, R. and RUSH, A. J. (2000). Flexible treatment strategies in chronic disease: Clinical and research implications. *Biol. Psychiatry* **48** 605–614.
LIPKOVICH, I., DMITRIENKO, A., DENNE, J. and ENAS, G. (2011). Subgroup identification based on differential effect search—a recursive partitioning method for establishing response to treatment in patient subpopulations. *Stat. Med.* **30** 2601–2621. MR2815438 https://doi.org/10.1002/sim.4289
LIU, Y., WANG, Y., KOSOROK, M. R., ZHAO, Y. and ZENG, D. (2018). Augmented outcome-weighted learning for estimating optimal dynamic treatment regiments. *Stat. Med.* **37** 3776–3788. MR3869154 https://doi.org/10.1002/sim.7844
LU, W., ZHANG, H. H. and ZENG, D. (2013). Variable selection for optimal treatment decision. *Stat. Methods Med. Res.* **22** 493–504. MR3190671 https://doi.org/10.1177/0962280211428383
LUCKETT, D. J., LABER, E. B., KAHKOSKA, A. R., MAAHS, D. M., MAYER-DAVIS, E. and KOSOROK, M. R. (2020). Estimating dynamic treatment regimes in mobile health using V-learning. *J. Amer. Statist. Assoc.* **115** 692–706. MR4107673 https://doi.org/10.1080/01621459.2018.1537919
MARIN, F., GONZALEZ-CONEJERO, R., CAPRANZANO, P., BASS, T. A., ROLDAN, V. and ANGIOLILLO, D. J. (2009). Pharmacogenetics in cardiovascular antithrombotic therapy. *J. Am. Coll. Cardiol.* **54** 1041–1057.
MOODIE, E. E. M., DEAN, N. and SUN, Y. R. (2014). Q-learning: Flexible learning about useful utilities. *Stat. Biosci.* **6** 223–243.
MOODIE, E. E. M., RICHARDSON, T. S. and STEPHENS, D. A. (2007). Demystifying optimal dynamic treatment regimes. *Biometrics* **63** 447–455. MR2370803 https://doi.org/10.1111/j.1541-0420.2006.00686.x
MURPHY, S. A. (2003). Optimal dynamic treatment regimes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 331–366. MR1983752 https://doi.org/10.1111/1467-9868.00389
MURPHY, S. A. (2005). A generalization error for Q-learning. *J. Mach. Learn. Res.* **6** 1073–1097. MR2249849
MURPHY, S. A., VAN DER LAAN, M. J. and ROBINS, J. M. (2001). Marginal mean models for dynamic regimes. *J. Amer. Statist. Assoc.* **96** 1410–1423. MR1946586 https://doi.org/10.1198/016214501753382327
MURPHY, S. A., OSLIN, D. W., RUSH, A. J. and ZHU, J. (2007). Methodological challenges in constructing effective treatment sequences for chronic psychiatric disorders. *Neuropsychoarmacology* **32** 257–262.
OH, E. J., QIAN, M. and CHEUNG, Y. K. (2022). Supplement to "Generalization error bounds of dynamic treatment regimes in penalized regression-based learning." https://doi.org/10.1214/22-AOS2171SUPP
OH, E. J., QIAN, M., CHEUNG, K. and MOHR, D. C. (2020). Building health application recommender system using partially penalized regression. In *Statistical Modeling in Biomedical Research* 105–123. Springer, Berlin.
PINEAU, J., BELLERNARE, M. G., RUSH, A. J., GHIZARU, A. and MURPHY, S. A. (2007). Constructing evidence-based treatment strategies using methods from computer science. *Drug Alcohol Depend.* **88** S52–S60.
PIPER, W. E., BOROTO, D. R., JOYCE, A. S., MCCALLUM, M. and AZIM, H. F. A. (1995). Pattern of alliance and outcome in short-term individual psychotherapy. *Psychotherapy* **32** 639–647.
QI, Z. and LIU, Y. (2018). D-learning to estimate optimal individual treatment rules. *Electron. J. Stat.* **12** 3601–3638. MR3870507 https://doi.org/10.1214/18-ejs1480
QIAN, M. and MURPHY, S. A. (2011). Performance guarantees for individualized treatment rules. *Ann. Statist.* **39** 1180–1210. MR2816351 https://doi.org/10.1214/10-AOS864
ROBINS, J. M. (1989). The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In *Health Service Research Methodology*: *A Focus on AIDS* 113–159.
ROBINS, J. M. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. In *Proceedings of the Biopharmaceutical Section* 24–33. Am. Statist. Assoc., Alexandria.

ROBINS, J. M. (1997). Causal inference from complex longitudinal data. In *Latent Variable Modeling and Applications to Causality* (*Los Angeles*, *CA*, 1994). *Lect. Notes Stat.* **120** 69–117. Springer, New York. MR1601279 https://doi.org/10.1007/978-1-4612-1842-5_4

ROBINS, J. M. (1998). Marginal structural models. In *Proceedings of the American Statistical Association. Section on Bayesian Statistical Science* 1–10.

ROBINS, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium in Biostatistics. Lect. Notes Stat.* **179** 189–326. Springer, New York. MR2129402 https://doi.org/10.1007/978-1-4419-9076-1_11

ROBINS, J., ORELLANA, L. and ROTNITZKY, A. (2008). Estimation and extrapolation of optimal treatment and testing strategies. *Stat. Med.* **27** 4678–4721. MR2528576 https://doi.org/10.1002/sim.3301

RUDIN, C. and ERTEKIN, Ş. (2018). Learning customized and optimized lists of rules with mathematical programming. *Math. Program. Comput.* **10** 659–702. MR3863707 https://doi.org/10.1007/s12532-018-0143-8

SCHULTE, P. J., TSIATIS, A. A., LABER, E. B. and DAVIDIAN, M. (2014). *Q*- and *A*-learning methods for estimating optimal dynamic treatment regimes. *Statist. Sci.* **29** 640–661. MR3300363 https://doi.org/10.1214/13-STS450

SHI, C., SONG, R. and LU, W. (2016). Robust learning for optimal treatment decision with NP-dimensionality. *Electron. J. Stat.* **10** 2894–2921. MR3557316 https://doi.org/10.1214/16-EJS1178

SHI, C., FAN, A., SONG, R. and LU, W. (2018). High-dimensional *A*-learning for optimal dynamic treatment regimes. *Ann. Statist.* **46** 925–957. MR3797992 https://doi.org/10.1214/17-AOS1570

SONG, R., WANG, W., ZENG, D. and KOSOROK, M. R. (2015). Penalized *Q*-learning for dynamic treatment regimens. *Statist. Sinica* **25** 901–920. MR3409730

SU, X., ZHOU, T., YAN, X., FAN, J. and YANG, S. (2008). Interaction trees with censored survival data. *Int. J. Biostat.* **4** Art. 2, 28. MR2383729 https://doi.org/10.2202/1557-4679.1071

TIAN, L., ALIZADEH, A. A., GENTLES, A. J. and TIBSHIRANI, R. (2014). A simple method for estimating interactions between a treatment and a large number of covariates. *J. Amer. Statist. Assoc.* **109** 1517–1532. MR3293607 https://doi.org/10.1080/01621459.2014.951443

TIBSHIRANI, R. (2011). Regression shrinkage and selection via the lasso: A retrospective. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 273–282. MR2815776 https://doi.org/10.1111/j.1467-9868.2011.00771.x

TSYBAKOV, A. B. (2004). Optimal aggregation of classifiers in statistical learning. *Ann. Statist.* **32** 135–166. MR2051002 https://doi.org/10.1214/aos/1079120131

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

WALLACE, M. P. and MOODIE, E. E. M. (2015). Doubly-robust dynamic treatment regimen estimation via weighted least squares. *Biometrics* **71** 636–644. MR3402599 https://doi.org/10.1111/biom.12306

WATKINS, C. J. (1989). Learning from delayed rewards. Ph.D. thesis, Univ. Cambridge England.

ZHANG, B. and ZHANG, M. (2018). C-learning: A new classification framework to estimate optimal dynamic treatment regimes. *Biometrics* **74** 891–899. MR3860710 https://doi.org/10.1111/biom.12836

ZHANG, B., TSIATIS, A. A., LABER, E. B. and DAVIDIAN, M. (2013). Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika* **100** 681–694. MR3094445 https://doi.org/10.1093/biomet/ast014

ZHANG, Y., LABER, E. B., TSIATIS, A. and DAVIDIAN, M. (2015). Using decision lists to construct interpretable and parsimonious treatment regimes. *Biometrics* **71** 895–904. MR3436715 https://doi.org/10.1111/biom.12354

ZHANG, Y., LABER, E. B., DAVIDIAN, M. and TSIATIS, A. A. (2018). Interpretable dynamic treatment regimes. *J. Amer. Statist. Assoc.* **113** 1541–1549.

ZHAO, Y., ZENG, D., RUSH, A. J. and KOSOROK, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *J. Amer. Statist. Assoc.* **107** 1106–1118. MR3010898 https://doi.org/10.1080/01621459.2012.695674

ZHAO, Y.-Q., ZENG, D., LABER, E. B. and KOSOROK, M. R. (2015). New statistical learning methods for estimating optimal dynamic treatment regimes. *J. Amer. Statist. Assoc.* **110** 583–598. MR3367249 https://doi.org/10.1080/01621459.2014.937488

ZHAO, Y.-Q., ZHU, R., CHEN, G. and ZHENG, Y. (2020). Constructing dynamic treatment regimes with shared parameters for censored data. *Stat. Med.* **39** 1250–1263. MR4098488 https://doi.org/10.1002/sim.8473

ZHOU, X., MAYER-HAMBLETT, N., KHAN, U. and KOSOROK, M. R. (2017). Residual weighted learning for estimating individualized treatment rules. *J. Amer. Statist. Assoc.* **112** 169–187. MR3646564 https://doi.org/10.1080/01621459.2015.1093947

ZHU, R., ZENG, D. and KOSOROK, M. R. (2015). Reinforcement learning trees. *J. Amer. Statist. Assoc.* **110** 1770–1784. MR3449072 https://doi.org/10.1080/01621459.2015.1036994

ZHU, W., ZENG, D. and SONG, R. (2019). Proper inference for value function in high-dimensional Q-learning for dynamic treatment regimes. *J. Amer. Statist. Assoc.* **114** 1404–1417. MR4011788 https://doi.org/10.1080/01621459.2018.1506341

ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279469 https://doi.org/10.1198/016214506000000735

ZOU, H. and ZHANG, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. *Ann. Statist.* **37** 1733–1751. MR2533470 https://doi.org/10.1214/08-AOS625

The Annals of Statistics
2022, Vol. 50, No. 4, 2072–2095
https://doi.org/10.1214/22-AOS2176

A CLT FOR SECOND DIFFERENCE ESTIMATORS WITH AN APPLICATION TO VOLATILITY AND INTENSITY

BY EMIL A. STOLTENBERG[1,a], PER A. MYKLAND[2,b] AND LAN ZHANG[3,c]

[1]*Department of Data Science and Analytics, BI Norwegian Business School,* [a]*emilstoltenberg@gmail.com*
[2]*Department of Statistics, The University of Chicago,* [b]*mykland@pascal.uchicago.edu*
[3]*Department of Finance, University of Illinois at Chicago,* [c]*lanzhang@uic.edu*

In this paper, we introduce a general method for estimating the quadratic covariation of one or more spot parameter processes associated with continuous time semimartingales, and present a central limit theorem that has this class of estimators as one of its applications. The class of estimators we introduce, that we call Two-Scales Quadratic Covariation (TSQC) estimators, is based on sums of increments of second differences of the observed processes, and the intervals over which the differences are computed are rolling and overlapping. This latter feature lets us take full advantage of the data, and, by sufficiency considerations, ought to outperform estimators that are based on only one partition of the observational window. Moreover, a two-scales approach is employed to deal with asymptotic bias terms in a systematic manner, thus automatically giving consistent estimators without having to work out the form of the bias term on a case-to-case basis. We highlight the versatility of our central limit theorem by applying it to a novel leverage effect estimator that does not belong to the class of TSQC estimators. The principal empirical motivation for the present study is that the discrete times at which a continuous time semimartingale is observed might depend on features of the observable process other than its level, such as its spot-volatility process. As an application of the TSQC estimators, we therefore show how it may be used to estimate the quadratic covariation between the spot-volatility process and the intensity process of the observation times, when both of these are taken to be semimartingales. The finite sample properties of this estimator are studied by way of a simulation experiment, and we also apply this estimator in an empirical analysis of the Apple stock. Our analysis of the Apple stock indicates a rather strong correlation between the spot volatility process of the log-prices process and the times at which this stock is traded and hence observed.

REFERENCES

AÏT-SAHALIA, Y. and JACOD, J. (2014). *High-Frequency Financial Econometrics*. Princeton Univ. Press, Princeton.

AÏT-SAHALIA, Y., MYKLAND, P. A. and ZHANG, L. (2011). Ultra high frequency volatility estimation with dependent microstructure noise. *J. Econometrics* **160** 160–175. MR2745875 https://doi.org/10.1016/j.jeconom.2010.03.028

AÏT-SAHALIA, Y., FAN, J., LAEVEN, R. J. A., WANG, C. D. and YANG, X. (2017). Estimation of the continuous and discontinuous leverage effects. *J. Amer. Statist. Assoc.* **112** 1744–1758. MR3750896 https://doi.org/10.1080/01621459.2016.1240082

ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. and LABYS, P. (2001). The distribution of realized exchange rate volatility. *J. Amer. Statist. Assoc.* **96** 42–55. MR1952727 https://doi.org/10.1198/016214501750332965

MSC2020 subject classifications. Primary 60F05, 60G44, 62M09, 62M10; secondary 60G42, 62G20, 62P20.

Key words and phrases. Asynchronous times, central limit theorem, consistency, convergence rates, counting processes, endogenous observation times, high frequency, intensity, irregular times, microstructure, observed asymptotic variance, overlapping intervals, rolling intervals, sufficiency, two-scales estimation.

BARNDORFF-NIELSEN, O. E. and SHEPHARD, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 253–280. MR1904704 https://doi.org/10.1111/1467-9868.00336

BARNDORFF-NIELSEN, O. E. and SHEPHARD, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica* **72** 885–925. MR2051439 https://doi.org/10.1111/j.1468-0262.2004.00515.x

BIBINGER, M. and MYKLAND, P. A. (2016). Inference for multi-dimensional high-frequency data with an application to conditional independence testing. *Scand. J. Stat.* **43** 1078–1102. MR3573677 https://doi.org/10.1111/sjos.12230

BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd ed. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR1324786

BOLLERSLEV, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31** 307–327. MR0853051 https://doi.org/10.1016/0304-4076(86)90063-1

CHAN, N. H. and WEI, C. Z. (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *Ann. Statist.* **15** 1050–1063. MR0902245 https://doi.org/10.1214/aos/1176350492

COX, J. C., INGERSOLL, J. E. JR. and ROSS, S. A. (1985). A theory of the term structure of interest rates. *Econometrica* **53** 385–407. MR0785475 https://doi.org/10.2307/1911242

ENGLE, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50** 987–1007. MR0666121 https://doi.org/10.2307/1912773

ENGLE, R. F. (1995). *ARCH*: *Selected Readings*. Oxford Univ. Press, New York, NY.

FELLER, W. (1951). Two singular diffusion problems. *Ann. of Math.* (2) **54** 173–182. MR0054814 https://doi.org/10.2307/1969318

FUKASAWA, M. and ROSENBAUM, M. (2012). Central limit theorems for realized volatility under hitting times of an irregular grid. *Stochastic Process. Appl.* **122** 3901–3920. MR2971719 https://doi.org/10.1016/j.spa.2012.08.005

HAYASHI, T., JACOD, J. and YOSHIDA, N. (2011). Irregular sampling and central limit theorems for power variations: The continuous case. *Ann. Inst. Henri Poincaré Probab. Stat.* **47** 1197–1218. MR2884231 https://doi.org/10.1214/11-AIHP432

JACOD, J. (1997). On continuous conditional Gaussian martingales and stable convergence in law. In *Séminaire de Probabilités*, *XXXI*. *Lecture Notes in Math.* **1655** 232–246. Springer, Berlin. MR1478732 https://doi.org/10.1007/BFb0119308

JACOD, J., LI, Y. and ZHENG, X. (2017). Statistical properties of microstructure noise. *Econometrica* **85** 1133–1174. MR3681766 https://doi.org/10.3982/ECTA13085

JACOD, J., LI, Y. and ZHENG, X. (2019). Estimating the integrated volatility with tick observations. *J. Econometrics* **208** 80–100. MR3906962 https://doi.org/10.1016/j.jeconom.2018.09.006

JACOD, J. and PROTTER, P. (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.* **26** 267–307. MR1617049 https://doi.org/10.1214/aop/1022855419

JACOD, J. and PROTTER, P. (2012). *Discretization of Processes*. *Stochastic Modelling and Applied Probability* **67**. Springer, Heidelberg. MR2859096 https://doi.org/10.1007/978-3-642-24127-7

JACOD, J. and SHIRYAEV, A. N. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **288**. Springer, Berlin. MR1943877 https://doi.org/10.1007/978-3-662-05265-5

JACOD, J. and TODOROV, V. (2010). Do price and volatility jump together? *Ann. Appl. Probab.* **20** 1425–1469. MR2676944 https://doi.org/10.1214/09-AAP654

KALNINA, I. and XIU, D. (2017). Nonparametric estimation of the leverage effect: A trade-off between robustness and efficiency. *J. Amer. Statist. Assoc.* **112** 384–396. MR3646579 https://doi.org/10.1080/01621459.2016.1141687

LI, J., TODOROV, V. and TAUCHEN, G. (2017a). Jump regressions. *Econometrica* **85** 173–195. MR3611769 https://doi.org/10.3982/ECTA12962

LI, J., TODOROV, V. and TAUCHEN, G. (2017b). Robust jump regressions. *J. Amer. Statist. Assoc.* **112** 332–341. MR3646575 https://doi.org/10.1080/01621459.2016.1138866

LI, Y., ZHANG, Z. and ZHENG, X. (2013). Volatility inference in the presence of both endogenous time and microstructure noise. *Stochastic Process. Appl.* **123** 2696–2727. MR3054542 https://doi.org/10.1016/j.spa.2013.04.002

LI, Y., MYKLAND, P. A., RENAULT, E., ZHANG, L. and ZHENG, X. (2014). Realized volatility when sampling times are possibly endogenous. *Econometric Theory* **30** 580–605. MR3205607 https://doi.org/10.1017/S0266466613000418

MYKLAND, P. A. and ZHANG, L. (2006). ANOVA for diffusions and Itô processes. *Ann. Statist.* **34** 1931–1963. MR2283722 https://doi.org/10.1214/009053606000000452

MYKLAND, P. A. and ZHANG, L. (2008). Inference for volatility-type objects and implications for hedging. *Stat. Interface* **1** 255–278. MR2476743 https://doi.org/10.4310/SII.2008.v1.n2.a4

MYKLAND, P. A. and ZHANG, L. (2009). Inference for continuous semimartingales observed at high frequency. *Econometrica* **77** 1403–1445. MR2561071 https://doi.org/10.3982/ECTA7417

MYKLAND, P. A. and ZHANG, L. (2012). The econometrics of high-frequency data. In *Statistical Methods for Stochastic Differential Equations*. *Monogr. Statist. Appl. Probab.* **124** 109–190. CRC Press, Boca Raton, FL. MR2976983 https://doi.org/10.1201/b12126-3

MYKLAND, P. A. and ZHANG, L. (2017a). Assessment of uncertainty in high frequency data: The observed asymptotic variance. *Econometrica* **85** 197–231. MR3611770 https://doi.org/10.3982/ECTA12501

MYKLAND, P. A. and ZHANG, L. (2017b). Supplement to "Assessment of uncertainty in high frequency data: The observed asymptotic variance". *Econometrica* **85**. Available at https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA12501.

MYKLAND, P. A., ZHANG, L. and CHEN, D. (2019). The algebra of two scales estimation, and the S-TSRV: High frequency estimation that is robust to sampling times. *J. Econometrics* **208** 101–119. MR3906963 https://doi.org/10.1016/j.jeconom.2018.09.007

PHILLIPS, P. C. B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika* **74** 535–547. MR0909357 https://doi.org/10.1093/biomet/74.3.535

POTIRON, Y. and MYKLAND, P. A. (2017). Estimation of integrated quadratic covariation with endogenous sampling times. *J. Econometrics* **197** 20–41. MR3598643 https://doi.org/10.1016/j.jeconom.2016.10.004

PROTTER, P. E. (2004). *Stochastic Integration and Differential Equations*, 2nd ed. *Applications of Mathematics* (*New York*) **21**. Springer, Berlin. MR2020294

REISS, M., TODOROV, V. and TAUCHEN, G. (2015). Nonparametric test for a constant beta between Itô semi-martingales based on high-frequency data. *Stochastic Process. Appl.* **125** 2955–2988. MR3343284 https://doi.org/10.1016/j.spa.2015.02.008

RENAULT, E. and WERKER, B. J. M. (2011). Causality effects in return volatility measures with random times. *J. Econometrics* **160** 272–279. MR2745883 https://doi.org/10.1016/j.jeconom.2010.03.036

STOLTENBERG, E. A., MYKLAND, P. A. and ZHANG, L. (2022). Supplement to "A CLT for second difference estimators with an application to volatility and intensity." https://doi.org/10.1214/22-AOS2176SUPP

VETTER, M. (2015). Estimation of integrated volatility of volatility with applications to goodness-of-fit testing. *Bernoulli* **21** 2393–2418. MR3378471 https://doi.org/10.3150/14-BEJ648

WANG, C. D. and MYKLAND, P. A. (2014). The estimation of leverage effect with high-frequency data. *J. Amer. Statist. Assoc.* **109** 197–215. MR3180557 https://doi.org/10.1080/01621459.2013.864189

ZHANG, L. (2006). Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli* **12** 1019–1043. MR2274854 https://doi.org/10.3150/bj/1165269149

ZHANG, L. (2012). Implied and realized volatility: Empirical model selection. *Ann. Finance* **8** 259–275. MR2922798 https://doi.org/10.1007/s10436-010-0168-0

ZHANG, L., MYKLAND, P. A. and AÏT-SAHALIA, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.* **100** 1394–1411. MR2236450 https://doi.org/10.1198/016214505000000169

The Annals of Statistics
2022, Vol. 50, No. 4, 2096–2126
https://doi.org/10.1214/22-AOS2178

SHARP OPTIMAL RECOVERY IN THE TWO COMPONENT GAUSSIAN MIXTURE MODEL

BY MOHAMED NDAOUD[a]

Department of Decisions Sciences, ESSEC Business School, [a]*ndaoud@essec.edu*

In this paper, we study the problem of clustering in the Two component Gaussian mixture model when the centers are separated by some $\Delta > 0$. We present a nonasymptotic lower bound for the corresponding minimax Hamming risk improving on existing results. We also propose an optimal, efficient and adaptive procedure that is minimax rate optimal. The rate optimality is moreover sharp in the asymptotics when the sample size goes to infinity. Our procedure is based on a variant of Lloyd's iterations initialized by a spectral method. As a consequence of nonasymptotic results, we find a sharp phase transition for the problem of exact recovery in the Gaussian mixture model. We prove that the phase transition occurs around the critical threshold $\bar{\Delta}$ given by

$$\bar{\Delta}^2 = \sigma^2\left(1 + \sqrt{1 + \frac{2p}{n \log n}}\right) \log n.$$

REFERENCES

[1] ABBE, E. (2017). Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** Paper No. 177, 86. 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., FAN, J. and WANG, K. (2020). An ℓ_p theory of PCA and spectral clustering. arXiv preprint arXiv:2006.14062.

[4] 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

[5] 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

[6] BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2012). The singular values and vectors of low rank perturbations of large rectangular random matrices. *J. Multivariate Anal.* **111** 120–135. MR2944410 https://doi.org/10.1016/j.jmva.2012.04.019

[7] BUTUCEA, C., NDAOUD, M., STEPANOVA, N. A. and TSYBAKOV, A. B. (2018). Variable selection with Hamming loss. *Ann. Statist.* **46** 1837–1875. MR3845003 https://doi.org/10.1214/17-AOS1572

[8] CHEN, X. and YANG, Y. (2021). Cutoff for exact recovery of Gaussian mixture models. *IEEE Trans. Inf. Theory* **67** 4223–4238. MR4289373 https://doi.org/10.1109/TIT.2021.3063155

[9] FEI, Y. and CHEN, Y. (2018). Hidden integrality of sdp relaxation for sub-gaussian mixture models. arXiv preprint arXiv:1803.06510.

[10] FELDMAN, V., PERKINS, W. and VEMPALA, S. (2015). Subsampled power iteration: A unified algorithm for block models and planted csp's. In *Advances in Neural Information Processing Systems* 2836–2844.

[11] FLORESCU, L. and PERKINS, W. (2016). Spectral thresholds in the bipartite stochastic block model. In *Conference on Learning Theory* 943–959.

[12] GAO, C., MA, Z., ZHANG, A. Y. and ZHOU, H. H. (2018). Community detection in degree-corrected block models. *Ann. Statist.* **46** 2153–2185. MR3845014 https://doi.org/10.1214/17-AOS1615

[13] GIRAUD, C. and VERZELEN, N. (2018). Partial recovery bounds for clustering with the relaxed K-means. *Math. Stat. Learn.* **1** 317–374. MR4059724

MSC2020 subject classifications. Primary 62H30; secondary 62C20, 62F07.
Key words and phrases. Gaussian mixtures, sharp recovery, Lloyd's algorithm.

[14] 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

[15] HSU, D., KAKADE, S. M. and ZHANG, T. (2012). A tail inequality for quadratic forms of subgaussian random vectors. *Electron. Commun. Probab.* **17** no. 52, 6. MR2994877 https://doi.org/10.1214/ECP.v17-2079

[16] KLUSOWSKI, J. M. and BRINDA, W. (2016). Statistical guarantees for estimating the centers of a two-component gaussian mixture by EM. arXiv preprint arXiv:1608.02280.

[17] LEHMANN, E. L. and CASELLA, G. (1998). *Theory of Point Estimation*, 2nd ed. *Springer Texts in Statistics*. Springer, New York. MR1639875

[18] LLOYD, S. P. (1982). Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **28** 129–137. MR0651807 https://doi.org/10.1109/TIT.1982.1056489

[19] 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

[20] LU, Y. and ZHOU, H. H. (2016). Statistical and computational guarantees of Lloyd's algorithm and its variants. arXiv preprint arXiv:1612.02099.

[21] MIXON, D. G., VILLAR, S. and WARD, R. (2017). Clustering subgaussian mixtures by semidefinite programming. *Inf. Inference* **6** 389–415. MR3764529 https://doi.org/10.1093/imaiai/iax001

[22] NDAOUD, M. (2019). Interplay of minimax estimation and minimax support recovery under sparsity. In *Algorithmic Learning Theory* 2019. *Proc. Mach. Learn. Res.* (*PMLR*) **98** 647–668, 22. Proceedings of Machine Learning Research PMLR, [place of publication not identified]. MR3932863

[23] 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

[24] ROYER, M. (2017). Adaptive clustering through semidefinite programming. In *Advances in Neural Information Processing Systems* 1795–1803.

[25] VEMPALA, S. and WANG, G. (2004). A spectral algorithm for learning mixture models. *J. Comput. System Sci.* **68** 841–860. MR2059647 https://doi.org/10.1016/j.jcss.2003.11.008

[26] 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. With a foreword by Sara van de Geer. MR3837109 https://doi.org/10.1017/9781108231596

[27] WU, Y. and ZHOU, H. H. (2021). Randomly initialized EM algorithm for two-component Gaussian mixture achieves near optimality in $O(\sqrt{n})$ iterations. *Math. Stat. Learn.* **4** 143–220. MR4383733 https://doi.org/10.4171/msl/29

[28] XIA, D. and ZHOU, F. (2019). The sup-norm perturbation of HOSVD and low rank tensor denoising. *J. Mach. Learn. Res.* **20** Paper No. 61, 42. MR3960915

[29] ZHANG, A. R., CAI, T. T. and WU, Y. (2022). Heteroskedastic PCA: Algorithm, optimality, and applications. *Ann. Statist.* **50** 53–80. MR4382008 https://doi.org/10.1214/21-aos2074

The Annals of Statistics
2022, Vol. 50, No. 4, 2127–2156
https://doi.org/10.1214/22-AOS2180

PRECISE STATISTICAL ANALYSIS OF CLASSIFICATION ACCURACIES FOR ADVERSARIAL TRAINING

BY ADEL JAVANMARD[1,a] AND MAHDI SOLTANOLKOTABI[2,b]

[1]*Department of Data Sciences and Operations, University of Southern California,* [a]*ajavanma@usc.edu*
[2]*Department of Electrical and Computer Engineering, University of Southern California,* [b]*soltanol@usc.edu*

Despite the wide empirical success of modern machine learning algorithms and models in a multitude of applications, they are known to be highly susceptible to seemingly small indiscernible perturbations to the input data known as *adversarial attacks*. A variety of recent adversarial training procedures have been proposed to remedy this issue. Despite the success of such procedures at increasing accuracy on adversarially perturbed inputs or *robust accuracy*, these techniques often reduce accuracy on natural unperturbed inputs or *standard accuracy*. Complicating matters further, the effect and trend of adversarial training procedures on standard and robust accuracy is rather counter intuitive and radically dependent on a variety of factors including the perceived form of the perturbation during training, size/quality of data, model overparameterization, etc. In this paper, we focus on binary classification problems where the data is generated according to the mixture of two Gaussians with general anisotropic covariance matrices and derive a precise characterization of the standard and robust accuracy for a class of minimax adversarially trained models. We consider a general norm-based adversarial model, where the adversary can add perturbations of bounded ℓ_p norm to each input data, for an arbitrary $p \geq 1$. Our comprehensive analysis allows us to theoretically explain several intriguing empirical phenomena and provide a precise understanding of the role of different problem parameters on standard and robust accuracies.

REFERENCES

[1] AMELUNXEN, D., LOTZ, M., MCCOY, M. B. and TROPP, J. A. (2014). Living on the edge: Phase transitions in convex programs with random data. *Inf. Inference* **3** 224–294. MR3311453 https://doi.org/10.1093/imaiai/iau005

[2] BARTL, D., DRAPEAU, S., OBLOJ, J. and WIESEL, J. (2020). Robust uncertainty sensitivity analysis. ArXiv preprint. Available at arXiv:2006.12022.

[3] BAYATI, M. and MONTANARI, A. (2012). The LASSO risk for Gaussian matrices. *IEEE Trans. Inf. Theory* **58** 1997–2017. MR2951312 https://doi.org/10.1109/TIT.2011.2174612

[4] BELKIN, M., HSU, D., MA, S. and MANDAL, S. (2019). Reconciling modern machine-learning practice and the classical bias-variance trade-off. *Proc. Natl. Acad. Sci. USA* **116** 15849–15854. MR3997901 https://doi.org/10.1073/pnas.1903070116

[5] BELKIN, M., MA, S. and MANDAL, S. (2018). To understand deep learning we need to understand kernel learning. In *International Conference on Machine Learning* 541–549.

[6] BU, Z., KLUSOWSKI, J. M., RUSH, C. and SU, W. J. (2021). Algorithmic analysis and statistical estimation of SLOPE via approximate message passing. *IEEE Trans. Inf. Theory* **67** 506–537. MR4231969 https://doi.org/10.1109/TIT.2020.3025272

[7] CANDÈS, E. J. and SUR, P. (2020). The phase transition for the existence of the maximum likelihood estimate in high-dimensional logistic regression. *Ann. Statist.* **48** 27–42. MR4065151 https://doi.org/10.1214/18-AOS1789

[8] CELENTANO, M. and MONTANARI, A. (2022). Fundamental barriers to high-dimensional regression with convex penalties. *Ann. Statist.* **50** 170–196. MR4382013 https://doi.org/10.1214/21-aos2100

[9] DANIELY, A. (2017). Sgd learns the conjugate kernel class of the network. In *Advances in Neural Information Processing Systems* 2422–2430.

MSC2020 subject classifications. Primary 62F12, 62E20; secondary 62J12.
Key words and phrases. Precise high-dimensional asymptotics, adversarial training, binary classification.

[10] DANIELY, A., FROSTIG, R. and SINGER, Y. (2016). Toward deeper understanding of neural networks: The power of initialization and a dual view on expressivity. In *Advances in Neural Information Processing Systems* 2253–2261.
[11] DENG, Z., KAMMOUN, A. and THRAMPOULIDIS, C. (2019). A model of double descent for high-dimensional binary linear classification. ArXiv preprint. Available at arXiv:1911.05822.
[12] DESHPANDE, Y. and MONTANARI, A. (2016). Sparse PCA via covariance thresholding. *J. Mach. Learn. Res.* **17** Paper No. 141. MR3555032
[13] DOBRIBAN, E., HASSANI, H., HONG, D. and ROBEY, A. (2020). Provable tradeoffs in adversarially robust classification. ArXiv preprint. Available at arXiv:2006.05161.
[14] 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
[15] DONOHO, D. and MONTANARI, A. (2016). High dimensional robust M-estimation: Asymptotic variance via approximate message passing. *Probab. Theory Related Fields* **166** 935–969. MR3568043 https://doi.org/10.1007/s00440-015-0675-z
[16] DONOHO, D. L., JAVANMARD, A. and MONTANARI, A. (2013). Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing. *IEEE Trans. Inf. Theory* **59** 7434–7464. MR3124654 https://doi.org/10.1109/TIT.2013.2274513
[17] DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2011). The noise-sensitivity phase transition in compressed sensing. *IEEE Trans. Inf. Theory* **57** 6920–6941. MR2882271 https://doi.org/10.1109/TIT.2011.2165823
[18] EL KAROUI, N. (2018). On the impact of predictor geometry on the performance on high-dimensional ridge-regularized generalized robust regression estimators. *Probab. Theory Related Fields* **170** 95–175. MR3748322 https://doi.org/10.1007/s00440-016-0754-9
[19] GOODFELLOW, I. J., SHLENS, J. and SZEGEDY, C. (2015). Explaining and harnessing adversarial examples. In 3*rd International Conference on Learning Representations*, *ICLR* 2015, *San Diego*, *CA*, *USA*, *May* 7–9, 2015, *Conference Track Proceedings*.
[20] GORDON, Y. (1988). On Milman's inequality and random subspaces which escape through a mesh in $\mathbf{R}^n$. In *Geometric Aspects of Functional Analysis* (1986/87). *Lecture Notes in Math.* **1317** 84–106. Springer, Berlin. MR0950977 https://doi.org/10.1007/BFb0081737
[21] GRANT, M., BOYD, S. and YE, Y. (2006). Disciplined convex programming. In *Global Optimization. Nonconvex Optim. Appl.* **84** 155–210. Springer, New York. MR2206954 https://doi.org/10.1007/0-387-30528-9_7
[22] HASSANI, H. and JAVANMARD, A. (2022). The curse of overparametrization in adversarial training: Precise analysis of robust generalization for random features regression. ArXiv preprint. Available at arXiv:2201.05149.
[23] HASTIE, T., MONTANARI, A., ROSSET, S. and TIBSHIRANI, R. J. (2022). Surprises in high-dimensional ridgeless least squares interpolation. *Ann. Statist.* **50** 949–986. MR4404925 https://doi.org/10.1214/21-aos2133
[24] HU, H. and LU, Y. M. (2019). Asymptotics and optimal designs of slope for sparse linear regression. ArXiv preprint. Available at arXiv:1903.11582.
[25] HUANG, H. (2017). Asymptotic behavior of support vector machine for spiked population model. *J. Mach. Learn. Res.* **18** Paper No. 45. MR3655310
[26] JACOT, A., GABRIEL, F. and HONGLER, C. (2018). Neural tangent kernel: Convergence and generalization in neural networks. In *Advances in Neural Information Processing Systems* 8571–8580.
[27] 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
[28] JAVANMARD, A. and SOLTANOLKOTABI, M. (2022). Supplement to "Precise statistical analysis of classification accuracies for adversarial training." https://doi.org/10.1214/22-AOS2180SUPP
[29] JAVANMARD, A., SOLTANOLKOTABI, M. and HASSANI, H. Precise tradeoffs in adversarial training for linear regression. In *Proceedings of Machine Learning Research, Conference of Learning Theory (COLT)* 125 2034–2078.
[30] JI, Z. and TELGARSKY, M. (2018). Risk and parameter convergence of logistic regression. ArXiv preprint. Available at arXiv:1803.07300.
[31] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961 https://doi.org/10.1214/aos/1009210544
[32] KAMMOUN, A. and ALOUINI, M.-S. (2021). On the precise error analysis of support vector machines. *IEEE Open Journal of Signal Processing* **2** 99–118.
[33] KAROUI, N. E. (2013). Asymptotic behavior of unregularized and ridge-regularized high-dimensional robust regression estimators: Rigorous results. ArXiv preprint. Available at arXiv:1311.2445.

[34] KURAKIN, A., GOODFELLOW, I. and BENGIO, S. (2016). Adversarial machine learning at scale. ArXiv preprint. Available at arXiv:1611.01236.
[35] LAI, L. and BAYRAKTAR, E. (2020). On the adversarial robustness of robust estimators. *IEEE Trans. Inf. Theory* **66** 5097–5109. MR4130664 https://doi.org/10.1109/TIT.2020.2985966
[36] LI, Y. and LIANG, Y. (2018). Learning overparameterized neural networks via stochastic gradient descent on structured data. *NeurIPS*.
[37] LIANG, T. and SUR, P. (2020). A precise high-dimensional asymptotic theory for boosting and min-l1-norm interpolated classifiers. ArXiv preprint. Available at arXiv:2002.01586.
[38] LOLAS, P. (2020). Regularization in high-dimensional regression and classification via random matrix theory. ArXiv preprint. Available at arXiv:2003.13723.
[39] LYU, K. and LI, J. (2019). Gradient descent maximizes the margin of homogeneous neural networks. ArXiv preprint. Available at arXiv:1906.05890.
[40] MADRY, A., MAKELOV, A., SCHMIDT, L., TSIPRAS, D. and VLADU, A. (2018). Towards deep learning models resistant to adversarial attacks. In 6*th International Conference on Learning Representations*, *ICLR* 2018, *Vancouver*, *BC*, *Canada*, *April* 30–*May* 3, 2018, *Conference Track Proceedings*.
[41] MAI, X., LIAO, Z. and COUILLET, R. (2019). A large scale analysis of logistic regression: Asymptotic performance and new insights. In *ICASSP* 2019-2019 *IEEE International Conference on Acoustics*, *Speech and Signal Processing* (*ICASSP*) 3357–3361. IEEE, New York.
[42] MEHRABI, M., JAVANMARD, A., ROSSI, R. A., RAO, A. and MAI, T. (2021). Fundamental tradeoffs in distributionally adversarial training. In *Proceedings of the* 38*th International Conference on Machine Learning* **139** 7544–7554. PMLR.
[43] MEI, S. and MONTANARI, A. (2022). The generalization error of random features regression: Precise asymptotics and the double descent curve. *Comm. Pure Appl. Math.* **75** 667–766. MR4400901
[44] MIGNACCO, F., KRZAKALA, F., LU, Y. M. and ZDEBOROVÁ, L. (2020). The role of regularization in classification of high-dimensional noisy gaussian mixture. ArXiv preprint. Available at arXiv:2002.11544.
[45] MIN, Y., CHEN, L. and KARBASI, A. (2020). The curious case of adversarially robust models: More data can help, double descend, or hurt generalization. ArXiv preprint. Available at arXiv:2002.11080.
[46] MIOLANE, L. and MONTANARI, A. (2021). The distribution of the Lasso: Uniform control over sparse balls and adaptive parameter tuning. *Ann. Statist.* **49** 2313–2335. MR4319252 https://doi.org/10.1214/20-aos2038
[47] MONTANARI, A., RUAN, F., SOHN, Y. and YAN, J. (2019). The generalization error of max-margin linear classifiers: High-dimensional asymptotics in the overparametrized regime. ArXiv preprint. Available at arXiv:1911.01544.
[48] OYMAK, S., RECHT, B. and SOLTANOLKOTABI, M. (2018). Sharp time-data tradeoffs for linear inverse problems. *IEEE Trans. Inf. Theory* **64** 4129–4158. MR3809731 https://doi.org/10.1109/TIT.2017.2773497
[49] OYMAK, S., THRAMPOULIDIS, C. and HASSIBI, B. (2013). The squared-error of generalized lasso: A precise analysis. ArXiv preprint. Available at arXiv:1311.0830.
[50] PYDI, M. S. and JOG, V. (2021). Adversarial risk via optimal transport and optimal couplings. *IEEE Trans. Inf. Theory* **67** 6031–6052. MR4345049 https://doi.org/10.1109/tit.2021.3100107
[51] RAGHUNATHAN, A., STEINHARDT, J. and LIANG, P. (2018). Certified defenses against adversarial examples. In 6*th International Conference on Learning Representations*, *ICLR* 2018, *Vancouver*, *BC*, *Canada*, *April* 30–*May* 3, 2018, *Conference Track Proceedings*.
[52] RAGHUNATHAN, A., XIE, S. M., YANG, F., DUCHI, J. C. and LIANG, P. (2019). Adversarial training can hurt generalization. ArXiv preprint. Available at arXiv:1906.06032.
[53] RAHIMI, A. and RECHT, B. (2007). Random features for large-scale kernel machines. *Adv. Neural Inf. Process. Syst.* **20** 1177–1184.
[54] SALEHI, F., ABBASI, E. and HASSIBI, B. (2019). The impact of regularization on high-dimensional logistic regression. ArXiv preprint. Available at arXiv:1906.03761.
[55] SCHMIDT, L., SANTURKAR, S., TSIPRAS, D., TALWAR, K. and MADRY, A. (2018). Adversarially robust generalization requires more data. In *Advances in Neural Information Processing Systems* 31: *Annual Conference on Neural Information Processing Systems* 2018, *NeurIPS* 2018, 3–8 *December* 2018, *Montréal*, *Canada* 5019–5031.
[56] STAIB, M. and JEGELKA, S. (2017). Distributionally robust deep learning as a generalization of adversarial training. In *NIPS Workshop on Machine Learning and Computer Security*.
[57] STOJNIC, M. (2009). Various thresholds for ℓ_1-optimization in compressed sensing. ArXiv preprint. Available at arXiv:0907.3666.
[58] STOJNIC, M. (2013). A framework to characterize performance of lasso algorithms. ArXiv preprint. Available at arXiv:1303.7291.

[59] SUR, P. and CANDÈS, E. J. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. *Proc. Natl. Acad. Sci. USA* **116** 14516–14525. MR3984492 https://doi.org/10.1073/pnas.1810420116

[60] TAHERI, H., PEDARSANI, R. and THRAMPOULIDIS, C. (2021). Sharp guarantees and optimal performance for inference in binary and Gaussian-mixture models. *Entropy* **23** Paper No. 178. MR4224548 https://doi.org/10.3390/e23020178

[61] THRAMPOULIDIS, C., ABBASI, E. and HASSIBI, B. (2018). Precise error analysis of regularized M-estimators in high dimensions. *IEEE Trans. Inf. Theory* **64** 5592–5628. MR3832326 https://doi.org/10.1109/TIT.2018.2840720

[62] THRAMPOULIDIS, C., OYMAK, S. and HASSIBI, B. (2015). Regularized linear regression: A precise analysis of the estimation error. In *Proceedings of the* 28*th Conference on Learning Theory* 1683–1709.

[63] TSIPRAS, D., SANTURKAR, S., ENGSTROM, L., TURNER, A. and MADRY, A. (2018). Robustness may be at odds with accuracy. ArXiv preprint. Available at arXiv:1805.12152.

[64] WANG, S., WENG, H. and MALEKI, A. (2022). Does SLOPE outperform bridge regression? *Inf. Inference* **11** 1–54. MR4409197 https://doi.org/10.1093/imaiai/iaab025

[65] WONG, E. and KOLTER, J. Z. (2018). Provable defenses against adversarial examples via the convex outer adversarial polytope. In *Proceedings of the* 35*th International Conference on Machine Learning*, *ICML* 2018, *Stockholmsmässan*, *Stockholm*, *Sweden*, *July* 10–15, 2018 5283–5292.

[66] ZHANG, H., YU, Y., JIAO, J., XING, E. P., GHAOUI, L. E. and JORDAN, M. I. (2019). Theoretically principled trade-off between robustness and accuracy. In *Proceedings of the* 36*th International Conference on Machine Learning*, *ICML* 2019, 9–15 *June* 2019, *Long Beach*, *California*, *USA* 7472–7482.

The Annals of Statistics
2022, Vol. 50, No. 4, 2157–2178
https://doi.org/10.1214/22-AOS2181

EXACT MINIMAX RISK FOR LINEAR LEAST SQUARES, AND THE LOWER TAIL OF SAMPLE COVARIANCE MATRICES

BY JAOUAD MOURTADA[a]

CREST, ENSAE, Institut Polytechnique de Paris, [a]*jaouad.mourtada@ensae.fr*

We consider random-design linear prediction and related questions on the lower tail of random matrices. It is known that, under boundedness constraints, the minimax risk is of order d/n in dimension d with n samples. Here, we study the minimax expected excess risk over the full linear class, depending on the distribution of covariates. First, the least squares estimator is exactly minimax optimal in the well-specified case, for every distribution of covariates. We express the minimax risk in terms of the distribution of statistical leverage scores of individual samples, and deduce a minimax lower bound of $d/(n-d+1)$ for any covariate distribution, nearly matching the risk for Gaussian design. We then obtain sharp nonasymptotic upper bounds for covariates that satisfy a "small ball"-type regularity condition in both well-specified and misspecified cases.

Our main technical contribution is the study of the lower tail of the smallest singular value of empirical covariance matrices at small values. We establish a lower bound on this lower tail, valid for any distribution in dimension $d \geq 2$, together with a matching upper bound under a necessary regularity condition. Our proof relies on the PAC-Bayes technique for controlling empirical processes, and extends an analysis of Oliveira devoted to a different part of the lower tail.

REFERENCES

[1] ADAMCZAK, R., LITVAK, A. E., PAJOR, A. and TOMCZAK-JAEGERMANN, N. (2010). Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. *J. Amer. Math. Soc.* **23** 535–561. MR2601042 https://doi.org/10.1090/S0894-0347-09-00650-X

[2] ANDERSON, G. W., GUIONNET, A. and ZEITOUNI, O. (2010). *An Introduction to Random Matrices*. *Cambridge Studies in Advanced Mathematics* **118**. Cambridge Univ. Press, Cambridge. MR2760897

[3] ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. *Wiley Publications in Statistics*. Wiley, New York. MR0091588

[4] AUDIBERT, J.-Y. and CATONI, O. (2010). Linear regression through PAC-Bayesian truncation. Preprint. Available at arXiv:1010.0072.

[5] AUDIBERT, J.-Y. and CATONI, O. (2011). Robust linear least squares regression. *Ann. Statist.* **39** 2766–2794. MR2906886 https://doi.org/10.1214/11-AOS918

[6] AZOURY, K. S. and WARMUTH, M. K. (2001). Relative loss bounds for on-line density estimation with the exponential family of distributions. *Mach. Learn.* **43** 211–246.

[7] 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

[8] BARTLETT, P. L., KOOLEN, W. M., MALEK, A., TAKIMOTO, E. and WARMUTH, M. K. (2015). Minimax fixed-design linear regression. In *Proc. 28th Conference on Learning Theory* 226–239.

[9] BHATIA, R. (2007). *Positive Definite Matrices*. *Princeton Series in Applied Mathematics*. Princeton Univ. Press, Princeton, NJ. MR2284176

[10] 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

MSC2020 subject classifications. Primary 62J05; secondary 60B20, 62C20.

Key words and phrases. Least squares, covariance matrices, lower bounds, decision theory, statistical learning theory, anticoncentration.

[11] BOYD, S. and VANDENBERGHE, L. (2004). *Convex Optimization*. Cambridge Univ. Press, Cambridge. MR2061575 https://doi.org/10.1017/CBO9780511804441
[12] BREIMAN, L. and FREEDMAN, D. (1983). How many variables should be entered in a regression equation? *J. Amer. Statist. Assoc.* **78** 131–136. MR0696857
[13] BUNEA, F., TSYBAKOV, A. B. and WEGKAMP, M. H. (2007). Aggregation for Gaussian regression. *Ann. Statist.* **35** 1674–1697. MR2351101 https://doi.org/10.1214/009053606000001587
[14] CAPONNETTO, A. and DE VITO, E. (2007). Optimal rates for the regularized least-squares algorithm. *Found. Comput. Math.* **7** 331–368. MR2335249 https://doi.org/10.1007/s10208-006-0196-8
[15] CATONI, O. (2004). *Statistical Learning Theory and Stochastic Optimization. Lecture Notes in Math.* **1851**. Springer, Berlin. MR2163920 https://doi.org/10.1007/b99352
[16] CATONI, O. (2007). *Pac-Bayesian Supervised Classification: The Thermodynamics of Statistical Learning. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **56**. IMS, Beachwood, OH. MR2483528
[17] CHATTERJEE, S. and HADI, A. S. (1988). *Sensitivity Analysis in Linear Regression. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Wiley, New York. MR0939610 https://doi.org/10.1002/9780470316764
[18] CUCKER, F. and SMALE, S. (2002). Best choices for regularization parameters in learning theory: On the bias-variance problem. *Found. Comput. Math.* **2** 413–428. MR1930945 https://doi.org/10.1007/s102080010030
[19] CUCKER, F. and SMALE, S. (2002). On the mathematical foundations of learning. *Bull. Amer. Math. Soc. (N.S.)* **39** 1–49. MR1864085 https://doi.org/10.1090/S0273-0979-01-00923-5
[20] DE VITO, E., CAPONNETTO, A. and ROSASCO, L. (2005). Model selection for regularized least-squares algorithm in learning theory. *Found. Comput. Math.* **5** 59–85. MR2125691 https://doi.org/10.1007/s10208-004-0134-1
[21] DICKER, L. H. (2016). Ridge regression and asymptotic minimax estimation over spheres of growing dimension. *Bernoulli* **22** 1–37. MR3449775 https://doi.org/10.3150/14-BEJ609
[22] DOBRIBAN, E. and WAGER, S. (2018). High-dimensional asymptotics of prediction: Ridge regression and classification. *Ann. Statist.* **46** 247–279. MR3766952 https://doi.org/10.1214/17-AOS1549
[23] DONOHO, D. and MONTANARI, A. (2016). High dimensional robust M-estimation: Asymptotic variance via approximate message passing. *Probab. Theory Related Fields* **166** 935–969. MR3568043 https://doi.org/10.1007/s00440-015-0675-z
[24] EDELMAN, A. (1988). Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.* **9** 543–560. MR0964668 https://doi.org/10.1137/0609045
[25] EL KAROUI, N. (2018). On the impact of predictor geometry on the performance on high-dimensional ridge-regularized generalized robust regression estimators. *Probab. Theory Related Fields* **170** 95–175. MR3748322 https://doi.org/10.1007/s00440-016-0754-9
[26] EL KAROUI, N. (2013). Asymptotic behavior of unregularized and ridge-regularized high-dimensional robust regression estimators: Rigorous results. Available at arXiv:1311.2445.
[27] EL KAROUI, N. and KÖSTERS, H. (2011). Geometric sensitivity of random matrix results: Consequences for shrinkage estimators of covariance and related statistical methods. Preprint. Available at arXiv:1105.1404.
[28] FOSTER, D. P. (1991). Prediction in the worst case. *Ann. Statist.* **19** 1084–1090. MR1105864 https://doi.org/10.1214/aos/1176348140
[29] GYÖRFI, L., KOHLER, M., KRZYŻAK, A. and WALK, H. (2002). *A Distribution-Free Theory of Nonparametric Regression. Springer Series in Statistics*. Springer, New York. MR1920390 https://doi.org/10.1007/b97848
[30] HOAGLIN, D. C. and WELSCH, R. E. (1978). The hat matrix in regression and ANOVA. *Amer. Statist.* **32** 17–22.
[31] HOERL, A. E. (1962). Application of ridge analysis to regression problems. *Chemical Engineering Progress* **58** 54–59.
[32] HORN, R. A. and JOHNSON, C. R. (1990). *Matrix Analysis*. Cambridge Univ. Press, Cambridge. MR1084815
[33] HSU, D., KAKADE, S. M. and ZHANG, T. (2014). Random design analysis of ridge regression. *Found. Comput. Math.* **14** 569–600. MR3201956 https://doi.org/10.1007/s10208-014-9192-1
[34] HSU, D. and SABATO, S. (2016). Loss minimization and parameter estimation with heavy tails. *J. Mach. Learn. Res.* **17** Paper No. 18. MR3491112
[35] HUBER, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821. MR0356373
[36] HUBER, P. J. (1981). *Robust Statistics. Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0606374

[37] JOHNSTONE, I. M. (2019). Gaussian Estimation: Sequence and Wavelet Models. Draft version, September 16, 2019.
[38] 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
[39] KOLTCHINSKII, V. and MENDELSON, S. (2015). Bounding the smallest singular value of a random matrix without concentration. *Int. Math. Res. Not. IMRN* **23** 12991–13008. MR3431642 https://doi.org/10.1093/imrn/rnv096
[40] LANGFORD, J. and SHAWE-TAYLOR, J. (2003). PAC-Bayes & margins. In *Advances in Neural Information Processing Systems* 15 439–446.
[41] LECUÉ, G. and MENDELSON, S. (2016). Performance of empirical risk minimization in linear aggregation. *Bernoulli* **22** 1520–1534. MR3474824 https://doi.org/10.3150/15-BEJ701
[42] 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
[43] LEHMANN, E. L. and CASELLA, G. (1998). *Theory of Point Estimation*, 2nd ed. *Springer Texts in Statistics*. Springer, New York. MR1639875
[44] LÖWNER, K. (1934). Über monotone Matrixfunktionen. *Math. Z.* **38** 177–216. MR1545446 https://doi.org/10.1007/BF01170633
[45] LUGOSI, G. and MENDELSON, S. (2019). Mean estimation and regression under heavy-tailed distributions: A survey. *Found. Comput. Math.* **19** 1145–1190. MR4017683 https://doi.org/10.1007/s10208-019-09427-x
[46] MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb.* (*N.S.*) **72** 507–536. MR0208649
[47] MCALLESTER, D. A. (1999). Some PAC-Bayesian theorems. *Mach. Learn.* **37** 355–363.
[48] MCALLESTER, D. A. (2003). PAC-Bayesian stochastic model selection. *Mach. Learn.* **51** 5–21.
[49] MENDELSON, S. (2015). Learning without concentration. *J. ACM* **62** Art. 21. MR3367000 https://doi.org/10.1145/2699439
[50] 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
[51] MOURTADA, J. (2022). Supplement to "Exact minimax risk for linear least squares, and the lower tail of sample covariance matrices." https://doi.org/10.1214/22-AOS2181SUPP
[52] NEMIROVSKI, A. (2000). Topics in non-parametric statistics. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1998). *Lecture Notes in Math.* **1738** 85–277. Springer, Berlin. MR1775640
[53] NGUYEN, H. H. and VU, V. H. (2013). Small ball probability, inverse theorems, and applications. In *Erdös Centennial. Bolyai Soc. Math. Stud.* **25** 409–463. János Bolyai Math. Soc., Budapest. MR3203607 https://doi.org/10.1007/978-3-642-39286-3_16
[54] 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
[55] RASKUTTI, G. and MAHONEY, M. W. (2016). A statistical perspective on randomized sketching for ordinary least-squares. *J. Mach. Learn. Res.* **17** Paper No. 214. MR3595148
[56] RINALDO, A., WASSERMAN, L. and G'SELL, M. (2019). Bootstrapping and sample splitting for high-dimensional, assumption-lean inference. *Ann. Statist.* **47** 3438–3469. MR4025748 https://doi.org/10.1214/18-AOS1784
[57] ROGOZIN, B. A. (1987). An estimate for the maximum of the convolution of bounded densities. *Teor. Veroyatn. Primen.* **32** 53–61. MR0890930
[58] RUDELSON, M. and VERSHYNIN, R. (2008). The Littlewood–Offord problem and invertibility of random matrices. *Adv. Math.* **218** 600–633. MR2407948 https://doi.org/10.1016/j.aim.2008.01.010
[59] RUDELSON, M. and VERSHYNIN, R. (2009). Smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.* **62** 1707–1739. MR2569075 https://doi.org/10.1002/cpa.20294
[60] RUDELSON, M. and VERSHYNIN, R. (2010). Non-asymptotic theory of random matrices: Extreme singular values. In *Proceedings of the International Congress of Mathematicians. Volume III* 1576–1602. Hindustan Book Agency, New Delhi. MR2827856
[61] RUDELSON, M. and VERSHYNIN, R. (2015). Small ball probabilities for linear images of high-dimensional distributions. *Int. Math. Res. Not. IMRN* **19** 9594–9617. MR3431603 https://doi.org/10.1093/imrn/rnu243
[62] SHAMIR, O. (2015). The sample complexity of learning linear predictors with the squared loss. *J. Mach. Learn. Res.* **16** 3475–3486. MR3450546
[63] SMALE, S. and ZHOU, D.-X. (2007). Learning theory estimates via integral operators and their approximations. *Constr. Approx.* **26** 153–172. MR2327597 https://doi.org/10.1007/s00365-006-0659-y

[64] SRIVASTAVA, N. and VERSHYNIN, R. (2013). Covariance estimation for distributions with $2+\varepsilon$ moments. *Ann. Probab.* **41** 3081–3111. MR3127875 https://doi.org/10.1214/12-AOP760

[65] STEIN, C. (1960). Multiple regression. In *Contributions to Probability and Statistics*: *Essays in Honor of Harold Hotelling*. Stanford University Press.

[66] STEINWART, I., HUSH, D. and SCOVEL, C. (2009). Optimal rates for regularized least squares regression. In *Proc. 22nd Conference on Learning Theory* 79–93.

[67] TAO, T. (2012). *Topics in Random Matrix Theory*. *Graduate Studies in Mathematics* **132**. Amer. Math. Soc., Providence, RI. MR2906465 https://doi.org/10.1090/gsm/132

[68] TAO, T. and VU, V. (2009). From the Littlewood–Offord problem to the circular law: Universality of the spectral distribution of random matrices. *Bull. Amer. Math. Soc.* (*N.S.*) **46** 377–396. MR2507275 https://doi.org/10.1090/S0273-0979-09-01252-X

[69] TAO, T. and VU, V. H. (2009). Inverse Littlewood–Offord theorems and the condition number of random discrete matrices. *Ann. of Math.* (2) **169** 595–632. MR2480613 https://doi.org/10.4007/annals.2009.169.595

[70] 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

[71] TIKHONOV, A. N. (1963). Solution of incorrectly formulated problems and the regularization method. *Sov. Math., Dokl.* **4** 1035–1038.

[72] TSYBAKOV, A. B. (2003). Optimal rates of aggregation. In *Learning Theory and Kernel Machines*. *Lecture Notes in Artificial Intelligence* 303–313. Springer, Berlin.

[73] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. MR2724359 https://doi.org/10.1007/b13794

[74] VAN DE GEER, S. and MURO, A. (2014). On higher order isotropy conditions and lower bounds for sparse quadratic forms. *Electron. J. Stat.* **8** 3031–3061. MR3301300 https://doi.org/10.1214/15-EJS983

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

[76] 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

[77] VOVK, V. (2001). Competitive on-line statistics. *Int. Stat. Rev.* **69** 213–248.

[78] WU, Y. and VERDÚ, S. (2012). Optimal phase transitions in compressed sensing. *IEEE Trans. Inf. Theory* **58** 6241–6263. MR2982661 https://doi.org/10.1109/TIT.2012.2205894

[79] YASKOV, P. (2014). Lower bounds on the smallest eigenvalue of a sample covariance matrix. *Electron. Commun. Probab.* **19** no. 83. MR3291620 https://doi.org/10.1214/ECP.v19-3807

[80] YASKOV, P. (2015). Sharp lower bounds on the least singular value of a random matrix without the fourth moment condition. *Electron. Commun. Probab.* **20** no. 44. MR3358966 https://doi.org/10.1214/ECP.v20-4089

The Annals of Statistics
2022, Vol. 50, No. 4, 2179–2204
https://doi.org/10.1214/22-AOS2182

STOCHASTIC CONTINUUM-ARMED BANDITS WITH ADDITIVE MODELS: MINIMAX REGRETS AND ADAPTIVE ALGORITHM

BY T. TONY CAI[a] AND HONGMING PU[b]

Department of Statistics and Data Science, The Wharton School, University of Pennsylvania, [a]*tcai@wharton.upenn.edu,* [b]*hpu@wharton.upenn.edu*

We consider d-dimensional stochastic continuum-armed bandits with the expected reward function being additive β-Hölder with sparsity s for $0 < \beta < \infty$ and $1 \le s \le d$. The rate of convergence $\tilde{O}(s \cdot T^{\frac{\beta+1}{2\beta+1}})$ for the minimax regret is established where T is the number of rounds. In particular, the minimax regret does not depend on d and is linear in s. A novel algorithm is proposed and is shown to be rate-optimal, up to a logarithmic factor of T.

The problem of adaptivity is also studied. A lower bound on the cost of adaptation to the smoothness is obtained and the result implies that adaptation for free is impossible in general without further structural assumptions. We then consider adaptive additive SCAB under an additional self-similarity assumption. An adaptive procedure is constructed and is shown to simultaneously achieve the minimax regret for a range of smoothness levels.

REFERENCES

ABBASI-YADKORI, Y., PÁL, D. and SZEPESVÁRI, C. (2011). Improved algorithms for linear stochastic bandits. In *NIPS* **11** 2312–2320.

AGARWAL, A., FOSTER, D. P., HSU, D., KAKADE, S. M. and RAKHLIN, A. (2013). Stochastic convex optimization with bandit feedback. *SIAM J. Optim.* **23** 213–240. MR3033105 https://doi.org/10.1137/110850827

AGRAWAL, R. (1995). The continuum-armed bandit problem. *SIAM J. Control Optim.* **33** 1926–1951. MR1358102 https://doi.org/10.1137/S0363012992237273

AGRAWAL, S., AVADHANULA, V., GOYAL, V. and ZEEVI, A. (2019). MNL-Bandit: A dynamic learning approach to assortment selection. *Oper. Res.* **67** 1453–1485. MR4014580 https://doi.org/10.1287/opre.2018.1832

AUDIBERT, J.-Y. and BUBECK, S. (2010). Regret bounds and minimax policies under partial monitoring. *J. Mach. Learn. Res.* **11** 2785–2836. MR2738783

AUDIBERT, J.-Y. and TSYBAKOV, A. B. (2007). Fast learning rates for plug-in classifiers. *Ann. Statist.* **35** 608–633. MR2336861 https://doi.org/10.1214/009053606000001217

AUER, P. (2002). Using confidence bounds for exploitation-exploration trade-offs. *J. Mach. Learn. Res.* **3** 397–422. MR1984023 https://doi.org/10.1162/153244303321897663

AUER, P., CESA-BIANCHI, N. and FISCHER, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Mach. Learn.* **47** 235–256.

AUER, P., ORTNER, R. and SZEPESVÁRI, C. (2007). Improved rates for the stochastic continuum-armed bandit problem. In *Learning Theory. Lecture Notes in Computer Science* **4539** 454–468. Springer, Berlin. MR2397605 https://doi.org/10.1007/978-3-540-72927-3_33

AUER, P., CESA-BIANCHI, N., FREUND, Y. and SCHAPIRE, R. E. (2002). The nonstochastic multiarmed bandit problem. *SIAM J. Comput.* **32** 48–77.

AWERBUCH, B. and KLEINBERG, R. (2008). Online linear optimization and adaptive routing. *J. Comput. System Sci.* **74** 97–114. MR2364184 https://doi.org/10.1016/j.jcss.2007.04.016

BANKS, J. S. and SUNDARAM, R. K. (1992). Denumerable-armed bandits. *Econometrica* **60** 1071–1096. MR1180234 https://doi.org/10.2307/2951539

BLUM, A., BURCH, C. and KALAI, A. (1999). Finely-competitive paging. In 40*th Annual Symposium on Foundations of Computer Science* (*New York*, 1999) 450–457. IEEE Computer Soc., Los Alamitos, CA. MR1917584 https://doi.org/10.1109/SFFCS.1999.814617

MSC2020 subject classifications. Primary 62G08; secondary 62L12.

Key words and phrases. Adaptivity, communication constraints, additive model, bandits, curse of dimensionality, minimax lower bound, optimal rate of convergence, regret, self-similarity.

BRESLER, G., CHEN, G. H. and SHAH, D. (2014). A latent source model for online collaborative filtering. *Adv. Neural Inf. Process. Syst.* **27** 3347–3355.
BUBECK, S., MUNOS, R., STOLTZ, G. and SZEPESVÁRI, C. (2011). $\mathcal{X}$-armed bandits. *J. Mach. Learn. Res.* **12** 1655–1695. MR2813150
BULL, A. D. (2015). Adaptive-treed bandits. *Bernoulli* **21** 2289–2307. MR3378467 https://doi.org/10.3150/14-BEJ644
CAI, T. T. and GUO, Z. (2017). Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *Ann. Statist.* **45** 615–646. MR3650395 https://doi.org/10.1214/16-AOS1461
CAI, T. T. and LOW, M. G. (2004). An adaptation theory for nonparametric confidence intervals. *Ann. Statist.* **32** 1805–1840. MR2102494 https://doi.org/10.1214/009053604000000049
CAI, T. T., LOW, M. G. and XIA, Y. (2013). Adaptive confidence intervals for regression functions under shape constraints. *Ann. Statist.* **41** 722–750. MR3099119 https://doi.org/10.1214/12-AOS1068
CAI, T. T and PU, H. (2022). Supplement to "Stochastic continuum-armed bandits with additive models: Minimax regrets and adaptive algorithm." https://doi.org/10.1214/22-AOS2182SUPP
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
COMBES, R. and PROUTIERE, A. (2014). Unimodal bandits: Regret lower bounds and optimal algorithms. In *International Conference on Machine Learning* 521–529. PMLR.
COPE, E. W. (2009). Regret and convergence bounds for a class of continuum-armed bandit problems. *IEEE Trans. Automat. Control* **54** 1243–1253. MR2532613 https://doi.org/10.1109/TAC.2009.2019797
DANI, V., HAYES, T. P. and KAKADE, S. M. (2008). Stochastic linear optimization under bandit feedback.
DELBRIDGE, I., BINDEL, D. and WILSON, A. G. (2020). Randomly projected additive Gaussian processes for regression. In *Proceedings of the* 37*th International Conference on Machine Learning* (H. D. III and A. Singh, eds.). *Proceedings of Machine Learning Research* **119** 2453–2463. PMLR.
DEN BOER, A. V. (2015). Dynamic pricing and learning: Historical origins, current research, and new directions. *Surv. Oper. Res. Manag. Sci.* **20** 1–18. MR3352577 https://doi.org/10.1016/j.sorms.2015.03.001
DUDIK, M., HSU, D., KALE, S., KARAMPATZIAKIS, N., LANGFORD, J., REYZIN, L. and ZHANG, T. (2011). Efficient optimal learning for contextual bandits. arXiv preprint. Available at arXiv:1106.2369.
GINÉ, E. and NICKL, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38** 1122–1170. MR2604707 https://doi.org/10.1214/09-AOS738
GUR, Y., MOMENI, A. and WAGER, S. (2021). Smoothness-adaptive contextual bandits. Available at SSRN.
GYÖRFI, L., KOHLER, M., KRZYZAK, A. and WALK, H. (2006). *A Distribution-Free Theory of Nonparametric Regression*. Springer, Berlin.
HADIJI, H. (2019). Polynomial cost of adaptation for x-armed bandits. In *Proceedings of the* 33*rd International Conference on Neural Information Processing Systems* 1029–1038.
HAO, B., LATTIMORE, T. and WANG, M. (2020). High-dimensional sparse linear bandits. arXiv preprint. Available at arXiv:2011.04020.
HAZAN, E. and KALE, S. (2011). Better algorithms for benign bandits. *J. Mach. Learn. Res.* **12** 1287–1311. MR2804601
HAZAN, E. and MEGIDDO, N. (2007). Online learning with prior knowledge. In *Learning Theory. Lecture Notes in Computer Science* **4539** 499–513. Springer, Berlin. MR2397608 https://doi.org/10.1007/978-3-540-72927-3_36
HU, Y., KALLUS, N. and MAO, X. (2020). Smooth contextual bandits: Bridging the parametric and non-differentiable regret regimes. In *Conference on Learning Theory* 2007–2010. PMLR.
KAKADE, S. M., KALAI, A. T. and LIGETT, K. (2009). Playing games with approximation algorithms. *SIAM J. Comput.* **39** 1088–1106. MR2538851 https://doi.org/10.1137/070701704
KANDASAMY, K., SCHNEIDER, J. and PÓCZOS, B. (2015). High dimensional Bayesian optimisation and bandits via additive models. In *International Conference on Machine Learning* 295–304. PMLR.
KLEINBERG, R. (2004). Nearly tight bounds for the continuum-armed bandit problem. *Adv. Neural Inf. Process. Syst.* **17** 697–704.
KLEINBERG, R., SLIVKINS, A. and UPFAL, E. (2008). Multi-armed bandits in metric spaces. In *STOC'*08 681–690. ACM, New York. MR2582691 https://doi.org/10.1145/1374376.1374475
KLEINBERG, R., SLIVKINS, A. and UPFAL, E. (2019). Bandits and experts in metric spaces. *J. ACM* **66** 30. MR3962354 https://doi.org/10.1145/3299873
LAI, T. L. and ROBBINS, H. (1985). Asymptotically efficient adaptive allocation rules. *Adv. in Appl. Math.* **6** 4–22. MR0776826 https://doi.org/10.1016/0196-8858(85)90002-8
LATTIMORE, T. and SZEPESVÁRI, C. (2020). *Bandit Algorithms*. Cambridge Univ. Press, Cambridge.
LINTON, O. and NIELSEN, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82** 93–100. MR1332841 https://doi.org/10.1093/biomet/82.1.93

LIU, Y., WANG, Y. and SINGH, A. (2021). Smooth bandit optimization: Generalization to Hölder space. In *International Conference on Artificial Intelligence and Statistics* 2206–2214. PMLR.

LOCATELLI, A. and CARPENTIER, A. (2018). Adaptivity to smoothness in x-armed bandits. In *Conference on Learning Theory* 1463–1492. PMLR.

LU, T., PÁL, D. and PÁL, M. (2009). Showing relevant ads via context multi-armed bandits Technical report.

MAILLARD, O.-A. and MUNOS, R. (2010). Online learning in adversarial Lipschitz environments. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* 305–320. Springer, Berlin.

MCMAHAN, H. B. and BLUM, A. (2004). Online geometric optimization in the bandit setting against an adaptive adversary. In *Learning Theory*. *Lecture Notes in Computer Science* **3120** 109–123. Springer, Berlin. MR2177904 https://doi.org/10.1007/978-3-540-27819-1_8

MINSKER, S. (2013). Estimation of extreme values and associated level sets of a regression function via selective sampling. In *Conference on Learning Theory* 105–121. PMLR.

PANDEY, S., AGARWAL, D., CHAKRABARTI, D. and JOSIFOVSKI, V. (2007). Bandits for taxonomies: A model-based approach. In *Proceedings of the* 2007 *SIAM International Conference on Data Mining* 216–227. SIAM, Philadelphia.

PEARSON, L. M. and BERRY, D. A. (1981). Optimal designs for two-stage clinical trials with dichotomous responses Technical report, Univ. Minnesota.

PICARD, D. and TRIBOULEY, K. (2000). Adaptive confidence interval for pointwise curve estimation. *Ann. Statist*. **28** 298–335. MR1762913 https://doi.org/10.1214/aos/1016120374

RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2011). Minimax rates of estimation for high-dimensional linear regression over ℓ_q-balls. *IEEE Trans. Inf. Theory* **57** 6976–6994. MR2882274 https://doi.org/10.1109/TIT.2011.2165799

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

RIGOLLET, P. and ZEEVI, A. (2010). Nonparametric bandits with covariates. arXiv preprint. Available at arXiv:1003.1630.

ROBBINS, H. (1952). Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc*. **58** 527–535. MR0050246 https://doi.org/10.1090/S0002-9904-1952-09620-8

ROLLAND, P., SCARLETT, J., BOGUNOVIC, I. and CEVHER, V. (2018). High-dimensional Bayesian optimization via additive models with overlapping groups. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics* (A. Storkey and F. Perez-Cruz, eds.). *Proceedings of Machine Learning Research* **84** 298–307. PMLR.

RUSMEVICHIENTONG, P. and TSITSIKLIS, J. N. (2010). Linearly parameterized bandits. *Math. Oper. Res*. **35** 395–411. MR2674726 https://doi.org/10.1287/moor.1100.0446

SINGH, S. (2021). Continuum-armed bandits: A function space perspective. In *International Conference on Artificial Intelligence and Statistics* 2620–2628. PMLR.

SLIVKINS, A. (2011). Contextual bandits with similarity information. In *Proceedings of the* 24*th Annual Conference on Learning Theory*. *JMLR Workshop and Conference Proceedings*. 679–702.

SLIVKINS, A. and VAUGHAN, J. W. (2014). Online decision making in crowdsourcing markets: Theoretical challenges. *ACM SIGecom Exch*. **12** 4–23.

STONE, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist*. **13** 689–705. MR0790566 https://doi.org/10.1214/aos/1176349548

WANG, Y., BALAKRISHNAN, S. and SINGH, A. (2019). Optimization of smooth functions with noisy observations: Local minimax rates. *IEEE Trans. Inf. Theory* **65** 7350–7366. MR4030889 https://doi.org/10.1109/TIT.2019.2921985

YUAN, M. and ZHOU, D.-X. (2016). Minimax optimal rates of estimation in high dimensional additive models. *Ann. Statist*. **44** 2564–2593. MR3576554 https://doi.org/10.1214/15-AOS1422

ZHAO, P. and LAI, L. (2021). Optimal stochastic nonconvex optimization with bandit feedback. arXiv preprint. Available at arXiv:2103.16082.

The Annals of Statistics
2022, Vol. 50, No. 4, 2205–2230
https://doi.org/10.1214/22-AOS2183

ASYMPTOTIC INDEPENDENCE OF SPIKED EIGENVALUES AND LINEAR SPECTRAL STATISTICS FOR LARGE SAMPLE COVARIANCE MATRICES

BY ZHIXIANG ZHANG[1,a], SHURONG ZHENG[2,c], GUANGMING PAN[1,b] AND PING-SHOU ZHONG[3,d]

[1]*School of Physical and Mathematical Sciences, Nanyang Technological University,* [a]*S160011@e.ntu.edu.sg,* [b]*GMPAN@ntu.edu.sg*

[2]*School of Mathematics and Statistics, Northeast Normal University,* [c]*zhengsr@nenu.edu.cn*

[3]*Department of Mathematics, Statistics, and Computer Science, University of Illinois Chicago,* [d]*pszhong@uic.edu*

We consider general high-dimensional spiked sample covariance models and show that their leading sample spiked eigenvalues and their linear spectral statistics are asymptotically independent when the sample size and dimension are proportional to each other. As a byproduct, we also establish the central limit theorem of the leading sample spiked eigenvalues by removing the block diagonal assumption on the population covariance matrix, which is commonly needed in the literature. Moreover, we propose consistent estimators of the L_4 norm of the spiked population eigenvectors. Based on these results, we develop a new statistic to test the equality of two spiked population covariance matrices. Numerical studies show that the new test procedure is more powerful than some existing methods.

REFERENCES

[1] ANDERSON, T. W. (1962). An introduction to multivariate statistical analysis. Technical Report. Wiley, New York.

[2] BAI, Z. and DING, X. (2012). Estimation of spiked eigenvalues in spiked models. *Random Matrices Theory Appl.* **1** 1150011. MR2934717 https://doi.org/10.1142/S2010326311500110

[3] 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

[4] BAI, Z. and YAO, J. (2008). Central limit theorems for eigenvalues in a spiked population model. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 447–474. MR2451053 https://doi.org/10.1214/07-AIHP118

[5] BAI, Z. and YAO, J. (2012). On sample eigenvalues in a generalized spiked population model. *J. Multivariate Anal.* **106** 167–177. MR2887686 https://doi.org/10.1016/j.jmva.2011.10.009

[6] BAI, Z. D., MIAO, B. Q. and PAN, G. M. (2007). On asymptotics of eigenvectors of large sample covariance matrix. *Ann. Probab.* **35** 1532–1572. MR2330979 https://doi.org/10.1214/009117906000001079

[7] BAI, Z. D. and SILVERSTEIN, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.* **26** 316–345. MR1617051 https://doi.org/10.1214/aop/1022855421

[8] BAI, Z. D. and SILVERSTEIN, J. W. (1999). Exact separation of eigenvalues of large-dimensional sample covariance matrices. *Ann. Probab.* **27** 1536–1555. MR1733159 https://doi.org/10.1214/aop/1022677458

[9] BAI, Z. D. and SILVERSTEIN, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** 553–605. MR2040792 https://doi.org/10.1214/aop/1078415845

[10] BAIK, J., LEE, J. O. and WU, H. (2018). Ferromagnetic to paramagnetic transition in spherical spin glass. *J. Stat. Phys.* **173** 1484–1522. MR3878351 https://doi.org/10.1007/s10955-018-2150-6

[11] BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. MR2279680 https://doi.org/10.1016/j.jmva.2005.08.003

[12] BILLINGSLEY, P. (1995). *Probability and Measure*. Wiley, New York.

MSC2020 subject classifications. Primary 62E20; secondary 62H15.

Key words and phrases. Leading spiked eigenvalues, sample covariance matrix, linear spectral statistics, central limit theorem.

[13] Cai, T., Liu, W. and Xia, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. Amer. Statist. Assoc.* **108** 265–277. MR3174618 https://doi.org/10.1080/01621459.2012.758041

[14] Cai, T. T., Han, X. and Pan, G. (2020). Limiting laws for divergent spiked eigenvalues and largest nonspiked eigenvalue of sample covariance matrices. *Ann. Statist.* **48** 1255–1280. MR4124322 https://doi.org/10.1214/18-AOS1798

[15] Ding, X. (2017). Asymptotics of empirical eigen-structure for high dimensional sample covariance matrices of general form. arXiv preprint. Available at arXiv:1708.06296.

[16] Jiang, D. and Bai, Z. (2021). Generalized four moment theorem and an application to CLT for spiked eigenvalues of high-dimensional covariance matrices. *Bernoulli* **27** 274–294. MR4177370 https://doi.org/10.3150/20-BEJ1237

[17] Jiang, D. and Bai, Z. (2021). Partial generalized four moment theorem revisited. *Bernoulli* **27** 2337–2352. MR4303885 https://doi.org/10.3150/20-BEJ1310

[18] Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961 https://doi.org/10.1214/aos/1009210544

[19] Jung, S. and Marron, J. S. (2009). PCA consistency in high dimension, low sample size context. *Ann. Statist.* **37** 4104–4130. MR2572454 https://doi.org/10.1214/09-AOS709

[20] 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

[21] 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

[22] Mestre, X. (2008). Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Trans. Inf. Theory* **54** 5113–5129. MR2589886 https://doi.org/10.1109/TIT.2008.929938

[23] Najim, J. and Yao, J. (2016). Gaussian fluctuations for linear spectral statistics of large random covariance matrices. *Ann. Appl. Probab.* **26** 1837–1887. MR3513608 https://doi.org/10.1214/15-AAP1135

[24] Pan, G. (2014). Comparison between two types of large sample covariance matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** 655–677. MR3189088 https://doi.org/10.1214/12-AIHP506

[25] Pan, G. M. and Zhou, W. (2008). Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. *Ann. Appl. Probab.* **18** 1232–1270. MR2418244 https://doi.org/10.1214/07-AAP477

[26] Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17** 1617–1642. MR2399865

[27] Paul, D. and Aue, A. (2014). Random matrix theory in statistics: A review. *J. Statist. Plann. Inference* **150** 1–29. MR3206718 https://doi.org/10.1016/j.jspi.2013.09.005

[28] Schmidt, M., Böhm, D., Von Törne, C., Steiner, E., Puhl, A., Pilch, H., Lehr, H.-A., Hengstler, J. G., Kölbl, H. et al. (2008). The humoral immune system has a key prognostic impact in node-negative breast cancer. *Cancer Res.* **68** 5405–5413.

[29] Wang, W. and Fan, J. (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Ann. Statist.* **45** 1342–1374. MR3662457 https://doi.org/10.1214/16-AOS1487

[30] Yang, Q. and Pan, G. (2017). Weighted statistic in detecting faint and sparse alternatives for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **112** 188–200. MR3646565 https://doi.org/10.1080/01621459.2015.1122602

[31] Yao, J., Zheng, S. and Bai, Z. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis*. *Cambridge Series in Statistical and Probabilistic Mathematics* **39**. Cambridge Univ. Press, New York. MR3468554 https://doi.org/10.1017/CBO9781107588080

[32] Yata, K. and Aoshima, M. (2012). Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations. *J. Multivariate Anal.* **105** 193–215. MR2877512 https://doi.org/10.1016/j.jmva.2011.09.002

[33] Zhang, Z., Zheng, S., Pan, G. and Zhong, P.-S. (2022). Supplement to "Asymptotic independence of spiked eigenvalues and linear spectral statistics for large sample covariance matrices." https://doi.org/10.1214/22-AOS2183SUPP

[34] Zheng, S., Bai, Z. and Yao, J. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *Ann. Statist.* **43** 546–591. MR3316190 https://doi.org/10.1214/14-AOS1292

The Annals of Statistics
2022, Vol. 50, No. 4, 2231–2255
https://doi.org/10.1214/22-AOS2185

LEARNING MIXTURES OF PERMUTATIONS: GROUPS OF PAIRWISE COMPARISONS AND COMBINATORIAL METHOD OF MOMENTS

BY CHENG MAO[1,a] AND YIHONG WU[2,b]

[1]*School of Mathematics, Georgia Institute of Technology,* [a]*cheng.mao@math.gatech.edu*
[2]*Department of Statistics and Data Science, Yale University,* [b]*yihong.wu@yale.edu*

In applications such as rank aggregation, mixture models for permutations are frequently used when the population exhibits heterogeneity. In this work, we study the widely used Mallows mixture model. In the high-dimensional setting, we propose a polynomial-time algorithm that learns a Mallows mixture of permutations on n elements with the optimal sample complexity that is proportional to $\log n$, improving upon previous results that scale polynomially with n. In the high-noise regime, we characterize the optimal dependency of the sample complexity on the noise parameter. Both objectives are accomplished by first studying demixing permutations under a noiseless query model using groups of pairwise comparisons, which can be viewed as moments of the mixing distribution, and then extending these results to the noisy Mallows model by simulating the noiseless oracle.

REFERENCES

[1] AWASTHI, P., BLUM, A., SHEFFET, O. and VIJAYARAGHAVAN, A. (2014). Learning mixtures of ranking models. In *Advances in Neural Information Processing Systems* 2609–2617.

[2] BALTRUNAS, L., MAKCINSKAS, T. and RICCI, F. (2010). Group recommendations with rank aggregation and collaborative filtering. In *Proceedings of the Fourth ACM Conference on Recommender Systems* 119–126.

[3] BORDA, J. C. (1781). Mémoire sur les élections au scrutin. *Histoire de l'Academie Royale des Sciences Pour*.

[4] BRAVERMAN, M. and MOSSEL, E. (2009). Sorting from noisy information. Preprint. Available at arXiv:0910.1191.

[5] BUSA-FEKETE, R., FOTAKIS, D., SZÖRÉNYI, B. and ZAMPETAKIS, M. (2019). Optimal learning of Mallows block model. In *Proceedings of the Thirty-Second Conference on Learning Theory* (A. Beygelzimer and D. Hsu, eds.) **99** 529–532.

[6] BUSA-FEKETE, R., HÜLLERMEIER, E. and SZÖRÉNYI, B. (2014). Preference-based rank elicitation using statistical models: The case of Mallows. In *Proceedings of the* 31*st International Conference on Machine Learning* **32** II–1071. JMLR.org.

[7] BUSSE, L. M., ORBANZ, P. and BUHMANN, J. M. (2007). Cluster analysis of heterogeneous rank data. In *Proceedings of the* 24*th International Conference on Machine Learning* 113–120. ACM, New York.

[8] CARAGIANNIS, I., PROCACCIA, A. D. and SHAH, N. (2013). When do noisy votes reveal the truth? In *Proceedings of the Fourteenth ACM Conference on Electronic Commerce* 143–160.

[9] CHIERICHETTI, F., DASGUPTA, A., KUMAR, R. and LATTANZI, S. (2015). On learning mixture models for permutations. In *ITCS'*15*—Proceedings of the* 6*th Innovations in Theoretical Computer Science* 85–92. ACM, New York. MR3418998

[10] CONDORCET, M. J. (1785). *Essai sur L'application de L'analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix*.

[11] DE, A., O'DONNELL, R. and SERVEDIO, R. (2018). Learning sparse mixtures of rankings from noisy information. Preprint. Available at arXiv:1811.01216.

[12] DOIGNON, J.-P., PEKEČ, A. and REGENWETTER, M. (2004). The repeated insertion model for rankings: Missing link between two subset choice models. *Psychometrika* **69** 33–54. MR2272438 https://doi.org/10.1007/BF02295838

MSC2020 subject classifications. Primary 62F07, 62H30; secondary 68P10.

Key words and phrases. Mixture of permutations, Mallows model, rank aggregation, group of pairwise comparisons, method of moments.

[13] Doss, N., Wu, Y., Yang, P. and Zhou, H. H. (2020). Optimal estimation of high-dimensional Gaussian mixtures. Preprint. Available at arXiv:2002.05818.

[14] Dwork, C., Kumar, R., Naor, M. and Sivakumar, D. (2001). Rank aggregation methods for the web. In *Proceedings of the* 10*th International Conference on World Wide Web* 613–622.

[15] Fagin, R., Kumar, R. and Sivakumar, D. (2003). Efficient similarity search and classification via rank aggregation. In *Proceedings of the* 2003 *ACM SIGMOD International Conference on Management of Data* 301–312.

[16] Fligner, M. A. and Verducci, J. S. (1986). Distance based ranking models. *J. Roy. Statist. Soc. Ser. B* **48** 359–369. MR0876847

[17] Gormley, I. C. and Murphy, T. B. (2008). Exploring voting blocs within the Irish electorate: A mixture modeling approach. *J. Amer. Statist. Assoc.* **103** 1014–1027. MR2528824 https://doi.org/10.1198/016214507000001049

[18] Gormley, I. C. and Murphy, T. B. (2008). A mixture of experts model for rank data with applications in election studies. *Ann. Appl. Stat.* **2** 1452–1477. MR2655667 https://doi.org/10.1214/08-AOAS178

[19] Heinrich, P. and Kahn, J. (2018). Strong identifiability and optimal minimax rates for finite mixture estimation. *Ann. Statist.* **46** 2844–2870. MR3851757 https://doi.org/10.1214/17-AOS1641

[20] Irurozki, E., Calvo, B. and Lozano, J. A. (2019). Mallows and generalized Mallows model for matchings. *Bernoulli* **25** 1160–1188. MR3920369 https://doi.org/10.3150/17-bej1017

[21] Jordan, M. I. and Jacobs, R. A. (1994). Hierarchical mixtures of experts and the EM algorithm. *Neural Comput.* **6** 181–214.

[22] Korba, A., Clémençon, S. and Sibony, E. (2017). A learning theory of ranking aggregation. In *Artificial Intelligence and Statistics* 1001–1010.

[23] Liu, A. X. and Moitra, A. (2018). Efficiently learning mixtures of Mallows models. In 59*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2018 627–638. IEEE Computer Soc., Los Alamitos, CA. MR3899628 https://doi.org/10.1109/FOCS.2018.00066

[24] Liu, Y.-T., Liu, T.-Y., Qin, T., Ma, Z.-M. and Li, H. (2007). Supervised rank aggregation. In *Proceedings of the* 16*th International Conference on World Wide Web* 481–490.

[25] Lu, T. and Boutilier, C. (2011). Learning Mallows models with pairwise preferences. In *Proceedings of the* 28*th International Conference on International Conference on Machine Learning* 145–152.

[26] Lu, T. and Boutilier, C. (2014). Effective sampling and learning for Mallows models with pairwise-preference data. *J. Mach. Learn. Res.* **15** 3783–3829. MR3317212

[27] Mallows, C. L. (1957). Non-null ranking models. I. *Biometrika* **44** 114–130. MR0087267 https://doi.org/10.1093/biomet/44.1-2.114

[28] Mao, C. and Wu, Y. (2022). Supplement to "Learning mixtures of permutations: Groups of pairwise comparisons and combinatorial method of moments." https://doi.org/10.1214/22-AOS2185SUPP

[29] Marden, J. I. (1995). *Analyzing and Modeling Rank Data. Monographs on Statistics and Applied Probability* **64**. CRC Press, London. MR1346107

[30] Meilă, M. and Chen, H. (2010). Dirichlet process mixtures of generalized Mallows models. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence* 358–367.

[31] Meilă, M., Phadnis, K., Patterson, A. and Bilmes, J. (2007). Consensus ranking under the exponential model. In *Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence* 285–294.

[32] Moitra, A. and Valiant, G. (2010). Settling the polynomial learnability of mixtures of Gaussians. In 2010 *IEEE* 51*st Annual Symposium on Foundations of Computer Science—FOCS* 2010 93–102. IEEE Computer Soc., Los Alamitos, CA. MR3024779

[33] Murphy, T. B. and Martin, D. (2003). Mixtures of distance-based models for ranking data. *Comput. Statist. Data Anal.* **41** 645–655. MR1973732 https://doi.org/10.1016/S0167-9473(02)00165-2

[34] Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philos. Trans. R. Soc. Lond. A* **185** 71–110.

[35] Wu, Y. and Yang, P. (2020). Optimal estimation of Gaussian mixtures via denoised method of moments. *Ann. Statist.* **48** 1981–2007. MR4134783 https://doi.org/10.1214/19-AOS1873

[36] Zagier, D. (1992). Realizability of a model in infinite statistics. *Comm. Math. Phys.* **147** 199–210. MR1171767

[37] Zhao, Z., Piech, P. and Xia, L. (2016). Learning mixtures of Plackett–Luce models. In *International Conference on Machine Learning* 2906–2914.

The Annals of Statistics
2022, Vol. 50, No. 4, 2256–2283
https://doi.org/10.1214/22-AOS2186

GENERALIZED RESILIENCE AND ROBUST STATISTICS

BY BANGHUA ZHU[1,a], JIANTAO JIAO[1,b] AND JACOB STEINHARDT[2,c]

[1]*Department of EECS, University of California, Berkeley,* [a]*banghua@berkeley.edu,* [b]*jiantao@berkeley.edu*
[2]*Department of Statistics, University of California, Berkeley,* [c]*jsteinhardt@berkeley.edu*

Robust statistics traditionally focuses on outliers, or perturbations in total variation distance. However, a dataset could be maliciously corrupted in many other ways, such as systematic measurement errors and missing covariates. We consider corruption in either TV or Wasserstein distance, and show that robust estimation is possible whenever the true population distribution satisfies a property called *generalized resilience*, which holds under moment or hypercontractive conditions. For TV corruption model, our finite-sample analysis improves over previous results for mean estimation with bounded kth moment, linear regression, and joint mean and covariance estimation. For W_1 corruption, we provide the first finite-sample guarantees for second moment estimation and linear regression.

Technically, our robust estimators are a generalization of minimum distance (MD) functionals, which project the corrupted distribution onto a given set of well-behaved distributions. The error of these MD functionals is bounded by a certain modulus of continuity, and we provide a systematic method for upper bounding this modulus for the class of generalized resilient distributions, which usually gives sharp population-level results and good finite-sample guarantees.

REFERENCES

ADROVER, J. and YOHAI, V. (2002). Projection estimates of multivariate location. *Ann. Statist.* **30** 1760–1781. MR1969449 https://doi.org/10.1214/aos/1043351256

BAKSHI, A. and PRASAD, A. (2021). Robust linear regression: Optimal rates in polynomial time. In *STOC '21—Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing* 102–115. ACM, New York. MR4398827

BATENI, A.-H. and DALALYAN, A. S. (2019). Minimax rates in outlier-robust estimation of discrete models. ArXiv preprint. Available at arXiv:1902.04650.

BEN-TAL, A., EL GHAOUI, L. and NEMIROVSKI, A. (2009). *Robust Optimization*. *Princeton Series in Applied Mathematics*. Princeton Univ. Press, Princeton, NJ. MR2546839 https://doi.org/10.1515/9781400831050

BORWEIN, J. M. and LEWIS, A. S. (2006). *Convex Analysis and Nonlinear Optimization*: *Theory and Examples*, 2nd ed. *CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC* **3**. Springer, New York. MR2184742 https://doi.org/10.1007/978-0-387-31256-9

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

CHEN, M., GAO, C. and REN, Z. (2016). A general decision theory for Huber's ϵ-contamination model. *Electron. J. Stat.* **10** 3752–3774. MR3579675 https://doi.org/10.1214/16-EJS1216

CHEN, M., GAO, C. and REN, Z. (2018). Robust covariance and scatter matrix estimation under Huber's contamination model. *Ann. Statist.* **46** 1932–1960. MR3845006 https://doi.org/10.1214/17-AOS1607

CHEN, Z. and TYLER, D. E. (2002). The influence function and maximum bias of Tukey's median. *Ann. Statist.* **30** 1737–1759. MR1969448 https://doi.org/10.1214/aos/1043351255

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

MSC2020 subject classifications. Primary 62F35; secondary 62G35.

Key words and phrases. Robust statistics, minimum distance functional, total variation distance perturbation, Wasserstein distance perturbation.

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 Computer Soc., Los Alamitos, CA. MR3734219 https://doi.org/10.1109/FOCS.2017.16

DIAKONIKOLAS, I., KANE, D. M. and STEWART, A. (2018). Learning geometric concepts with nasty noise. In *STOC'*18*—Proceedings of the* 50*th Annual ACM SIGACT Symposium on Theory of Computing* 1061–1073. ACM, New York. MR3826317 https://doi.org/10.1145/3188745.3188754

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

DIAKONIKOLAS, I., KAMATH, G., KANE, D. M., LI, J., MOITRA, A. and STEWART, A. (2017). Being robust (in high dimensions) can be practical. In *Proceedings of the* 34*th International Conference on Machine Learning-Volume* 70 999–1008. JMLR.org.

DIAKONIKOLAS, I., KAMATH, G., KANE, D. M., LI, J., STEINHARDT, J. and STEWART, A. (2018a). Sever: A robust meta-algorithm for stochastic optimization. ArXiv preprint. Available at arXiv:1803.02815.

DIAKONIKOLAS, I., KAMATH, G., KANE, D. M., LI, J., MOITRA, A. and STEWART, A. (2018b). Robustly learning a Gaussian: Getting optimal error, efficiently. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms* 2683–2702. SIAM, Philadelphia, PA. MR3775959 https://doi.org/10.1137/1.9781611975031.171

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

DONOHO, D. L. (1982). Breakdown properties of multivariate location estimators. Technical report, Harvard Univ., Boston.

DONOHO, D. L. and LIU, R. C. (1988a). The "automatic" robustness of minimum distance functionals. *Ann. Statist.* **16** 552–586. MR0947562 https://doi.org/10.1214/aos/1176350820

DONOHO, D. L. and LIU, R. C. (1988b). Pathologies of some minimum distance estimators. *Ann. Statist.* **16** 587–608. MR0947563 https://doi.org/10.1214/aos/1176350821

DUDLEY, R. M. (1968). The speed of mean Glivenko–Cantelli convergence. *Ann. Math. Stat.* **40** 40–50. MR0236977 https://doi.org/10.1214/aoms/1177697802

DUDLEY, R. M. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **6** 899–929. MR0512411

GAO, C. (2020). Robust regression via mutivariate regression depth. *Bernoulli* **26** 1139–1170. MR4058363 https://doi.org/10.3150/19-BEJ1144

GAO, C., YAO, Y. and ZHU, W. (2020). Generative adversarial nets for robust scatter estimation: A proper scoring rule perspective. *J. Mach. Learn. Res.* **21** Paper No. 160. MR4209446

GOLUB, G. H. and VAN LOAN, C. F. (1980). An analysis of the total least squares problem. *SIAM J. Numer. Anal.* **17** 883–893. MR0595451 https://doi.org/10.1137/0717073

HOPKINS, S. B. and LI, J. (2019). How hard is robust mean estimation? ArXiv preprint. Available at arXiv:1903.07870.

HUBER, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821. MR0356373

HUBER, P. J. (2011). *Robust Statistics*. Springer, Berlin.

KLIVANS, A., KOTHARI, P. K. and MEKA, R. (2018). Efficient algorithms for outlier-robust regression. ArXiv preprint. Available at arXiv:1803.03241.

KOTHARI, P. K. and STEURER, D. (2017). Outlier-robust moment-estimation via sum-of-squares. ArXiv preprint. Available at arXiv:1711.11581.

KOWALSKI, M. (2020). Bias estimation and coordinate conversion in 3-D sensor measurements.

LAI, K. A., RAO, A. B. and VEMPALA, S. (2016). Agnostic estimation of mean and covariance. In 57*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2016 665–674. IEEE Computer Soc., Los Alamitos, CA. MR3631029

LEUNG, J., DUDAS, J., CHAPMAN, G. H., KOREN, I. and KOREN, Z. (2007). Quantitative analysis of in-field defects in image sensor arrays. In 22*nd IEEE International Symposium on Defect and Fault-Tolerance in VLSI Systems* (*DFT* 2007) 526–534. IEEE, New York.

LIU, L., SHEN, Y., LI, T. and CARAMANIS, C. (2018). High dimensional robust sparse regression. ArXiv preprint. Available at arXiv:1805.11643.

LUKEŠ, J., MALÝ, J., NETUKA, I. and SPURNÝ, J. (2010). *Integral Representation Theory*: *Applications to Convexity*, *Banach Spaces and Potential Theory*. *De Gruyter Studies in Mathematics* **35**. de Gruyter, Berlin. MR2589994

MARSHALL, A. W., OLKIN, I. and ARNOLD, B. C. (1979). *Inequalities*: *Theory of Majorization and Its Applications* **143**. Springer.

MASSEY JR., F. J. (1951). The Kolmogorov–Smirnov test for goodness of fit. *J. Amer. Statist. Assoc.* **46** 68–78.
PENOT, J.-P. (2013). *Calculus Without Derivatives. Graduate Texts in Mathematics* **266**. Springer, New York. MR2986672 https://doi.org/10.1007/978-1-4614-4538-8
PRASAD, A., BALAKRISHNAN, S. and RAVIKUMAR, P. (2019). A unified approach to robust mean estimation. ArXiv preprint. Available at arXiv:1907.00927.
STEINHARDT, J. (2018). Robust Learning: Information Theory and Algorithms. Ph.D. thesis, Stanford Univ.
STEINHARDT, J., CHARIKAR, M. and VALIANT, G. (2017). Resilience: A criterion for learning in the presence of arbitrary outliers. ArXiv preprint. Available at arXiv:1703.04940.
STEINHARDT, J., CHARIKAR, M. and VALIANT, G. (2018). Resilience: A criterion for learning in the presence of arbitrary outliers. In 9*th Innovations in Theoretical Computer Science. LIPIcs. Leibniz Int. Proc. Inform.* **94** Art. No. 45. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR3761781
STEINHARDT, J., KOH, P. W. W. and LIANG, P. S. (2017). Certified defenses for data poisoning attacks. In *Advances in Neural Information Processing Systems* 3517–3529.
VAPNIK, V. N. and CHERVONENKIS, A. Y. (2015). On the uniform convergence of relative frequencies of events to their probabilities. In *Measures of Complexity* 11–30. Springer, Cham. MR3408730
ZHU, B., JIAO, J. and STEINHARDT, J. (2021). Robust estimation via generalized quasi-gradients. *Inf. Inference* iaab018. https://doi.org/10.1093/imaiai/iaab018
ZHU, B., JIAO, J. and JORDAN, M. (2022). Robust estimation for nonparametric families via generative adversarial networks. ArXiv preprint. Available at arXiv:2202.01269.
ZHU, B., JIAO, J. and STEINHARDT, J. (2022). Supplement to "Generalized resilience and robust statistics." https://doi.org/10.1214/22-AOS2186SUPP

The Annals of Statistics
2022, Vol. 50, No. 4, 2284–2305
https://doi.org/10.1214/22-AOS2188

NONREGULAR AND MINIMAX ESTIMATION OF INDIVIDUALIZED THRESHOLDS IN HIGH DIMENSION WITH BINARY RESPONSES

BY HUIJIE FENG[1,a], YANG NING[1,b] AND JIWEI ZHAO[2,c]

[1]*Department of Statistics and Data Science, Cornell University,* [a]*hf279@cornell.edu,* [b]*yn265@cornell.edu*
[2]*Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison,* [c]*jiwei.zhao@wisc.edu*

Given a large number of covariates $\boldsymbol{Z}$, we consider the estimation of a high-dimensional parameter $\boldsymbol{\theta}$ in an individualized linear threshold $\boldsymbol{\theta}^T\boldsymbol{Z}$ for a continuous variable X, which minimizes the disagreement between $\operatorname{sign}(X - \boldsymbol{\theta}^T\boldsymbol{Z})$ and a binary response Y. While the problem can be formulated into the M-estimation framework, minimizing the corresponding empirical risk function is computationally intractable due to discontinuity of the sign function. Moreover, estimating $\boldsymbol{\theta}$ even in the fixed-dimensional setting is known as a nonregular problem leading to nonstandard asymptotic theory. To tackle the computational and theoretical challenges in the estimation of the high-dimensional parameter $\boldsymbol{\theta}$, we propose an empirical risk minimization approach based on a regularized smoothed non-convex loss function. The Fisher consistency of the proposed method is guaranteed as the bandwidth of the smoothed loss is shrunk to 0. Statistically, we show that the finite sample error bound for estimating $\boldsymbol{\theta}$ in ℓ_2 norm is $(s \log d/n)^{\beta/(2\beta+1)}$, where d is the dimension of $\boldsymbol{\theta}$, s is the sparsity level, n is the sample size and β is the smoothness of the conditional density of X given the response Y and the covariates $\boldsymbol{Z}$. The convergence rate is nonstandard and slower than that in the classical Lasso problems. Furthermore, we prove that the resulting estimator is minimax rate optimal up to a logarithmic factor. The Lepski's method is developed to achieve the adaption to the unknown sparsity s or smoothness β. Computationally, an efficient path-following algorithm is proposed to compute the solution path. We show that this algorithm achieves geometric rate of convergence for computing the whole path. Finally, we evaluate the finite sample performance of the proposed estimator in simulation studies and a real data analysis from the ChAMP (Chondral Lesions And Meniscus Procedures) Trial.

REFERENCES

[1] AGARWAL, A., NEGAHBAN, S. and WAINWRIGHT, M. J. (2010). Fast global convergence rates of gradient methods for high-dimensional statistical recovery. In *Advances in Neural Information Processing Systems* 23 (J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel and A. Culotta, eds.) 37–45. Curran Associates, Inc.

[2] BARTLETT, P. L., JORDAN, M. I. and MCAULIFFE, J. D. (2006). Convexity, classification, and risk bounds. *J. Amer. Statist. Assoc.* **101** 138–156. MR2268032 https://doi.org/10.1198/016214505000000907

[3] BELLEC, P. C., LECUÉ, G. and TSYBAKOV, A. B. (2018). Slope meets Lasso: Improved oracle bounds and optimality. *Ann. Statist.* **46** 3603–3642. MR3852663 https://doi.org/10.1214/17-AOS1670

[4] 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

[5] BIRGÉ, L. (2001). An alternative point of view on Lepski's method. In *State of the Art in Probability and Statistics* (*Leiden*, 1999). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **36** 113–133. IMS, Beachwood, OH. MR1836557 https://doi.org/10.1214/lnms/1215090065

MSC2020 subject classifications. Primary 62Fxx, 62C20.

Key words and phrases. High-dimensional statistics, nonstandard asymptotics, non-convex optimization, minimax optimality, adaptivity, kernel method.

[6] BISSON, L. J., KLUCZYNSKI, M. A., WIND, W. M., FINEBERG, M. S., BERNAS, G. A., RAUH, M. A., MARZO, J. M., ZHOU, Z. and ZHAO, J. (2017). Patient outcomes after observation versus debridement of unstable chondral lesions during partial meniscectomy: The chondral lesions and meniscus procedures (ChAMP) randomized controlled trial. *J. Bone Jt. Surg.* **99** 1078.

[7] BOUFOUNOS, P. T. and BARANIUK, R. G. (2008). 1-bit compressive sensing. In 2008 42*nd Annual Conference on Information Sciences and Systems* 16–21. Available at https://doi.org/10.1109/CISS.2008.4558487

[8] 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

[9] CAI, T. T., LOW, M. and MA, Z. (2014). Adaptive confidence bands for nonparametric regression functions. *J. Amer. Statist. Assoc.* **109** 1054–1070. MR3265680 https://doi.org/10.1080/01621459.2013.879260

[10] CANDES, E. J. and TAO, T. (2005). Decoding by linear programming. *IEEE Trans. Inf. Theory* **51** 4203–4215. MR2243152 https://doi.org/10.1109/TIT.2005.858979

[11] EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499. MR2060166 https://doi.org/10.1214/009053604000000067

[12] FENG, H., NING, Y. and ZHAO, J. (2022). Supplement to "Nonregular and minimax estimation of individualized thresholds in high dimension with binary responses." https://doi.org/10.1214/22-AOS2188SUPP

[13] GINÉ, E. and NICKL, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38** 1122–1170. MR2604707 https://doi.org/10.1214/09-AOS738

[14] 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

[15] HOROWITZ, J. L. (1992). A smoothed maximum score estimator for the binary response model. *Econometrica* **60** 505–531. MR1162997 https://doi.org/10.2307/2951582

[16] KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219. MR1041391 https://doi.org/10.1214/aos/1176347498

[17] LEPSKIĬ, O. V. (1990). A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatn. Primen.* **35** 459–470. MR1091202 https://doi.org/10.1137/1135065

[18] LEPSKIĬ, O. V. (1992). Asymptotically minimax adaptive estimation. II. Schemes without optimal adaptation. Adaptive estimates. *Teor. Veroyatn. Primen.* **37** 468–481. MR1214353 https://doi.org/10.1137/1137095

[19] LOH, P.-L. and WAINWRIGHT, M. J. (2015). Regularized M-estimators with nonconvexity: Statistical and algorithmic theory for local optima. *J. Mach. Learn. Res.* **16** 559–616. MR3335800

[20] LUGOSI, G. and VAYATIS, N. (2004). On the Bayes-risk consistency of regularized boosting methods. *Ann. Statist.* **32** 30–55. MR2051000 https://doi.org/10.1214/aos/1079120129

[21] MAMMEN, E. and TSYBAKOV, A. B. (1999). Smooth discrimination analysis. *Ann. Statist.* **27** 1808–1829. MR1765618 https://doi.org/10.1214/aos/1017939240

[22] MANSKI, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *J. Econometrics* **3** 205–228. MR0436905 https://doi.org/10.1016/0304-4076(75)90032-9

[23] MANSKI, C. F. (1985). Semiparametric analysis of discrete response. Asymptotic properties of the maximum score estimator. *J. Econometrics* **27** 313–333. MR0788628 https://doi.org/10.1016/0304-4076(85)90009-0

[24] MUKHERJEE, D., BANERJEE, M. and RITOV, Y. (2021). Optimal linear discriminators for the discrete choice model in growing dimensions. *Ann. Statist.* **49** 3324–3357. MR4352532 https://doi.org/10.1214/21-aos2085

[25] NEGAHBAN, S. N., RAVIKUMAR, P., WAINWRIGHT, M. J. and YU, B. (2012). A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers. *Statist. Sci.* **27** 538–557. MR3025133 https://doi.org/10.1214/12-STS400

[26] NESTEROV, YU. (2013). Gradient methods for minimizing composite functions. *Math. Program.* **140** 125–161. MR3071865 https://doi.org/10.1007/s10107-012-0629-5

[27] PARK, M. Y. and HASTIE, T. (2007). L_1-regularization path algorithm for generalized linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 659–677. MR2370074 https://doi.org/10.1111/j.1467-9868.2007.00607.x

[28] QIU, X., ZENG, D. and WANG, Y. (2018). Estimation and evaluation of linear individualized treatment rules to guarantee performance. *Biometrics* **74** 517–528. MR3825338 https://doi.org/10.1111/biom.12773

[29] RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2011). Minimax rates of estimation for high-dimensional linear regression over ℓ_q-balls. *IEEE Trans. Inf. Theory* **57** 6976–6994. MR2882274 https://doi.org/10.1109/TIT.2011.2165799

[30] SHEN, X., TSENG, G. C., ZHANG, X. and WONG, W. H. (2003). On ψ-learning. *J. Amer. Statist. Assoc.* **98** 724–734. MR2011686 https://doi.org/10.1198/016214503000000639

[31] SU, W. and CANDÈS, E. (2016). SLOPE is adaptive to unknown sparsity and asymptotically minimax. *Ann. Statist.* **44** 1038–1068. MR3485953 https://doi.org/10.1214/15-AOS1397

[32] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. MR2724359 https://doi.org/10.1007/b13794

[33] 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

[34] WANG, Z., LIU, H. and ZHANG, T. (2014). Optimal computational and statistical rates of convergence for sparse nonconvex learning problems. *Ann. Statist.* **42** 2164–2201. MR3269977 https://doi.org/10.1214/14-AOS1238

[35] WU, Y. and LIU, Y. (2007). Robust truncated hinge loss support vector machines. *J. Amer. Statist. Assoc.* **102** 974–983. MR2411659 https://doi.org/10.1198/016214507000000617

[36] XIAO, L. and ZHANG, T. (2013). A proximal-gradient homotopy method for the sparse least-squares problem. *SIAM J. Optim.* **23** 1062–1091. MR3057332 https://doi.org/10.1137/120869997

[37] XU, T., WANG, J. and FANG, Y. (2014). A model-free estimation for the covariate-adjusted Youden index and its associated cut-point. *Stat. Med.* **33** 4963–4974. MR3276512 https://doi.org/10.1002/sim.6290

[38] YOUDEN, W. J. (1950). Index for rating diagnostic tests. *Cancer* **3** 32–35.

[39] ZHANG, T. (2004). Statistical behavior and consistency of classification methods based on convex risk minimization. *Ann. Statist.* **32** 56–85. MR2051001 https://doi.org/10.1214/aos/1079120130

[40] ZHANG, X., WU, Y., WANG, L. and LI, R. (2016). Variable selection for support vector machines in moderately high dimensions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 53–76. MR3453646 https://doi.org/10.1111/rssb.12100

[41] ZHAO, Y., ZENG, D., RUSH, A. J. and KOSOROK, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *J. Amer. Statist. Assoc.* **107** 1106–1118. MR3010898 https://doi.org/10.1080/01621459.2012.695674

The Annals of Statistics
2022, Vol. 50, No. 4, 2306–2333
https://doi.org/10.1214/22-AOS2190

ON THE ROBUSTNESS OF MINIMUM NORM INTERPOLATORS AND REGULARIZED EMPIRICAL RISK MINIMIZERS

BY GEOFFREY CHINOT[a], MATTHIAS LÖFFLER[b] AND SARA VAN DE GEER[c]

Seminar for Statistics, Department of Mathematics, ETH Zürich, [a]*geoffrey.chinot@stat.math.ethz.ch,*
[b]*matthias.loeffler@stat.math.ethz.ch,* [c]*sara.vandegeer@stat.math.ethz.ch*

This article develops a general theory for minimum norm interpolating estimators and regularized empirical risk minimizers (RERM) in linear models in the presence of additive, potentially adversarial, errors. In particular, no conditions on the errors are imposed. A quantitative bound for the prediction error is given, relating it to the Rademacher complexity of the covariates, the norm of the minimum norm interpolator of the errors and the size of the subdifferential around the true parameter.

The general theory is illustrated for Gaussian features and several norms: The ℓ_1, ℓ_2, group Lasso and nuclear norms. In case of sparsity or low-rank inducing norms, minimum norm interpolators and RERM yield a prediction error of the order of the average noise level, provided that the overparameterization is at least a logarithmic factor larger than the number of samples and that, in case of RERM, the regularization parameter is small enough.

Lower bounds that show near optimality of the results complement the analysis.

REFERENCES

[1] ARORA, S., COHEN, N., HU, W. and LUO, Y. (2019). Implicit regularization in deep matrix factorization. In *Neural Information Processing Systems* (*NeurIPS*).

[2] BARTLETT, P. L., LONG, P. M., LUGOSI, G. and TSIGLER, A. (2020). Benign overfitting in linear regression. *Proc. Natl. Acad. Sci. USA* **117** 30063–30070. MR4263288 https://doi.org/10.1073/pnas.1907378117

[3] BELKIN, M., HSU, D., MA, S. and MANDAL, S. (2019). Reconciling modern machine-learning practice and the classical bias-variance trade-off. *Proc. Natl. Acad. Sci. USA* **116** 15849–15854. MR3997901 https://doi.org/10.1073/pnas.1903070116

[4] BELLEC, P. C. (2019). Localized Gaussian width of M-convex hulls with applications to Lasso and convex aggregation. *Bernoulli* **25** 3016–3040. MR4003572 https://doi.org/10.3150/18-BEJ1078

[5] BENNING, M. and BURGER, M. (2018). Modern regularization methods for inverse problems. *Acta Numer.* **27** 1–111. MR3826506 https://doi.org/10.1017/s0962492918000016

[6] 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

[7] 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

[8] BOYD, S. and VANDENBERGHE, L. (2004). *Convex Optimization*. Cambridge Univ. Press, Cambridge. MR2061575 https://doi.org/10.1017/CBO9780511804441

[9] BRUGIAPAGLIA, S. and ADCOCK, B. (2018). Robustness to unknown error in sparse regularization. *IEEE Trans. Inf. Theory* **64** 6638–6661. MR3860752 https://doi.org/10.1109/TIT.2017.2788445

[10] CAI, T. T., LIANG, T. and RAKHLIN, A. (2016). Geometric inference for general high-dimensional linear inverse problems. *Ann. Statist.* **44** 1536–1563. MR3519932 https://doi.org/10.1214/15-AOS1426

[11] CANDÈS, E. J., ROMBERG, J. K. and TAO, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.* **59** 1207–1223. MR2230846 https://doi.org/10.1002/cpa.20124

MSC2020 subject classifications. Primary 62J05; secondary 65F45.

Key words and phrases. Sparse linear regression, regularization, basis pursuit, trace regression, interpolation, minimum norm interpolation.

[12] CANDÈS, E. J. and TAO, T. (2010). The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inf. Theory* **56** 2053–2080. MR2723472 https://doi.org/10.1109/TIT.2010.2044061

[13] CASELLA, G. (1980). Minimax ridge regression estimation. *Ann. Statist.* **8** 1036–1056. MR0585702

[14] CHANDRASEKARAN, V., RECHT, B., PARRILO, P. A. and WILLSKY, A. S. (2012). The convex geometry of linear inverse problems. *Found. Comput. Math.* **12** 805–849. MR2989474 https://doi.org/10.1007/s10208-012-9135-7

[15] CHATTERJI, N. S. and LONG, P. M. (2021). Finite-sample analysis of interpolating linear classifiers in the overparameterized regime. *J. Mach. Learn. Res.* **22** Paper No. 129, 30 pp. MR4279780

[16] CHEN, S. S., DONOHO, D. L. and SAUNDERS, M. A. (1998). Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **20** 33–61. MR1639094 https://doi.org/10.1137/S1064827596304010

[17] CHINOT, G., KUCHELMEISTER, F., LÖFFLER, M. and VAN DE GEER, S. (2021). AdaBoost and robust one-bit compressed sensing. Preprint. Available at arXiv:2105.02083.

[18] CHINOT, G. and LERASLE, M. (2020). On the robustness of the minimum ℓ_2 interpolator. Preprint. Available at arXiv:2003.05838.

[19] DAVIDSON, K. R. and SZAREK, S. J. (2001). Local operator theory, random matrices and Banach spaces. In *Handbook of the Geometry of Banach Spaces*, *Vol. I* 317–366. North-Holland, Amsterdam. MR1863696 https://doi.org/10.1016/S1874-5849(01)80010-3

[20] DEVORE, R., PETROVA, G. and WOJTASZCZYK, P. (2009). Instance-optimality in probability with an l_1-minimization decoder. *Appl. Comput. Harmon. Anal.* **27** 275–288. MR2559727 https://doi.org/10.1016/j.acha.2009.05.001

[21] DE LA PEÑA, V. H. and GINÉ, E. (2012). *Decoupling*: *From Dependence to Independence*. *Probability and Its Applications* (*New York*). Springer, New York. MR1666908 https://doi.org/10.1007/978-1-4612-0537-1

[22] FAZEL, M., HINDI, H. and BOYD, S. (2001). A rank minimization heuristic with application to minimum order system approximation. In *Proceedings of the American Control Conference*, 4734–4739. IEEE.

[23] FOUCART, S. (2014). Stability and robustness of ℓ_1-minimizations with Weibull matrices and redundant dictionaries. *Linear Algebra Appl.* **441** 4–21. MR3134335 https://doi.org/10.1016/j.laa.2012.10.003

[24] GAO, C., VAN DER VAART, A. W. and ZHOU, H. H. (2020). A general framework for Bayes structured linear models. *Ann. Statist.* **48** 2848–2878. MR4152123 https://doi.org/10.1214/19-AOS1909

[25] 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

[26] GUÉDON, O., KRAHMER, F., KÜMMERLE, C., MENDELSON, S. and RAUHUT, H. (2022). On the geometry of polytopes generated by heavy-tailed random vectors. *Commun. Contemp. Math.* **24** Paper No. 2150056, 31 pp. MR4400194 https://doi.org/10.1142/S0219199721500565

[27] HASTIE, T., MONTANARI, A., ROSSET, S. and TIBSHIRANI, R. J. (2022). Surprises in high-dimensional ridgeless least squares interpolation. *Ann. Statist.* **50** 949–986. MR4404925 https://doi.org/10.1214/21-aos2133

[28] HOERL, A. E. and KENNARD, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12** 55–67.

[29] JU, P., LIN, X. and LIU, J. (2020). Overfitting can be harmless for basis pursuit: Only to a degree. In *Advances in Neural Information Processing Systems* (*NeurIPS*) **33**.

[30] KOLTCHINSKII, V. and MENDELSON, S. (2015). Bounding the smallest singular value of a random matrix without concentration. *Int. Math. Res. Not. IMRN* **2015** 12991–13008. MR3431642 https://doi.org/10.1093/imrn/rnv096

[31] KRAHMER, F., KÜMMERLE, C. and MELNYK, O. (2020). On the robustness of noise-blind low-rank recovery from rank-one measurements. Preprint. Available at arXiv:2010.12402.

[32] KRAHMER, F., KÜMMERLE, C. and RAUHUT, H. (2018). A quotient property for matrices with heavy-tailed entries and its application to noise-blind compressed sensing. Preprint. Available at arXiv:1806.04261.

[33] LAURENT, B. and MASSART, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* **28** 1302–1338. MR1805785 https://doi.org/10.1214/aos/1015957395

[34] LECUÉ, G. and MENDELSON, S. (2017). Regularization and the small-ball method II: Complexity dependent error rates. *J. Mach. Learn. Res.* **18** Paper No. 146, 48 pp. MR3763780

[35] LECUÉ, G. and MENDELSON, S. (2018). Regularization and the small-ball method I: Sparse recovery. *Ann. Statist.* **46** 611–641. MR3782379 https://doi.org/10.1214/17-AOS1562

[36] LIANG, T. and RAKHLIN, A. (2020). Just interpolate: Kernel "ridgeless" regression can generalize. *Ann. Statist.* **48** 1329–1347. MR4124325 https://doi.org/10.1214/19-AOS1849

[37] LIANG, T. and RECHT, B. (2021). Interpolating classifiers make few mistakes. Preprint. Available at arXiv:2101.11815.

[38] LIANG, T. and SUR, P. (2022). A precise high-dimensional asymptotic theory for boosting and minimum-$\ell 1$-norm interpolated classifiers. *Ann. Statist.* **50** 1669–1695. MR4441136 https://doi.org/10.1214/22-aos2170

[39] LITVAK, A. E., PAJOR, A., RUDELSON, M. and TOMCZAK-JAEGERMANN, N. (2005). Smallest singular value of random matrices and geometry of random polytopes. *Adv. Math.* **195** 491–523. MR2146352 https://doi.org/10.1016/j.aim.2004.08.004

[40] LOUNICI, K., PONTIL, M., VAN DE GEER, S. and TSYBAKOV, A. B. (2011). Oracle inequalities and optimal inference under group sparsity. *Ann. Statist.* **39** 2164–2204. MR2893865 https://doi.org/10.1214/11-AOS896

[41] MEI, S. and MONTANARI, A. (2022). The generalization error of random features regression: Precise asymptotics and the double descent curve. *Comm. Pure Appl. Math.* **75** 667–766. MR4400901 https://doi.org/10.1002/cpa.22008

[42] MENDELSON, S. (2014). Learning without concentration. In *Conference on Learning Theory* (*COLT*) 25–39.

[43] MONTANARI, A., RUAN, F., SOHN, Y. and YAN, J. (2020). The generalization error of max-margin linear classifiers: High-dimensional asymptotics in the overparametrized regime. Preprint. Available at arXiv:1911.01544.

[44] MUTHUKUMAR, V., NARANG, A., SUBRAMANIAN, V., BELKIN, M., HSU, D. and SAHAI, A. (2021). Classification vs regression in overparameterized regimes: Does the loss function matter? *J. Mach. Learn. Res.* **22** Paper No. 222, 69 pp. MR4329801

[45] MUTHUKUMAR, V., VODRAHALLI, K., SUBRAMANIAN, V. and SAHAI, A. (2020). Harmless interpolation of noisy data in regression. *IEEE J. Sel. Areas Inf. Theory* **1** 67–83.

[46] RECHT, B. (2011). A simpler approach to matrix completion. *J. Mach. Learn. Res.* **12** 3413–3430. MR2877360

[47] RECHT, B., FAZEL, M. and PARRILO, P. A. (2010). Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.* **52** 471–501. MR2680543 https://doi.org/10.1137/070697835

[48] ROSSET, S., ZHU, J. and HASTIE, T. (2004). Boosting as a regularized path to a maximum margin classifier. *J. Mach. Learn. Res.* **5** 941–973. MR2248005

[49] SOUDRY, D., HOFFER, E., NACSON, M. S., GUNASEKAR, S. and SREBRO, N. (2018). The implicit bias of gradient descent on separable data. *J. Mach. Learn. Res.* **19** Paper No. 70, 57 pp. MR3899772

[50] TELGARSKY, M. (2013). Margins, shrinkage, and boosting. In *International Conference on Machine Learning* (*ICML*).

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

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

[53] WOJTASZCZYK, P. (2010). Stability and instance optimality for Gaussian measurements in compressed sensing. *Found. Comput. Math.* **10** 1–13. MR2591836 https://doi.org/10.1007/s10208-009-9046-4

[54] WYNER, A. J., OLSON, M., BLEICH, J. and MEASE, D. (2017). Explaining the success of AdaBoost and random forests as interpolating classifiers. *J. Mach. Learn. Res.* **18** Paper No. 48, 33 pp. MR3670493

[55] 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

[56] ZHANG, C., BENGIO, S., HARDT, M., RECHT, B. and VINYALS, O. (2017). Understanding deep learning requires rethinking generalization. In *International Conference on Learning Representations* (*ICLR*).

The Annals of Statistics
2022, Vol. 50, No. 4, 2334–2358
https://doi.org/10.1214/22-AOS2191

RANDOM GRAPH ASYMPTOTICS FOR TREATMENT EFFECT ESTIMATION UNDER NETWORK INTERFERENCE

BY SHUANGNING LI[1,a] AND STEFAN WAGER[2,b]

[1]*Department of Statistics, Stanford University,* [a]*lsn@stanford.edu*
[2]*Graduate School of Business, Stanford University,* [b]*swager@stanford.edu*

The network interference model for treatment effect estimation places experimental units at the vertices of an undirected exposure graph, such that treatment assigned to one unit may affect the outcome of another unit if and only if these two units are connected by an edge. This model has recently gained popularity as means of incorporating interference effects into the Neyman–Rubin potential outcomes framework; and several authors have considered estimation of various causal targets, including the direct and indirect effects of treatment. In this paper, we consider large-sample asymptotics for treatment effect estimation under network interference in a setting where the exposure graph is a random draw from a graphon. When targeting the direct effect, we establish a central limit theorem and find that—in our setting—popular estimators are considerably more accurate than existing results suggest. Meanwhile, when targeting the indirect effect, we leverage our generative assumptions to propose a consistent estimator in a setting where no other consistent estimators are currently available. Overall, our results highlight the promise of random graph asymptotics in understanding the practicality and limits of causal inference under network interference.

REFERENCES

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

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

ARONOW, P. M. and SAMII, C. (2017). Estimating average causal effects under general interference, with application to a social network experiment. *Ann. Appl. Stat.* **11** 1912–1947. MR3743283 https://doi.org/10.1214/16-AOAS1005

ATHEY, S., ECKLES, D. and IMBENS, G. W. (2018). Exact p-values for network interference. *J. Amer. Statist. Assoc.* **113** 230–240. MR3803460 https://doi.org/10.1080/01621459.2016.1241178

ATHREYA, A., FISHKIND, D. E., TANG, M., PRIEBE, C. E., PARK, Y., VOGELSTEIN, J. T., LEVIN, K., LYZINSKI, V., QIN, Y. et al. (2017). Statistical inference on random dot product graphs: A survey. *J. Mach. Learn. Res.* **18** Paper No. 226, 92 pp. MR3827114

BAIRD, S., BOHREN, J. A., MCINTOSH, C. and ÖZLER, B. (2018). Optimal design of experiments in the presence of interference. *Rev. Econ. Stat.* **100** 844–860.

BASSE, G. and FELLER, A. (2018). Analyzing two-stage experiments in the presence of interference. *J. Amer. Statist. Assoc.* **113** 41–55. MR3803438 https://doi.org/10.1080/01621459.2017.1323641

BASSE, G. W., FELLER, A. and TOULIS, P. (2019). Randomization tests of causal effects under interference. *Biometrika* **106** 487–494. MR3949317 https://doi.org/10.1093/biomet/asy072

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

CHATTERJEE, S. (2015). Matrix estimation by universal singular value thresholding. *Ann. Statist.* **43** 177–214. MR3285604 https://doi.org/10.1214/14-AOS1272

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

MSC2020 subject classifications. 62G20.
Key words and phrases. Causal inference, direct and indirect effects, graphon, potential outcome.

DUFLO, E., GREENSTONE, M., PANDE, R. and RYAN, N. (2013). Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India. *Q. J. Econ.* **128** 1499–1545.
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
HALLORAN, M. E. and STRUCHINER, C. J. (1995). Causal inference in infectious diseases. *Epidemiology* **6** 142–151.
HU, Y., LI, S. and WAGER, S. (2022). Average direct and indirect causal effects under interference. *Biometrika*.
HUDGENS, M. G. and HALLORAN, M. E. (2008). Toward causal inference with interference. *J. Amer. Statist. Assoc.* **103** 832–842. MR2435472 https://doi.org/10.1198/016214508000000292
IMBENS, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Rev. Econ. Stat.* **86** 4–29.
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
LEUNG, M. P. (2020). Treatment and spillover effects under network interference. *Rev. Econ. Stat.* **102** 368–380.
LI, S. and WAGER, S. (2022). Supplement to "Random graph asymptotics for treatment effect estimation under network interference." https://doi.org/10.1214/22-AOS2191SUPP
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
MANSKI, C. F. (2013). Identification of treatment response with social interactions. *Econom. J.* **16** S1–S23. MR3030060 https://doi.org/10.1111/j.1368-423X.2012.00368.x
NEYMAN, J. (1923). Sur les applications de la théorie des probabilités aux experiences agricoles: Essai des principes. *Rocz. Nauk Rol.* **10** 1–51.
PARISE, F. and OZDAGLAR, A. (2019). Graphon games. In *Proceedings of the* 2019 *ACM Conference on Economics and Computation* 457–458.
ROSS, N. (2011). Fundamentals of Stein's method. *Probab. Surv.* **8** 210–293. MR2861132 https://doi.org/10.1214/11-PS182
RUDELSON, M. and VERSHYNIN, R. (2013). Hanson–Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.* **18** no. 82, 9 pp. MR3125258 https://doi.org/10.1214/ECP.v18-2865
SÄVJE, F., ARONOW, P. M. and HUDGENS, M. G. (2021). Average treatment effects in the presence of unknown interference. *Ann. Statist.* **49** 673–701. MR4255103 https://doi.org/10.1214/20-aos1973
STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. MR0630098
TCHETGEN TCHETGEN, E. J. and VANDERWEELE, T. J. (2012). On causal inference in the presence of interference. *Stat. Methods Med. Res.* **21** 55–75. MR2867538 https://doi.org/10.1177/0962280210386779
VANDERWEELE, T. J. and TCHETGEN TCHETGEN, E. J. (2011). Effect partitioning under interference in two-stage randomized vaccine trials. *Statist. Probab. Lett.* **81** 861–869. MR2793754 https://doi.org/10.1016/j.spl.2011.02.019
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
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
ZHANG, Y., LEVINA, E. and ZHU, J. (2017). Estimating network edge probabilities by neighbourhood smoothing. *Biometrika* **104** 771–783. MR3737303 https://doi.org/10.1093/biomet/asx042

The Annals of Statistics
2022, Vol. 50, No. 4, 2359–2385
https://doi.org/10.1214/22-AOS2196

AN ℓ_p THEORY OF PCA AND SPECTRAL CLUSTERING

BY EMMANUEL ABBE[1,a], JIANQING FAN[2,b] AND KAIZHENG WANG[3,c]

[1]*Institute of Mathematics, EPFL,* [a]*emmanuel.abbe@epfl.ch*
[2]*Department of ORFE, Princeton University,* [b]*jqfan@princeton.edu*
[3]*Department of IEOR and Data Science Institute, Columbia University,* [c]*kaizheng.wang@columbia.edu*

Principal Component Analysis (PCA) is a powerful tool in statistics and machine learning. While existing study of PCA focuses on the recovery of principal components and their associated eigenvalues, there are few precise characterizations of individual principal component scores that yield low-dimensional embedding of samples. That hinders the analysis of various spectral methods. In this paper, we first develop an ℓ_p perturbation theory for a hollowed version of PCA in Hilbert spaces which provably improves upon the vanilla PCA in the presence of heteroscedastic noises. Through a novel ℓ_p analysis of eigenvectors, we investigate entrywise behaviors of principal component score vectors and show that they can be approximated by linear functionals of the Gram matrix in ℓ_p norm, which includes ℓ_2 and ℓ_∞ as special cases. For sub-Gaussian mixture models, the choice of p giving optimal bounds depends on the signal-to-noise ratio, which further yields optimality guarantees for spectral clustering. For contextual community detection, the ℓ_p theory leads to simple spectral algorithms that achieve the information threshold for exact recovery and the optimal misclassification rate.

REFERENCES

ABBE, E. (2017). Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** 177. 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., FAN, J. and WANG, K. (2022). Supplement to "An ℓ_p theory of PCA and spectral clustering." https://doi.org/10.1214/22-AOS2196SUPP

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

AMINI, A. A. and RAZAEE, Z. S. (2021). Concentration of kernel matrices with application to kernel spectral clustering. *Ann. Statist.* **49** 531–556. MR4206689 https://doi.org/10.1214/20-AOS1967

ANDERSON, T. W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Stat.* **34** 122–148. MR0145620 https://doi.org/10.1214/aoms/1177704248

ARONSZAJN, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68** 337–404. MR0051437 https://doi.org/10.2307/1990404

AWASTHI, P., BANDEIRA, A. S., CHARIKAR, M., KRISHNASWAMY, R., VILLAR, S. and WARD, R. (2015). Relax, no need to round: Integrality of clustering formulations. In *ITCS'15—Proceedings of the 6th Innovations in Theoretical Computer Science* 191–200. ACM, New York. MR3419009

AZIZYAN, M., SINGH, A. and WASSERMAN, L. (2013). Minimax theory for high-dimensional Gaussian mixtures with sparse mean separation. In *Advances in Neural Information Processing Systems* 2139–2147.

BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. MR2165575 https://doi.org/10.1214/009117905000000233

BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2012). The singular values and vectors of low rank perturbations of large rectangular random matrices. *J. Multivariate Anal.* **111** 120–135. MR2944410 https://doi.org/10.1016/j.jmva.2012.04.019

MSC2020 subject classifications. Primary 62H25; secondary 62H30, 60B20.

Key words and phrases. Principal component analysis, eigenvector perturbation, spectral clustering, mixture models, community detection, contextual network models, phase transitions.

BINKIEWICZ, N., VOGELSTEIN, J. T. and ROHE, K. (2017). Covariate-assisted spectral clustering. *Biometrika* **104** 361–377. MR3698259 https://doi.org/10.1093/biomet/asx008
BLANCHARD, G., BOUSQUET, O. and ZWALD, L. (2007). Statistical properties of kernel principal component analysis. *Mach. Learn.* **66** 259–294.
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
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
CANDÈS, E. J. and RECHT, B. (2009). Exact matrix completion via convex optimization. *Found. Comput. Math.* **9** 717–772. MR2565240 https://doi.org/10.1007/s10208-009-9045-5
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
CHATTERJEE, S. (2014). *Superconcentration and Related Topics*. *Springer Monographs in Mathematics*. Springer, Cham. MR3157205 https://doi.org/10.1007/978-3-319-03886-5
CHEN, X. and YANG, Y. (2021a). Hanson–Wright inequality in Hilbert spaces with application to K-means clustering for non-Euclidean data. *Bernoulli* **27** 586–614. MR4177382 https://doi.org/10.3150/20-BEJ1251
CHEN, X. and YANG, Y. (2021b). Cutoff for exact recovery of Gaussian mixture models. *IEEE Trans. Inf. Theory* **67** 4223–4238. MR4289373 https://doi.org/10.1109/TIT.2021.3063155
CHEN, Y., FAN, J., MA, C. and WANG, K. (2019). Spectral method and regularized MLE are both optimal for top-K ranking. *Ann. Statist.* **47** 2204–2235. MR3953449 https://doi.org/10.1214/18-AOS1745
CRISTIANINI, N. and SHAWE-TAYLOR, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press, Cambridge.
DAMLE, A. and SUN, Y. (2020). Uniform bounds for invariant subspace perturbations. *SIAM J. Matrix Anal. Appl.* **41** 1208–1236. MR4138209 https://doi.org/10.1137/19M1262760
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
DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38. MR0501537
DESHPANDE, Y., SEN, S., MONTANARI, A. and MOSSEL, E. (2018). Contextual stochastic block models. In *Advances in Neural Information Processing Systems* 8581–8593.
EL KAROUI, N. (2018). On the impact of predictor geometry on the performance on high-dimensional ridge-regularized generalized robust regression estimators. *Probab. Theory Related Fields* **170** 95–175. MR3748322 https://doi.org/10.1007/s00440-016-0754-9
ELDRIDGE, J., BELKIN, M. and WANG, Y. (2018). Unperturbed: Spectral analysis beyond Davis–Kahan. In *Algorithmic Learning Theory* 321–358. PMLR.
ERDŐS, L., SCHLEIN, B. and YAU, H.-T. (2009). Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Ann. Probab.* **37** 815–852. MR2537522 https://doi.org/10.1214/08-AOP421
FAN, J., WANG, W. and ZHONG, Y. (2019). An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. *J. Econometrics* **208** 5–22.
FEI, Y. and CHEN, Y. (2018). Hidden integrality of SDP relaxations for sub-Gaussian mixture models. In *Conference on Learning Theory* 1931–1965.
FEIGE, U. and OFEK, E. (2005). Spectral techniques applied to sparse random graphs. *Random Structures Algorithms* **27** 251–275. MR2155709 https://doi.org/10.1002/rsa.20089
GAO, C. and ZHANG, A. Y. (2022). Iterative algorithm for discrete structure recovery. *Ann. Statist.* **50** 1066–1094. MR4404929 https://doi.org/10.1214/21-aos2140
GIRAUD, C. and VERZELEN, N. (2019). Partial recovery bounds for clustering with the relaxed K-means. *Math. Statist. Learn.* **1** 317–374.
GROSS, D. (2011). Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inf. Theory* **57** 1548–1566. MR2815834 https://doi.org/10.1109/TIT.2011.2104999
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
HSU, D. and KAKADE, S. M. (2013). Learning mixtures of spherical Gaussians: Moment methods and spectral decompositions. In *ITCS'13—Proceedings of the* 2013 *ACM Conference on Innovations in Theoretical Computer Science* 11–19. ACM, New York. MR3385380
JAVANMARD, A. and MONTANARI, A. (2018). Debiasing the Lasso: Optimal sample size for Gaussian designs. *Ann. Statist.* **46** 2593–2622. MR3851749 https://doi.org/10.1214/17-AOS1630
JIN, J. and WANG, W. (2016). Influential features PCA for high dimensional clustering. *Ann. Statist.* **44** 2323–2359. MR3576543 https://doi.org/10.1214/15-AOS1423

JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961 https://doi.org/10.1214/aos/1009210544

JOHNSTONE, I. M. and LU, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* **104** 682–693. MR2751448 https://doi.org/10.1198/jasa.2009.0121

JUNG, S. and MARRON, J. S. (2009). PCA consistency in high dimension, low sample size context. *Ann. Statist.* **37** 4104–4130. MR2572454 https://doi.org/10.1214/09-AOS709

KOLTCHINSKII, V. and GINÉ, E. (2000). Random matrix approximation of spectra of integral operators. *Bernoulli* **6** 113–167. MR1781185 https://doi.org/10.2307/3318636

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

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

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

KUMAR, A. and KANNAN, R. (2010). Clustering with spectral norm and the k-means algorithm. In 2010 *IEEE 51st Annual Symposium on Foundations of Computer Science—FOCS* 2010 299–308. IEEE Computer Soc., Los Alamitos, CA. MR3025203

KUMAR, A., SABHARWAL, Y. and SEN, S. (2004). A simple linear time $(1+\varepsilon)$-approximation algorithm for k-means clustering in any dimensions. In 45*th Annual IEEE Symposium on Foundations of Computer Science* 454–462. IEEE, Los Alamitos.

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

LLOYD, S. P. (1982). Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **28** 129–137. MR0651807 https://doi.org/10.1109/TIT.1982.1056489

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

LU, Y. and ZHOU, H. H. (2016). Statistical and computational guarantees of Lloyd's algorithm and its variants. ArXiv preprint. Available at arXiv:1612.02099.

MA, Z. and MA, Z. (2017). Exploration of large networks with covariates via fast and universal latent space model fitting. ArXiv preprint. Available at arXiv:1705.02372.

MAO, X., SARKAR, P. and CHAKRABARTI, D. (2021). Estimating mixed memberships with sharp eigenvector deviations. *J. Amer. Statist. Assoc.* **116** 1928–1940. MR4353723 https://doi.org/10.1080/01621459.2020.1751645

MELE, A., HAO, L., CAPE, J. and PRIEBE, C. E. (2019). Spectral inference for large stochastic blockmodels with nodal covariates. ArXiv preprint. Available at arXiv:1908.06438.

MIXON, D. G., VILLAR, S. and WARD, R. (2017). Clustering subgaussian mixtures by semidefinite programming. *Inf. Inference* **6** 389–415. MR3764529 https://doi.org/10.1093/imaiai/iax001

MONTANARI, A. and SUN, N. (2018). Spectral algorithms for tensor completion. *Comm. Pure Appl. Math.* **71** 2381–2425. MR3862094 https://doi.org/10.1002/cpa.21748

NADLER, B. (2008). Finite sample approximation results for principal component analysis: A matrix perturbation approach. *Ann. Statist.* **36** 2791–2817. MR2485013 https://doi.org/10.1214/08-AOS618

NDAOUD, M. (2018). Sharp optimal recovery in the two component Gaussian mixture model. ArXiv preprint. Available at arXiv:1812.08078.

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

NG, A. Y., JORDAN, M. I. and WEISS, Y. (2002). On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems* 849–856.

O'ROURKE, S., VU, V. and WANG, K. (2018). Random perturbation of low rank matrices: Improving classical bounds. *Linear Algebra Appl.* **540** 26–59. MR3739989 https://doi.org/10.1016/j.laa.2017.11.014

PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17** 1617–1642. MR2399865

PEARSON, K. (1894). Contributions to the mathematical theory of evolution. *Philos. Trans. R. Soc. Lond. A* **185** 71–110.

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

ROYER, M. (2017). Adaptive clustering through semidefinite programming. In *Advances in Neural Information Processing Systems* 1795–1803.

SCHÖLKOPF, B., SMOLA, A. and MÜLLER, K.-R. (1997). Kernel principal component analysis. In *International Conference on Artificial Neural Networks* 583–588. Springer, Berlin.

SHI, J. and MALIK, J. (2000). Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **22** 888–905.

SRIVASTAVA, P. R., SARKAR, P. and HANASUSANTO, G. A. (2019). A robust spectral clustering algorithm for sub-Gaussian mixture models with outliers. ArXiv preprint. Available at arXiv:1912.07546.

STEWART, G. W. and SUN, J. G. (1990). *Matrix Perturbation Theory. Computer Science and Scientific Computing*. Academic Press, Boston, MA. MR1061154

VEMPALA, S. and WANG, G. (2004). A spectral algorithm for learning mixture models. *J. Comput. System Sci.* **68** 841–860. MR2059647 https://doi.org/10.1016/j.jcss.2003.11.008

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

WAHBA, G. (1965). A least squares estimate of satellite attitude. *SIAM Rev.* **7** 409–409.

WANG, K. (2019). Some compact notations for concentration inequalities and user-friendly results. ArXiv preprint. Available at arXiv:1912.13463.

WANG, W. and FAN, J. (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Ann. Statist.* **45** 1342–1374. MR3662457 https://doi.org/10.1214/16-AOS1487

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

WENG, H. and FENG, Y. (2016). Community detection with nodal information. ArXiv preprint. Available at arXiv:1610.09735.

YAN, B. and SARKAR, P. (2021). Covariate regularized community detection in sparse graphs. *J. Amer. Statist. Assoc.* **116** 734–745. MR4270020 https://doi.org/10.1080/01621459.2019.1706541

YEUNG, K. Y. and RUZZO, W. L. (2001). Principal component analysis for clustering gene expression data. *Bioinformatics* **17** 763–774.

ZHANG, A. R., CAI, T. T. and WU, Y. (2022). Heteroskedastic PCA: Algorithm, optimality, and applications. *Ann. Statist.* **50** 53–80. MR4382008 https://doi.org/10.1214/21-aos2074

ZHANG, Y., LEVINA, E. and ZHU, J. (2016). Community detection in networks with node features. *Electron. J. Stat.* **10** 3153–3178. MR3571965 https://doi.org/10.1214/16-EJS1206

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

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

ZWALD, L. and BLANCHARD, G. (2006). On the convergence of eigenspaces in kernel principal component analysis. In *Advances in Neural Information Processing Systems* 1649–1656.

The Annals of Statistics
2022, Vol. 50, No. 4, 2386–2415
https://doi.org/10.1214/22-AOS2197

ESTIMATION OF SMOOTH FUNCTIONALS IN HIGH-DIMENSIONAL MODELS: BOOTSTRAP CHAINS AND GAUSSIAN APPROXIMATION

BY VLADIMIR KOLTCHINSKII[a]

School of Mathematics, Georgia Institute of Technology, [a]*vlad@math.gatech.edu*

Let $X^{(n)}$ be an observation sampled from a distribution $P_\theta^{(n)}$ with an unknown parameter θ, θ being a vector in a Banach space E (most often, a high-dimensional space of dimension d). We study the problem of estimation of $f(\theta)$ for a functional $f : E \mapsto \mathbb{R}$ of some smoothness $s > 0$ based on an observation $X^{(n)} \sim P_\theta^{(n)}$. Assuming that there exists an estimator $\hat{\theta}_n = \hat{\theta}_n(X^{(n)})$ of parameter θ such that $\sqrt{n}(\hat{\theta}_n - \theta)$ is sufficiently close in distribution to a mean zero Gaussian random vector in E, we construct a functional $g : E \mapsto \mathbb{R}$ such that $g(\hat{\theta}_n)$ is an asymptotically normal estimator of $f(\theta)$ with $\sqrt{n}$ rate provided that $s > \frac{1}{1-\alpha}$ and $d \le n^\alpha$ for some $\alpha \in (0, 1)$. We also derive general upper bounds on Orlicz norm error rates for estimator $g(\hat{\theta})$ depending on smoothness s, dimension d, sample size n and the accuracy of normal approximation of $\sqrt{n}(\hat{\theta}_n - \theta)$. In particular, this approach yields asymptotically efficient estimators in high-dimensional log-concave exponential models.

REFERENCES

[1] ANASTASIOU, A. (2018). Assessing the multivariate normal approximation of the maximum likelihood estimator from high-dimensional, heterogeneous data. *Electron. J. Stat.* **12** 3794–3828. MR3881763 https://doi.org/10.1214/18-EJS1492

[2] ANASTASIOU, A. and GAUNT, R. E. (2021). Wasserstein distance error bounds for the multivariate normal approximation of the maximum likelihood estimator. *Electron. J. Stat.* **15** 5758–5810. MR4355697 https://doi.org/10.1214/21-ejs1920

[3] ARRAS, B. and HOUDRÉ, C. (2019). On Stein's method for multivariate self-decomposable laws. *Electron. J. Probab.* **24** 128. MR4029431 https://doi.org/10.1214/19-ejp378

[4] BENTKUS, V., BLOZNELIS, M. and GÖTZE, F. (1997). A Berry–Esséen bound for M-estimators. *Scand. J. Stat.* **24** 485–502. MR1615335 https://doi.org/10.1111/1467-9469.00076

[5] BHATTACHARYA, R. N. and RANGA RAO, R. (1976). *Normal Approximation and Asymptotic Expansions. Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0436272

[6] 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

[7] BICKEL, P. J. and RITOV, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhyā Ser. A* **50** 381–393. MR1065550

[8] BIRGÉ, L. and MASSART, P. (1995). Estimation of integral functionals of a density. *Ann. Statist.* **23** 11–29. MR1331653 https://doi.org/10.1214/aos/1176324452

[9] BOBKOV, S. G. and LEDOUX, M. (2000). From Brunn–Minkowski to Brascamp–Lieb and to logarithmic Sobolev inequalities. *Geom. Funct. Anal.* **10** 1028–1052. MR1800062 https://doi.org/10.1007/PL00001645

[10] CAI, T. T. and LOW, M. G. (2005). On adaptive estimation of linear functionals. *Ann. Statist.* **33** 2311–2343. MR2211088 https://doi.org/10.1214/009053605000000633

[11] CAI, T. T. and LOW, M. G. (2005). Nonquadratic estimators of a quadratic functional. *Ann. Statist.* **33** 2930–2956. MR2253108 https://doi.org/10.1214/009053605000000147

[12] ČENCOV, N. N. (1982). *Statistical Decision Rules and Optimal Inference. Translations of Mathematical Monographs* **53**. Amer. Math. Soc., Providence, RI. MR0645898

MSC2020 subject classifications. Primary 62H12; secondary 62G20, 62H25, 60B20.

Key words and phrases. Efficiency, smooth functionals, bootstrap chain, concentration inequalities, normal approximation.

[13] CHEN, Y. (2021). An almost constant lower bound of the isoperimetric coefficient in the KLS conjecture. *Geom. Funct. Anal.* **31** 34–61. MR4244847 https://doi.org/10.1007/s00039-021-00558-4
[14] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. MR3693963 https://doi.org/10.1214/16-AOP1113
[15] CHERNOZHUKOV, V., CHETVERIKOV, D., KATO, K. and KOIKE, Y. (2022). Improved central limit theorem and bootstrap approximation in high dimensions. *Ann. Statist.* To appear.
[16] CHERNOZHUKOV, V., CHETVERIKOV, D. and KOIKE, Y. Nearly optimal central limit theorem and bootstrap approximation in high dimensions. Available at arXiv:2012.09513.
[17] COLLIER, O., COMMINGES, L. and TSYBAKOV, A. B. (2017). Minimax estimation of linear and quadratic functionals on sparsity classes. *Ann. Statist.* **45** 923–958. MR3662444 https://doi.org/10.1214/15-AOS1432
[18] COURTADE, T. A., FATHI, M. and PANANJADY, A. (2019). Existence of Stein kernels under a spectral gap, and discrepancy bounds. *Ann. Inst. Henri Poincaré Probab. Stat.* **55** 777–790. MR3949953 https://doi.org/10.1214/18-aihp898
[19] DONOHO, D. and LIU, R. (1987). On minimax estimation of linear functionals Technical Report N 105, Department of Statistics, UC Berkeley.
[20] DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence. II, III. *Ann. Statist.* **19** 633–667. MR1105839 https://doi.org/10.1214/aos/1176348114
[21] DONOHO, D. L. and NUSSBAUM, M. (1990). Minimax quadratic estimation of a quadratic functional. *J. Complexity* **6** 290–323. MR1081043 https://doi.org/10.1016/0885-064X(90)90025-9
[22] ELDAN, R. (2013). Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geom. Funct. Anal.* **23** 532–569. MR3053755 https://doi.org/10.1007/s00039-013-0214-y
[23] 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
[24] FATHI, M. (2019). Higher order Stein Kernels for Gaussian approximation. *Studia Math.* **256** 241–258.
[25] GILL, R. D. and LEVIT, B. Y. (1995). Applications of the Van Trees inequality: A Bayesian Cramér–Rao bound. *Bernoulli* **1** 59–79. MR1354456 https://doi.org/10.2307/3318681
[26] GIRKO, V. L. (1987). An introduction to general statistical analysis. *Teor. Veroyatn. Primen.* **32** 252–265. MR0902754
[27] GIRKO, V. L. (1995). *Statistical Analysis of Observations of Increasing Dimension*. *Theory and Decision Library. Series B*: *Mathematical and Statistical Methods* **28**. Kluwer Academic, Dordrecht. MR1473719 https://doi.org/10.1007/978-94-015-8567-5
[28] HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. *Springer Series in Statistics*. Springer, New York. MR1145237 https://doi.org/10.1007/978-1-4612-4384-7
[29] HALL, P. and MARTIN, M. A. (1988). On bootstrap resampling and iteration. *Biometrika* **75** 661–671. MR0995110 https://doi.org/10.1093/biomet/75.4.661
[30] HAN, Y., JIAO, J. and MUKHERJEE, R. (2020). On estimation of L_r-norms in Gaussian white noise models. *Probab. Theory Related Fields* **177** 1243–1294. MR4126939 https://doi.org/10.1007/s00440-020-00982-x
[31] IBRAGIMOV, I. A. and HAS'MINSKIĬ, R. Z. (1981). *Statistical Estimation*: *Asymptotic Theory*. *Applications of Mathematics* **16**. Springer, New York-Berlin. MR0620321
[32] IBRAGIMOV, I. A., NEMIROVSKI, A. S. and KHASMINSKII, R. Z. (1987). Some problems of nonparametric estimation in Gaussian white noise. *Theory Probab. Appl.* **31** 391–406.
[33] JIAO, J. and HAN, Y. (2020). Bias correction with jackknife, bootstrap, and Taylor series. *IEEE Trans. Inf. Theory* **66** 4392–4418. MR4130622 https://doi.org/10.1109/TIT.2020.2969439
[34] KLEMELÄ, J. (2006). Sharp adaptive estimation of quadratic functionals. *Probab. Theory Related Fields* **134** 539–564. MR2214904 https://doi.org/10.1007/s00440-005-0447-2
[35] KOIKE, Y. (2021). Notes on the dimension dependence in high-dimensional central limit theorems for hyperrectangles. *Jpn. J. Stat. Data Sci.* **4** 257–297. MR4273258 https://doi.org/10.1007/s42081-020-00096-7
[36] KOLTCHINSKII, V. (2018). Asymptotic efficiency in high-dimensional covariance estimation. In *Proc. ICM* 2018 **3** 2891–2912, Rio de Janeiro.
[37] KOLTCHINSKII, V. (2021). Asymptotically efficient estimation of smooth functionals of covariance operators. *J. Eur. Math. Soc.* (*JEMS*) **23** 765–843. MR4210724 https://doi.org/10.4171/jems/1023
[38] KOLTCHINSKII, V. (2022). Supplement to "Estimation of smooth functionals in high-dimensional models: Bootstrap chains and Gaussian approximation." https://doi.org/10.1214/22-AOS2197SUPP
[39] KOLTCHINSKII, V., LÖFFLER, M. and NICKL, R. (2020). Efficient estimation of linear functionals of principal components. *Ann. Statist.* **48** 464–490. MR4065170 https://doi.org/10.1214/19-AOS1816

[40] KOLTCHINSKII, V. and ZHILOVA, M. (2021). Efficient estimation of smooth functionals in Gaussian shift models. *Ann. Inst. Henri Poincaré Probab. Stat.* **57** 351–386. MR4255178 https://doi.org/10.1214/20-aihp1081

[41] KOLTCHINSKII, V. and ZHILOVA, M. (2021). Estimation of smooth functionals in normal models: Bias reduction and asymptotic efficiency. *Ann. Statist.* **49** 2577–2610. MR4338376 https://doi.org/10.1214/20-aos2047

[42] KOLTCHINSKII, V. and ZHILOVA, M. (2021). Estimation of smooth functionals of location parameter in Gaussian and Poincaré random shift models. *Sankhya A* **83** 569–596. MR4284575 https://doi.org/10.1007/s13171-020-00232-1

[43] KUCHIBHOTLA, A. K., MUKHERJEE, S. and BANERJEE, D. (2021). High-dimensional CLT: Improvements, non-uniform extensions and large deviations. *Bernoulli* **27** 192–217. MR4177366 https://doi.org/10.3150/20-BEJ1233

[44] LAURENT, B. (1996). Efficient estimation of integral functionals of a density. *Ann. Statist.* **24** 659–681. MR1394981 https://doi.org/10.1214/aos/1032894458

[45] LEE, Y.-T. and VEMPALA, S. (2017). *Eldan's Stochastic Localization and the KLS Hyperplane Conjecture*: *An Improved Lower Bound for Expansion*. 58*th Annual IEEE Symposium on Foundations of Computer Science FOCM* 2017.

[46] LEPSKI, O., NEMIROVSKI, A. and SPOKOINY, V. (1999). On estimation of the L_r norm of a regression function. *Probab. Theory Related Fields* **113** 221–253. MR1670867 https://doi.org/10.1007/s004409970006

[47] LEVIT, B. (1975). On the efficiency of a class of non-parametric estimates. *Theory Probab. Appl.* **20** 723–740.

[48] LEVIT, B. (1978). Asymptotically efficient estimation of nonlinear functionals. *Probl. Pereda. Inf.* (*Probl. Inf. Transm.*) **14** 65–72.

[49] MILMAN, E. (2009). On the role of convexity in isoperimetry, spectral gap and concentration. *Invent. Math.* **177** 1–43. MR2507637 https://doi.org/10.1007/s00222-009-0175-9

[50] MUKHERJEE, R., NEWEY, W. and ROBINS, J. (2017). Semiparametric Efficient Empirical Higher Order Influence Function Estimators. Available at arXiv:1705.07577.

[51] NEMIROVSKI, A. (1990). On necessary conditions for the efficient estimation of functionals of a nonparametric signal which is observed in white noise. *Theory Probab. Appl.* **35** 94–103.

[52] NEMIROVSKI, A. (2000). *Topics in Non-parametric Statistics*. *Ecole D'Ete de Probabilités de Saint-Flour. Lecture Notes in Math*. **1738**. Springer, New York.

[53] PAULAUSKAS, V. and RAČKAUSKAS, A. (1989). *Approximation Theory in the Central Limit Theorem*. *Exact Results in Banach Spaces*. *Mathematics and Its Applications* (*Soviet Series*) **32**. Kluwer Academic, Dordrecht. MR1015294 https://doi.org/10.1007/978-94-011-7798-6

[54] PFANZAGL, J. (1971). The Berry–Esseen bound for minimum contrast estimates. *Metrika* **17** 82–91. MR0295467 https://doi.org/10.1007/BF02613813

[55] PINELIS, I. (2017). Optimal-order uniform and nonuniform bounds on the rate of convergence to normality for maximum likelihood estimators. *Electron. J. Stat.* **11** 1160–1179. MR3634332 https://doi.org/10.1214/17-EJS1264

[56] PORTNOY, S. (1986). On the central limit theorem in $\mathbf{R}^p$ when $p \to \infty$. *Probab. Theory Related Fields* **73** 571–583. MR0863546 https://doi.org/10.1007/BF00324853

[57] PORTNOY, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Ann. Statist.* **16** 356–366. MR0924876 https://doi.org/10.1214/aos/1176350710

[58] RIO, E. (2009). Upper bounds for minimal distances in the central limit theorem. *Ann. Inst. Henri Poincaré Probab. Stat.* **45** 802–817. MR2548505 https://doi.org/10.1214/08-AIHP187

[59] ROBINS, J., LI, L., TCHETGEN, E. and VAN DER VAART, A. (2008). Higher order influence functions and minimax estimation of nonlinear functionals. In *Probability and Statistics*: *Essays in Honor of David A. Freedman*. *Inst. Math. Stat.* (*IMS*) *Collect.* **2** 335–421. IMS, Beachwood, OH. MR2459958 https://doi.org/10.1214/193940307000000527

[60] ROBINS, J. M., LI, L., TCHETGEN, E. T. and VAN DER VAART, A. (2016). Asymptotic normality of quadratic estimators. *Stochastic Process. Appl.* **126** 3733–3759. MR3565475 https://doi.org/10.1016/j.spa.2016.04.005

[61] SENATOV, V. (1998). *Normal Approximation*: *New Results*, *Methods and Problems*. VSP, Utrecht, The Netherlands.

[62] VAN DER VAART, A. (2014). Higher order tangent spaces and influence functions. *Statist. Sci.* **29** 679–686. MR3300365 https://doi.org/10.1214/14-STS478

[63] VILLANI, C. (2009). *Optimal Transport*. *Old and New*. Springer, Berlin.

[64] ZOLOTAREV, V. M. (1976). Metric distances in spaces of random variables and of their distributions. *Mat. Sb.* (*N.S.*) **101(143)** 416–454. MR0467869

The Annals of Statistics
2022, Vol. 50, No. 4, 2416–2442
https://doi.org/10.1214/22-AOS2198

A MINIMAX FRAMEWORK FOR QUANTIFYING RISK-FAIRNESS TRADE-OFF IN REGRESSION

BY EVGENII CHZHEN[1,a] AND NICOLAS SCHREUDER[2,b]

[1]*Laboratoire de mathématiques d'Orsay, CNRS, Université Paris-Saclay,* [a]*evgenii.chzhen@cnrs.fr*
[2]*DIBRIS, MaLGa, Università di Genova,* [b]*schreuder.nicolas@gmail.com*

We propose a theoretical framework for the problem of learning a real-valued function which meets fairness requirements. This framework is built upon the notion of α-relative (fairness) improvement of the regression function which we introduce using the theory of optimal transport. Setting $\alpha = 0$ corresponds to the regression problem under the Demographic Parity constraint, while $\alpha = 1$ corresponds to the classical regression problem without any constraints. For $\alpha \in (0, 1)$ the proposed framework allows to continuously interpolate between these two extreme cases and to study partially fair predictors. Within this framework, we precisely quantify the cost in risk induced by the introduction of the fairness constraint. We put forward a statistical minimax setup and derive a general problem-dependent lower bound on the risk of any estimator satisfying α-relative improvement constraint. We illustrate our framework on a model of linear regression with Gaussian design and systematic group-dependent bias, deriving matching (up to absolute constants) upper and lower bounds on the minimax risk under the introduced constraint. We provide a general post-processing strategy which enjoys fairness, risk guarantees and can be applied on top of any black-box algorithm. Finally, we perform a simulation study of the linear model and numerical experiments of benchmark data, validating our theoretical contributions.

REFERENCES

AGARWAL, A., DUDIK, M. and WU, Z. S. (2019). Fair regression: Quantitative definitions and reduction-based algorithms. In *International Conference on Machine Learning*.

AGARWAL, A., BEYGELZIMER, A., DUDÍK, M., LANGFORD, J. and WALLACH, H. (2018). A reductions approach to fair classification. In *International Conference on Machine Learning*.

AGUEH, M. and CARLIER, G. (2011). Barycenters in the Wasserstein space. *SIAM J. Math. Anal.* **43** 904–924. MR2801182 https://doi.org/10.1137/100805741

AUDIBERT, J.-Y. and CATONI, O. (2011). Robust linear least squares regression. *Ann. Statist.* **39** 2766–2794. MR2906886 https://doi.org/10.1214/11-AOS918

BAHARLOUEI, S., NOUIEHED, M., BEIRAMI, A. and RAZAVIYAYN, M. (2019). Rényi fair inference. Preprint. Available at arXiv:1906.12005.

BARBER, R. F., CANDÈS, E. J., RAMDAS, A. and TIBSHIRANI, R. J. (2021). The limits of distribution-free conditional predictive inference. *Inf. Inference* **10** 455–482. MR4270755 https://doi.org/10.1093/imaiai/iaaa017

BAROCAS, S., HARDT, M. and NARAYANAN, A. (2019). *Fairness and Machine Learning*. fairmlbook.org.

BELLEC, P. C. (2017). Optimal exponential bounds for aggregation of density estimators. *Bernoulli* **23** 219–248. MR3556772 https://doi.org/10.3150/15-BEJ742

BERK, R., HEIDARI, H., JABBARI, S., JOSEPH, M., KEARNS, M., MORGENSTERN, J., NEEL, S. and ROTH, A. (2017). A convex framework for fair regression. In *Fairness, Accountability, and Transparency in Machine Learning*.

BERTSIMAS, D., FARIAS, V. and TRICHAKIS, N. (2012). On the efficiency-fairness trade-off. *Manage. Sci.* **58** 2234–2250.

BOBKOV, S. and LEDOUX, M. (2019). One-dimensional empirical measures, order statistics, and Kantorovich transport distances. *Mem. Amer. Math. Soc.* **261** v+126. MR4028181 https://doi.org/10.1090/memo/1259

MSC2020 subject classifications. Primary 62A99; secondary 62J05, 62G08, 62G30, 62C20.

Key words and phrases. Algorithmic fairness, risk-fairness trade-off, regressions, demographic parity, least-squares, optimal transport, minimax analysis, statistical learning, Wasserstein barycenter, optimal transport, Pareto optimality.

CALDERS, T., KAMIRAN, F. and PECHENIZKIY, M. (2009). Building classifiers with independency constraints. In *IEEE International Conference on Data Mining*.

CALDERS, T., KARIM, A., KAMIRAN, F., ALI, W. and ZHANG, X. (2013). Controlling attribute effect in linear regression. In *IEEE International Conference on Data Mining*.

CATONI, O. (2004). *Statistical Learning Theory and Stochastic Optimization*. *Lecture Notes in Math*. **1851**. Springer, Berlin. MR2163920 https://doi.org/10.1007/b99352

CHEN, M., GAO, C. and REN, Z. (2016). A general decision theory for Huber's ϵ-contamination model. *Electron. J. Stat*. **10** 3752–3774. MR3579675 https://doi.org/10.1214/16-EJS1216

CHEN, M., GAO, C. and REN, Z. (2018). Robust covariance and scatter matrix estimation under Huber's contamination model. *Ann. Statist*. **46** 1932–1960. MR3845006 https://doi.org/10.1214/17-AOS1607

CHIAPPA, S., JIANG, R., STEPLETON, T., PACCHIANO, A., JIANG, H. and ASLANIDES, J. (2020). A general approach to fairness with optimal transport. In *AAAI*.

CHZHEN, E. and SCHREUDER, N. (2022). Supplement to "A minimax framework for quantifying risk-fairness trade-off in regression." https://doi.org/10.1214/22-AOS2198SUPP

CHZHEN, E., DENIS, C., HEBIRI, M., ONETO, L. and PONTIL, M. (2020a). Fair regression with Wasserstein barycenters. In *Advances in Neural Information Processing Systems*.

CHZHEN, E., DENIS, C., HEBIRI, M., ONETO, L. and PONTIL, M. (2020b). Fair regression via plug-in estimator and recalibration with statistical guarantees. In *Advances in Neural Information Processing Systems*.

DEL BARRIO, E., GORDALIZA, P. and LOUBES, J. (2020). Review of mathematical frameworks for fairness in machine learning. Preprint. Available at arXiv:2005.13755.

DONINI, M., ONETO, L., BEN-DAVID, S., SHAWE-TAYLOR, J. S. and PONTIL, M. (2018). Empirical risk minimization under fairness constraints. In *Neural Information Processing Systems*.

DWORK, C., HARDT, M., PITASSI, T., REINGOLD, O. and ZEMEL, R. (2012). Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference* 214–226. ACM, New York. MR3388391

FITZSIMONS, J., ALI, A. A., OSBORNE, M. and ROBERTS, S. (2018). Equality constrained decision trees: For the algorithmic enforcement of group fairness. Preprint. Available at arXiv:1810.05041.

FITZSIMONS, J., AL ALI, A., OSBORNE, M. and ROBERTS, S. (2019). A general framework for fair regression. *Entropy* **21** Paper No. 741, 15 pp. MR4003044 https://doi.org/10.3390/e21080741

GANGBO, W. and ŚWIĘCH, A. (1998). Optimal maps for the multidimensional Monge–Kantorovich problem. *Comm. Pure Appl. Math*. **51** 23–45. MR1486630 https://doi.org/10.1002/(SICI)1097-0312(199801)51:1<23::AID-CPA2>3.0.CO;2-H

GORDALIZA, P., DEL BARRIO, E., FABRICE, G. and LOUBES, J. M. (2019). Obtaining fairness using optimal transport theory. In *International Conference on Machine Learning*.

GYÖRFI, L., KOHLER, M., KRZYŻAK, A. and WALK, H. (2006). *A Distribution-Free Theory of Nonparametric Regression*. *Springer Series in Statistics*. Springer, New York. MR1920390 https://doi.org/10.1007/b97848

HAAS, C. (2019). The price of fairness—A framework to explore trade-offs in algorithmic fairness. Preprint.

HARDT, M., PRICE, E. and SREBRO, N. (2016). Equality of opportunity in supervised learning. In *Neural Information Processing Systems*.

HSU, D., KAKADE, S. and ZHANG, T. (2012). Random design analysis of ridge regression. In *Conference on Learning Theory* 9.1–9.24.

JIANG, R., PACCHIANO, A., STEPLETON, T., JIANG, H. and CHIAPPA, S. (2020). Wasserstein fair classification. In *Uncertainty in Artificial Intelligence Conference*.

KEARNS, M., NEEL, S., ROTH, A. and WU, Z. S. (2018). Preventing fairness gerrymandering: Auditing and learning for subgroup fairness. In *Proceedings of the 35th International Conference on Machine Learning* (J. Dy and A. Krause, eds.). *Proceedings of Machine Learning Research* **80** 2564–2572. PMLR.

KERKYACHARIAN, G., TSYBAKOV, A. B., TEMLYAKOV, V., PICARD, D. and KOLTCHINSKII, V. (2014). Optimal exponential bounds on the accuracy of classification. *Constr. Approx*. **39** 421–444. MR3207667 https://doi.org/10.1007/s00365-014-9229-3

KÖEPPEN, M., YOSHIDA, K. and OHNISHI, K. (2014). Evolving fair linear regression for the representation of human-drawn regression lines. In 2014 *International Conference on Intelligent Networking and Collaborative Systems* 296–303.

KOMIYAMA, J. and SHIMAO, H. (2017). Two-stage algorithm for fairness-aware machine learning. Preprint. Available at arXiv:1710.04924.

KOMIYAMA, J., TAKEDA, A., HONDA, J. and SHIMAO, H. (2018). Nonconvex optimization for regression with fairness constraints. In *International Conference on Machine Learning*.

LE GOUIC, T. and LOUBES, J.-M. (2017). Existence and consistency of Wasserstein barycenters. *Probab. Theory Related Fields* **168** 901–917. MR3663634 https://doi.org/10.1007/s00440-016-0727-z

LE GOUIC, T., LOUBES, J. and RIGOLLET, P. (2020). Projection to fairness in statistical learning. Preprint. Available at arXiv:2005.11720.

LEI, J. and WASSERMAN, L. (2014). Distribution-free prediction bands for non-parametric regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 71–96. MR3153934 https://doi.org/10.1111/rssb.12021

LEI, J., G'SELL, M., RINALDO, A., TIBSHIRANI, R. J. and WASSERMAN, L. (2018). Distribution-free predictive inference for regression. *J. Amer. Statist. Assoc.* **113** 1094–1111. MR3862342 https://doi.org/10.1080/01621459.2017.1307116

LIPTON, Z., CHOULDECHOVA, A. and MCAULEY, J. (2018). Does mitigating ML's impact disparity require treatment disparity? In *Advances in Neural Information Processing Systems* 8136–8146.

MADRAS, D., CREAGER, E., PITASSI, T. and ZEMEL, R. (2018). Learning adversarially fair and transferable representations. In *International Conference on Machine Learning* 3384–3393.

MARY, J., CALAUZÈNES, C. and EL KAROUI, N. (2019). Fairness-aware learning for continuous attributes and treatments. In *International Conference on Machine Learning* 4382–4391.

MEHRABI, N., MORSTATTER, F., SAXENA, N., LERMAN, K. and GALSTYAN, A. (2019). A survey on bias and fairness in machine learning. Preprint. Available at arXiv:1908.09635.

MOURTADA, J. (2019). Exact minimax risk for linear least squares, and the lower tail of sample covariance matrices. Preprint. Available at arXiv:1912.10754.

NEMIROVSKI, A. (2000). Topics in non-parametric statistics. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1998). *Lecture Notes in Math.* **1738** 85–277. Springer, Berlin. MR1775640

OLFAT, M., SLOAN, S., HESPANHOL, P., PORTER, W., VASUDEVAN, R. and ASWANI, A. (2020). Covariance-robust dynamic watermarking. Preprint. Available at arXiv:2003.13908.

ONETO, L. and CHIAPPA, S. (2020). Fairness in machine learning. In *Recent Trends in Learning from Data* 155–196. Springer, Cham.

ONETO, L., DONINI, M. and PONTIL, M. (2020). General fair empirical risk minimization. In 2020 *International Joint Conference on Neural Networks* (*IJCNN*) 1–8. IEEE.

ONETO, L., DONINI, M., PONTIL, M. and MAURER, A. (2020). Learning fair and transferable representations with theoretical guarantees. In 2020 *IEEE 7th International Conference on Data Science and Advanced Analytics* (*DSAA*) 30–39. IEEE.

OSBORNE, M. J. and RUBINSTEIN, A. (1994). *A Course in Game Theory*. MIT Press, Cambridge, MA. MR1301776

PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A. et al. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12** 2825–2830. MR2854348

PÉREZ-SUAY, A., LAPARRA, V., MATEO-GARCÍA, G., MUÑOZ-MARÍ, J., GÓMEZ-CHOVA, L. and CAMPS-VALLS, G. (2017). Fair kernel learning. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*.

PLEČKO, D. and MEINSHAUSEN, N. (2020). Fair data adaptation with quantile preservation. *J. Mach. Learn. Res.* **21** Paper No. 242, 44 pp. MR4209528

QUADRIANTO, N. and SHARMANSKA, V. (2017). Recycling privileged learning and distribution matching for fairness. In *Advances in Neural Information Processing Systems* 677–688.

RAFF, E., SYLVESTER, J. and MILLS, S. (2018). Fair forests: Regularized tree induction to minimize model bias. In *AAAI/ACM Conference on AI, Ethics, and Society*.

SABATO, S. and YOM-TOV, E. (2020). Bounding the fairness and accuracy of classifiers from population statistics. In *International Conference on Machine Learning* 8316–8325. PMLR.

STEINBERG, D., REID, A. and O'CALLAGHAN, S. (2020). Fairness measures for regression via probabilistic classification. Preprint. Available at arXiv:2001.06089.

STEINBERG, D., REID, A., O'CALLAGHAN, S., LATTIMORE, F., MCCALMAN, L. and CAETANO, T. (2020). Fast fair regression via efficient approximations of mutual information. Preprint. Available at arXiv:2002.06200.

STONE, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–620. MR0443204

TSYBAKOV, A. (2003). Optimal rates of aggregation. In *Learning Theory and Kernel Machines* 303–313. Springer, Berlin.

TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation. Springer Series in Statistics*. Springer, New York. MR2724359 https://doi.org/10.1007/b13794

VAN DER VAART, A. W. (2000). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

VOVK, V., GAMMERMAN, A. and SHAFER, G. (2005). *Algorithmic Learning in a Random World*. Springer, New York. MR2161220

WICK, M., PANDA, S. and TRISTAN, J. (2019). Unlocking fairness: A trade-off revisited. In *Advances in Neural Information Processing Systems* **32** 8783–8792. Curran Associates, Red Hook.

ZAFAR, M. B., VALERA, I., GOMEZ RODRIGUEZ, M. and GUMMADI, K. P. (2017). Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. In *International Conference on World Wide Web*.

ZINK, A. and ROSE, S. (2020). Fair regression for health care spending. *Biometrics* **76** 973–982.
ZLIOBAITE, I. (2015). On the relation between accuracy and fairness in binary classification. Preprint. Available at arXiv:1505.05723.

The Annals of Statistics
2022, Vol. 50, No. 4, 2443–2466
https://doi.org/10.1214/22-AOS2200

CONSISTENCY OF INVARIANCE-BASED RANDOMIZATION TESTS

BY EDGAR DOBRIBAN[a]

Department of Statistics and Data Science, The Wharton School, University of Pennsylvania, [a]*dobriban@wharton.upenn.edu*

Invariance-based randomization tests—such as permutation tests, rotation tests, or sign changes—are an important and widely used class of statistical methods. They allow drawing inferences under weak assumptions on the data distribution. Most work focuses on their type I error control properties, while their consistency properties are much less understood.

We develop a general framework to study the consistency of invariance-based randomization tests, assuming the data is drawn from a signal-plus-noise model. We allow the transforms (e.g., permutations or rotations) to be general compact topological groups, such as rotation groups, acting by linear group representations. We study test statistics with a generalized subadditivity property.

We apply our framework to a number of fundamental and highly important problems in statistics, including sparse vector detection, testing for low-rank matrices in noise, sparse detection in linear regression, and two-sample testing. Comparing with minimax lower bounds we develop, we find perhaps surprisingly that in some cases, randomization tests detect signals at the minimax optimal rate.

REFERENCES

ALBERS, W., BICKEL, P. J. and VAN ZWET, W. R. (1976). Asymptotic expansions for the power of distribution free tests in the one-sample problem. *Ann. Statist.* **4** 108–156. MR0391373

ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. *Wiley Publications in Statistics*. Wiley, New York. MR0091588

ANDERSON, M. J. and ROBINSON, J. (2001). Permutation tests for linear models. *Aust. N. Z. J. Stat.* **43** 75–88. MR1837497 https://doi.org/10.1111/1467-842X.00156

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

BARRY, W. T., NOBEL, A. B. and WRIGHT, F. A. (2005). Significance analysis of functional categories in gene expression studies: A structured permutation approach. *Bioinformatics* **21** 1943–1949.

BEDNORZ, W. and LATAŁA, R. (2013). On the suprema of Bernoulli processes. *C. R. Math. Acad. Sci. Paris* **351** 131–134. MR3038002 https://doi.org/10.1016/j.crma.2013.02.013

BERRY, K. J., JOHNSTON, J. E. and MIELKE, P. W. JR. (2014). *A Chronicle of Permutation Statistical Methods*. Springer, Cham. MR3289450 https://doi.org/10.1007/978-3-319-02744-9

BICKEL, P. J. and VAN ZWET, W. R. (1978). Asymptotic expansions for the power of distribution free tests in the two-sample problem. *Ann. Statist.* **6** 937–1004. MR0499567

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

BIRGÉ, L. (2001). An alternative point of view on Lepski's method. In *State of the Art in Probability and Statistics* (*Leiden*, 1999). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **36** 113–133. IMS, Beachwood, OH. MR1836557 https://doi.org/10.1214/lnms/1215090065

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

MSC2020 subject classifications. Primary 62G10; secondary 62G09, 62H15.

Key words and phrases. Randomization test, permutation test, nonparametric group invariance, sparse detection.

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
CHAMAKH, L., GOBET, E. and LIU, W. (2021). Orlicz norms and concentration inequalities for β-heavy tailed random variables.
CHAMAKH, L., GOBET, E. and SZABÓ, Z. (2020). Orlicz random Fourier features. *J. Mach. Learn. Res.* **21** 145. MR4138129
CHURCHILL, G. A. and DOERGE, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* **138** 963–971.
DAVID, H. A. (2008). The beginnings of randomization tests. *Amer. Statist.* **62** 70–72. MR2416900 https://doi.org/10.1198/000313008X269576
DIACONIS, P. (1988). *Group Representations in Probability and Statistics*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **11**. IMS, Hayward, CA. MR0964069
DOBRIBAN, E. (2020). Permutation methods for factor analysis and PCA. *Ann. Statist.* **48** 2824–2847. MR4152122 https://doi.org/10.1214/19-AOS1907
DOBRIBAN, E. (2022a). Code supplement for "Consistency of invariance-based randomization tests." https://doi.org/10.1214/22-AOS2200SUPPB.
DOBRIBAN, E. (2022b). Supplementary technical material for "Consistency of invariance-based randomization tests." https://doi.org/10.1214/22-AOS2200SUPPA.
DWASS, M. (1957). Modified randomization tests for nonparametric hypotheses. *Ann. Math. Stat.* **28** 181–187. MR0087280 https://doi.org/10.1214/aoms/1177707045
EATON, M. L. (1989). *Group Invariance Applications in Statistics*. *NSF-CBMS Regional Conference Series in Probability and Statistics*. IMS, Hayward, CA. MR1089423
EDEN, T. and YATES, F. (1933). On the validity of Fisher's z test when applied to an actual example of non-normal data. *J. Agric. Sci.* **23** 6–17.
EFRON, B. and TIBSHIRANI, R. (2007). On testing the significance of sets of genes. *Ann. Appl. Stat.* **1** 107–129. MR2393843 https://doi.org/10.1214/07-AOAS101
EKLUND, A., NICHOLS, T. E. and KNUTSSON, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proc. Natl. Acad. Sci. USA* **113** 7900–7905.
EPSTEIN, M. P., DUNCAN, R., JIANG, Y., CONNEELY, K. N., ALLEN, A. S. and SATTEN, G. A. (2012). A permutation procedure to correct for confounders in case-control studies, including tests of rare variation. *Am. J. Hum. Genet.* **91** 215–223.
ERNST, M. D. (2004). Permutation methods: A basis for exact inference. *Statist. Sci.* **19** 676–685. MR2185589 https://doi.org/10.1214/088342304000000396
FANG, K.-T., KOTZ, S. and NG, K. W. (2018). *Symmetric Multivariate and Related Distributions*. CRC Press/CRC, Boca Raton.
FANG, K. T. and ZHANG, Y. T. (1990). *Generalized Multivariate Analysis*. Springer, Berlin. MR1079542
FISHER, R. A. (1935). *The Design of Experiments*. Oliver & Boyd, Edinburgh.
FOLLAND, G. B. (2016). *A Course in Abstract Harmonic Analysis*, 2nd ed. *Textbooks in Mathematics*. CRC Press, Boca Raton, FL. MR3444405
FULTON, W. and HARRIS, J. (2013). *Representation Theory*: *A First Course*. Springer, Berlin.
GIRI, N. C. (1996). *Group Invariance in Statistical Inference*. World Scientific, River Edge, NJ. MR1626154 https://doi.org/10.1142/9789812831705
GOOD, P. I. (2006). *Permutation*, *Parametric*, *and Bootstrap Tests of Hypotheses*. Springer, Berlin.
GUPTA, A. K. and VARGA, T. (2012). *Elliptically Contoured Models in Statistics*. Springer, Berlin.
HALL, B. (2015). *Lie Groups*, *Lie Algebras*, *and Representations*: *An Elementary Introduction*, 2nd ed. *Graduate Texts in Mathematics* **222**. Springer, Cham. MR3331229 https://doi.org/10.1007/978-3-319-13467-3
HEMERIK, J. and GOEMAN, J. (2018a). Exact testing with random permutations. *TEST* **27** 811–825. MR3878362 https://doi.org/10.1007/s11749-017-0571-1
HEMERIK, J. and GOEMAN, J. J. (2018b). False discovery proportion estimation by permutations: Confidence for significance analysis of microarrays. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 137–155. MR3744715 https://doi.org/10.1111/rssb.12238
HEMERIK, J. and GOEMAN, J. J. (2020). Another look at the lady tasting tea and differences between permutation tests and randomisation tests. *Int. Stat. Rev.*. **89**. 367-381.
HEMERIK, J., GOEMAN, J. J. and FINOS, L. (2020). Robust testing in generalized linear models by sign flipping score contributions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 841–864. MR4112787 https://doi.org/10.1111/rssb.12369
HEMERIK, J., SOLARI, A. and GOEMAN, J. J. (2019). Permutation-based simultaneous confidence bounds for the false discovery proportion. *Biometrika* **106** 635–649. MR3992394 https://doi.org/10.1093/biomet/asz021
HOEFFDING, W. (1952). The large-sample power of tests based on permutations of observations. *Ann. Math. Stat.* **23** 169–192. MR0057521 https://doi.org/10.1214/aoms/1177729436

HONG, D., SHENG, Y. and DOBRIBAN, E. (2020). Selecting the number of components in PCA via random signflips. ArXiv preprint. Available at arXiv:2012.02985.
INGSTER, Y. I. (1987). Minimax testing of nonparametric hypotheses on a distribution density in the L_p metrics. *Theory Probab. Appl.* **31** 333–337.
INGSTER, Y. and SUSLINA, I. A. (2012). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models* Springer, Berlin.
JAMES, G. and LIEBECK, M. (2001). *Representations and Characters of Groups*, 2nd ed. Cambridge Univ. Press, New York. MR1864147 https://doi.org/10.1017/CBO9780511814532
JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961 https://doi.org/10.1214/aos/1009210544
JOHNSTONE, I. M. and ONATSKI, A. (2015). Testing in high-dimensional spiked models. *Annals of Statistics*. **48**. 1231–1254.
JOHNSTONE, I. M. and PAUL, D. (2018). PCA in high dimensions: An orientation. *Proc. IEEE* **106** 1277–1292. https://doi.org/10.1109/JPROC.2018.2846730
KALLENBERG, O. (2006). *Probabilistic Symmetries and Invariance Principles*. Springer, Berlin.
KENNEDY, P. E. (1995). Randomization tests in econometrics. *J. Bus. Econom. Statist.* **13** 85–94. MR1323048 https://doi.org/10.2307/1392523
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
KIM, I., BALAKRISHNAN, S. and WASSERMAN, L. (2020b). Minimax optimality of permutation tests. ArXiv preprint. Available at arXiv:2003.13208.
KNAPP, A. W. (2013). *Lie Groups Beyond an Introduction* Springer, Berlin.
KONING, N. W. and HEMERIK, J. (2022). Faster exact permutation testing: Using a representative subgroup. ArXiv preprint. Available at arXiv:2202.00967.
LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. *Springer Texts in Statistics*. Springer, New York. MR2135927
LEHMANN, E. L. and STEIN, C. (1949). On the theory of some nonparametric hypotheses. *Ann. Math. Stat.* **20** 28–45. MR0030168
LI, C.-K. and TSING, N.-K. (1988). Some isometries of rectangular complex matrices. *Linear Multilinear Algebra* **23** 47–53. MR0943768 https://doi.org/10.1080/03081088808817855
LUGOSI, G. and MENDELSON, S. (2019). Mean estimation and regression under heavy-tailed distributions: A survey. *Found. Comput. Math.* **19** 1145–1190. MR4017683 https://doi.org/10.1007/s10208-019-09427-x
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
MARCHINA, A. (2019). About the rate function in concentration inequalities for suprema of bounded empirical processes. *Stochastic Process. Appl.* **129** 3967–3980. MR3997668 https://doi.org/10.1016/j.spa.2018.11.010
MUIRHEAD, R. J. (2009). *Aspects of Multivariate Statistical Theory* Wiley, New York.
NARICI, L. and BECKENSTEIN, E. (2010). *Topological Vector Spaces*. CRC Press, Boca Raton.
ONGHENA, P. (2018). *Randomization, Masking, and Allocation Concealment*. Chapman and Hall/CRC.
PESARIN, F. (2001a). *Multivariate Permutation Tests: With Applications in Biostatistics*. Wiley, Chichester. MR1855501
PESARIN, F. (2015). Some elementary theory of permutation tests. *Comm. Statist. Theory Methods* **44** 4880–4892. MR3424814 https://doi.org/10.1080/03610926.2013.802350
PESARIN, F. and SALMASO, L. (2010a). *Permutation Tests for Complex Data: Theory, Applications and Software*. Wiley, New York.
PESARIN, F. and SALMASO, L. (2010). Finite-sample consistency of combination-based permutation tests with application to repeated measures designs. *J. Nonparametr. Stat.* **22** 669–684. MR2682214 https://doi.org/10.1080/10485250902807407
PESARIN, F. and SALMASO, L. (2012). A review and some new results on permutation testing for multivariate problems. *Stat. Comput.* **22** 639–646. MR2865041 https://doi.org/10.1007/s11222-011-9261-0
PESARIN, F. and SALMASO, L. (2013). On the weak consistency of permutation tests. *Comm. Statist. Simulation Comput.* **42** 1368–1379.
PETROV, V. (2012). *Sums of Independent Random Variables* Springer, Berlin.
PITMAN, E. J. (1937). Significance tests which may be applied to samples from any populations. *Suppl. J. R. Stat. Soc.* **4** 119–130.
PITMAN, E. J. G. (1939). Tests of hypotheses concerning location and scale parameters. *Biometrika* **31** 200–215. MR0000382 https://doi.org/10.1093/biomet/31.1-2.200
POLYANSKIY, Y. (2019). Information Theoretic Methods in Statistics and Computer Science.

PURCELL, S., NEALE, B., TODD-BROWN, K., THOMAS, L., FERREIRA, M. A., BENDER, D., MALLER, J., SKLAR, P., DE BAKKER, P. I. et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81** 559–575.

RIO, E. (2017). About the constants in the Fuk-Nagaev inequalities. *Electron. Commun. Probab.* **22** 28. MR3652041 https://doi.org/10.1214/17-ECP57

SERRE, J.-P. (1977). *Linear Representations of Finite Groups. Graduate Texts in Mathematics* Springer, New York. MR0450380

SIMPSON, S. L., LYDAY, R. G., HAYASAKA, S., MARSH, A. P. and LAURIENTI, P. J. (2013). A permutation testing framework to compare groups of brain networks. *Front. Comput. Neurosci.* **7** 171. https://doi.org/10.3389/fncom.2013.00171

SOUTHWORTH, L. K., KIM, S. K. and OWEN, A. B. (2009). Properties of balanced permutations. *J. Comput. Biol.* **16** 625–638. MR2511824 https://doi.org/10.1089/cmb.2008.0144

SUBRAMANIAN, A., TAMAYO, P., MOOTHA, V. K., MUKHERJEE, S., EBERT, B. L., GILLETTE, M. A., PAULOVICH, A., POMEROY, S. L., GOLUB, T. R. et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102** 15545–15550.

TALAGRAND, M. (2014). *Upper and Lower Bounds for Stochastic Processes*: *Modern Methods and Classical Problems. Ergebnisse der Mathematik und Ihrer Grenzgebiete*. 3. *Folge. A Series of Modern Surveys in Mathematics* Springer, Heidelberg. MR3184689 https://doi.org/10.1007/978-3-642-54075-2

TUSHER, V. G., TIBSHIRANI, R. and CHU, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA* **98** 5116–5121.

VAN DE GEER, S. and LEDERER, J. (2013). The Bernstein–Orlicz norm and deviation inequalities. *Probab. Theory Related Fields* **157** 225–250. MR3101846 https://doi.org/10.1007/s00440-012-0455-y

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

WIJSMAN, R. A. (1990). *Invariant Measures on Groups and Their Use in Statistics. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **14**. IMS, Hayward, CA. MR1218397

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.

The Annals of Statistics
2022, Vol. 50, No. 4, 2467–2468
https://doi.org/10.1214/21-AOS2135
Main article: https://doi.org/10.1214/19-AOS1823
This research was funded, in whole or in part, by [EPSRC, EP/R034710/1].

ERRATUM: ASYMPTOTIC GENEALOGIES OF INTERACTING PARTICLE SYSTEMS WITH AN APPLICATION TO SEQUENTIAL MONTE CARLO

BY JERE KOSKELA[a], PAUL A. JENKINS[b], ADAM M. JOHANSEN[c] AND DARIO SPANÒ[d]

Department of Statistics, University of Warwick, [a]*j.koskela@warwick.ac.uk,* [b]*p.jenkins@warwick.ac.uk,*
[c]*a.m.johansen@warwick.ac.uk,* [d]*d.spano@warwick.ac.uk*

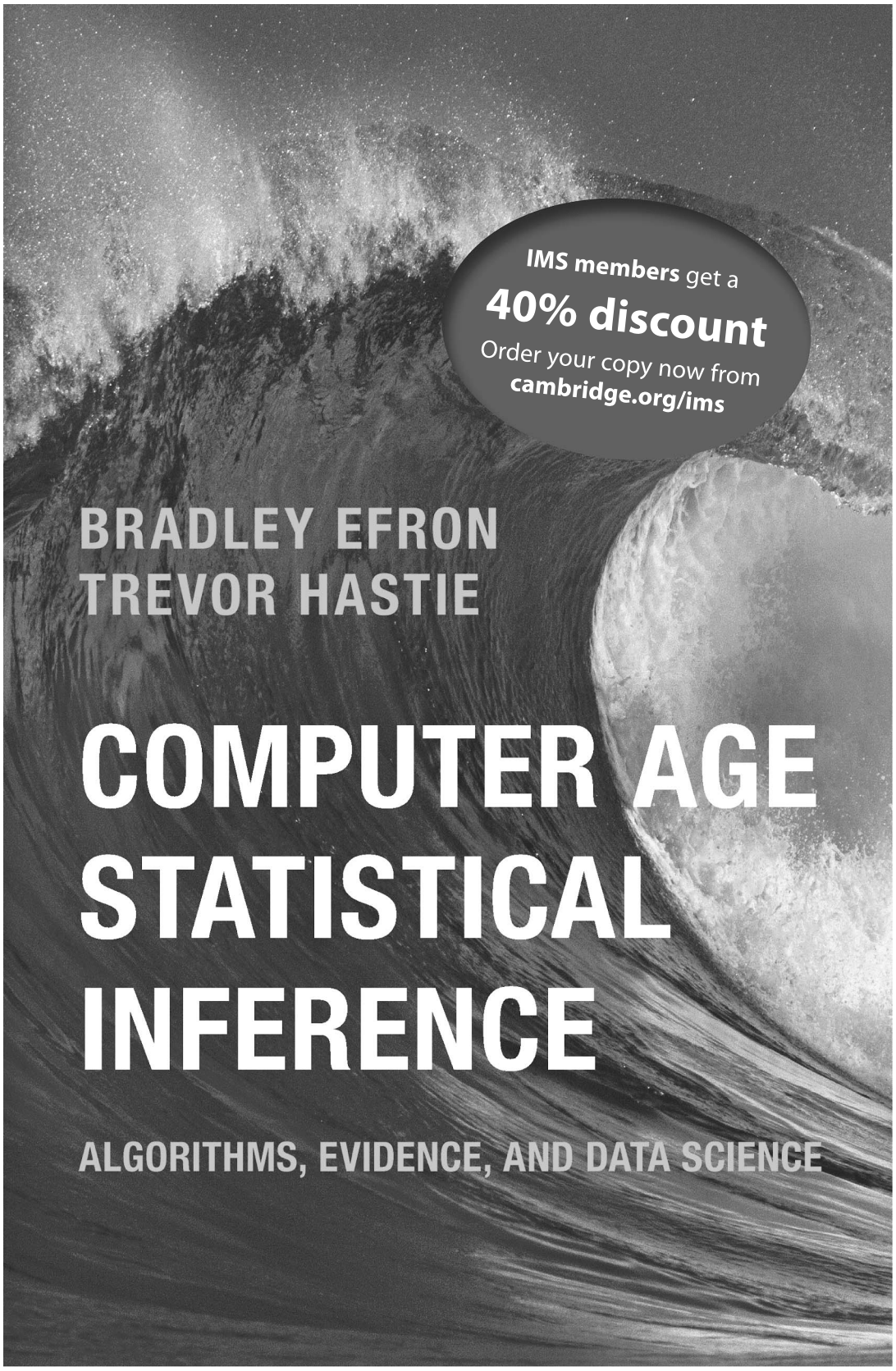
REFERENCES

KOSKELA, J., JENKINS, P. A., JOHANSEN, A. M. and SPANÒ, D. (2020). Asymptotic genealogies of interacting particle systems with an application to sequential Monte Carlo. *Ann. Statist.* **48** 560–583. MR4065174 https://doi.org/10.1214/19-AOS1823

KOSKELA, J., JENKINS, P. A., JOHANSEN, A. M. and SPANÒ, D. (2022). Supplement to "Erratum: Asymptotic genealogies of interacting particle systems with an application to sequential Monte Carlo." https://doi.org/10.1214/21-AOS2135SUPP

MSC2020 subject classifications. Primary 60E15; secondary 60G99, 62E20.

Key words and phrases. Coalescent, finite-dimensional distributions, genealogy, interacting particle system, sequential Monte Carlo.


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