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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

continued

Vol. 49, No. 1—February 2021

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*

Correction notes

Vol. 49, No. 1—February 2021

THE ANNALS OF STATISTICS Vol. 49, No. 1, pp. 1–614 February 2021

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

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

The Annals of Statistics **[ISSN 0090-5364 (print); ISSN 2168-8966 (online)],** Volume 49, Number 1, February 2021. Published bimonthly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, Ohio 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, 9650 Rockville Pike, Suite L 2310, Bethesda, Maryland 20814-3998, USA.

The Annals of Statistics
2021, Vol. 49, No. 1, 1–20
https://doi.org/10.1214/19-AOS1813

ON THE OPTIMALITY OF SLICED INVERSE REGRESSION IN HIGH DIMENSIONS

BY QIAN LIN[1], XINRAN LI[2], DONGMING HUANG[3] AND JUN S. LIU[4]

[1]*Center for Statistical Science, Department of Industrial Engineering, Tsinghua University, qianlin@tsinghua.edu.cn*
[2]*Department of Statistics, University of Illinois at Urbana-Champaign, xinranli@illinois.edu*
[3]*Department of Statistics and Applied Probability, National University of Singapore, stahd@nus.edu.sg*
[4]*Department of Statistics, Harvard University, jliu@stat.harvard.edu*

The central subspace of a pair of random variables $(y, \boldsymbol{x}) \in \mathbb{R}^{p+1}$ is the minimal subspace $\mathcal{S}$ such that $y \perp\!\!\!\perp \boldsymbol{x} | P_{\mathcal{S}}\boldsymbol{x}$. In this paper, we consider the minimax rate of estimating the central space under the multiple index model $y = f(\boldsymbol{\beta}_1^\tau \boldsymbol{x}, \boldsymbol{\beta}_2^\tau \boldsymbol{x}, \dots, \boldsymbol{\beta}_d^\tau \boldsymbol{x}, \epsilon)$ with at most s active predictors, where $\boldsymbol{x} \sim N(0, \boldsymbol{\Sigma})$ for some class of $\boldsymbol{\Sigma}$. We first introduce a large class of models depending on the smallest nonzero eigenvalue λ of $\mathrm{var}(\mathbb{E}[\boldsymbol{x}|y])$, over which we show that an aggregated estimator based on the SIR procedure converges at rate $d \wedge ((sd + s\log(ep/s))/(n\lambda))$. We then show that this rate is optimal in two scenarios, the single index models and the multiple index models with fixed central dimension d and fixed λ. By assuming a technical conjecture, we can show that this rate is also optimal for multiple index models with bounded dimension of the central space.

REFERENCES

ABRAMOVICH, F., BENJAMINI, Y., DONOHO, D. L. and JOHNSTONE, I. M. (2006). Adapting to unknown sparsity by controlling the false discovery rate. *Ann. Statist.* **34** 584–653. MR2281879 https://doi.org/10.1214/009053606000000074

AMINI, A. A. and WAINWRIGHT, M. J. (2008). High-dimensional analysis of semidefinite relaxations for sparse principal components. In *Information Theory*, 2008. *ISIT* 2008. *IEEE International Symposium on* 2454–2458. IEEE.

BERTHET, Q. and RIGOLLET, P. (2013). Computational lower bounds for sparse pca. Preprint. Available at arXiv:1304.0828.

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

BIRNBAUM, A., JOHNSTONE, I. M., NADLER, B. and PAUL, D. (2013). Minimax bounds for sparse PCA with noisy high-dimensional data. *Ann. Statist.* **41** 1055–1084. MR3113803 https://doi.org/10.1214/12-AOS1014

CAI, T. T., MA, Z. and WU, Y. (2013). Sparse PCA: Optimal rates and adaptive estimation. *Ann. Statist.* **41** 3074–3110. MR3161458 https://doi.org/10.1214/13-AOS1178

CAI, T. T. and ZHOU, H. H. (2012). Optimal rates of convergence for sparse covariance matrix estimation. *Ann. Statist.* **40** 2389–2420. MR3097607 https://doi.org/10.1214/12-AOS998

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

CHEN, C.-H. and LI, K.-C. (1998). Can SIR be as popular as multiple linear regression? *Statist. Sinica* **8** 289–316. MR1624402

COOK, R. D., FORZANI, L. and ROTHMAN, A. J. (2012). Estimating sufficient reductions of the predictors in abundant high-dimensional regressions. *Ann. Statist.* **40** 353–384. MR3014310 https://doi.org/10.1214/11-AOS962

DENNIS COOK, R. (1998). *Regression Graphics*: *Ideas for Studying Regressions Through Graphics*. *Wiley Series in Probability and Statistics*: *Probability and Statistics*. Wiley, New York. MR1645673 https://doi.org/10.1002/9780470316931

MSC2020 subject classifications. Primary 62J02; secondary 62H25.
Key words and phrases. Sufficient dimension reduction, optimal rates, sliced inverse regression, semidefinite positive programming.

DENNIS COOK, R. (2000). Save: A method for dimension reduction and graphics in regression. *Comm. Statist. Theory Methods* **29** 2109–2121.

DENNIS COOK, R. and WEISBERG, S. (1991). Comment. *J. Amer. Statist. Assoc.* **86** 328–332.

DUAN, N. and LI, K.-C. (1991). Slicing regression: A link-free regression method. *Ann. Statist.* **19** 505–530. MR1105834 https://doi.org/10.1214/aos/1176348109

FERRÉ, L. (1998). Determining the dimension in sliced inverse regression and related methods. *J. Amer. Statist. Assoc.* **93** 132–140. MR1614604 https://doi.org/10.2307/2669610

HSING, T. and CARROLL, R. J. (1992). An asymptotic theory for sliced inverse regression. *Ann. Statist.* **20** 1040–1061. MR1165605 https://doi.org/10.1214/aos/1176348669

JOHNSTONE, I. M. and LU, A. Y. (2004). *Sparse Principal Components Analysis*.

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

LI, K.-C. (1991). Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.* **86** 316–342. MR1137117

LI, K.-C. (2000). High dimensional data analysis via the SIR/PHD approach.

LI, L. (2007). Sparse sufficient dimension reduction. *Biometrika* **94** 603–613. MR2410011 https://doi.org/10.1093/biomet/asm044

LI, L. and NACHTSHEIM, C. J. (2006). Sparse sliced inverse regression. *Technometrics* **48** 503–510. MR2328619 https://doi.org/10.1198/004017006000000129

LI, B. and WANG, S. (2007). On directional regression for dimension reduction. *J. Amer. Statist. Assoc.* **102** 997–1008. MR2354409 https://doi.org/10.1198/016214507000000536

LIN, Q., ZHAO, Z. and LIU, J. S. (2018). On consistency and sparsity for sliced inverse regression in high dimensions. *Ann. Statist.* **46** 580–610. MR3782378 https://doi.org/10.1214/17-AOS1561

LIN, Q., ZHAO, Z. and LIU, J. S. (2019). Sparse sliced inverse regression via Lasso. *J. Amer. Statist. Assoc.* **114** 1726–1739. MR4047295 https://doi.org/10.1080/01621459.2018.1520115

LIN, Q., LI, X., HUANG, D. and LIU, J. S. (2021). Supplement to "On the optimality of sliced inverse regression in high dimensions." https://doi.org/10.1214/19-AOS1813SUPP

NEYKOV, M., LIN, Q. and LIU, J. S. (2016). Signed support recovery for single index models in high-dimensions. *Ann. Math. Sci. Appl.* **1** 379–426. MR3876481 https://doi.org/10.4310/amsa.2016.v1.n2.a5

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

SCHOTT, J. R. (1994). Determining the dimensionality in sliced inverse regression. *J. Amer. Statist. Assoc.* **89** 141–148. MR1266291

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

VU, V. Q. and LEI, J. (2012). Minimax rates of estimation for sparse pca in high dimensions. Preprint. Available at arXiv:1202.0786.

YANG, S. S. (1977). General distribution theory of the concomitants of order statistics. *Ann. Statist.* **5** 996–1002. MR0501519

ZHU, L., MIAO, B. and PENG, H. (2006). On sliced inverse regression with high-dimensional covariates. *J. Amer. Statist. Assoc.* **101** 630–643. MR2281245 https://doi.org/10.1198/016214505000001285

ZOU, H. and HASTIE, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 301–320. MR2137327 https://doi.org/10.1111/j.1467-9868.2005.00503.x

The Annals of Statistics
2021, Vol. 49, No. 1, 21–48
https://doi.org/10.1214/19-AOS1831

ASYMPTOTIC OPTIMALITY IN STOCHASTIC OPTIMIZATION

BY JOHN C. DUCHI* AND FENG RUAN†

Department of Statistics, Stanford University, *jduchi@stanford.edu; †fengruan@stanford.edu*

We study local complexity measures for stochastic convex optimization problems, providing a local minimax theory analogous to that of Hájek and Le Cam for classical statistical problems. We give complementary optimality results, developing fully online methods that adaptively achieve optimal convergence guarantees. Our results provide function-specific lower bounds and convergence results that make precise a correspondence between statistical difficulty and the geometric notion of tilt-stability from optimization. As part of this development, we show how variants of Nesterov's dual averaging—a stochastic gradient-based procedure—guarantee finite time identification of constraints in optimization problems, while stochastic gradient procedures fail. Additionally, we highlight a gap between problems with linear and nonlinear constraints: standard stochastic-gradient-based procedures are suboptimal even for the simplest nonlinear constraints, necessitating the development of asymptotically optimal Riemannian stochastic gradient methods.

REFERENCES

[1] ABSIL, P.-A., BAKER, C. G. and GALLIVAN, K. A. (2007). Trust-region methods on Riemannian manifolds. *Found. Comput. Math.* **7** 303–330. MR2335248 https://doi.org/10.1007/s10208-005-0179-9

[2] ABSIL, P.-A., MAHONY, R. and SEPULCHRE, R. (2008). *Optimization Algorithms on Matrix Manifolds*. Princeton Univ. Press, Princeton, NJ. MR2364186 https://doi.org/10.1515/9781400830244

[3] AGARWAL, A., BARTLETT, P. L., RAVIKUMAR, P. and WAINWRIGHT, M. J. (2012). Information-theoretic lower bounds on the oracle complexity of stochastic convex optimization. *IEEE Trans. Inf. Theory* **58** 3235–3249. MR2952543 https://doi.org/10.1109/TIT.2011.2182178

[4] ALI, S. M. and SILVEY, S. D. (1966). A general class of coefficients of divergence of one distribution from another. *J. Roy. Statist. Soc. Ser. B* **28** 131–142. MR0196777

[5] BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York. Reprint of the 1993 original. MR1623559

[6] BIRGÉ, L. (1983). Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsch. Verw. Gebiete* **65** 181–237. MR0722129 https://doi.org/10.1007/BF00532480

[7] BONNABEL, S. (2013). Stochastic gradient descent on Riemannian manifolds. *IEEE Trans. Automat. Control* **58** 2217–2229. MR3101606 https://doi.org/10.1109/TAC.2013.2254619

[8] BONNANS, J. F. and SHAPIRO, A. (1998). Optimization problems with perturbations: A guided tour. *SIAM Rev.* **40** 228–264. MR1624098 https://doi.org/10.1137/S0036144596302644

[9] BOTTOU, L. and BOUSQUET, O. (2007). The tradeoffs of large scale learning. In *Advances in Neural Information Processing Systems* 20.

[10] BOUMAL, N. (2014). Optimization and estimation on manifolds. Ph.D. thesis, Univ. Catholique de Louvain.

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

[12] BURKE, J. V. and MORÉ, J. J. (1994). Exposing constraints. *SIAM J. Optim.* **4** 573–595. MR1287817 https://doi.org/10.1137/0804032

[13] CSISZÁR, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.* **2** 299–318. MR0219345

[14] DEFAZIO, A., BACH, F. and SAGA, S. L.-J. (2014). A fast incremental gradient method with support for non-strongly convex composite objectives. In *Advances in Neural Information Processing Systems* 27.

[15] DEMBO, A. (2016). Lecture notes on probability theory: Stanford statistics 310. Accessed October 1, 2016. http://statweb.stanford.edu/~adembo/stat-310b/lnotes.pdf.

MSC2020 subject classifications. 62F10, 62F12, 68Q25.

Key words and phrases. Local asymptotic minimax theory, convex analysis, stochastic gradients, manifold identification.

[16] DONOHO, D. L. and LIU, R. C. (1987). Geometrizing rates of convergence I. Technical Report 137, Dept. Statistics, Univ. California, Berkeley.

[17] DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence II. *Ann. Statist.* **19** 633–667. MR1105839 https://doi.org/10.1214/aos/1176348114

[18] DONTCHEV, A. L. and ROCKAFELLAR, R. T. (2014). *Implicit Functions and Solution Mappings: A View from Variational Analysis*, 2nd ed. *Springer Series in Operations Research and Financial Engineering*. Springer, New York. MR3288139 https://doi.org/10.1007/978-1-4939-1037-3

[19] DRUSVYATSKIY, D. and LEWIS, A. S. (2013). Tilt stability, uniform quadratic growth, and strong metric regularity of the subdifferential. *SIAM J. Optim.* **23** 256–267. MR3033107 https://doi.org/10.1137/120876551

[20] DUCHI, J., HAZAN, E. and SINGER, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.* **12** 2121–2159. MR2825422

[21] DUCHI, J. C. and RUAN, F. (2021). Supplement to "Asymptotic optimality in stochastic optimization." https://doi.org/10.1214/19-AOS1831SUPP

[22] ERMOLIEV, Y. (1969). On the stochastic quasi-gradient method and stochastic quasi-Feyer sequences. *Kibernetika* **2** 72–83.

[23] ERMOLIEV, Y. (1983). Stochastic quasigradient methods and their application to system optimization. *Stochastics* **9** 1–36. MR0703846 https://doi.org/10.1080/17442508308833246

[24] FABIAN, V. (1973). Asymptotically efficient stochastic approximation; the RM case. *Ann. Statist.* **1** 486–495. MR0381189

[25] HARE, W. L. and LEWIS, A. S. (2004). Identifying active constraints via partial smoothness and prox-regularity. *J. Convex Anal.* **11** 251–266. MR2158904

[26] HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2722294 https://doi.org/10.1007/978-0-387-84858-7

[27] HIRIART-URRUTY, J.-B. and LEMARÉCHAL, C. (1993). *Convex Analysis and Minimization Algorithms I & II*. Springer, Berlin. MR1261420

[28] IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1981). *Statistical Estimation: Asymptotic Theory*. Springer.

[29] JOHNSON, R. and ZHANG, T. (2013). Accelerating stochastic gradient descent using predictive variance reduction. In *Advances in Neural Information Processing Systems* 26.

[30] KUSHNER, H. J. (1974). Stochastic approximation algorithms for constrained optimization problems. *Ann. Statist.* **2** 713–723. MR0365955

[31] 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

[32] LE ROUX, N., SCHMIDT, M. and BACH, F. (2012). A stochastic gradient method with an exponential convergence rate for finite training sets. In *Advances in Neural Information Processing Systems* 25.

[33] LEE, S. and WRIGHT, S. J. (2012). Manifold identification in dual averaging for regularized stochastic online learning. *J. Mach. Learn. Res.* **13** 1705–1744. MR2956341

[34] LIN, H., MAIRAL, J. and HARCHAOUI, Z. (2015). A universal catalyst for first-order optimization. In *Advances in Neural Information Processing Systems* 3384–3392.

[35] NEMIROVSKI, A., JUDITSKY, A., LAN, G. and SHAPIRO, A. (2008). Robust stochastic approximation approach to stochastic programming. *SIAM J. Optim.* **19** 1574–1609. MR2486041 https://doi.org/10.1137/070704277

[36] NEMIROVSKY, A. S. and YUDIN, D. B. (1983). *Problem Complexity and Method Efficiency in Optimization. Wiley-Interscience Series in Discrete Mathematics*. Wiley, New York. Translated from the Russian and with a preface by E. R. Dawson. MR0702836

[37] NESTEROV, Y. (2009). Primal-dual subgradient methods for convex problems. *Math. Program.* **120** 221–259. MR2496434 https://doi.org/10.1007/s10107-007-0149-x

[38] NOCEDAL, J. and WRIGHT, S. J. (2006). *Numerical Optimization*, 2nd ed. *Springer Series in Operations Research and Financial Engineering*. Springer, New York. MR2244940

[39] POLIQUIN, R. A. and ROCKAFELLAR, R. T. (1998). Tilt stability of a local minimum. *SIAM J. Optim.* **8** 287–299. MR1618790 https://doi.org/10.1137/S1052623496309296

[40] POLYAK, B. T. and JUDITSKY, A. B. (1992). Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.* **30** 838–855. MR1167814 https://doi.org/10.1137/0330046

[41] ROBBINS, H. and MONRO, S. (1951). A stochastic approximation method. *Ann. Math. Stat.* **22** 400–407. MR0042668 https://doi.org/10.1214/aoms/1177729586

[42] ROBBINS, H. and SIEGMUND, D. (1971). A convergence theorem for non negative almost supermartingales and some applications. In *Optimizing Methods in Statistics* (*Proc. Sympos., Ohio State Univ., Columbus, Ohio*, 1971) 233–257. MR0343355

[43] ROCKAFELLAR, R. T. (1970). *Convex Analysis*. *Princeton Mathematical Series* **28**. Princeton Univ. Press, Princeton, NJ. MR0274683

[44] SHALEV-SWHARTZ, S. and SREBRO, N. (2008). SVM optimization: Inverse dependence on training set size. In *Proceedings of the* 25*th International Conference on Machine Learning*.

[45] SHAPIRO, A. (1988). Sensitivity analysis of nonlinear programs and differentiability properties of metric projections. *SIAM J. Control Optim*. **26** 628–645. MR0937676 https://doi.org/10.1137/0326037

[46] SHAPIRO, A. (1989). Asymptotic properties of statistical estimators in stochastic programming. *Ann. Statist*. **17** 841–858. MR0994271 https://doi.org/10.1214/aos/1176347146

[47] SHAPIRO, A., DENTCHEVA, D. and RUSZCZYŃSKI, A. (2009). *Lectures on Stochastic Programming*: *Modeling and Theory*. *MPS/SIAM Series on Optimization* **9**. SIAM, Philadelphia, PA; Mathematical Programming Society (MPS), Philadelphia, PA. MR2562798 https://doi.org/10.1137/1.9780898718751

[48] STEIN, C. (1956). Efficient nonparametric testing and estimation. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, *Vol. I* 187–195. Univ. California Press, Berkeley and Los Angeles. MR0084921

[49] TRIPURANENI, N., STERN, M., JIN, C., REGIER, J. and JORDAN, M. I. (2017). Stochastic cubic regularization for fast nonconvex optimization. arXiv:1711.02838 [cs.LG].

[50] VAN DER VAART, A. W. (1997). Superefficiency. In *Festschrift for Lucien Le Cam* 397–410. Springer, New York. MR1462961

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

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

[53] VENTER, J. H. (1967). An extension of the Robbins–Monro procedure. *Ann. Math. Stat*. **38** 181–190. MR0205396 https://doi.org/10.1214/aoms/1177699069

[54] WALK, H. (1983/84). Stochastic iteration for a constrained optimization problem. *Seq. Anal*. **2** 369–385. MR0752415 https://doi.org/10.1080/07474948408836045

[55] WRIGHT, S. J. (1993). Identifiable surfaces in constrained optimization. *SIAM J. Control Optim*. **31** 1063–1079. MR1227547 https://doi.org/10.1137/0331048

[56] XIAO, L. (2010). Dual averaging methods for regularized stochastic learning and online optimization. *J. Mach. Learn. Res*. **11** 2543–2596. MR2738777

[57] ZHU, Y., CHATTERJEE, S., DUCHI, J. and LAFFERTY, J. (2016). Local minimax complexity of stochastic convex optimization. In *Advances in Neural Information Processing Systems* 29.

[58] ZINKEVICH, M. (2003). Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the Twentieth International Conference on Machine Learning*.

The Annals of Statistics
2021, Vol. 49, No. 1, 49–75
https://doi.org/10.1214/19-AOS1908

CONCORDANCE AND VALUE INFORMATION CRITERIA FOR OPTIMAL TREATMENT DECISION

BY CHENGCHUN SHI[1], RUI SONG[2,*] AND WENBIN LU[2,†]

[1]*Department of Statistics, London School of Economics, C.Shi7@lse.ac.uk*
[2]*Department of Statistics, North Carolina State University, *rsong@ncsu.edu; †wlu4@ncsu.edu*

Personalized medicine is a medical procedure that receives considerable scientific and commercial attention. The goal of personalized medicine is to assign the optimal treatment regime for each individual patient, according to his/her personal prognostic information. When there are a large number of pretreatment variables, it is crucial to identify those important variables that are necessary for treatment decision making. In this paper, we study two information criteria: the concordance and value information criteria, for variable selection in optimal treatment decision making. We consider both fixed-p and high dimensional settings, and show our information criteria are consistent in model/tuning parameter selection. We further apply our information criteria to four estimation approaches, including robust learning, concordance-assisted learning, penalized A-learning and sparse concordance-assisted learning, and demonstrate the empirical performance of our methods by simulations.

REFERENCES

AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (*Tsahkadsor*, 1971) 267–281. Akadémiai Kiadó, Budapest. MR0483125

ARCONES, M. A. (1995). A Bernstein-type inequality for U-statistics and U-processes. *Statist*. *Probab*. *Lett*. **22** 239–247. MR1323145 https://doi.org/10.1016/0167-7152(94)00072-G

CANDÈS, E. and TAO, T. (2007). Rejoinder: "The Dantzig selector: Statistical estimation when p is much larger than n" [Ann. Statist. **35** (2007), no. 6, 2313–2351; MR2382644]. *Ann*. *Statist*. **35** 2392–2404. MR2382651 https://doi.org/10.1214/009053607000000532

CHAKRABORTY, B., MURPHY, S. and STRECHER, V. (2010). Inference for non-regular parameters in optimal dynamic treatment regimes. *Stat*. *Methods Med*. *Res*. **19** 317–343. MR2757118 https://doi.org/10.1177/0962280209105013

CLÉMENÇON, S., LUGOSI, G. and VAYATIS, N. (2008). Ranking and empirical minimization of U-statistics. *Ann*. *Statist*. **36** 844–874. MR2396817 https://doi.org/10.1214/009052607000000910

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

FAN, J. and LV, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statist*. *Sinica* **20** 101–148. MR2640659

FAN, J. and LV, J. (2011). Nonconcave penalized likelihood with NP-dimensionality. *IEEE Trans*. *Inf*. *Theory* **57** 5467–5484. MR2849368 https://doi.org/10.1109/TIT.2011.2158486

FAN, Y. and TANG, C. Y. (2013). Tuning parameter selection in high dimensional penalized likelihood. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **75** 531–552. MR3065478 https://doi.org/10.1111/rssb.12001

FAN, C., LU, W., SONG, R. and ZHOU, Y. (2017). Concordance-assisted learning for estimating optimal individualized treatment regimes. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **79** 1565–1582. MR3731676 https://doi.org/10.1111/rssb.12216

GAI, Y., ZHU, L. and LIN, L. (2013). Model selection consistency of Dantzig selector. *Statist*. *Sinica* **23** 615–634. MR3086649

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

MSC2020 subject classifications. 62E99.
Key words and phrases. Concordance and value information criteria, optimal treatment regime, tuning parameter selection, variable selection.

LI, H., REN, C. and LI, L. (2014). U-processes and preference learning. *Neural Comput*. **26** 2896–2924. MR3223194 https://doi.org/10.1162/NECO_a_00674

LIANG, S., LU, W., SONG, R. and WANG, L. (2017). Sparse concordance-assisted learning for optimal treatment decision. *J. Mach. Learn. Res*. **18** Paper No. 202, 26. MR3827090

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

LUEDTKE, A. R. and VAN DER LAAN, M. J. (2016). Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Ann. Statist*. **44** 713–742. MR3476615 https://doi.org/10.1214/15-AOS1384

MEBANE, W. R. JR., SEKHON, J. S. et al. (2011). Genetic optimization using derivatives: The rgenoud package for R. *J. Stat. Softw*. **42** 1–26.

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

NOLAN, D. and POLLARD, D. (1987). U-processes: Rates of convergence. *Ann. Statist*. **15** 780–799. MR0888439 https://doi.org/10.1214/aos/1176350374

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., HERNAN, M. A. and BRUMBACK, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11** 550–560.

SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist*. **6** 461–464. MR0468014

SHERMAN, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica* **61** 123–137. MR1201705 https://doi.org/10.2307/2951780

SHERMAN, R. P. (1994). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Ann. Statist*. **22** 439–459. MR1272092 https://doi.org/10.1214/aos/1176325377

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

SHI, C., SONG, R. and LU, W. (2021). Supplement to “Concordance and value information criteria for optimal treatment decision.” https://doi.org/10.1214/19-AOS1908SUPP

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

WATKINS, C. J. C. H. and DAYAN, P. (1992). Q-learning. *Mach. Learn*. **8** 279–292.

ZHANG, B., TSIATIS, A. A., LABER, E. B. and DAVIDIAN, M. (2012). A robust method for estimating optimal treatment regimes. *Biometrics* **68** 1010–1018. MR3040007 https://doi.org/10.1111/j.1541-0420.2012.01763.x

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, X., WU, Y., WANG, L. and LI, R. (2016). A consistent information criterion for support vector machines in diverging model spaces. *J. Mach. Learn. Res*. **17** Paper No. 16, 26. MR3491110 https://doi.org/10.1016/j.jocs.2016.07.001

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

The Annals of Statistics
2021, Vol. 49, No. 1, 76–99
https://doi.org/10.1214/20-AOS1942

STATISTICALLY OPTIMAL AND COMPUTATIONALLY EFFICIENT LOW RANK TENSOR COMPLETION FROM NOISY ENTRIES

BY DONG XIA[1], MING YUAN[2] AND CUN-HUI ZHANG[3]

[1]*Department of Mathematics, Hong Kong University of Science and Technology, ming.yuan@columbia.edu*
[2]*Department of Statistics, Columbia University, ming.yuan@columbia.edu*
[3]*Department of Statistics and Biostatistics, Rutgers University, czhang@stat.rutgers.edu*

In this article, we develop methods for estimating a low rank tensor from noisy observations on a subset of its entries to achieve both statistical and computational efficiencies. There have been a lot of recent interests in this problem of noisy tensor completion. Much of the attention has been focused on the fundamental computational challenges often associated with problems involving higher order tensors, yet very little is known about their statistical performance. To fill in this void, in this article, we characterize the fundamental statistical limits of noisy tensor completion by establishing minimax optimal rates of convergence for estimating a kth order low rank tensor under the general ℓ_p $(1 \le p \le 2)$ norm which suggest significant room for improvement over the existing approaches. Furthermore, we propose a polynomial-time computable estimating procedure based upon power iteration and a second-order spectral initialization that achieves the optimal rates of convergence. Our method is fairly easy to implement and numerical experiments are presented to further demonstrate the practical merits of our estimator.

REFERENCES

[1] ANANDKUMAR, A., GE, R., HSU, D., KAKADE, S. M. and TELGARSKY, M. (2014). Tensor decompositions for learning latent variable models. *J. Mach. Learn. Res.* **15** 2773–2832. MR3270750

[2] AUFFINGER, A. and BEN AROUS, G. (2013). Complexity of random smooth functions on the high-dimensional sphere. *Ann. Probab.* **41** 4214–4247. MR3161473 https://doi.org/10.1214/13-AOP862

[3] AUFFINGER, A., BEN AROUS, G. and ČERNÝ, J. (2013). Random matrices and complexity of spin glasses. *Comm. Pure Appl. Math.* **66** 165–201. MR2999295 https://doi.org/10.1002/cpa.21422

[4] BARAK, B. and MOITRA, A. (2016). Noisy tensor completion via the sum-of-squares hierarchy. In *29th Annual Conference on Learning Theory* 417–445.

[5] CANDES, E. J. and PLAN, Y. (2010). Matrix completion with noise. *Proc. IEEE* **98** 925–936.

[6] CHAGANTY, A. T. and LIANG, P. (2013). Spectral experts for estimating mixtures of linear regressions. In *ICML* (3) 1040–1048.

[7] COCOSCO, C. A., KOLLOKIAN, V., KWAN, R. K.-S., PIKE, G. B. and EVANS, A. C. (1997). Brainweb: Online interface to a 3D MRI simulated brain database. In *NeuroImage* Citeseer.

[8] COHEN, S. and COLLINS, M. (2012). Tensor decomposition for fast parsing with latent-variable PCFGS. In *Advances in Neural Information Processing Systems*.

[9] DE LATHAUWER, L., DE MOOR, B. and VANDEWALLE, J. (2000). A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.* **21** 1253–1278. MR1780272 https://doi.org/10.1137/S0895479896305696

[10] DE SILVA, V. and LIM, L.-H. (2008). Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM J. Matrix Anal. Appl.* **30** 1084–1127. MR2447444 https://doi.org/10.1137/06066518X

[11] HILLAR, C. J. and LIM, L.-H. (2013). Most tensor problems are NP-hard. *J. ACM* **60** 45. MR3144915 https://doi.org/10.1145/2512329

[12] JAIN, P. and OH, S. (2014). Provable tensor factorization with missing data. In *Advances in Neural Information Processing Systems* 1431–1439.

MSC2020 subject classifications. 60K35.
Key words and phrases. Tensor completion, low-rank, power iteration, spectral initialization, minimax optimality.

[13] KESHAVAN, R. H., MONTANARI, A. and OH, S. (2010). Matrix completion from noisy entries. *J. Mach. Learn. Res.* **11** 2057–2078. MR2678022

[14] KLOPP, O. (2014). Noisy low-rank matrix completion with general sampling distribution. *Bernoulli* **20** 282–303. MR3160583 https://doi.org/10.3150/12-BEJ486

[15] KOLDA, T. G. and BADER, B. W. (2009). Tensor decompositions and applications. *SIAM Rev.* **51** 455–500. MR2535056 https://doi.org/10.1137/07070111X

[16] KOLTCHINSKII, V., LOUNICI, K. and TSYBAKOV, A. B. (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *Ann. Statist.* **39** 2302–2329. MR2906869 https://doi.org/10.1214/11-AOS894

[17] KREIMER, N., STANTON, A. and SACCHI, M. D. (2013). Tensor completion based on nuclear norm minimization for 5d seismic data reconstruction. *Geophysics* **78** V273–V284.

[18] KWAN, R. K.-S., EVANS, A. C. and PIKE, G. B. (1996). An extensible MRI simulator for post-processing evaluation. In *Visualization in Biomedical Computing* 135–140. Springer, Berlin.

[19] LI, N. and LI, B. (2010). Tensor completion for on-board compression of hyperspectral images. In 17*th IEEE International Conference on Image Processing* (*ICIP*) 517–520.

[20] LIM, L.-H. and COMON, P. (2010). Multiarray signal processing: Tensor decomposition meets compressed sensing. *C. R., Méc.* **338** 311–320.

[21] LIU, J., MUSIALSKI, P., WONKA, P. and YE, J. (2013). Tensor completion for estimating missing values in visual data. *IEEE Trans. Pattern Anal. Mach. Intell.* **35** 208–220.

[22] 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

[23] NION, D. and SIDIROPOULOS, N. D. (2010). Tensor algebra and multidimensional harmonic retrieval in signal processing for MIMO radar. *IEEE Trans. Signal Process.* **58** 5693–5705. MR2789612 https://doi.org/10.1109/TSP.2010.2058802

[24] ROHDE, A. and TSYBAKOV, A. B. (2011). Estimation of high-dimensional low-rank matrices. *Ann. Statist.* **39** 887–930. MR2816342 https://doi.org/10.1214/10-AOS860

[25] SEMERCI, O., HAO, N., KILMER, M. E. and MILLER, E. L. (2014). Tensor-based formulation and nuclear norm regularization for multienergy computed tomography. *IEEE Trans. Image Process.* **23** 1678–1693. MR3191324 https://doi.org/10.1109/TIP.2014.2305840

[26] USCHMAJEW, A. (2012). Local convergence of the alternating least squares algorithm for canonical tensor approximation. *SIAM J. Matrix Anal. Appl.* **33** 639–652. MR2970223 https://doi.org/10.1137/110843587

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

[28] XIA, D. and YUAN, M. (2019). On polynomial time methods for exact low-rank tensor completion. *Found. Comput. Math.* **19** 1265–1313. MR4029842 https://doi.org/10.1007/s10208-018-09408-6

[29] XIA, D., YUAN, M. and ZHANG, C.-H (2021). Supplement to "Statistically optimal and computationally efficient low rank tensor completion from noisy entries." https://doi.org/10.1214/20-AOS1942SUPP

[30] XIE, W., ZHU, F., JIANG, J., LIM, E.-P. and TOPICSKETCH, K. W. (2016). Real-time bursty topic detection from Twitter. *IEEE Trans. Knowl. Data Eng.* **28** 2216–2229.

[31] XU, Y. and YIN, W. (2013). A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM J. Imaging Sci.* **6** 1758–1789. MR3105787 https://doi.org/10.1137/120887795

[32] YUAN, M. and ZHANG, C.-H. (2016). On tensor completion via nuclear norm minimization. *Found. Comput. Math.* **16** 1031–1068. MR3529132 https://doi.org/10.1007/s10208-015-9269-5

[33] YUAN, M. and ZHANG, C.-H. (2017). Incoherent tensor norms and their applications in higher order tensor completion. *IEEE Trans. Inf. Theory* **63** 6753–6766. MR3707566 https://doi.org/10.1109/TIT.2017.2724549

The Annals of Statistics
2021, Vol. 49, No. 1, 100–128
https://doi.org/10.1214/20-AOS1949

TRANSFER LEARNING FOR NONPARAMETRIC CLASSIFICATION: MINIMAX RATE AND ADAPTIVE CLASSIFIER

BY T. TONY CAI[*] AND HONGJI WEI[†]

Department of Statistics, The Wharton School, University of Pennsylvania, [*]tcai@wharton.upenn.edu; [†]hongjiw@wharton.upenn.edu

Human learners have the natural ability to use knowledge gained in one setting for learning in a different but related setting. This ability to transfer knowledge from one task to another is essential for effective learning. In this paper, we study transfer learning in the context of nonparametric classification based on observations from different distributions under the posterior drift model, which is a general framework and arises in many practical problems.

We first establish the minimax rate of convergence and construct a rate-optimal two-sample weighted K-NN classifier. The results characterize precisely the contribution of the observations from the source distribution to the classification task under the target distribution. A data-driven adaptive classifier is then proposed and is shown to simultaneously attain within a logarithmic factor of the optimal rate over a large collection of parameter spaces. Simulation studies and real data applications are carried out where the numerical results further illustrate the theoretical analysis. Extensions to the case of multiple source distributions are also considered.

REFERENCES

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

BEN-DAVID, S., BLITZER, J., CRAMMER, K. and PEREIRA, F. (2007). Analysis of representations for domain adaptation. In *Advances in Neural Information Processing Systems* 137–144.

BLITZER, J., CRAMMER, K., KULESZA, A., PEREIRA, F. and WORTMAN, J. (2008). Learning bounds for domain adaptation. In *Advances in Neural Information Processing Systems* 129–136.

CAI, T. T. and WEI, H. (2020). Supplement to "Transfer learning for nonparametric classification: Minimax rate and adaptive classifier." https://doi.org/10.1214/20-AOS1949SUPP

CHOI, K., FAZEKAS, G., SANDLER, M. and CHO, K. (2017). Transfer learning for music classification and regression tasks. Preprint. Availabe at arXiv:1703.09179.

COVER, T. M. and HART, P. E. (1967). Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **13** 21–27.

CRAMMER, K., KEARNS, M. and WORTMAN, J. (2006). Learning from data of variable quality. In *Advances in Neural Information Processing Systems* 219–226.

GADAT, S., KLEIN, T. and MARTEAU, C. (2016). Classification in general finite dimensional spaces with the k-nearest neighbor rule. *Ann. Statist.* **44** 982–1009. MR3485951 https://doi.org/10.1214/15-AOS1395

GAMA, J., ŽLIOBAITĖ, I., BIFET, A., PECHENIZKIY, M. and BOUCHACHIA, A. (2014). A survey on concept drift adaptation. *ACM Comput. Surv.* **46** Art. ID 44.

GONG, B., SHI, Y., SHA, F. and GRAUMAN, K. (2012). Geodesic flow kernel for unsupervised domain adaptation. In 2012 *IEEE Conference on Computer Vision and Pattern Recognition* 2066–2073. IEEE, Piscataway, NJ.

GYŐRFI, L. (1978). On the rate of convergence of nearest neighbor rules. *IEEE Trans. Inf. Theory* **24** 509–512. MR0501595 https://doi.org/10.1109/TIT.1978.1055898

HUANG, J.-T., LI, J., YU, D., DENG, L. and GONG, Y. (2013). Cross-language knowledge transfer using multilingual deep neural network with shared hidden layers. In 2013 *IEEE International Conference on Acoustics, Speech and Signal Processing* 7304–7308. IEEE, Piscataway, NJ.

MSC2020 subject classifications. Primary 62F30; secondary 62B10, 62F12.
Key words and phrases. Adaptivity, classification, domain adaptation, minimax rate, transfer learning.

JOHNSON, B. A. and IIZUKA, K. (2016). Integrating OpenStreetMap crowdsourced data and Landsat time-series imagery for rapid land use/land cover (LULC) mapping: Case study of the Laguna de Bay area of the Philippines. *Appl. Geogr.* **67** 140–149.

KARGER, D. R., OH, S. and SHAH, D. (2011). Budget-optimal crowdsourcing using low-rank matrix approximations. In 2011 49*th Annual Allerton Conference on Communication, Control, and Computing (Allerton)* 284–291. IEEE, Piscataway, NJ.

KPOTUFE, S. and MARTINET, G. (2018). Marginal singularity, and the benefits of labels in covariate-shift. Preprint. Availabe at arXiv:1803.01833.

LEE, S.-I., CHATALBASHEV, V., VICKREY, D. and KOLLER, D. (2007). Learning a meta-level prior for feature relevance from multiple related tasks. In *Proceedings of the* 24*th International Conference on Machine Learning* 489–496. ACM, New York.

LEPSKI, O. V. and SPOKOINY, V. G. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.* **25** 2512–2546. MR1604408 https://doi.org/10.1214/aos/1030741083

LEPSKI, O. V. (1991). On a problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** 454–466.

LEPSKI, O. V. (1992). Asymptotically minimax adaptive estimation. I: Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.* **36** 682–697.

LEPSKI, O. V. (1993). Asymptotically minimax adaptive estimation. II. Schemes without optimal adaptation: Adaptive estimators. *Theory Probab. Appl.* **37** 433–448.

MANSOUR, Y., MOHRI, M. and ROSTAMIZADEH, A. (2009). Domain adaptation: Learning bounds and algorithms. In *Proceedings of the* 22*nd Annual Conference on Learning Theory* (*COLT* 2009), Montréal, Canada.

MENON, A., VAN ROOYEN, B., ONG, C. S. and WILLIAMSON, B. (2015). Learning from corrupted binary labels via class-probability estimation. In *International Conference on Machine Learning* 125–134.

PAN, S. J. and YANG, Q. (2010). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **22** 1345–1359.

SAMWORTH, R. J. (2012). Optimal weighted nearest neighbour classifiers. *Ann. Statist.* **40** 2733–2763. MR3097618 https://doi.org/10.1214/12-AOS1049

SCOTT, C. (2019). A generalized Neyman–Pearson criterion for optimal domain adaption. In *Algorithmic Learning Theory* 2019. *Proc. Mach. Learn. Res.* (*PMLR*) **98** 738–761. MR3932867

SHIMODAIRA, H. (2000). Improving predictive inference under covariate shift by weighting the log-likelihood function. *J. Statist. Plann. Inference* **90** 227–244. MR1795598 https://doi.org/10.1016/S0378-3758(00)00115-4

SUGIYAMA, M., NAKAJIMA, S., KASHIMA, H., BUENAU, P. V. and KAWANABE, M. (2008). Direct importance estimation with model selection and its application to covariate shift adaptation. In *Advances in Neural Information Processing Systems* 1433–1440.

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

TSYMBAL, A. (2004). The problem of concept drift: Definitions and related work. Computer Science Department, Trinity College Dublin.

TZENG, E., HOFFMAN, J., SAENKO, K. and DARRELL, T. (2017). Adversarial discriminative domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 7167–7176.

VAN ROOYEN, B. and WILLIAMSON, R. C. (2017). A theory of learning with corrupted labels. *J. Mach. Learn. Res.* **18** Art. ID 228. MR3845527

WEISS, K., KHOSHGOFTAAR, T. M. and WANG, D. (2016). A survey of transfer learning. *J. Big Data* **3** Art. ID 9.

YAO, Y. and DORETTO, G. (2010). Boosting for transfer learning with multiple sources. In 2010 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 1855–1862. IEEE, Piscataway, NJ.

YUEN, M.-C., KING, I. and LEUNG, K.-S. (2011). A survey of crowdsourcing systems. In 2011 *IEEE Third International Conference on Privacy, Security, Risk and Trust and* 2011 *IEEE Third International Conference on Social Computing* 766–773. IEEE, Piscataway, NJ.

ZHANG, Y., CHEN, X., ZHOU, D. and JORDAN, M. I. (2014). Spectral methods meet EM: A provably optimal algorithm for crowdsourcing. In *Advances in Neural Information Processing Systems* 1260–1268.

The Annals of Statistics
2021, Vol. 49, No. 1, 129–153
https://doi.org/10.1214/20-AOS1950

ADAPTATION IN MULTIVARIATE LOG-CONCAVE DENSITY ESTIMATION

BY OLIVER Y. FENG[1,*], ADITYANAND GUNTUBOYINA[2], ARLENE K. H. KIM[3] AND RICHARD J. SAMWORTH[1,†]

[1]*Statistical Laboratory, University of Cambridge, [*]oyf20@cam.ac.uk; [†]r.samworth@statslab.cam.ac.uk*
[2]*Department of Statistics, University of California, Berkeley, aditya@stat.berkeley.edu*
[3]*Department of Statistics, Korea University, arlenent@korea.ac.kr*

We study the adaptation properties of the multivariate log-concave maximum likelihood estimator over three subclasses of log-concave densities. The first consists of densities with polyhedral support whose logarithms are piecewise affine. The complexity of such densities f can be measured in terms of the sum $\Gamma(f)$ of the numbers of facets of the subdomains in the polyhedral subdivision of the support induced by f. Given n independent observations from a d-dimensional log-concave density with $d \in \{2, 3\}$, we prove a sharp oracle inequality, which in particular implies that the Kullback–Leibler risk of the log-concave maximum likelihood estimator for such densities is bounded above by $\Gamma(f)/n$, up to a polylogarithmic factor. Thus, the rate can be essentially parametric, even in this multivariate setting. For the second type of adaptation, we consider densities that are bounded away from zero on a polytopal support; we show that up to polylogarithmic factors, the log-concave maximum likelihood estimator attains the rate $n^{-4/7}$ when $d = 3$, which is faster than the worst-case rate of $n^{-1/2}$. Finally, our third type of subclass consists of densities whose contours are well separated; these new classes are constructed to be affine invariant and turn out to contain a wide variety of densities, including those that satisfy Hölder regularity conditions. Here, we prove another sharp oracle inequality, which reveals in particular that the log-concave maximum likelihood estimator attains a risk bound of order $n^{-\min(\frac{\beta+3}{\beta+7}, \frac{4}{7})}$ when $d = 3$ over the class of β-Hölder log-concave densities with $\beta > 1$, again up to a polylogarithmic factor.

REFERENCES

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

BALABDAOUI, F., RUFIBACH, K. and WELLNER, J. A. (2009). Limit distribution theory for maximum likelihood estimation of a log-concave density. *Ann. Statist.* **37** 1299–1331. MR2509075 https://doi.org/10.1214/08-AOS609

BARAUD, Y. and BIRGÉ, L. (2016). Rates of convergence of rho-estimators for sets of densities satisfying shape constraints. *Stochastic Process. Appl.* **12** 3888–3912.

BELLEC, P. C. (2018). Sharp oracle inequalities for least squares estimators in shape restricted regression. *Ann. Statist.* **46** 745–780. MR3782383 https://doi.org/10.1214/17-AOS1566

BIRGÉ, L. (1987). Estimating a density under order restrictions: Nonasymptotic minimax risk. *Ann. Statist.* **15** 995–1012. MR0902241 https://doi.org/10.1214/aos/1176350488

BRUNEL, V.-E. (2013). Adaptive estimation of convex polytopes and convex sets from noisy data. *Electron. J. Stat.* **7** 1301–1327. MR3063609 https://doi.org/10.1214/13-EJS804

BRUNS, W. and GUBELADZE, J. (2009). *Polytopes, Rings, and K-Theory. Springer Monographs in Mathematics.* Springer, Dordrecht. MR2508056 https://doi.org/10.1007/b105283

CAI, T. T. and LOW, M. G. (2015). A framework for estimation of convex functions. *Statist. Sinica* **25** 423–456. MR3379081

MSC2020 subject classifications. 62G07, 62G20.
Key words and phrases. Multivariate adaptation, bracketing entropy, log-concavity, contour separation, maximum likelihood estimation.

CARPENTER, T., DIAKONIKOLAS, I., SIDIROPOULOS, A. and STEWART, A. (2018). Near-optimal sample complexity bounds for maximum likelihood estimation of multivariate log-concave densities. *Proc. Mach. Learn. Res.* **75** 1–29.

CHATTERJEE, S. (2014). A new perspective on least squares under convex constraint. *Ann. Statist.* **42** 2340–2381. MR3269982 https://doi.org/10.1214/14-AOS1254

CHATTERJEE, S., GUNTUBOYINA, A. and SEN, B. (2015). On risk bounds in isotonic and other shape restricted regression problems. *Ann. Statist.* **43** 1774–1800. MR3357878 https://doi.org/10.1214/15-AOS1324

CHATTERJEE, S., GUNTUBOYINA, A. and SEN, B. (2018). On matrix estimation under monotonicity constraints. *Bernoulli* **24** 1072–1100. MR3706788 https://doi.org/10.3150/16-BEJ865

CHATTERJEE, S. and LAFFERTY, J. (2019). Adaptive risk bounds in unimodal regression. *Bernoulli* **25** 1–25. MR3892309 https://doi.org/10.3150/16-bej922

CHEN, Y. and SAMWORTH, R. J. (2013). Smoothed log-concave maximum likelihood estimation with applications. *Statist. Sinica* **23** 1373–1398. MR3114718

CHEN, Y. and SAMWORTH, R. J. (2016). Generalized additive and index models with shape constraints. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 729–754. MR3534348 https://doi.org/10.1111/rssb.12137

DENG, H. and ZHANG, C.-H. (2018). Isotonic regression in multi-dimensional spaces and graphs. Preprint. Available at arXiv:1812.08944.

DOSS, C. R. and WELLNER, J. A. (2016). Global rates of convergence of the MLEs of log-concave and s-concave densities. *Ann. Statist.* **44** 954–981. MR3485950 https://doi.org/10.1214/15-AOS1394

DÜMBGEN, L. and RUFIBACH, K. (2009). Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency. *Bernoulli* **15** 40–68. MR2546798 https://doi.org/10.3150/08-BEJ141

DÜMBGEN, L., SAMWORTH, R. and SCHUHMACHER, D. (2011). Approximation by log-concave distributions, with applications to regression. *Ann. Statist.* **39** 702–730. MR2816336 https://doi.org/10.1214/10-AOS853

GARDNER, R. J., KIDERLEN, M. and MILANFAR, P. (2006). Convergence of algorithms for reconstructing convex bodies and directional measures. *Ann. Statist.* **34** 1331–1374. MR2278360 https://doi.org/10.1214/009053606000000335

GOLDENSHLUGER, A. and LEPSKI, O. (2014). On adaptive minimax density estimation on R^d. *Probab. Theory Related Fields* **159** 479–543. MR3230001 https://doi.org/10.1007/s00440-013-0512-1

GROENEBOOM, P. and HENDRICKX, K. (2018). Current status linear regression. *Ann. Statist.* **46** 1415–1444. MR3819105 https://doi.org/10.1214/17-AOS1589

GROENEBOOM, P. and JONGBLOED, G. (2014). *Nonparametric Estimation Under Shape Constraints: Estimators, Algorithms and Asymptotics. Cambridge Series in Statistical and Probabilistic Mathematics* **38**. Cambridge Univ. Press, New York. MR3445293 https://doi.org/10.1017/CBO9781139020893

GUNTUBOYINA, A. (2012). Optimal rates of convergence for convex set estimation from support functions. *Ann. Statist.* **40** 385–411. MR3014311 https://doi.org/10.1214/11-AOS959

GUNTUBOYINA, A. and SEN, B. (2013). Covering numbers for convex functions. *IEEE Trans. Inf. Theory* **59** 1957–1965. MR3043776 https://doi.org/10.1109/TIT.2012.2235172

HAN, Q. (2019). Global empirical risk minimizers with "shape constraints" are rate optimal in general dimensions. Preprint. Available at arXiv:1905.12823v1.

HAN, Q. and WELLNER, J. A. (2016). Multivariate convex regression: Global risk bounds and adaptation. Preprint. Available at arXiv:1601.06844.

HAN, Q., WANG, T., CHATTERJEE, S. and SAMWORTH, R. J. (2019). Isotonic regression in general dimensions. *Ann. Statist.* **47** 2440–2471. MR3988762 https://doi.org/10.1214/18-AOS1753

JANKOWSKI, H. (2014). Convergence of linear functionals of the Grenander estimator under misspecification. *Ann. Statist.* **42** 625–653. MR3210981 https://doi.org/10.1214/13-AOS1196

FENG, O. Y., GUNTUBOYINA, A., KIM, A. K. and SAMWORTH, R. J. (2021). Supplement to "Adaptation in multivariate log-concave density estimation." https://doi.org/10.1214/20-AOS1950SUPP

KIM, A. K. H., GUNTUBOYINA, A. and SAMWORTH, R. J. (2018). Adaptation in log-concave density estimation. *Ann. Statist.* **46** 2279–2306. MR3845018 https://doi.org/10.1214/17-AOS1619

KIM, A. K. H. and SAMWORTH, R. J. (2016). Global rates of convergence in log-concave density estimation. *Ann. Statist.* **44** 2756–2779. MR3576560 https://doi.org/10.1214/16-AOS1480

KOENKER, R. and MIZERA, I. (2014). Convex optimization, shape constraints, compound decisions, and empirical Bayes rules. *J. Amer. Statist. Assoc.* **109** 674–685. MR3223742 https://doi.org/10.1080/01621459.2013.869224

KUR, G., DAGAN, Y. and RAKHLIN, A. (2019). Optimality of maximum likelihood for log-concave density estimation and bounded convex regression. Preprint. Available at arXiv:1903.05315.

MAZUMDER, R., CHOUDHURY, A., IYENGAR, G. and SEN, B. (2019). A computational framework for multivariate convex regression and its variants. *J. Amer. Statist. Assoc.* **114** 318–331. MR3941257 https://doi.org/10.1080/01621459.2017.1407771

SAMWORTH, R. J. (2018). Recent progress in log-concave density estimation. *Statist. Sci.* **33** 493–509. MR3881205 https://doi.org/10.1214/18-STS666

SAMWORTH, R. J. and YUAN, M. (2012). Independent component analysis via nonparametric maximum likelihood estimation. *Ann. Statist.* **40** 2973–3002. MR3097966 https://doi.org/10.1214/12-AOS1060

SAUMARD, A. and WELLNER, J. A. (2014). Log-concavity and strong log-concavity: A review. *Stat. Surv.* **8** 45–114. MR3290441 https://doi.org/10.1214/14-SS107

SCHNEIDER, R. (2014). *Convex Bodies: The Brunn–Minkowski Theory*, 2nd ed. *Encyclopedia of Mathematics and Its Applications* **151**. Cambridge Univ. Press, Cambridge. MR3155183

SEREGIN, A. and WELLNER, J. A. (2010). Nonparametric estimation of multivariate convex-transformed densities. *Ann. Statist.* **38** 3751–3781. MR2766867 https://doi.org/10.1214/10-AOS840

SHAH, N. B., BALAKRISHNAN, S., GUNTUBOYINA, A. and WAINWRIGHT, M. J. (2017). Stochastically transitive models for pairwise comparisons: Statistical and computational issues. *IEEE Trans. Inf. Theory* **63** 934–959. MR3604649 https://doi.org/10.1109/TIT.2016.2634418

SOLOFF, J. A., GUNTUBOYINA, A. and PITMAN, J. (2019). Distribution-free properties of isotonic regression. *Electron. J. Stat.* **13** 3243–3253. MR4010598 https://doi.org/10.1214/19-ejs1594

VAN DE GEER, S. A. (2000). *Empirical Processes in M-Estimation*. Cambridge Univ. Press, Cambridge.

WALTHER, G. (2009). Inference and modeling with log-concave distributions. *Statist. Sci.* **24** 319–327. MR2757433 https://doi.org/10.1214/09-STS303

WIEACKER, J. A. (1987). *Einige Probleme der Polyedrischen Approximation*. Diplomarbeit, Freiburg.

XU, M., CHEN, M. and LAFFERTY, J. (2016). Faithful variable screening for high-dimensional convex regression. *Ann. Statist.* **44** 2624–2660. MR3576556 https://doi.org/10.1214/15-AOS1425

XU, M. and SAMWORTH, R. J. (2019). High-dimensional nonparametric density estimation via symmetry and shape constraints. Preprint. Available at arXiv:1903.06092.

ZHANG, C.-H. (2002). Risk bounds in isotonic regression. *Ann. Statist.* **30** 528–555. MR1902898 https://doi.org/10.1214/aos/1021379864

The Annals of Statistics
2021, Vol. 49, No. 1, 154–181
https://doi.org/10.1214/20-AOS1951

ASYMPTOTICALLY INDEPENDENT U-STATISTICS IN HIGH-DIMENSIONAL TESTING

BY YINQIU HE[1,*], GONGJUN XU[1,†], CHONG WU[2] AND WEI PAN[3]

[1]*Department of Statistics, University of Michigan,* [*]*yqhe@umich.edu;* [†]*gongjun@umich.edu*
[2]*Department of Statistics, Florida State University,* *cwu3@fsu.edu*
[3]*Division of Biostatistics, School of Public Health, University of Minnesota,* *panxx014@umn.edu*

Many high-dimensional hypothesis tests aim to globally examine marginal or low-dimensional features of a high-dimensional joint distribution, such as testing of mean vectors, covariance matrices and regression coefficients. This paper constructs a family of U-statistics as unbiased estimators of the ℓ_p-norms of those features. We show that under the null hypothesis, the U-statistics of different finite orders are asymptotically independent and normally distributed. Moreover, they are also asymptotically independent with the maximum-type test statistic, whose limiting distribution is an extreme value distribution. Based on the asymptotic independence property, we propose an adaptive testing procedure which combines p-values computed from the U-statistics of different orders. We further establish power analysis results and show that the proposed adaptive procedure maintains high power against various alternatives.

REFERENCES

[1] ANDERSON, T. W. (2009). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.
[2] ALZHEIMER'S ASSOCIATION (2018). 2018 Alzheimer's disease facts and figures. *Alzheimer's Dement*. **14** 367–429.
[3] BAI, Z., JIANG, D., YAO, J.-F. and ZHENG, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.* **37** 3822–3840. MR2572444 https://doi.org/10.1214/09-AOS694
[4] BAI, Z. and SARANADASA, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6** 311–329. MR1399305
[5] BICKEL, P. J. and LEVINA, E. (2008). Regularized estimation of large covariance matrices. *Ann. Statist.* **36** 199–227. MR2387969 https://doi.org/10.1214/009053607000000758
[6] 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
[7] CAI, T. T. (2017). Global testing and large-scale multiple testing for high-dimensional covariance structures. *Annu. Rev. Stat. Appl.* **4** 423–446. https://doi.org/10.1146/annurev-statistics-060116-053754
[8] CAI, T. T. and JIANG, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39** 1496–1525. MR2850210 https://doi.org/10.1214/11-AOS879
[9] CAI, T. T., LIU, W. and XIA, Y. (2014). Two-sample test of high dimensional means under dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 349–372. MR3164870 https://doi.org/10.1111/rssb.12034
[10] CAI, T. T. and MA, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** 2359–2388. MR3160557 https://doi.org/10.3150/12-BEJ455
[11] CHEN, S. X., LI, J. and ZHONG, P.-S. (2019). Two-sample and ANOVA tests for high dimensional means. *Ann. Statist.* **47** 1443–1474. MR3911118 https://doi.org/10.1214/18-AOS1720
[12] CHEN, S. X. and QIN, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. MR2604697 https://doi.org/10.1214/09-AOS716
[13] CHEN, S. X., ZHANG, L.-X. and ZHONG, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** 810–819. MR2724863 https://doi.org/10.1198/jasa.2010.tm09560
[14] CHEN, X. (2018). Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications. *Ann. Statist.* **46** 642–678. MR3782380 https://doi.org/10.1214/17-AOS1563

MSC2020 subject classifications. 62F03, 62F05.
Key words and phrases. High-dimensional hypothesis test, U-statistics, adaptive testing.

[15] COLANTUONI, C., LIPSKA, B. K., YE, T., HYDE, T. M., TAO, R., LEEK, J. T., COLANTUONI, E. A., ELKAHLOUN, A. G., HERMAN, M. M. et al. (2011). Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* **478** 519.

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

[17] DONOHO, D. and JIN, J. (2015). Higher criticism for large-scale inference, especially for rare and weak effects. *Statist. Sci.* **30** 1–25. MR3317751 https://doi.org/10.1214/14-STS506

[18] FAN, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *J. Amer. Statist. Assoc.* **91** 674–688. MR1395735 https://doi.org/10.2307/2291663

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

[20] FAN, J., LIAO, Y. and YAO, J. (2015). Power enhancement in high-dimensional cross-sectional tests. *Econometrica* **83** 1497–1541. MR3384226 https://doi.org/10.3982/ECTA12749

[21] FAN, J., LV, J. and QI, L. (2011). Sparse high dimensional models in economics. *Ann. Rev. Econ.* **3** 291–317. https://doi.org/10.1146/annurev-economics-061109-080451

[22] FRAHM, G. (2004). Generalized elliptical distributions: Theory and applications. Ph.D. thesis, Univ. Köln.

[23] FRISTON, K. J. (2009). Modalities, modes, and models in functional neuroimaging. *Science* **326** 399–403. https://doi.org/10.1126/science.1174521

[24] GAETAN, C. and GUYON, X. (2010). *Spatial Statistics and Modeling*. *Springer Series in Statistics*. Springer, New York. MR2569034 https://doi.org/10.1007/978-0-387-92257-7

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

[26] GREGORY, K. B., CARROLL, R. J., BALADANDAYUTHAPANI, V. and LAHIRI, S. N. (2015). A two-sample test for equality of means in high dimension. *J. Amer. Statist. Assoc.* **110** 837–849. MR3367268 https://doi.org/10.1080/01621459.2014.934826

[27] HALL, P. and JIN, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.* **38** 1686–1732. MR2662357 https://doi.org/10.1214/09-AOS764

[28] HE, Y., XU, G., WU, C. and PAN, W. (2021). Supplement to "Asymptotically independent U-statistics in high-dimensional testing." https://doi.org/10.1214/20-AOS1951SUPP

[29] HEINIG, M., PETRETTO, E., WALLACE, C., BOTTOLO, L., ROTIVAL, M., LU, H., LI, Y., SARWAR, R., LANGLEY, S. R. et al. (2010). A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk. *Nature* **467** 460.

[30] HO, H.-C. and HSING, T. (1996). On the asymptotic joint distribution of the sum and maximum of stationary normal random variables. *J. Appl. Probab.* **33** 138–145. MR1371961 https://doi.org/10.2307/3215271

[31] HO, H.-C. and MCCORMICK, W. P. (1999). Asymptotic distribution of sum and maximum for Gaussian processes. *J. Appl. Probab.* **36** 1031–1044. MR1742148 https://doi.org/10.1239/jap/1032374753

[32] HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* **19** 293–325. MR0026294 https://doi.org/10.1214/aoms/1177730196

[33] HSING, T. (1995). A note on the asymptotic independence of the sum and maximum of strongly mixing stationary random variables. *Ann. Probab.* **23** 938–947. MR1334178

[34] JAMES, B., JAMES, K. and QI, Y. (2007). Limit distribution of the sum and maximum from multivariate Gaussian sequences. *J. Multivariate Anal.* **98** 517–532. MR2293012 https://doi.org/10.1016/j.jmva.2006.06.009

[35] JANSEN, I. E., SAVAGE, J. E., WATANABE, K., BRYOIS, J., WILLIAMS, D. M. et al. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51** 404–413.

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

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

[38] 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

[39] KANEHISA, M., GOTO, S., FURUMICHI, M., TANABE, M. and HIRAKAWA, M. (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **38** D355–D360. https://doi.org/10.1093/nar/gkp896

[40] KIM, J., ZHANG, Y. and PAN, W. (2016). Powerful and adaptive testing for multi-trait and multi-SNP associations with GWAS and sequencing data. *Genetics* **203** 715–731.

[41] LAN, W., LUO, R., TSAI, C.-L., WANG, H. and YANG, Y. (2015). Testing the diagonality of a large covariance matrix in a regression setting. *J. Bus. Econom. Statist.* **33** 76–86. MR3303743 https://doi.org/10.1080/07350015.2014.923317

[42] LEDOIT, O. and WOLF, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30** 1081–1102. MR1926169 https://doi.org/10.1214/aos/1031689018

[43] LEUNG, D. and DRTON, M. (2018). Testing independence in high dimensions with sums of rank correlations. *Ann. Statist.* **46** 280–307. MR3766953 https://doi.org/10.1214/17-AOS1550

[44] LI, D. and XUE, L. (2015). Joint limiting laws for high-dimensional independence tests. ArXiv E-prints.

[45] 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

[46] LIU, W.-D., LIN, Z. and SHAO, Q.-M. (2008). The asymptotic distribution and Berry–Esseen bound of a new test for independence in high dimension with an application to stochastic optimization. *Ann. Appl. Probab.* **18** 2337–2366. MR2474539 https://doi.org/10.1214/08-AAP527

[47] MANOLIO, T. A., COLLINS, F. S., COX, N. J., GOLDSTEIN, D. B., HINDORFF, L. A. et al. (2009). Finding the missing heritability of complex diseases. *Nature* **461** 747.

[48] MCCORMICK, W. P. and QI, Y. (2000). Asymptotic distribution for the sum and maximum of Gaussian processes. *J. Appl. Probab.* **37** 958–971. MR1808861 https://doi.org/10.1239/jap/1014843076

[49] MOSTELLER, F. and FISHER, R. A. (1948). Questions and answers. *Amer. Statist.* **2** 30–31.

[50] MUIRHEAD, R. J. (2009). *Aspects of Multivariate Statistical Theory*. Wiley, New York.

[51] PAINDAVEINE, D. and VAN BEVER, G. (2014). Inference on the shape of elliptical distributions based on the MCD. *J. Multivariate Anal.* **129** 125–144. MR3215984 https://doi.org/10.1016/j.jmva.2014.04.013

[52] PAN, W., KIM, J., ZHANG, Y., SHEN, X. and WEI, P. (2014). A powerful and adaptive association test for rare variants. *Genetics* **197** 1081–1095.

[53] PÉCHÉ, S. (2009). Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probab. Theory Related Fields* **143** 481–516. MR2475670 https://doi.org/10.1007/s00440-007-0133-7

[54] PENG, Z. and NADARAJAH, S. (2003). On the joint limiting distribution of sums and maxima of stationary normal sequence. *Theory Probab. Appl.* **47** 706–709.

[55] PHAM, T. D. and TRAN, L. T. (1985). Some mixing properties of time series models. *Stochastic Process. Appl.* **19** 297–303. MR0787587 https://doi.org/10.1016/0304-4149(85)90031-6

[56] PRINCE, M., BRYCE, R., ALBANESE, E., WIMO, A., RIBEIRO, W. and FERRI, C. P. (2013). The global prevalence of dementia: A systematic review and metaanalysis. *Alzheimer's Dement.* **9** 63–75.

[57] SCHOTT, J. R. (2007). A test for the equality of covariance matrices when the dimension is large relative to the sample sizes. *Comput. Statist. Data Anal.* **51** 6535–6542. MR2408613 https://doi.org/10.1016/j.csda.2007.03.004

[58] SHAO, Q.-M. and ZHOU, W.-X. (2014). Necessary and sufficient conditions for the asymptotic distributions of coherence of ultra-high dimensional random matrices. *Ann. Probab.* **42** 623–648. MR3178469 https://doi.org/10.1214/13-AOP837

[59] SOSHNIKOV, A. (2002). A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Stat. Phys.* **108** 1033–1056. MR1933444 https://doi.org/10.1023/A:1019739414239

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

[61] SRIVASTAVA, M. S. and YANAGIHARA, H. (2010). Testing the equality of several covariance matrices with fewer observations than the dimension. *J. Multivariate Anal.* **101** 1319–1329. MR2609494 https://doi.org/10.1016/j.jmva.2009.12.010

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

[63] STOREY, J. D. and TIBSHIRANI, R. (2003). Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100** 9440–9445. MR1994856 https://doi.org/10.1073/pnas.1530509100

[64] WANG, L., JIA, P., WOLFINGER, R. D., CHEN, X. and ZHAO, Z. (2011). Gene set analysis of genome-wide association studies: Methodological issues and perspectives. *Genomics* **98** 1–8.

[65] WU, C., XU, G. and PAN, W. (2019). An adaptive test on high-dimensional parameters in generalized linear models. *Statist. Sinica* **29** 2163–2186. MR3970351

[66] XIA, Y., CAI, T. and CAI, T. T. (2015). Testing differential networks with applications to the detection of gene-gene interactions. *Biometrika* **102** 247–266. MR3371002 https://doi.org/10.1093/biomet/asu074

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

[68] Xu, J., Murphy, S. L., Kochanek, K. D., Bastian, B. and Arias, E. (2018). Deaths: Final data for 2016. *Natl*. *Vital Stat*. *Rep*. **67** 1–76.

[69] Xu, Z., Xu, G. and Pan, W. (2017). Adaptive testing for association between two random vectors in moderate to high dimensions. *Genet*. *Epidemiol*. **41** 599–609.

[70] 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

[71] Yu, K., Li, Q., Bergen, A. W., Pfeiffer, R. M., Rosenberg, P. S., Caporaso, N., Kraft, P. and Chatterjee, N. (2009). Pathway analysis by adaptive combination of p-values. *Genet*. *Epidemiol*. **33** 700–709.

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

[73] Zhong, P.-S., Chen, S. X. and Xu, M. (2013). Tests alternative to higher criticism for high-dimensional means under sparsity and column-wise dependence. *Ann*. *Statist*. **41** 2820–2851. MR3161449 https://doi.org/10.1214/13-AOS1168

The Annals of Statistics
2021, Vol. 49, No. 1, 182–202
https://doi.org/10.1214/20-AOS1952

FREQUENTIST VALIDITY OF BAYESIAN LIMITS

BY B. J. K. KLEIJN[1]

[1] *Korteweg–de Vries Institute for Mathematics, University of Amsterdam, B.Kleijn@uva.nl*

To the frequentist who computes posteriors, not all priors are useful asymptotically: in this paper, a Bayesian perspective on test sequences is proposed and Schwartz's Kullback–Leibler condition is generalised to widen the range of frequentist applications of posterior convergence. With *Bayesian tests* and a weakened form of contiguity termed *remote contiguity*, we prove simple and fully general frequentist theorems, for posterior consistency and rates of convergence, for consistency of posterior odds in model selection, and for conversion of sequences of credible sets into sequences of confidence sets with asymptotic coverage one. For frequentist uncertainty quantification, this means that a prior inducing remote contiguity allows one to enlarge credible sets of calculated, simulated or approximated posteriors to obtain asymptotically consistent confidence sets.

REFERENCES

[1] 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

[2] BARRON, A. (1988). The exponential convergence of posterior probabilities with implications for Bayes estimators of density functions. Technical Report 7, Dept. Statistics, Univ. Illinois.

[3] BARRON, A., SCHERVISH, M. J. and WASSERMAN, L. (1999). The consistency of posterior distributions in nonparametric problems. *Ann*. *Statist*. **27** 536–561. MR1714718 https://doi.org/10.1214/aos/1018031206

[4] BAYARRI, M. J. and BERGER, J. O. (2004). The interplay of Bayesian and frequentist analysis. *Statist*. *Sci*. **19** 58–80. MR2082147 https://doi.org/10.1214/088342304000000116

[5] BIRGÉ, L. (1983). Approximation dans les espaces métriques et théorie de l'estimation. *Z*. *Wahrsch*. *Verw*. *Gebiete* **65** 181–237. MR0722129 https://doi.org/10.1007/BF00532480

[6] BIRGÉ, L. and MASSART, P. (1997). From model selection to adaptive estimation. In *Festschrift for Lucien Le Cam* 55–87. Springer, New York. MR1462939

[7] BIRGÉ, L. and MASSART, P. (2001). Gaussian model selection. *J*. *Eur*. *Math*. *Soc*. (*JEMS*) **3** 203–268. MR1848946 https://doi.org/10.1007/s100970100031

[8] BREIMAN, L., LE CAM, L. and SCHWARTZ, L. (1964). Consistent estimates and zero–one sets. *Ann*. *Math*. *Stat*. **35** 157–161. MR0161413 https://doi.org/10.1214/aoms/1177703737

[9] 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

[10] COX, D. D. (1993). An analysis of Bayesian inference for nonparametric regression. *Ann*. *Statist*. **21** 903–923. MR1232525 https://doi.org/10.1214/aos/1176349157

[11] DOOB, J. L. (1949). Application of the theory of martingales. In *Le Calcul des Probabilités et Ses Applications*. *Colloques Internationaux du Centre National de la Recherche Scientifique* **13** 23–27. Centre National de la Recherche Scientifique, Paris. MR0033460

[12] FALK, M., PADOAN, S. and RIZZELLI, S. (2019). Strong convergence of multivariate maxima. Preprint. Available at arXiv:1903.10596.

[13] FREEDMAN, D. (1999). On the Bernstein–von Mises theorem with infinite-dimensional parameters. *Ann*. *Statist*. **27** 1119–1140. MR1740119 https://doi.org/10.1214/aos/1017938917

[14] GHOSAL, S., GHOSH, J. K. and VAN DER VAART, A. W. (2000). Convergence rates of posterior distributions. *Ann*. *Statist*. **28** 500–531. MR1790007 https://doi.org/10.1214/aos/1016218228

MSC2020 subject classifications. Primary 62G05, 62G20; secondary 62G10, 62C10, 62B15.
Key words and phrases. Contiguity, frequentist consistency, posterior consistency, posterior rate of convergence, posterior odds, model selection, credible set, coverage, uncertainty quantification.

[15] 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

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

[17] GREENWOOD, P. E. and SHIRYAYEV, A. N. (1985). *Contiguity and the Statistical Invariance Principle*. *Stochastics Monographs* **1**. Gordon & Breach, New York. MR0822226

[18] HENGARTNER, N. W. and STARK, P. B. (1995). Finite-sample confidence envelopes for shape-restricted densities. *Ann*. *Statist*. **23** 525–550. MR1332580 https://doi.org/10.1214/aos/1176324534

[19] JOHNSON, V. E. and ROSSELL, D. (2010). On the use of non-local prior densities in Bayesian hypothesis tests. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **72** 143–170. MR2830762 https://doi.org/10.1111/j.1467-9868.2009.00730.x

[20] KLEIJN, B. The frequentist theory of Bayesian statistics, Ch. 8, Springer, Berlin (in preparation).

[21] KLEIJN, B. and VAN WAAIJ, J. (2018). Recovery and confidence sets of communities in a sparse stochastic block model. Preprint. Available at arXiv:1810.09533.

[22] KLEIJN, B. J. K. (2021). Supplement to "Frequentist validity of Bayesian limits." https://doi.org/10.1214/20-AOS1952SUPP

[23] KLEIJN, B. J. K. and ZHAO, Y. Y. (2019). Criteria for posterior consistency and convergence at a rate. *Electron*. *J*. *Stat*. **13** 4709–4742. MR4033683 https://doi.org/10.1214/19-EJS1633

[24] LE CAM, L. (1960). Locally asymptotically normal families of distributions. Certain approximations to families of distributions and their use in the theory of estimation and testing hypotheses. *Univ*. *Calif*. *Publ*. *Statist*. **3** 37–98. MR0126903

[25] LE CAM, L. (1972). Limits of experiments. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (*Univ*. *California*, *Berkeley*, *Calif*., 1970/1971), *Vol*. *I*: *Theory of Statistics* 245–261. Univ. California Press, Berkeley, CA. MR0415819

[26] LE CAM, L. (1979). An inequality concerning Bayes estimates. Preprint, Univ. California, Berkeley, CA.

[27] LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. *Springer Series in Statistics*. Springer, New York. MR0856411 https://doi.org/10.1007/978-1-4612-4946-7

[28] LE CAM, L. and YANG, G. L. (1988). On the preservation of local asymptotic normality under information loss. *Ann*. *Statist*. **16** 483–520. MR0947559 https://doi.org/10.1214/aos/1176350817

[29] LE CAM, L. and YANG, G. L. (1990). *Asymptotics in Statistics*: *Some Basic Concepts*. *Springer Series in Statistics*. Springer, New York. MR1066869 https://doi.org/10.1007/978-1-4684-0377-0

[30] LECAM, L. (1973). Convergence of estimates under dimensionality restrictions. *Ann*. *Statist*. **1** 38–53. MR0334381

[31] LOW, M. G. (1997). On nonparametric confidence intervals. *Ann*. *Statist*. **25** 2547–2554. MR1604412 https://doi.org/10.1214/aos/1030741084

[32] MOSSEL, E., NEEMAN, J. and SLY, A. (2016). Consistency thresholds for the planted bisection model. *Electron*. *J*. *Probab*. **21** Art. ID 21. MR3485363 https://doi.org/10.1214/16-EJP4185

[33] PADOAN, S. and RIZZELLI, S. (2019). Strong consistency of nonparametric Bayesian inferential methods for multivariate max-stable distributions. Preprint. Available at arXiv:1904.00245.

[34] SCHWARTZ, L. (1965). On Bayes procedures. *Z*. *Wahrsch*. *Verw*. *Gebiete* **4** 10–26. MR0184378 https://doi.org/10.1007/BF00535479

[35] SHEN, X. and WASSERMAN, L. (2001). Rates of convergence of posterior distributions. *Ann*. *Statist*. **29** 687–714. MR1865337 https://doi.org/10.1214/aos/1009210686

[36] SZABÓ, B., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2015). Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Ann*. *Statist*. **43** 1391–1428. MR3357861 https://doi.org/10.1214/14-AOS1270

[37] TAYLOR, J. and TIBSHIRANI, R. J. (2015). Statistical learning and selective inference. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **112** 7629–7634. MR3371123 https://doi.org/10.1073/pnas.1507583112

[38] TOKDAR, S. T., ZHU, Y. M. and GHOSH, J. K. (2010). Bayesian density regression with logistic Gaussian process and subspace projection. *Bayesian Anal*. **5** 319–344. MR2719655 https://doi.org/10.1214/10-BA605

[39] TORGERSEN, E. (1991). *Comparison of Statistical Experiments*. *Encyclopedia of Mathematics and Its Applications* **36**. Cambridge Univ. Press, Cambridge. MR1104437 https://doi.org/10.1017/CBO9780511666353

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

[41] WALKER, S. (2004). New approaches to Bayesian consistency. *Ann*. *Statist*. **32** 2028–2043. MR2102501 https://doi.org/10.1214/009053604000000409

[42] WALKER, S. G., LIJOI, A. and PRÜNSTER, I. (2007). On rates of convergence for posterior distributions in infinite-dimensional models. *Ann. Statist.* **35** 738–746. MR2336866 https://doi.org/10.1214/009053606000001361

[43] WASSERMAN, L. (2000). Bayesian model selection and model averaging. *J. Math. Psych.* **44** 92–107. MR1770003 https://doi.org/10.1006/jmps.1999.1278

[44] WONG, W. H. and SHEN, X. (1995). Probability inequalities for likelihood ratios and convergence rates of sieve MLEs. *Ann. Statist.* **23** 339–362. MR1332570 https://doi.org/10.1214/aos/1176324524

The Annals of Statistics
2021, Vol. 49, No. 1, 203–232
https://doi.org/10.1214/20-AOS1953

OPTIMAL CHANGE POINT DETECTION AND LOCALIZATION IN SPARSE DYNAMIC NETWORKS

BY DAREN WANG[1], YI YU[2] AND ALESSANDRO RINALDO[3]

[1]*Department of Statistics, University of Chicago, darenw@uchicago.edu*
[2]*Department of Statistics, University of Warwick, yi.yu.2@warwick.ac.uk*
[3]*Department of Statistics and Data Science, Carnegie Mellon University, arinaldo@cmu.edu*

We study the problem of change point localization in dynamic networks models. We assume that we observe a sequence of independent adjacency matrices of the same size, each corresponding to a realization of an unknown inhomogeneous Bernoulli model. The underlying distribution of the adjacency matrices are piecewise constant, and may change over a subset of the time points, called change points. We are concerned with recovering the unknown number and positions of the change points. In our model setting, we allow for all the model parameters to change with the total number of time points, including the network size, the minimal spacing between consecutive change points, the magnitude of the smallest change and the degree of sparsity of the networks. We first identify a region of impossibility in the space of the model parameters such that no change point estimator is provably consistent if the data are generated according to parameters falling in that region. We propose a computationally-simple algorithm for network change point localization, called network binary segmentation, that relies on weighted averages of the adjacency matrices. We show that network binary segmentation is consistent over a range of the model parameters that nearly cover the complement of the impossibility region, thus demonstrating the existence of a phase transition for the problem at hand. Next, we devise a more sophisticated algorithm based on singular value thresholding, called local refinement, that delivers more accurate estimates of the change point locations. Under appropriate conditions, local refinement guarantees a minimax optimal rate for network change point localization while remaining computationally feasible.

REFERENCES

ANASTASIOU, A. and FRYZLEWICZ, P. (2019). Detecting multiple generalized change-points by isolating single ones. Preprint. Available at arXiv:1901.10852.

BARABÁSI, A.-L. and ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. MR2091634 https://doi.org/10.1126/science.286.5439.509

BARANOWSKI, R., CHEN, Y. and FRYZLEWICZ, P. (2019). Narrowest-over-threshold detection of multiple change points and change-point-like features. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 649–672. MR3961502 https://doi.org/10.1111/rssb.12322

BHATTACHARJEE, M., BANERJEE, M. and MICHAILIDIS, G. (2018). Change point estimation in a dynamic stochastic block model. Preprint. Available at arXiv:1812.03090.

BHATTACHARYYA, S. and CHATTERJEE, S. (2017). Spectral clustering for dynamic stochastic block model. Technical report, Working paper.

BOCCALETTI, S., BIANCONI, G., CRIADO, R., DEL GENIO, C. I., GÓMEZ-GARDEÑES, J., ROMANCE, M., SENDIÑA-NADAL, I., WANG, Z. and ZANIN, M. (2014). The structure and dynamics of multilayer networks. *Phys. Rep.* **544** 1–122. MR3270140 https://doi.org/10.1016/j.physrep.2014.07.001

CARRINGTON, P. J., SCOTT, J. and WASSERMAN, S. (2005). *Models and Methods in Social Network Analysis* **28**. Cambridge Univ. Press, Cambridge.

MSC2020 subject classifications. Primary 62M10; secondary 91B84.
Key words and phrases. Change point detection, low-rank networks, stochastic block model, minimax optimality.

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

CHEN, K. and LEI, J. (2018). Network cross-validation for determining the number of communities in network data. *J. Amer. Statist. Assoc.* **113** 241–251. MR3803461 https://doi.org/10.1080/01621459.2016.1246365

CHO, H. (2016). Change-point detection in panel data via double CUSUM statistic. *Electron. J. Stat.* **10** 2000–2038. MR3522667 https://doi.org/10.1214/16-EJS1155

CHO, H. and FRYZLEWICZ, P. (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 475–507. MR3310536 https://doi.org/10.1111/rssb.12079

CHU, L. and CHEN, H. (2017). Asymptotic distribution-free change-point detection for modern data. Preprint. Available at arXiv:1707.00167.

CRANE, H. (2015). Time-varying network models. *Bernoulli* **21** 1670–1696. MR3352057 https://doi.org/10.3150/14-BEJ617

CRIBBEN, I. and YU, Y. (2017). Estimating whole-brain dynamics by using spectral clustering. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **66** 607–627. MR3632344 https://doi.org/10.1111/rssc.12169

EICHINGER, B. and KIRCH, C. (2018). A MOSUM procedure for the estimation of multiple random change points. *Bernoulli* **24** 526–564. MR3706768 https://doi.org/10.3150/16-BEJ887

ERDŐS, P. and RÉNYI, A. (1959). On random graphs. I. *Publ. Math. Debrecen* **6** 290–297. MR0120167

FRYZLEWICZ, P. (2014). Wild binary segmentation for multiple change-point detection. *Ann. Statist.* **42** 2243–2281. MR3269979 https://doi.org/10.1214/14-AOS1245

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

GOLDENBERG, A., ZHENG, A. X., FIENBERG, S. E. and AIROLDI, E. M. (2010). A survey of statistical network models. In *Foundations and Trends® in Machine Learning* 129–233.

HO, Q., SONG, L. and XING, E. (2011). Evolving cluster mixed-membership blockmodel for time-evolving networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics* 342–350.

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

KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* (3) **83** 016107. MR2788206 https://doi.org/10.1103/PhysRevE.83.016107

KESHAVARZ, H., MICHAILIDIS, G. and ATCHADE, Y. (2018). Sequential change-point detection in high-dimensional Gaussian graphical models. Preprint. Available at arXiv:1806.07870.

KOLACZYK, E. D. (2017). *Topics at the Frontier of Statistics and Network Analysis*: (*Re*)*Visiting the Foundations*. *SemStat Elements*. Cambridge Univ. Press, Cambridge. MR3702038 https://doi.org/10.1017/9781108290159

LIU, F., CHOI, D., XIE, L. and ROEDER, K. (2018). Global spectral clustering in dynamic networks. *Proc. Natl. Acad. Sci. USA* **115** 927–932. MR3763702 https://doi.org/10.1073/pnas.1718449115

LOH, P.-L. and WAINWRIGHT, M. J. (2013). Regularized M-estimators with nonconvexity: Statistical and algorithmic theory for local optima. In *Advances in Neural Information Processing Systems* 476–484.

MATIAS, C. and MIELE, V. (2017). Statistical clustering of temporal networks through a dynamic stochastic block model. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1119–1141. MR3689311 https://doi.org/10.1111/rssb.12200

MUKHERJEE, S. S. (2018). On some inference problems for networks. Ph.D. thesis. MR3885569

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

PENSKY, M. (2019). Dynamic network models and graphon estimation. *Ann. Statist.* **47** 2378–2403. MR3953455 https://doi.org/10.1214/18-AOS1751

PENSKY, M. and ZHANG, T. (2019). Spectral clustering in the dynamic stochastic block model. *Electron. J. Stat.* **13** 678–709. MR3914178 https://doi.org/10.1214/19-ejs1533

SEWELL, D. K. and CHEN, Y. (2015). Latent space models for dynamic networks. *J. Amer. Statist. Assoc.* **110** 1646–1657. MR3449061 https://doi.org/10.1080/01621459.2014.988214

SNIJDERS, T. A. B. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *J. Soc. Struct.* **3** 1–40.

TANG, M., PARK, Y., LEE, N. H. and PRIEBE, C. E. (2013). Attribute fusion in a latent process model for time series of graphs. *IEEE Trans. Signal Process.* **61** 1721–1732. MR3038385 https://doi.org/10.1109/TSP.2013.2243445

VENKATRAMAN, E. S. (1992). Consistency results in multiple change-point problems. Ph.D. thesis. MR2687536

VOSTRIKOVA, L. J. (1981). Discovery of "discord" in multidimensional random processes. *Dokl. Akad. Nauk SSSR* **259** 270–274. MR0625215

WANG, T. and SAMWORTH, R. J. (2018). High dimensional change point estimation via sparse projection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 57–83. MR3744712 https://doi.org/10.1111/rssb.12243

WANG, D., YU, Y. and RINALDO, A. (2017). Optimal covariance change point detection in high dimension. Preprint. Available at arXiv:1712.09912.

WANG, D., YU, Y. and RINALDO, A. (2020). Univariate mean change point detection: Penalization, CUSUM and optimality. *Electron. J. Stat.* **14** 1917–1961. MR4091859

WANG, D., YU, Y. and RINALDO, A. (2021). Supplement to "Optimal change point detection and localization in sparse dynamic networks." https://doi.org/10.1214/20-AOS1953SUPP

WANG, H., TANG, M., PARK, Y. and PRIEBE, C. E. (2014). Locality statistics for anomaly detection in time series of graphs. *IEEE Trans. Signal Process.* **62** 703–717. MR3160307 https://doi.org/10.1109/TSP.2013.2294594

XU, K. (2015). Stochastic block transition models for dynamic networks. In *Artificial Intelligence and Statistics* 1079–1087.

XU, J. (2018). Rates of convergence of spectral methods for graphon estimation. In *International Conference on Machine Learning* 5429–5438.

XU, K. S. and HERO, A. O. (2014). Dynamic stochastic blockmodels for time-evolving social networks. *IEEE J. Sel. Top. Signal Process.* **8** 552–562.

XU, A. and ZHENG, X. (2009). Dynamic social network analysis using latent space model and an integrated clustering algorithm. In *Dependable, Autonomic and Secure Computing*, 2009. *DASC'*09. *Eighth IEEE International Conference on* 620–625.

YOUNG, S. J. and SCHEINERMAN, E. R. (2007). Random dot product graph models for social networks. In *Algorithms and Models for the Web-Graph. Lecture Notes in Computer Science* **4863** 138–149. Springer, Berlin. MR2504912 https://doi.org/10.1007/978-3-540-77004-6_11

YU, B. (1997). Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam* 423–435. Springer, New York. MR1462963

ZHANG, Y., WAINWRIGHT, M. J. and DUCHI, J. C. (2012). Communication-efficient algorithms for statistical optimization. In *Advances in Neural Information Processing Systems* 1502–1510.

ZHAO, Z., CHEN, L. and LIN, L. (2019). Change-point detection in dynamic networks via graphon estimation. Preprint. Available at arXiv:1908.01823.

The Annals of Statistics
2021, Vol. 49, No. 1, 233–254
https://doi.org/10.1214/20-AOS1954

CONVERGENCE OF COVARIANCE AND SPECTRAL DENSITY ESTIMATES FOR HIGH-DIMENSIONAL LOCALLY STATIONARY PROCESSES

BY DANNA ZHANG[1] AND WEI BIAO WU[2]

[1]*Department of Mathematics, University of California, San Diego, daz076@ucsd.edu*
[2]*Department of Statistics, University of Chicago, wbwu@galton.uchicago.edu*

Covariances and spectral density functions play a fundamental role in the theory of time series. There is a well-developed asymptotic theory for their estimates for low-dimensional stationary processes. For high-dimensional non-stationary processes, however, many important problems on their asymptotic behaviors are still unanswered. This paper presents a systematic asymptotic theory for the estimates of time-varying second-order statistics for a general class of high-dimensional locally stationary processes. Using the framework of functional dependence measure, we derive convergence rates of the estimates which depend on the sample size T, the dimension p, the moment condition and the dependence of the underlying processes.

REFERENCES

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

ALEXOPOULOS, C. and GOLDSMAN, D. (2004). To batch or not to batch? *ACM Trans. Model. Comput. Simul.* **14** 76–114.

AN, H. Z., CHEN, Z. G. and HANNAN, E. J. (1982). Autocorrelation, autoregression and autoregressive approximation. *Ann. Statist.* **10** 926–936. MR0663443

ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York. MR0283939

BACCALÁ, L. A. and SAMESHIMA, K. (2001). Partial directed coherence: A new concept in neural structure determination. *Biol. Cybernet.* **84** 463–474.

BACH, F. R. and JORDAN, M. I. (2004). Learning graphical models for stationary time series. *IEEE Trans. Signal Process.* **52** 2189–2199. MR2085580 https://doi.org/10.1109/TSP.2004.831032

BARIGOZZI, M., HALLIN, M., SOCCORSI, S. and VON SACHS, R. (2019). Time-varying general dynamic factor models and the measurement of financial connectedness. Available at SSRN 3329445.

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

BICKEL, P. J. and LEVINA, E. (2008). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. MR2485008 https://doi.org/10.1214/08-AOS600

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

BLINOWSKA, K. J. (2011). Review of the methods of determination of directed connectivity from multichannel data. *Med. Biol. Eng. Comput.* **49** 521–529.

BRILLINGER, D. R. (1975). *Time Series: Data Analysis and Theory*. Holt, Rinehart, and Winston, New York. MR0443257

BRILLINGER, D. R. (1996). Remarks concerning graphical models for time series and point processes. *Revista de Econometria* **16** 23.

BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR1093459 https://doi.org/10.1007/978-1-4419-0320-4

BRYC, W. and DEMBO, A. (1997). Large deviations for quadratic functionals of Gaussian processes. *J. Theoret. Probab.* **10** 307–332. MR1455147 https://doi.org/10.1023/A:1022656331883

BÜHLMANN, P. (2002). Bootstraps for time series. *Statist. Sci.* **17** 52–72. MR1910074 https://doi.org/10.1214/ss/1023798998

MSC2020 subject classifications. Primary 62M10; secondary 62M15.

Key words and phrases. High-dimensional time series, locally stationary processes, second-order statistics, convergence rate, Hanson–Wright-type inequalities.

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

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

CHEN, X., XU, M. and WU, W. B. (2013). Covariance and precision matrix estimation for high-dimensional time series. *Ann. Statist.* **41** 2994–3021. MR3161455 https://doi.org/10.1214/13-AOS1182

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

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

DAHLHAUS, R. (2000b). Graphical interaction models for multivariate time series. *Metrika* **51** 157–172. MR1790930 https://doi.org/10.1007/s001840000055

DAHLHAUS, R. (2012). Locally stationary processes. *Handb. Statist.* **30** 351–412.

DAHLHAUS, R. and POLONIK, W. (2009). Empirical spectral processes for locally stationary time series. *Bernoulli* **15** 1–39. MR2546797 https://doi.org/10.3150/08-BEJ137

DAHLHAUS, R., RICHTER, S. and WU, W. B. (2019). Towards a general theory for nonlinear locally stationary processes. *Bernoulli* **25** 1013–1044. MR3920364 https://doi.org/10.3150/17-bej1011

DAHLHAUS, R. and SUBBA RAO, S. (2006). Statistical inference for time-varying ARCH processes. *Ann. Statist.* **34** 1075–1114. MR2278352 https://doi.org/10.1214/009053606000000227

DAHLHAUS, R. and SUBBA RAO, S. (2007). A recursive online algorithm for the estimation of time-varying ARCH parameters. *Bernoulli* **13** 389–422. MR2331257 https://doi.org/10.3150/07-BEJ5009

EICHLER, M. (2007). Granger causality and path diagrams for multivariate time series. *J. Econometrics* **137** 334–353. MR2354948 https://doi.org/10.1016/j.jeconom.2005.06.032

EICHLER, M. (2012). Graphical modelling of multivariate time series. *Probab. Theory Related Fields* **153** 233–268. MR2925574 https://doi.org/10.1007/s00440-011-0345-8

EICHLER, M., DAHLHAUS, R. and SANDKÜHLER, J. (2003). Partial correlation analysis for the identification of synaptic connections. *Biol. Cybernet.* **89** 289–302. https://doi.org/10.1007/s00422-003-0400-3

FIECAS, M., LENG, C., LIU, W. and YU, Y. (2019). Spectral analysis of high-dimensional time series. *Electron. J. Stat.* **13** 4079–4101. MR4017528 https://doi.org/10.1214/19-EJS1621

FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Rev. Econ. Stat.* **82** 540–554.

FRIED, R. and DIDELEZ, V. (2003). Decomposability and selection of graphical models for multivariate time series. *Biometrika* **90** 251–267. MR1986645 https://doi.org/10.1093/biomet/90.2.251

FRYZLEWICZ, P. and SUBBA RAO, S. (2011). Mixing properties of ARCH and time-varying ARCH processes. *Bernoulli* **17** 320–346. MR2797994 https://doi.org/10.3150/10-BEJ270

GATHER, U., IMHOFF, M. and FRIED, R. (2002). Graphical models for multivariate time series from intensive care monitoring. *Stat. Med.* **21** 2685–2701.

GIURCANU, M. and SPOKOINY, V. (2004). Confidence estimation of the covariance function of stationary and locally stationary processes. *Statist. Decisions* **22** 283–300. MR2158265 https://doi.org/10.1524/stnd.22.4.283.64315

GRENIER, Y. (1983). Time-dependent ARMA modeling of nonstationary signals. *IEEE Trans. Acoust. Speech Signal Process.* **31** 899–911.

HAFNER, C. M. and LINTON, O. (2010). Efficient estimation of a multivariate multiplicative volatility model. *J. Econometrics* **159** 55–73. MR2720844 https://doi.org/10.1016/j.jeconom.2010.04.007

HANNAN, E. J. (1974). The uniform convergence of autocovariances. *Ann. Statist.* **2** 803–806. MR0365961

HANNAN, E. J. and DEISTLER, M. (1988). *The Statistical Theory of Linear Systems*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York. MR0940698

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

HOOTEN, M. B. and WIKLE, C. K. (2008). A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian collared-dove. *Environ. Ecol. Stat.* **15** 59–70. MR2412679 https://doi.org/10.1007/s10651-007-0040-1

JACQUIER, E., POLSON, N. G. and ROSSI, P. E. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *J. Econometrics* **122** 185–212. MR2083256 https://doi.org/10.1016/j.jeconom.2003.09.001

JIRAK, M. (2011). On the maximum of covariance estimators. *J. Multivariate Anal.* **102** 1032–1046. MR2793874 https://doi.org/10.1016/j.jmva.2011.02.003

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

KONDRASHOV, D., KRAVTSOV, S., ROBERTSON, A. W. and GHIL, M. (2005). A hierarchy of data-based ENSO models. *J. Climate* **18** 4425–4444.

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

LENNARTZ, C., SCHIEFER, J., ROTTER, S., HENNIG, J. and LEVAN, P. (2018). Sparse estimation of resting-state effective connectivity from fMRI cross-spectra. *Front*. *Neurosci*. **12** 287. https://doi.org/10.3389/fnins.2018.00287

LINDQUIST, M. A., XU, Y., NEBEL, M. B. and CAFFO, B. S. (2014). Evaluating dynamic bivariate correlations in resting-state fMRI: A comparison study and a new approach. *NeuroImage* **101** 531–546.

LIU, C., GAETZ, W. and ZHU, H. (2010). Estimation of time-varying coherence and its application in understanding brain functional connectivity. *EURASIP Journal on Advances in Signal Processing* **2010**.

MAJDA, A. J., ABRAMOV, R. V. and GROTE, M. J. (2005). *Information Theory and Stochastics for Multiscale Nonlinear Systems*. *CRM Monograph Series* **25**. Amer. Math. Soc., Providence, RI. MR2166171

MALLAT, S., PAPANICOLAOU, G. and ZHANG, Z. (1998). Adaptive covariance estimation of locally stationary processes. *Ann*. *Statist*. **26** 1–47. MR1611808 https://doi.org/10.1214/aos/1030563977

MEDKOUR, T., WALDEN, A. T. and BURGESS, A. (2009). Graphical modelling for brain connectivity via partial coherence. *J. Neurosci*. *Methods* **180** 374–383.

MOULINES, E., PRIOURET, P. and ROUEFF, F. (2005). On recursive estimation for time varying autoregressive processes. *Ann*. *Statist*. **33** 2610–2654. MR2253097 https://doi.org/10.1214/009053605000000624

NASON, G. P., VON SACHS, R. and KROISANDT, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R*. *Stat*. *Soc*. *Ser. B*. *Stat*. *Methodol*. **62** 271–292. MR1749539 https://doi.org/10.1111/1467-9868.00231

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

OMBAO, H. and VAN BELLEGEM, S. (2008). Evolutionary coherence of nonstationary signals. *IEEE Trans*. *Signal Process*. **56** 2259–2266. MR2516630 https://doi.org/10.1109/TSP.2007.914341

OMBAO, H., VON SACHS, R. and GUO, W. (2005). SLEX analysis of multivariate nonstationary time series. *J. Amer. Statist*. *Assoc*. **100** 519–531. MR2160556 https://doi.org/10.1198/016214504000001448

OMBAO, H. C., RAZ, J. A., VON SACHS, R. and MALOW, B. A. (2001). Automatic statistical analysis of bivariate nonstationary time series. *J. Amer. Statist*. *Assoc*. **96** 543–560. MR1946424 https://doi.org/10.1198/016214501753168244

PAPARODITIS, E. and POLITIS, D. N. (2012). Nonlinear spectral density estimation: Thresholding the correlogram. *J. Time Series Anal*. **33** 386–397. MR2915091 https://doi.org/10.1111/j.1467-9892.2011.00771.x

PARK, T., ECKLEY, I. A. and OMBAO, H. C. (2014). Estimating time-evolving partial coherence between signals via multivariate locally stationary wavelet processes. *IEEE Trans*. *Signal Process*. **62** 5240–5250. MR3268108 https://doi.org/10.1109/TSP.2014.2343937

POLITIS, D. N. (2011). Higher-order accurate, positive semidefinite estimation of large-sample covariance and spectral density matrices. *Econometric Theory* **27** 703–744. MR2822363 https://doi.org/10.1017/S0266466610000484

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

POLITIS, D. N. and ROMANO, J. P. (1999). Multivariate density estimation with general flat-top kernels of infinite order. *J. Multivariate Anal*. **68** 1–25. MR1668848 https://doi.org/10.1006/jmva.1998.1774

POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling*. *Springer Series in Statistics*. Springer, New York. MR1707286 https://doi.org/10.1007/978-1-4612-1554-7

PRADO, R., WEST, M. and KRYSTAL, A. D. (2001). Multichannel electroencephalographic analyses via dynamic regression models with time-varying lag–lead structure. *J. R*. *Stat*. *Soc*. *Ser. C*. *Appl*. *Stat*. **50** 95–109.

PRIESTLEY, M. B. (1981). *Spectral Analysis and Time Series*. *Vol*. 1. Academic Press, London. MR0628735

PRIESTLEY, M. B. (1988). *Nonlinear and Nonstationary Time Series Analysis*. Academic Press, London. MR0991969

PRIESTLEY, M. B. and TONG, H. (1973). On the analysis of bivariate non-stationary processes. *J. Roy. Statist*. *Soc*. *Ser. B* **35** 153–166, 179–188. MR0336946

ROSENBLATT, M. (1971). *Markov Processes*. *Structure and Asymptotic Behavior*. Springer, New York. MR0329037

ROSENBLATT, M. (1985). *Stationary Sequences and Random Fields*. Birkhäuser, Inc., Boston, MA. MR0885090 https://doi.org/10.1007/978-1-4612-5156-9

RUDELSON, M. and VERSHYNIN, R. (2013). Hanson–Wright inequality and sub-Gaussian concentration. *Electron*. *Commun*. *Probab*. **18** no. 82, 9. MR3125258 https://doi.org/10.1214/ECP.v18-2865

SALVADOR, R., SUCKLING, J., SCHWARZBAUER, C. and BULLMORE, E. (2005). Undirected graphs of frequency-dependent functional connectivity in whole brain networks. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **360** 937–946.

SANDERSON, J., FRYZLEWICZ, P. and JONES, M. W. (2010). Estimating linear dependence between non-stationary time series using the locally stationary wavelet model. *Biometrika* **97** 435–446. MR2650749 https://doi.org/10.1093/biomet/asq007

SIMPSON, S. L., BOWMAN, F. D. and LAURIENTI, P. J. (2013). Analyzing complex functional brain networks: Fusing statistics and network science to understand the brain. *Stat. Surv.* **7** 1–36. MR3161730 https://doi.org/10.1214/13-SS103

SUBBA RAO, T. (1970). The fitting of non-stationary time-series models with time-dependent parameters. *J. Roy. Statist. Soc. Ser. B* **32** 312–322. MR0269065

SUN, Y., LI, Y., KUCEYESKI, A. and BASU, S. (2018). Large spectral density matrix estimation by thresholding. Preprint. Available at arXiv:1812.00532.

TIMMER, J., LAUK, M., HÄUSSLER, S., RADT, V., KÖSTER, B., HELLWIG, B., GUSCHLBAUER, B., LÜCKING, C. H., EICHLER, M. et al. (2000). Cross-spectral analysis of tremor time series. *Internat. J. Bifur. Chaos Appl. Sci. Engrg.* **10** 2595–2610.

TONG, H. (1990). *Nonlinear Time Series: A Dynamical System Approach. Oxford Statistical Science Series* **6**. The Clarendon Press, Oxford University Press, New York. MR1079320

TSAY, R. S. (2005). *Analysis of Financial Time Series*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR2162112 https://doi.org/10.1002/0471746193

VOGT, M. (2012). Nonparametric regression for locally stationary time series. *Ann. Statist.* **40** 2601–2633. MR3097614 https://doi.org/10.1214/12-AOS1043

WIENER, N. (1958). *Nonlinear Problems in Random Theory. Technology Press Research Monographs*. Wiley, New York. MR0100912

WIKLE, C. K. and HOOTEN, M. B. (2010). A general science-based framework for dynamical spatio-temporal models. *TEST* **19** 417–451. MR2745992 https://doi.org/10.1007/s11749-010-0209-z

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

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

WU, W. B. and SHAO, X. (2004). Limit theorems for iterated random functions. *J. Appl. Probab.* **41** 425–436. MR2052582 https://doi.org/10.1239/jap/1082999076

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

XIAO, H. and WU, W. B. (2014). Portmanteau test and simultaneous inference for serial covariances. *Statist. Sinica* **24** 577–599. MR3235390

ZANI, M. (2002). Large deviations for quadratic forms of locally stationary processes. *J. Multivariate Anal.* **81** 205–228. MR1906377 https://doi.org/10.1006/jmva.2001.2003

ZHANG, D. and WU, W. B. (2017). Gaussian approximation for high dimensional time series. *Ann. Statist.* **45** 1895–1919. MR3718156 https://doi.org/10.1214/16-AOS1512

ZHANG, D. and WU, W. B (2021). Supplement to "Convergence of covariance and spectral density estimates for high-dimensional locally stationary processes." https://doi.org/10.1214/20-AOS1954SUPP.

ZHOU, Z. (2010). Nonparametric inference of quantile curves for nonstationary time series. *Ann. Statist.* **38** 2187–2217. MR2676887 https://doi.org/10.1214/09-AOS769

ZHOU, Z. and WU, W. B. (2009). Local linear quantile estimation for nonstationary time series. *Ann. Statist.* **37** 2696–2729. MR2541444 https://doi.org/10.1214/08-AOS636

The Annals of Statistics
2021, Vol. 49, No. 1, 255–271
https://doi.org/10.1214/20-AOS1955

WORDLENGTH ENUMERATOR FOR FRACTIONAL FACTORIAL DESIGNS

BY YU TANG[1] AND HONGQUAN XU[2]

[1] *School of Mathematical Sciences, Soochow University, ytang@suda.edu.cn*
[2] *Department of Statistics, University of California, Los Angeles, hqxu@stat.ucla.edu*

While the minimum aberration criterion is popular for selecting good designs with qualitative factors under an ANOVA model, the minimum β-aberration criterion is more suitable for selecting designs with quantitative factors under a polynomial model. In this paper, we propose the concept of wordlength enumerator to unify these two criteria. The wordlength enumerator is defined as an average similarity of contrasts among all possible pairs of runs. The wordlength enumerator is easy and fast to compute, and can be used to compare and rank designs efficiently. Based on the wordlength enumerator, we develop simple and fast methods for calculating both the generalized wordlength pattern and the β-wordlength pattern. We further obtain a lower bound of the wordlength enumerator for three-level designs and characterize the combinatorial structure of designs achieving the lower bound. Finally, we propose two methods for constructing supersaturated designs that have both generalized minimum aberration and minimum β-aberration.

REFERENCES

BETH, T., JUNGNICKEL, D. and LENZ, H. (1999). *Design Theory*, 2nd ed. Cambridge Univ. Press, Cambridge. MR0890103

BIRD, E. M. and STREET, D. J. (2018). Complete enumeration of all geometrically nonisomorphic three-level orthogonal arrays in 18 runs. *Australas. J. Combin.* **71** 336–350. MR3801269

CHENG, C. S. (2014). *Theory of Factorial Design: Single- and Multi-Stratum Experiments*. CRC Press/CRC, Boca Raton, FL.

CHENG, S.-W. and WU, C. F. J. (2001). Factor screening and response surface exploration. *Statist. Sinica* **11** 553–604. MR1863152

CHENG, S.-W. and YE, K. Q. (2004). Geometric isomorphism and minimum aberration for factorial designs with quantitative factors. *Ann. Statist.* **32** 2168–2185. MR2102507 https://doi.org/10.1214/009053604000000599

CHIPMAN, H., HAMADA, M. and WU, C. F. J. (1997). A Bayesian variable-selection approach for analyzing designed experiments with complex aliasing. *Technometrics* **39** 372–381.

DE LAUNEY, W. (1986). A survey of generalised Hadamard matrices and difference matrices $D(k, \lambda; G)$ with large k. *Util. Math.* **30** 5–29. MR0864807

DRAPER, N. R. and SMITH, H. (1998). *Applied Regression Analysis*. Wiley, New York. MR0211537

GEORGIOU, S. D. (2014). Supersaturated designs: A review of their construction and analysis. *J. Statist. Plann. Inference* **144** 92–109. MR3131365 https://doi.org/10.1016/j.jspi.2012.09.014

HUANG, C., LIN, D. K. J. and LIU, M.-Q. (2012). An optimality criterion for supersaturated designs with quantitative factors. *J. Statist. Plann. Inference* **142** 1780–1788. MR2903388 https://doi.org/10.1016/j.jspi.2012.02.003

JOSEPH, V. R. and DELANEY, J. D. (2007). Functionally induced priors for the analysis of experiments. *Technometrics* **49** 1–11. MR2345447 https://doi.org/10.1198/004017006000000372

LIN, D. K. J. (1993). A new class of supersaturated designs. *Technometrics* **35** 28–31.

LIN, C.-Y. (2014). Optimal blocked orthogonal arrays. *J. Statist. Plann. Inference* **145** 139–147. MR3125355 https://doi.org/10.1016/j.jspi.2013.08.016

LIN, C.-Y. and CHENG, S.-W. (2012). Isomorphism examination based on the count vector. *Statist. Sinica* **22** 1253–1272. MR2987491 https://doi.org/10.5705/ss.2010.246

LIN, C.-Y., YANG, P. and CHENG, S.-W. (2017). Minimum contamination and β-aberration criteria for screening quantitative factors. *Statist. Sinica* **27** 607–623. MR3674688

MSC2020 subject classifications. 62K15.

Key words and phrases. Factor screening, generalized Hadamard matrix, generalized minimum aberration, generalized wordlength pattern, supersaturated design.

MOON, H., DEAN, A. M. and SANTNER, T. J. (2012). Two-stage sensitivity-based group screening in computer experiments. *Technometrics* **54** 376–387. MR3006387 https://doi.org/10.1080/00401706.2012.725994

MUKERJEE, R. and WU, C. F. J. (1995). On the existence of saturated and nearly saturated asymmetrical orthogonal arrays. *Ann*. *Statist*. **23** 2102–2115. MR1389867 https://doi.org/10.1214/aos/1034713649

MUKERJEE, R. and WU, C. F. J. (2006). *A Modern Theory of Factorial Designs*. *Springer Series in Statistics*. Springer, New York. MR2230487

PANG, F. and LIU, M.-Q. (2011). Geometric isomorphism check for symmetric factorial designs. *J*. *Complexity* **27** 441–448. MR2805530 https://doi.org/10.1016/j.jco.2011.04.001

SABBAGHI, A., DASGUPTA, T. and WU, C. F. J. (2014). Indicator functions and the algebra of the linear-quadratic parameterization. *Biometrika* **101** 351–363. MR3215352 https://doi.org/10.1093/biomet/ast070

SLOANE, N. J. A. (2019). Homepage of "A library of orthogonal arrays." Available at http://neilsloane.com/oadir/.

SUEN, C., DAS, A. and MIDHA, C. K. (2013). Optimal fractional factorial designs and their construction. *J*. *Statist*. *Plann*. *Inference* **143** 1828–1834. MR3082237 https://doi.org/10.1016/j.jspi.2013.05.004

SUN, F., LIN, D. K. J. and LIU, M.-Q. (2011). On construction of optimal mixed-level supersaturated designs. *Ann*. *Statist*. **39** 1310–1333. MR2816356 https://doi.org/10.1214/11-AOS877

TANG, B. and DENG, L.-Y. (1999). Minimum G_2-aberration for nonregular fractional factorial designs. *Ann*. *Statist*. **27** 1914–1926. MR1765622 https://doi.org/10.1214/aos/1017939244

TANG, Y. and XU, H. (2014). Permuting regular fractional factorial designs for screening quantitative factors. *Biometrika* **101** 333–350. MR3215351 https://doi.org/10.1093/biomet/ast073

WU, C.-F. J. (1993). Construction of supersaturated designs through partially aliased interactions. *Biometrika* **80** 661–669. MR1248029 https://doi.org/10.1093/biomet/80.3.661

WU, C. F. J. and HAMADA, M. S. (2009). *Experiments*: *Planning*, *Analysis*, *and Optimization*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2583259

XU, H. (2003). Minimum moment aberration for nonregular designs and supersaturated designs. *Statist*. *Sinica* **13** 691–708. MR1997169

XU, H., PHOA, F. K. H. and WONG, W. K. (2009). Recent developments in nonregular fractional factorial designs. *Stat*. *Surv*. **3** 18–46. MR2520978 https://doi.org/10.1214/08-SS040

XU, H. and WU, C. F. J. (2001). Generalized minimum aberration for asymmetrical fractional factorial designs. *Ann*. *Statist*. **29** 1066–1077. MR1869240 https://doi.org/10.1214/aos/1013699993

XU, H. and WU, C. F. J. (2005). Construction of optimal multi-level supersaturated designs. *Ann*. *Statist*. **33** 2811–2836. MR2253103 https://doi.org/10.1214/009053605000000688

YANG, P. and LIN, C.-Y. (2017). Optimal split-plot orthogonal arrays. *Aust*. *N*. *Z*. *J*. *Stat*. **59** 81–94. MR3635168 https://doi.org/10.1111/anzs.12178

YUAN, M., JOSEPH, V. R. and LIN, Y. (2007). An efficient variable selection approach for analyzing designed experiments. *Technometrics* **49** 430–439. MR2414515 https://doi.org/10.1198/004017007000000173

ZHOU, Y. and XU, H. (2015). Space-filling properties of good lattice point sets. *Biometrika* **102** 959–966. MR3431565 https://doi.org/10.1093/biomet/asv044

The Annals of Statistics
2021, Vol. 49, No. 1, 272–294
https://doi.org/10.1214/20-AOS1956

ESTIMATING MINIMUM EFFECT WITH OUTLIER SELECTION

BY ALEXANDRA CARPENTIER[1], SYLVAIN DELATTRE[2], ETIENNE ROQUAIN[3] AND NICOLAS VERZELEN[4]

[1]*Otto-von-Guericke-Universität Magdeburg, Fakultät für Mathematik (FMA), Institut für Mathematische Stochastik (IMST), alexandra.carpentier@ovgu.de*

[2]*Laboratoire de Probabilités Statistique et Modélisation, Université de Paris, CNRS, sylvain.delattre@univ-paris-diderot.fr*

[3]*Laboratoire de Probabilités Statistique et Modélisation, Sorbonne Université, CNRS, etienne.roquain@upmc.fr*

[4]*INRAE, SUPAGRO, Univ. Montpellier, UMR 729, MISTEA, nicolas.verzelen@inrae.fr*

We introduce one-sided versions of Huber's contamination model, in which corrupted samples tend to take larger values than uncorrupted ones. Two intertwined problems are addressed: estimation of the mean of the uncorrupted samples (minimum effect) and selection of the corrupted samples (outliers). Regarding estimation of the minimum effect, we derive the minimax risks and introduce estimators that are adaptive with respect to the unknown number of contaminations. The optimal convergence rates differ from the ones in the classical Huber contamination model. This fact uncovers the effect of the one-sided structural assumption of the contaminations. As for the problem of selecting the outliers, we formulate the problem in a multiple testing framework for which the location and scaling of the null hypotheses are unknown. We rigorously prove that estimating the null hypothesis while maintaining a theoretical guarantee on the amount of the falsely selected outliers is possible, both through false discovery rate (FDR) and through post hoc bounds. As a by-product, we address a long-standing open issue on FDR control under equi-correlation, which reinforces the interest of removing dependency in such a setting.

REFERENCES

[1] ARIAS-CASTRO, E. and CHEN, S. (2017). Distribution-free multiple testing. *Electron. J. Stat.* **11** 1983–2001. MR3651021 https://doi.org/10.1214/17-EJS1277

[2] BARAUD, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli* **8** 577–606. MR1935648

[3] BARBER, R. F. and CANDÈS, E. J. (2015). Controlling the false discovery rate via knockoffs. *Ann. Statist.* **43** 2055–2085. MR3375876 https://doi.org/10.1214/15-AOS1337

[4] 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

[5] BENJAMINI, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188. MR1869245 https://doi.org/10.1214/aos/1013699998

[6] BLANCHARD, G., NEUVIAL, P. and ROQUAIN, E. (2020). Post hoc confidence bounds on false positives using reference families. *Ann. Statist.* **40** 1281–1303. MR4124323 https://doi.org/10.1214/19-AOS1847

[7] 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

[8] CAI, T. T. and JIN, J. (2010). Optimal rates of convergence for estimating the null density and proportion of nonnull effects in large-scale multiple testing. *Ann. Statist.* **38** 100–145. MR2589318 https://doi.org/10.1214/09-AOS696

[9] 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

MSC2020 subject classifications. Primary 62G10; secondary 62C20.
Key words and phrases. Contamination, equicorrelation, false discovery rate, Hermite polynomials, minimax rate, moment matching, multiple testing, post hoc, selective inference, sparsity.

[10] CAI, T. T. and LOW, M. G. (2011). Testing composite hypotheses, Hermite polynomials and optimal estimation of a nonsmooth functional. *Ann*. *Statist*. **39** 1012–1041. MR2816346 https://doi.org/10.1214/10-AOS849

[11] CARPENTIER, A., DELATTRE, S., ROQUAIN, E. and VERZELEN, N. (2021). Supplement to "Estimating minimum effect with outlier selection." https://doi.org/10.1214/20-AOS1956SUPP

[12] CARPENTIER, A. and KIM, A. K. H. (2015). Adaptive and minimax optimal estimation of the tail coefficient. *Statist*. *Sinica* **25** 1133–1144. MR3410301

[13] CARPENTIER, A. and VERZELEN, N. (2019). Adaptive estimation of the sparsity in the Gaussian vector model. *Ann*. *Statist*. **47** 93–126. MR3909928 https://doi.org/10.1214/17-AOS1680

[14] 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

[15] 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

[16] COLLIER, O., COMMINGES, L. and TSYBAKOV, A. B. (2018). On estimation of nonsmooth functionals of sparse normal means. Preprint. Available at arXiv:1805.10791.

[17] COLLIER, O., COMMINGES, L., TSYBAKOV, A. B. and VERZELEN, N. (2018). Optimal adaptive estimation of linear functionals under sparsity. *Ann*. *Statist*. **46** 3130–3150. MR3851767 https://doi.org/10.1214/17-AOS1653

[18] DELATTRE, S. and ROQUAIN, E. (2011). On the false discovery proportion convergence under Gaussian equi-correlation. *Statist*. *Probab. Lett*. **81** 111–115. MR2740072 https://doi.org/10.1016/j.spl.2010.09.025

[19] DELATTRE, S. and ROQUAIN, E. (2015). New procedures controlling the false discovery proportion via Romano–Wolf's heuristic. *Ann*. *Statist*. **43** 1141–1177. MR3346700 https://doi.org/10.1214/14-AOS1302

[20] DELATTRE, S. and ROQUAIN, E. (2016). On empirical distribution function of high-dimensional Gaussian vector components with an application to multiple testing. *Bernoulli* **22** 302–324. MR3449784 https://doi.org/10.3150/14-BEJ659

[21] 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*. *Proceedings of Machine Learning Research* **70** 999–1008.

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

[23] 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

[24] 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

[25] EFRON, B. (2004). Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. *J. Amer. Statist*. *Assoc*. **99** 96–104. MR2054289 https://doi.org/10.1198/016214504000000089

[26] EFRON, B. (2007). Correlation and large-scale simultaneous significance testing. *J. Amer. Statist*. *Assoc*. **102** 93–103. MR2293302 https://doi.org/10.1198/016214506000001211

[27] EFRON, B. (2007). Doing thousands of hypothesis tests at the same time. *Metron* **LXV** 3–21.

[28] EFRON, B. (2009). Empirical Bayes estimates for large-scale prediction problems. *J. Amer. Statist*. *Assoc*. **104** 1015–1028. MR2562003 https://doi.org/10.1198/jasa.2009.tm08523

[29] EFRON, B. (2010). Correlated z-values and the accuracy of large-scale statistical estimates. *J. Amer. Statist*. *Assoc*. **105** 1042–1055. MR2752597 https://doi.org/10.1198/jasa.2010.tm09129

[30] FAN, J. and HAN, X. (2017). Estimation of the false discovery proportion with unknown dependence. *J. R. Stat*. *Soc*. *Ser. B*. *Stat*. *Methodol*. **79** 1143–1164. MR3689312 https://doi.org/10.1111/rssb.12204

[31] FAN, J., HAN, X. and GU, W. (2012). Estimating false discovery proportion under arbitrary covariance dependence. *J. Amer. Statist*. *Assoc*. **107** 1019–1035. MR3010887 https://doi.org/10.1080/01621459.2012.720478

[32] FINNER, H., DICKHAUS, T. and ROTERS, M. (2007). Dependency and false discovery rate: Asymptotics. *Ann*. *Statist*. **35** 1432–1455. MR2351092 https://doi.org/10.1214/009053607000000046

[33] FRIGUET, C., KLOAREG, M. and CAUSEUR, D. (2009). A factor model approach to multiple testing under dependence. *J. Amer. Statist*. *Assoc*. **104** 1406–1415. MR2750571 https://doi.org/10.1198/jasa.2009.tm08332

[34] GAVRILOV, Y., BENJAMINI, Y. and SARKAR, S. K. (2009). An adaptive step-down procedure with proven FDR control under independence. *Ann*. *Statist*. **37** 619–629. MR2502645 https://doi.org/10.1214/07-AOS586

[35] GENOVESE, C. and WASSERMAN, L. (2004). A stochastic process approach to false discovery control. *Ann. Statist.* **32** 1035–1061. MR2065197 https://doi.org/10.1214/009053604000000283

[36] GENOVESE, C. R. and WASSERMAN, L. (2006). Exceedance control of the false discovery proportion. *J. Amer. Statist. Assoc.* **101** 1408–1417. MR2279468 https://doi.org/10.1198/016214506000000339

[37] GOEMAN, J. J. and SOLARI, A. (2011). Multiple testing for exploratory research. *Statist. Sci.* **26** 584–597. MR2951390 https://doi.org/10.1214/11-STS356

[38] GOLDENSHLUGER, A. and LEPSKI, O. (2011). Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39** 1608–1632. MR2850214 https://doi.org/10.1214/11-AOS883

[39] GUO, W., HE, L. and SARKAR, S. K. (2014). Further results on controlling the false discovery proportion. *Ann. Statist.* **42** 1070–1101. MR3210996 https://doi.org/10.1214/14-AOS1214

[40] HAN, Y., JIAO, J. and WEISSMAN, T. (2016). Minimax estimation of KL divergence between discrete distributions. Preprint. Available at arXiv:1605.09124.

[41] HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Stat.* **35** 73–101. MR0161415 https://doi.org/10.1214/aoms/1177703732

[42] HUBER, P. J. (2011). Robust statistics. In *International Encyclopedia of Statistical Science* 1248–1251. Springer, Berlin.

[43] IBRAGIMOV, I. A. and KHASMINSKII, R. Z. (1985). On nonparametric estimation of the value of a linear functional in Gaussian white noise. *Theory Probab. Appl.* **29** 18–32.

[44] IGNATIADIS, N. and HUBER, W. (2017). Covariate powered cross-weighted multiple testing. Preprint. Available at arXiv:1701.05179.

[45] INGSTER, YU. I. and SUSLINA, I. A. (2012). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*. *Lecture Notes in Statistics* **169**. Springer, New York. MR1991446 https://doi.org/10.1007/978-0-387-21580-8

[46] JIAO, J., HAN, Y. and WEISSMAN, T. (2016). Minimax estimation of the L_1 distance. In 2016 *IEEE International Symposium on Information Theory* (*ISIT*) 750–754.

[47] JIN, J. (2008). Proportion of non-zero normal means: Universal oracle equivalences and uniformly consistent estimators. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 461–493. MR2420411 https://doi.org/10.1111/j.1467-9868.2007.00645.x

[48] JIN, J. and CAI, T. T. (2007). Estimating the null and the proportional of nonnull effects in large-scale multiple comparisons. *J. Amer. Statist. Assoc.* **102** 495–506. MR2325113 https://doi.org/10.1198/016214507000000167

[49] JUDITSKY, A. and NEMIROVSKI, A. (2002). On nonparametric tests of positivity/monotonicity/convexity. *Ann. Statist.* **30** 498–527. MR1902897 https://doi.org/10.1214/aos/1021379863

[50] JUREČKOVÁ, J., SEN, P. K. and PICEK, J. (2012). *Methodology in Robust and Nonparametric Statistics*. CRC Press, Boca Raton, FL. MR2963549

[51] KORN, E. L., TROENDLE, J. F., MCSHANE, L. M. and SIMON, R. (2004). Controlling the number of false discoveries: Application to high-dimensional genomic data. *J. Statist. Plann. Inference* **124** 379–398. MR2080371 https://doi.org/10.1016/S0378-3758(03)00211-8

[52] LACOUR, C. and MASSART, P. (2016). Minimal penalty for Goldenshluger–Lepski method. *Stochastic Process. Appl.* **126** 3774–3789. MR3565477 https://doi.org/10.1016/j.spa.2016.04.015

[53] LANCASTER, T. (2000). The incidental parameter problem since 1948. *J. Econometrics* **95** 391–413. MR1752336 https://doi.org/10.1016/S0304-4076(99)00044-5

[54] LEEK, J. T. and STOREY, J. D. (2008). A general framework for multiple testing dependence. *Proc. Natl. Acad. Sci. USA* **105** 18718–18723. https://doi.org/10.1073/pnas.0808709105

[55] 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

[56] 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

[57] 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

[58] NEYMAN, J. and SCOTT, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16** 1–32. MR0025113 https://doi.org/10.2307/1914288

[59] RABINOVICH, M., RAMDAS, A., JORDAN, M. I. and WAINWRIGHT, M. J. (2020). Optimal rates and tradeoffs in multiple testing. *Statist. Sinica.* **30** 741–762. https://doi.org/10.5705/ss.202017.0468

[60] ROMANO, J. P., SHAIKH, A. M. and WOLF, M. (2008). Control of the false discovery rate under dependence using the bootstrap and subsampling. *TEST* **17** 417–442. MR2470085 https://doi.org/10.1007/s11749-008-0126-6

[61] ROMANO, J. P. and WOLF, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *J. Amer. Statist. Assoc.* **100** 94–108. MR2156821 https://doi.org/10.1198/016214504000000539

[62] ROMANO, J. P. and WOLF, M. (2007). Control of generalized error rates in multiple testing. *Ann. Statist.* **35** 1378–1408. MR2351090 https://doi.org/10.1214/009053606000001622

[63] ROQUAIN, E. and VAN DE WIEL, M. A. (2009). Optimal weighting for false discovery rate control. *Electron. J. Stat.* **3** 678–711. MR2521216 https://doi.org/10.1214/09-EJS430

[64] SARKAR, S. K. (2008). Rejoinder: On methods controlling the false discovery rate [MR2551809; MR2551810; MR2551811]. *Sankhyā* **70** 183–185. MR2551812

[65] SPOKOINY, V. G. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.* **24** 2477–2498. MR1425962 https://doi.org/10.1214/aos/1032181163

[66] VERZELEN, N. (2012). Minimax risks for sparse regressions: Ultra-high dimensional phenomenons. *Electron. J. Stat.* **6** 38–90. MR2879672 https://doi.org/10.1214/12-EJS666

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

[68] WU, Y. and YANG, P. (2019). Chebyshev polynomials, moment matching, and optimal estimation of the unseen. *Ann. Statist.* **47** 857–883. MR3909953 https://doi.org/10.1214/17-AOS1665

The Annals of Statistics
2021, Vol. 49, No. 1, 295–320
https://doi.org/10.1214/20-AOS1957

MULTIPLE BLOCK SIZES AND OVERLAPPING BLOCKS FOR MULTIVARIATE TIME SERIES EXTREMES

BY NAN ZOU[1,*], STANISLAV VOLGUSHEV[1,†] AND AXEL BÜCHER[2]

[1]*Department of Statistical Sciences, University of Toronto*, [*]*nan.zou@utoronto.ca;* [†]*stanislav.volgushev@utoronto.ca*
[2]*Mathematisches Institut, Heinrich-Heine-Universität Düsseldorf, axel.buecher@hhu.de*

Block maxima methods constitute a fundamental part of the statistical toolbox in extreme value analysis. However, most of the corresponding theory is derived under the simplifying assumption that block maxima are independent observations from a genuine extreme value distribution. In practice, however, block sizes are finite and observations from different blocks are dependent. Theory respecting the latter complications is not well developed, and, in the multivariate case, has only recently been established for disjoint blocks of a single block size. We show that using overlapping blocks instead of disjoint blocks leads to a uniform improvement in the asymptotic variance of the multivariate empirical distribution function of rescaled block maxima and any smooth functionals thereof (such as the empirical copula), without any sacrifice in the asymptotic bias. We further derive functional central limit theorems for multivariate empirical distribution functions and empirical copulas that are uniform in the block size parameter, which seems to be the first result of this kind for estimators based on block maxima in general. The theory allows for various aggregation schemes over multiple block sizes, leading to substantial improvements over the single block length case and opens the door to further methodology developments. In particular, we consider bias correction procedures that can improve the convergence rates of extreme-value estimators and shed some new light on estimation of the second-order parameter when the main purpose is bias correction.

REFERENCES

BALKEMA, A. A. and RESNICK, S. I. (1977). Max-infinite divisibility. *J. Appl. Probab.* **14** 309–319. MR0438425 https://doi.org/10.1017/s002190020010498x

BEIRLANT, J., GOEGEBEUR, Y., TEUGELS, J. and SEGERS, J. (2004). *Statistics of Extremes: Theory and Applications. Wiley Series in Probability and Statistics*. Wiley, Chichester. MR2108013 https://doi.org/10.1002/0470012382

BEIRLANT, J., ESCOBAR-BACH, M., GOEGEBEUR, Y. and GUILLOU, A. (2016). Bias-corrected estimation of stable tail dependence function. *J. Multivariate Anal.* **143** 453–466. MR3431445 https://doi.org/10.1016/j.jmva.2015.10.006

BERBEE, H. C. P. (1979). *Random Walks with Stationary Increments and Renewal Theory. Mathematical Centre Tracts* **112**. Mathematisch Centrum, Amsterdam. MR0547109

BERGHAUS, B. and BÜCHER, A. (2018). Weak convergence of a pseudo maximum likelihood estimator for the extremal index. *Ann. Statist.* **46** 2307–2335. MR3845019 https://doi.org/10.1214/17-AOS1621

BÜCHER, A. and SEGERS, J. (2014). Extreme value copula estimation based on block maxima of a multivariate stationary time series. *Extremes* **17** 495–528. MR3252823 https://doi.org/10.1007/s10687-014-0195-8

BÜCHER, A. and SEGERS, J. (2018a). Inference for heavy tailed stationary time series based on sliding blocks. *Electron. J. Stat.* **12** 1098–1125. MR3780041 https://doi.org/10.1214/18-EJS1415

BÜCHER, A. and SEGERS, J. (2018b). Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series. *Bernoulli* **24** 1427–1462. MR3706798 https://doi.org/10.3150/16-BEJ903

MSC2020 subject classifications. Primary 62G32, 62E20; secondary 60G70, 62G20.
Key words and phrases. Extreme-value copula, second-order condition, bias correction, mixing coefficients, block maxima, empirical process.

BÜCHER, A., VOLGUSHEV, S. and ZOU, N. (2019). On second order conditions in the multivariate block maxima and peak over threshold method. *J. Multivariate Anal.* **173** 604–619. MR3959186 https://doi.org/10.1016/j.jmva.2019.04.011

CAN, S. U., EINMAHL, J. H. J., KHMALADZE, E. V. and LAEVEN, R. J. A. (2015). Asymptotically distribution-free goodness-of-fit testing for tail copulas. *Ann. Statist.* **43** 878–902. MR3325713 https://doi.org/10.1214/14-AOS1304

CAPÉRAÀ, P., FOUGÈRES, A.-L. and GENEST, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* **84** 567–577. MR1603985 https://doi.org/10.1093/biomet/84.3.567

CHARPENTIER, A. and SEGERS, J. (2009). Tails of multivariate Archimedean copulas. *J. Multivariate Anal.* **100** 1521–1537. MR2514145 https://doi.org/10.1016/j.jmva.2008.12.015

DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory: An Introduction. Springer Series in Operations Research and Financial Engineering*. Springer, New York. MR2234156 https://doi.org/10.1007/0-387-34471-3

DE HAAN, L., MERCADIER, C. and ZHOU, C. (2016). Adapting extreme value statistics to financial time series: Dealing with bias and serial dependence. *Finance Stoch.* **20** 321–354. MR3479324 https://doi.org/10.1007/s00780-015-0287-6

DE HAAN, L. and RESNICK, S. I. (1977). Limit theory for multivariate sample extremes. *Z. Wahrsch. Verw. Gebiete* **40** 317–337. MR0478290 https://doi.org/10.1007/BF00533086

DOMBRY, C. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli* **21** 420–436. MR3322325 https://doi.org/10.3150/13-BEJ573

DOMBRY, C. and FERREIRA, A. (2019). Maximum likelihood estimators based on the block maxima method. *Bernoulli* **25** 1690–1723. MR3961227 https://doi.org/10.3150/18-BEJ1032

DREES, H., DE HAAN, L. and RESNICK, S. (2000). How to make a Hill plot. *Ann. Statist.* **28** 254–274. MR1762911 https://doi.org/10.1214/aos/1016120372

DREES, H. and ROOTZÉN, H. (2010). Limit theorems for empirical processes of cluster functionals. *Ann. Statist.* **38** 2145–2186. MR2676886 https://doi.org/10.1214/09-AOS788

DUDLEY, R. M. (2002). *Real Analysis and Probability. Cambridge Studies in Advanced Mathematics* **74**. Cambridge Univ. Press, Cambridge. MR1932358 https://doi.org/10.1017/CBO9780511755347

EINMAHL, J. H. J., DE HAAN, L. and ZHOU, C. (2016). Statistics of heteroscedastic extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 31–51. MR3453645 https://doi.org/10.1111/rssb.12099

FERREIRA, A. and DE HAAN, L. (2015). On the block maxima method in extreme value theory: PWM estimators. *Ann. Statist.* **43** 276–298. MR3285607 https://doi.org/10.1214/14-AOS1280

FOUGÈRES, A.-L., DE HAAN, L. and MERCADIER, C. (2015). Bias correction in multivariate extremes. *Ann. Statist.* **43** 903–934. MR3325714 https://doi.org/10.1214/14-AOS1305

GENEST, C. and SEGERS, J. (2009). Rank-based inference for bivariate extreme-value copulas. *Ann. Statist.* **37** 2990–3022. MR2541453 https://doi.org/10.1214/08-AOS672

GENEST, C. and SEGERS, J. (2010). On the covariance of the asymptotic empirical copula process. *J. Multivariate Anal.* **101** 1837–1845. MR2651959 https://doi.org/10.1016/j.jmva.2010.03.018

GUDENDORF, G. and SEGERS, J. (2010). Extreme-value copulas. In *Copula Theory and Its Applications. Lect. Notes Stat. Proc.* **198** 127–145. Springer, Heidelberg. MR3051266 https://doi.org/10.1007/978-3-642-12465-5_6

GUMBEL, E. J. (1958). *Statistics of Extremes*. Columbia Univ. Press, New York. MR0096342

HSING, T. (1989). Extreme value theory for multivariate stationary sequences. *J. Multivariate Anal.* **29** 274–291. MR1004339 https://doi.org/10.1016/0047-259X(89)90028-6

HUANG, X. (1992). Statistics of bivariate extreme values. Ph.D. thesis, Tinbergen Institute Research Series, The Netherlands.

HÜSLER, J. (1990). Multivariate extreme values in stationary random sequences. *Stochastic Process. Appl.* **35** 99–108. MR1062586 https://doi.org/10.1016/0304-4149(90)90125-C

NAVEAU, P., GUILLOU, A., COOLEY, D. and DIEBOLT, J. (2009). Modelling pairwise dependence of maxima in space. *Biometrika* **96** 1–17. MR2482131 https://doi.org/10.1093/biomet/asp001

NORTHROP, P. J. (2015). An efficient semiparametric maxima estimator of the extremal index. *Extremes* **18** 585–603. MR3418769 https://doi.org/10.1007/s10687-015-0221-5

PICKANDS, J. III (1981). Multivariate extreme value distributions. *Bull. Int. Stat. Inst.* **49** 859–878, 894–902. MR0820979

RESNICK, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes. Applied Probability. A Series of the Applied Probability Trust* **4**. Springer, New York. MR0900810 https://doi.org/10.1007/978-0-387-75953-1

ROBERT, C. Y., SEGERS, J. and FERRO, C. A. T. (2009). A sliding blocks estimator for the extremal index. *Electron. J. Stat.* **3** 993–1020. MR2540849 https://doi.org/10.1214/08-EJS345

SCHMIDT, R. and STADTMÜLLER, U. (2006). Non-parametric estimation of tail dependence. *Scand. J. Stat.* **33** 307–335. MR2279645 https://doi.org/10.1111/j.1467-9469.2005.00483.x

SEGERS, J. (2012). Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli* **18** 764–782. MR2948900 https://doi.org/10.3150/11-BEJ387

SMITH, R. L. and WEISSMAN, I. (1994). Estimating the extremal index. *J. Roy. Statist. Soc. Ser. B* **56** 515–528. MR1278224

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

ZOU, N., VOLGUSHEV, S. and BÜCHER, A. (2021). Supplement to "Multiple block sizes and overlapping blocks for multivariate time series extremes." https://doi.org/10.1214/20-AOS1957SUPP

The Annals of Statistics
2021, Vol. 49, No. 1, 321–345
https://doi.org/10.1214/20-AOS1958

ESTIMATION OF LOW-RANK MATRICES VIA APPROXIMATE MESSAGE PASSING

BY ANDREA MONTANARI[1] AND RAMJI VENKATARAMANAN[2]

[1]*Department of Electrical Engineering and Department of Statistics, Stanford University, montanari@stanford.edu*
[2]*Department of Engineering, University of Cambridge, ramji.v@eng.cam.ac.uk*

Consider the problem of estimating a low-rank matrix when its entries are perturbed by Gaussian noise, a setting that is also known as "spiked model" or "deformed random matrix." If the empirical distribution of the entries of the spikes is known, optimal estimators that exploit this knowledge can substantially outperform simple spectral approaches. Recent work characterizes the asymptotic accuracy of Bayes-optimal estimators in the high-dimensional limit. In this paper, we present a practical algorithm that can achieve Bayes-optimal accuracy above the spectral threshold. A bold conjecture from statistical physics posits that no polynomial-time algorithm achieves optimal error below the same threshold (unless the best estimator is trivial).

Our approach uses Approximate Message Passing (AMP) in conjunction with a spectral initialization. AMP algorithms have proved successful in a variety of statistical estimation tasks, and are amenable to exact asymptotic analysis via state evolution. Unfortunately, state evolution is uninformative when the algorithm is initialized near an unstable fixed point, as often happens in low-rank matrix estimation problems. We develop a new analysis of AMP that allows for spectral initializations, and builds on a decoupling between the outlier eigenvectors and the bulk in the spiked random matrix model.

Our main theorem is general and applies beyond matrix estimation. However, we use it to derive detailed predictions for the problem of estimating a rank-one matrix in noise. Special cases of this problem are closely related—via universality arguments—to the network community detection problem for two asymmetric communities. For general rank-one models, we show that AMP can be used to construct confidence intervals and control false discovery rate.

We provide illustrations of the general methodology by considering the cases of sparse low-rank matrices and of block-constant low-rank matrices with symmetric blocks (we refer to the latter as to the "Gaussian block model").

REFERENCES

[1] ABBE, E. (2018). Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** Paper No. 177, 86. MR3827065

[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] 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

[4] 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

MSC2020 subject classifications. Primary 62H99; secondary 62F15, 62E20.
Key words and phrases. Low-rank matrix estimation, approximate message passing, spectral initialization.

[5] BARBIER, J., DIA, M., MACRIS, N., KRZAKALA, F., LESIEUR, T. and ZDEBOROVÁ, L. (2016). Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. In *Advances in Neural Information Processing Systems* 424–432.

[6] BAYATI, M., LELARGE, M. and MONTANARI, A. (2015). Universality in polytope phase transitions and message passing algorithms. *Ann. Appl. Probab.* **25** 753–822. MR3313755 https://doi.org/10.1214/14-AAP1010

[7] BAYATI, M. and MONTANARI, A. (2011). The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inf. Theory* **57** 764–785. MR2810285 https://doi.org/10.1109/TIT.2010.2094817

[8] BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** 494–521. MR2782201 https://doi.org/10.1016/j.aim.2011.02.007

[9] 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

[10] 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

[11] BERTHIER, R., MONTANARI, A. and NGUYEN, P.-M. (2020). State evolution for approximate message passing with non-separable functions. *Inf. Inference* **9** 33–79. MR4079177 https://doi.org/10.1093/imaiai/iay021

[12] BLEI, D. M., KUCUKELBIR, A. and MCAULIFFE, J. D. (2017). Variational inference: A review for statisticians. *J. Amer. Statist. Assoc.* **112** 859–877. MR3671776 https://doi.org/10.1080/01621459.2017.1285773

[13] BOLTHAUSEN, E. (2014). An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model. *Comm. Math. Phys.* **325** 333–366. MR3147441 https://doi.org/10.1007/s00220-013-1862-3

[14] CAPITAINE, M., DONATI-MARTIN, C. and FÉRAL, D. (2009). The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations. *Ann. Probab.* **37** 1–47. MR2489158 https://doi.org/10.1214/08-AOP394

[15] CHEN, Y. and CANDÈS, E. J. (2018). The projected power method: An efficient algorithm for joint alignment from pairwise differences. *Comm. Pure Appl. Math.* **71** 1648–1714. MR3847751 https://doi.org/10.1002/cpa.21760

[16] DESHPANDE, Y., ABBE, E. and MONTANARI, A. (2017). Asymptotic mutual information for the balanced binary stochastic block model. *Inf. Inference* **6** 125–170. MR3671474 https://doi.org/10.1093/imaiai/iaw017

[17] DESHPANDE, Y. and MONTANARI, A. (2014). Information-theoretically optimal sparse PCA. In *IEEE International Symposium on Information Theory* (*ISIT*) 2197–2201.

[18] DESHPANDE, Y. and MONTANARI, A. (2015). Finding hidden cliques of size $\sqrt{N/e}$ in nearly linear time. *Found. Comput. Math.* **15** 1069–1128. MR3371378 https://doi.org/10.1007/s10208-014-9215-y

[19] DONOHO, D. L., JOHNSTONE, I. and MONTANARI, A. (2013). Accurate prediction of phase transitions in compressed sensing via a connection to minimax denoising. *IEEE Trans. Inf. Theory* **59** 3396–3433. MR3061255 https://doi.org/10.1109/TIT.2013.2239356

[20] DONOHO, D. L. and JOHNSTONE, I. M. (1994). Minimax risk over l_p-balls for l_q-error. *Probab. Theory Related Fields* **99** 277–303. MR1278886 https://doi.org/10.1007/BF01199026

[21] DONOHO, D. L. and JOHNSTONE, I. M. (1998). Minimax estimation via wavelet shrinkage. *Ann. Statist.* **26** 879–921. MR1635414 https://doi.org/10.1214/aos/1024691081

[22] DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2009). Message passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci. USA* **106** 18914–18919.

[23] 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

[24] FÉRAL, D. and PÉCHÉ, S. (2007). The largest eigenvalue of rank one deformation of large Wigner matrices. *Comm. Math. Phys.* **272** 185–228. MR2291807 https://doi.org/10.1007/s00220-007-0209-3

[25] FLETCHER, A. K. and RANGAN, S. (2018). Iterative reconstruction of rank-one matrices in noise. *Inf. Inference* **7** 531–562. MR3858334 https://doi.org/10.1093/imaiai/iax014

[26] GAMERMAN, D. and LOPES, H. F. (2006). *Markov Chain Monte Carlo*: *Stochastic Simulation for Bayesian Inference*, 2nd ed. *Texts in Statistical Science Series*. CRC Press/CRC, Boca Raton, FL. MR2260716

[27] GUO, D., SHAMAI, S. and VERDÚ, S. (2005). Mutual information and minimum mean-square error in Gaussian channels. *IEEE Trans. Inf. Theory* **51** 1261–1282. MR2241490 https://doi.org/10.1109/TIT.2005.844072

[28] HOYLE, D. C. and RATTRAY, M. (2004). Principal-component-analysis eigenvalue spectra from data with symmetry-breaking structure. *Phys. Rev. E* **69** 026124.

[29] JAVANMARD, A. and MONTANARI, A. (2013). State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Inf. Inference* **2** 115–144. MR3311445 https://doi.org/10.1093/imaiai/iat004

[30] 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

[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] JOHNSTONE, I. M. (2007). High dimensional statistical inference and random matrices. In *International Congress of Mathematicians. Vol. I* 307–333. Eur. Math. Soc., Zürich. MR2334195 https://doi.org/10.4171/022-1/13

[33] 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

[34] JOURNÉE, M., NESTEROV, Y., RICHTÁRIK, P. and SEPULCHRE, R. (2010). Generalized power method for sparse principal component analysis. *J. Mach. Learn. Res.* **11** 517–553. MR2600619

[35] KABASHIMA, Y., KRZAKALA, F., MÉZARD, M., SAKATA, A. and ZDEBOROVÁ, L. (2016). Phase transitions and sample complexity in Bayes-optimal matrix factorization. *IEEE Trans. Inf. Theory* **62** 4228–4265. MR3515748 https://doi.org/10.1109/TIT.2016.2556702

[36] KNOWLES, A. and YIN, J. (2013). The isotropic semicircle law and deformation of Wigner matrices. *Comm. Pure Appl. Math.* **66** 1663–1750. MR3103909 https://doi.org/10.1002/cpa.21450

[37] KRZAKALA, F., XU, J. and ZDEBOROVÁ, L. (2016). Mutual information in rank-one matrix estimation. In *IEEE Information Theory Workshop* (*ITW*) 71–75.

[38] LEE, D. D. and SEUNG, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature* **401** 788–791. https://doi.org/10.1038/44565

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

[40] LESIEUR, T., KRZAKALA, F. and ZDEBOROVÁ, L. (2017). Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications. *J. Stat. Mech. Theory Exp.* **7** 073403, 86. MR3683819 https://doi.org/10.1088/1742-5468/aa7284

[41] MA, Z. (2013). Sparse principal component analysis and iterative thresholding. *Ann. Statist.* **41** 772–801. MR3099121 https://doi.org/10.1214/13-AOS1097

[42] MIOLANE, L. (2017). Fundamental limits of low-rank matrix estimation. arXiv:1702.00473.

[43] MONTANARI, A. and RICHARD, E. (2016). Non-negative principal component analysis: Message passing algorithms and sharp asymptotics. *IEEE Trans. Inf. Theory* **62** 1458–1484. MR3472260 https://doi.org/10.1109/TIT.2015.2457942

[44] MONTANARI, A. and VENKATARAMANAN, R. (2021). Supplement to "Estimation of low-rank matrices via approximate message passing." https://doi.org/10.1214/20-AOS1958SUPP

[45] MOORE, C. (2017). The computer science and physics of community detection: Landscapes, phase transitions, and hardness. *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS* **121** 26–61. MR3699594

[46] MOSSEL, E. and XU, J. (2016). Density evolution in the degree-correlated stochastic block model. In *Conference on Learning Theory* 1319–1356.

[47] PARKER, J. T., SCHNITER, P. and CEVHER, V. (2014). Bilinear generalized approximate message passing—Part I: Derivation. *IEEE Trans. Signal Process.* **62** 5839–5853. MR3281527 https://doi.org/10.1109/TSP.2014.2357776

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

[49] STOREY, J. D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 479–498. MR1924302 https://doi.org/10.1111/1467-9868.00346

[50] VILA, J., SCHNITER, P. and MEOLA, J. (2015). Hyperspectral unmixing via turbo bilinear approximate message passing. *IEEE Trans. Comput. Imaging* **1** 143–158. MR3423296 https://doi.org/10.1109/TCI.2015.2465161

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

[52] YUAN, X.-T. and ZHANG, T. (2013). Truncated power method for sparse eigenvalue problems. *J. Mach. Learn. Res.* **14** 899–925. MR3063614

[53] ZOU, H., HASTIE, T. and TIBSHIRANI, R. (2006). Sparse principal component analysis. *J. Comput. Graph. Statist.* **15** 265–286. MR2252527 https://doi.org/10.1198/106186006X113430

The Annals of Statistics
2021, Vol. 49, No. 1, 346–369
https://doi.org/10.1214/20-AOS1959

ASYMPTOTICS FOR SPHERICAL FUNCTIONAL AUTOREGRESSIONS

BY ALESSIA CAPONERA[1] AND DOMENICO MARINUCCI[2]

[1]*Department of Statistical Sciences, Sapienza University of Rome, alessia.caponera@uniroma1.it*
[2]*Department of Mathematics, University of Rome Tor Vergata, marinucc@mat.uniroma2.it*

In this paper, we investigate a class of spherical functional autoregressive processes, and we discuss the estimation of the corresponding autoregressive kernels. In particular, we first establish a consistency result (in mean-square and sup norm), then a quantitative central limit theorem (in Wasserstein distance), and finally a weak convergence result, under more restrictive regularity conditions. Our results are validated by a small numerical investigation.

REFERENCES

[1] AUE, A. and VAN DELFT, A. (2020). Testing for stationarity of functional time series in the frequency domain. *Ann. Statist.* **40** 2505–2547. MR4152111 https://doi.org/10.1214/19-AOS1895 To appear. Available at arXiv:1701.01741.

[2] BALDI, P., KERKYACHARIAN, G., MARINUCCI, D. and PICARD, D. (2009). Asymptotics for spherical needlets. *Ann. Statist.* **37** 1150–1171. MR2509070 https://doi.org/10.1214/08-AOS601

[3] BERG, C. and PORCU, E. (2017). From Schoenberg coefficients to Schoenberg functions. *Constr. Approx.* **45** 217–241. MR3619442 https://doi.org/10.1007/s00365-016-9323-9

[4] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. *Wiley Series in Probability and Statistics*: *Probability and Statistics*. Wiley, New York. MR1700749 https://doi.org/10.1002/9780470316962

[5] BOSQ, D. (2000). *Linear Processes in Function Spaces*: *Theory and Applications*. *Lecture Notes in Statistics* **149**. Springer, New York. MR1783138 https://doi.org/10.1007/978-1-4612-1154-9

[6] BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series*: *Theory and Methods*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR1093459 https://doi.org/10.1007/978-1-4419-0320-4

[7] CAPONERA, A. (2020). Statistical inference for spherical functional autoregressions Ph.D. thesis, Sapienza University of Rome.

[8] CAPONERA, A. and MARINUCCI, D. (2021). Supplement to "Asymptotics for spherical functional autoregressions." https://doi.org/10.1214/20-AOS1959SUPP

[9] CHENG, D., CAMMAROTA, V., FANTAYE, Y., MARINUCCI, D. and SCHWARTZMAN, A. (2020). Multiple testing of local maxima for detection of peaks on the (celestial) sphere. *Bernoulli* **26** 31–60. MR4036027 https://doi.org/10.3150/18-BEJ1068

[10] CLARKE DE LA CERDA, J., ALEGRÍA, A. and PORCU, E. (2018). Regularity properties and simulations of Gaussian random fields on the sphere cross time. *Electron. J. Stat.* **12** 399–426. MR3763911 https://doi.org/10.1214/18-EJS1393

[11] FAN, M., PAUL, D., LEE, T. C. M. and MATSUO, T. (2018). A multi-resolution model for non-Gaussian random fields on a sphere with application to ionospheric electrostatic potentials. *Ann. Appl. Stat.* **12** 459–489. MR3773401 https://doi.org/10.1214/17-AOAS1104

[12] FAN, M., PAUL, D., LEE, T. C. M. and MATSUO, T. (2018). Modeling tangential vector fields on a sphere. *J. Amer. Statist. Assoc.* **113** 1625–1636. MR3902234 https://doi.org/10.1080/01621459.2017.1356322

[13] GNEITING, T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli* **19** 1327–1349. MR3102554 https://doi.org/10.3150/12-BEJSP06

[14] GORSKI, K. M., HIVON, E., BANDAY, A. J., WANDELT, B. D., HANSEN, F. K., REINECKE, M. and BARTELMANN, M. (2005). HEALPix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere. *Astrophys. J.* **622** 759–771.

[15] GRADSHTEYN, I. S. and RYZHIK, I. M. (2015). *Table of Integrals*, *Series*, *and Products*, 8th ed. Elsevier/Academic Press, Amsterdam. MR3307944

MSC2020 subject classifications. Primary 62M15; secondary 60G15, 60F05, 62M40.
Key words and phrases. Spherical functional autoregressions, spherical harmonics, quantitative central limit theorem, Wasserstein distance, weak convergence.

[16] HÖRMANN, S., KOKOSZKA, P. and NISOL, G. (2018). Testing for periodicity in functional time series. *Ann. Statist.* **46** 2960–2984. MR3851761 https://doi.org/10.1214/17-AOS1645

[17] HSING, T. and EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators. Wiley Series in Probability and Statistics*. Wiley, Chichester. MR3379106 https://doi.org/10.1002/9781118762547

[18] JUN, M. (2014). Matérn-based nonstationary cross-covariance models for global processes. *J. Multivariate Anal.* **128** 134–146. MR3199833 https://doi.org/10.1016/j.jmva.2014.03.009

[19] KALNAY, E., KANAMITSU, M., KISTLER, R., COLLINS, W., DEAVEN, D., GANDIN, L., ZHU, Y. et al. (1996). The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77** 437–472.

[20] LANG, A. and SCHWAB, C. (2015). Isotropic Gaussian random fields on the sphere: Regularity, fast simulation and stochastic partial differential equations. *Ann. Appl. Probab.* **25** 3047–3094. MR3404631 https://doi.org/10.1214/14-AAP1067

[21] LEONENKO, N. N., TAQQU, M. S. and TERDIK, G. H. (2018). Estimation of the covariance function of Gaussian isotropic random fields on spheres, related Rosenblatt-type distributions and the cosmic variance problem. *Electron. J. Stat.* **12** 3114–3146. MR3857874 https://doi.org/10.1214/18-EJS1473

[22] MARINUCCI, D. and PECCATI, G. (2011). *Random Fields on the Sphere: Representation, Limit Theorems and Cosmological Applications. London Mathematical Society Lecture Note Series* **389**. Cambridge Univ. Press, Cambridge. MR2840154 https://doi.org/10.1017/CBO9780511751677

[23] NOURDIN, I. and PECCATI, G. (2009). Stein's method on Wiener chaos. *Probab. Theory Related Fields* **145** 75–118. MR2520122 https://doi.org/10.1007/s00440-008-0162-x

[24] NOURDIN, I. and PECCATI, G. (2012). *Normal Approximations with Malliavin Calculus: From Stein's Method to Universality. Cambridge Tracts in Mathematics* **192**. Cambridge Univ. Press, Cambridge. MR2962301 https://doi.org/10.1017/CBO9781139084659

[25] PANARETOS, V. M. and TAVAKOLI, S. (2013). Cramér–Karhunen–Loève representation and harmonic principal component analysis of functional time series. *Stochastic Process. Appl.* **123** 2779–2807. MR3054545 https://doi.org/10.1016/j.spa.2013.03.015

[26] PANARETOS, V. M. and TAVAKOLI, S. (2013). Fourier analysis of stationary time series in function space. *Ann. Statist.* **41** 568–603. MR3099114 https://doi.org/10.1214/13-AOS1086

[27] PLANCK COLLABORATION (2016). Planck 2015 results. I. Overview of products and scientific results. *Astron. Astrophys.* **594** Art. ID A1.

[28] PORCU, E., ALEGRIA, A. and FURRER, R. (2018). Modelling temporally evolving and spatially globally dependent data. *Int. Stat. Rev.* **86** 344–377. MR3852415 https://doi.org/10.1111/insr.12266

[29] PORCU, E., BEVILACQUA, M. and GENTON, M. G. (2016). Spatio-temporal covariance and cross-covariance functions of the great circle distance on a sphere. *J. Amer. Statist. Assoc.* **111** 888–898. MR3538713 https://doi.org/10.1080/01621459.2015.1072541

[30] ROBINSON, P. M. (1995). Log-periodogram regression of time series with long range dependence. *Ann. Statist.* **23** 1048–1072. MR1345214 https://doi.org/10.1214/aos/1176324636

[31] SZEGŐ, G. (1975). *Orthogonal Polynomials*, 4th ed. *Colloquium Publications* **XXIII**. Amer. Math. Soc., Providence, RI. MR0372517

[32] WIGMAN, I. (2010). Fluctuations of the nodal length of random spherical harmonics. *Comm. Math. Phys.* **298** 787–831. MR2670928 https://doi.org/10.1007/s00220-010-1078-8

The Annals of Statistics
2021, Vol. 49, No. 1, 370–392
https://doi.org/10.1214/20-AOS1960

SINGULAR VECTOR AND SINGULAR SUBSPACE DISTRIBUTION FOR THE MATRIX DENOISING MODEL

BY ZHIGANG BAO[1,*], XIUCAI DING[2] AND AND KE WANG[1,†]

[1]*Department of Mathematics, Hong Kong University of Science and Technology,* [*]*mazgbao@ust.hk;* [†]*kewang@ust.hk*
[2]*Department of Statistics, University of California, Davis, xcading@ucdavis.edu*

In this paper, we study the matrix denoising model $Y = S + X$, where S is a low rank deterministic signal matrix and X is a random noise matrix, and both are $M \times n$. In the scenario that M and n are comparably large and the signals are supercritical, we study the fluctuation of the outlier singular vectors of Y, under fully general assumptions on the structure of S and the distribution of X. More specifically, we derive the limiting distribution of angles between the principal singular vectors of Y and their deterministic counterparts, the singular vectors of S. Further, we also derive the distribution of the distance between the subspace spanned by the principal singular vectors of Y and that spanned by the singular vectors of S. It turns out that the limiting distributions depend on the structure of the singular vectors of S and the distribution of X, and thus they are nonuniversal. Statistical applications of our results to singular vector and singular subspace inferences are also discussed.

REFERENCES

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

[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] 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

[4] 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

[5] 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

[6] 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

[7] BAO, Z., DING, X., WANG, J. and WANG, K. (2019). Principal components of spiked covariance matrices in the supercritical regime. Available at arXiv:1907.12251.

[8] BAO, Z., DING, X. and WANG, K. (2021). Supplement to "Singular vector and singular subspace distribution for the matrix denoising model." https://doi.org/10.1214/20-AOS1960SUPP

[9] BAO, Z., HU, J., PAN, G. and ZHOU, W. (2019). Canonical correlation coefficients of high-dimensional Gaussian vectors: Finite rank case. *Ann. Statist.* **47** 612–640. MR3909944 https://doi.org/10.1214/18-AOS1704

[10] BAO, Z., PAN, G. and ZHOU, W. (2015). Universality for the largest eigenvalue of sample covariance matrices with general population. *Ann. Statist.* **43** 382–421. MR3311864 https://doi.org/10.1214/14-AOS1281

[11] BENAYCH-GEORGES, F., GUIONNET, A. and MAIDA, M. (2011). Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. *Electron. J. Probab.* **16** 1621–1662. MR2835249 https://doi.org/10.1214/EJP.v16-929

MSC2020 subject classifications. Primary 60B20, 62G10; secondary 62H10, 15B52, 62H25.
Key words and phrases. Random matrix, singular vector, singular subspace, matrix denoising model, signal-plus-noise model, nonuniversality.

[12] BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** 494–521. MR2782201 https://doi.org/10.1016/j.aim.2011.02.007

[13] 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

[14] BLOEMENDAL, A., ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2014). Isotropic local laws for sample covariance and generalized Wigner matrices. *Electron. J. Probab.* **19** no. 33, 53. MR3183577 https://doi.org/10.1214/ejp.v19-3054

[15] BLOEMENDAL, A., KNOWLES, A., YAU, H.-T. and YIN, J. (2016). On the principal components of sample covariance matrices. *Probab. Theory Related Fields* **164** 459–552. MR3449395 https://doi.org/10.1007/s00440-015-0616-x

[16] BLOEMENDAL, A. and VIRÁG, B. (2013). Limits of spiked random matrices I. *Probab. Theory Related Fields* **156** 795–825. MR3078286 https://doi.org/10.1007/s00440-012-0443-2

[17] BLOEMENDAL, A. and VIRÁG, B. (2016). Limits of spiked random matrices II. *Ann. Probab.* **44** 2726–2769. MR3531679 https://doi.org/10.1214/15-AOP1033

[18] 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

[19] CAPE, J., TANG, M. and PRIEBE, C. E. (2019). Signal-plus-noise matrix models: Eigenvector deviations and fluctuations. *Biometrika* **106** 243–250. MR3912394 https://doi.org/10.1093/biomet/asy070

[20] 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

[21] CAPITAINE, M. (2018). Limiting eigenvectors of outliers for spiked information-plus-noise type matrices. In *Séminaire de Probabilités XLIX. Lecture Notes in Math.* **2215** 119–164. Springer, Cham. MR3837103

[22] CAPITAINE, M. and DONATI-MARTIN, C. (2017). Spectrum of deformed random matrices and free probability. In *Advanced Topics in Random Matrices. Panor. Synthèses* **53** 151–190. Soc. Math. France, Paris. MR3792626

[23] CAPITAINE, M. and DONATI-MARTIN, C. (2018). Non universality of fluctuations of outlier eigenvectors for block diagonal deformations of Wigner matrices. Available at arXiv:1807.07773.

[24] CAPITAINE, M., DONATI-MARTIN, C. and FÉRAL, D. (2009). The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations. *Ann. Probab.* **37** 1–47. MR2489158 https://doi.org/10.1214/08-AOP394

[25] CAPITAINE, M., DONATI-MARTIN, C. and FÉRAL, D. (2012). Central limit theorems for eigenvalues of deformations of Wigner matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 107–133. MR2919200 https://doi.org/10.1214/10-AIHP410

[26] DING, X. (2020). High dimensional deformed rectangular matrices with applications in matrix denoising. *Bernoulli* **26** 387–417. MR4036038 https://doi.org/10.3150/19-BEJ1129

[27] DING, X. and YANG, F. (2019). Spiked separable covariance matrices and principal components. Available at arXiv:1905.13060.

[28] DONOHO, D. and GAVISH, M. (2014). Minimax risk of matrix denoising by singular value thresholding. *Ann. Statist.* **42** 2413–2440. MR3269984 https://doi.org/10.1214/14-AOS1257

[29] EL KAROUI, N. (2007). Tracy–Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.* **35** 663–714. MR2308592 https://doi.org/10.1214/009117906000000917

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

[31] FAN, J., SUN, Q., ZHOU, W.-X. and ZHU, Z. (2018). Principal component analysis for big data. Available at arXiv:1801.01602.

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

[33] FAN, J. and ZHONG, Y. (2018). Optimal subspace estimation using overidentifying vectors via generalized method of moments. Available at arXiv:1805.02826.

[34] FAN, Z., JOHNSTONE, I. and SUN, Y. (2018). Spiked covariances and principal components analysis in high-dimensional random effects models. Available at arXiv:1806.09529.

[35] FÉRAL, D. and PÉCHÉ, S. (2007). The largest eigenvalue of rank one deformation of large Wigner matrices. *Comm. Math. Phys.* **272** 185–228. MR2291807 https://doi.org/10.1007/s00220-007-0209-3

[36] GOLUB, G. H. and VAN LOAN, C. F. (2012). *Matrix Computations*, 3rd ed. *Johns Hopkins Studies in the Mathematical Sciences*. Johns Hopkins Univ. Press, Baltimore, MD.

[37] HACHEM, W., LOUBATON, P., MESTRE, X., NAJIM, J. and VALLET, P. (2013). A subspace estimator for fixed rank perturbations of large random matrices. *J. Multivariate Anal.* **114** 427–447. MR2993897 https://doi.org/10.1016/j.jmva.2012.08.006

[38] HECKEL, R., TSCHANNEN, M. and BÖLCSKEI, H. (2017). Dimensionality-reduced subspace clustering. *Inf. Inference* **6** 246–283. MR3764525 https://doi.org/10.1093/imaiai/iaw021

[39] HUANG, J., QIU, Q. and CALDERBANK, R. (2016). The role of principal angles in subspace classification. *IEEE Trans. Signal Process.* **64** 1933–1945. MR3470590 https://doi.org/10.1109/TSP.2015.2500889

[40] KAY, S. M. (1998). *Fundamentals of Statistical Signal Processing*, *Volume* 2: *Detection Theory*. Prentice-Hall, New York.

[41] KNOWLES, A. and YIN, J. (2013). The isotropic semicircle law and deformation of Wigner matrices. *Comm. Pure Appl. Math.* **66** 1663–1750. MR3103909 https://doi.org/10.1002/cpa.21450

[42] KNOWLES, A. and YIN, J. (2014). The outliers of a deformed Wigner matrix. *Ann. Probab.* **42** 1980–2031. MR3262497 https://doi.org/10.1214/13-AOP855

[43] KNOWLES, A. and YIN, J. (2017). Anisotropic local laws for random matrices. *Probab. Theory Related Fields* **169** 257–352. MR3704770 https://doi.org/10.1007/s00440-016-0730-4

[44] 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

[45] 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

[46] KOLTCHINSKII, V. and LOUNICI, K. (2017). Normal approximation and concentration of spectral projectors of sample covariance. *Ann. Statist.* **45** 121–157. MR3611488 https://doi.org/10.1214/16-AOS1437

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

[48] LEE, M., SHEN, H., HUANG, J. Z. and MARRON, J. S. (2010). Biclustering via sparse singular value decomposition. *Biometrics* **66** 1087–1095. MR2758496 https://doi.org/10.1111/j.1541-0420.2010.01392.x

[49] LEVIN, A. and NADLER, B. (2011). Natural image denoising: Optimality and inherent bounds. In 2011 *IEEE Conference On Computer Vision And Pattern Recognition* (*CVPR*) **40**.

[50] MARČENKO, V. A. and PASTUR, L. A. (1967). The spectrum of random matrices. *Teor. Funkciĭ Funkcional. Anal. i Priložen. Vyp.* **4** 122–145. MR0221550

[51] NADLER, B. and JOHNSTONE, I. M. (2011). On the distribution of Roy's Largest Root Test in MANOVA and in signal detection in noise. Technical report, Department of Statistics, Stanford University.

[52] 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

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

[54] PÉCHÉ, S. (2006). The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probab. Theory Related Fields* **134** 127–173. MR2221787 https://doi.org/10.1007/s00440-005-0466-z

[55] PETERFREUND, E. and GAVISH, M. (2018). Multidimensional scaling of noisy high dimensional data. Available at arXiv:1801.10229.

[56] PILLAI, N. S. and YIN, J. (2014). Universality of covariance matrices. *Ann. Appl. Probab.* **24** 935–1001. MR3199978 https://doi.org/10.1214/13-AAP939

[57] PIZZO, A., RENFREW, D. and SOSHNIKOV, A. (2013). On finite rank deformations of Wigner matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 64–94. MR3060148 https://doi.org/10.1214/11-AIHP459

[58] WANG, W., MIGUEL, A. C.-P. and LU, Z. (2011). A denoising view of matrix completion. *Adv. Neural Inf. Process. Syst.* **24** 334–342.

[59] XIA, D. (2019). Confidence region of singular subspaces for low-rank matrix regression. *IEEE Trans. Inf. Theory* **65** 7437–7459. MR4030894 https://doi.org/10.1109/TIT.2019.2924900

[60] YANG, D., MA, Z. and BUJA, A. (2014). A sparse singular value decomposition method for high-dimensional data. *J. Comput. Graph. Statist.* **23** 923–942. MR3270704 https://doi.org/10.1080/10618600.2013.858632

[61] YANG, D., MA, Z. and BUJA, A. (2016). Rate optimal denoising of simultaneously sparse and low rank matrices. *J. Mach. Learn. Res.* **17** Paper No. 92, 27. MR3543498

[62] ZAHERNIA, A., DEHGHANI, M. J. and JAVIDAN, R. (2011). MUSIC algorithm for DOA estimation using MIMO arrays. In 2011 6*th International Conference on Telecommunication Systems*, *Services*, *and Applications* (*TSSA*) 149–153.

[63] 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

The Annals of Statistics
2021, Vol. 49, No. 1, 393–410
https://doi.org/10.1214/20-AOS1961

ROBUST MULTIVARIATE MEAN ESTIMATION: THE OPTIMALITY OF TRIMMED MEAN

BY GÁBOR LUGOSI[1] AND SHAHAR MENDELSON[2]

[1]*ICREA; Department of Economics and Business, Pompeu Fabra University; Barcelona Graduate School of Economics* *gabor.lugosi@upf.edu*

[2]*Mathematical Sciences Institute, Australian National University,* *shahar.mendelson@gmail.com*

We consider the problem of estimating the mean of a random vector based on i.i.d. observations and adversarial contamination. We introduce a multivariate extension of the trimmed-mean estimator and show its optimal performance under minimal conditions.

REFERENCES

[1] BICKEL, P. J. (1965). On some robust estimates of location. *Ann. Math. Stat.* **36** 847–858. MR0177484 https://doi.org/10.1214/aoms/1177700058

[2] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford. MR3185193 https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

[3] 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

[4] 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

[5] CHERAPANAMJERI, Y., FLAMMARION, N. and BARTLETT, P. (2019). Fast mean estimation with sub-gaussian rates. Preprint. Available at arXiv:1902.01998.

[6] DALALYAN, A. and THOMPSON, P. (2019). Outlier-robust estimation of a sparse linear model using ℓ_1-penalized Huber's M-estimator. In *Advances in Neural Information Processing Systems* 13188–13198.

[7] DEPERSIN, J. and LECUÉ, G. (2019). Robust subgaussian estimation of a mean vector in nearly linear time. Preprint. Available at arXiv:1906.03058.

[8] DEVROYE, L., LERASLE, M., LUGOSI, G. and OLIVEIRA, R. I. (2016). Sub-Gaussian mean estimators. *Ann. Statist.* **44** 2695–2725. MR3576558 https://doi.org/10.1214/16-AOS1440

[9] DIAKONIKOLAS, I., KAMATH, G., KANE, D. M., LI, J., MOITRA, A. and STEWART, A. (2016). Robust estimators in high dimensions without the computational intractability. In 57*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2016 655–664. IEEE Computer Soc., Los Alamitos, CA. MR3631028 https://doi.org/10.1109/FOCS.2016.85

[10] 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* (*ICML* 2017).

[11] DIAKONIKOLAS, I., KAMATH, G., KANE, D. M., LI, J., MOITRA, A. and STEWART, A. (2018). 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

[12] GINÉ, E. and ZINN, J. (1984). Some limit theorems for empirical processes. *Ann. Probab.* **12** 929–998. MR0757767

[13] GOLDENSHLUGER, A. and NEMIROVSKI, A. (1997). On spatially adaptive estimation of nonparametric regression. *Math. Methods Statist.* **6** 135–170. MR1466625

[14] HOPKINS, S. B. (2020). Sub-Gaussian mean estimation in polynomial time. *Ann. Statist.* To appear.

[15] HUBER, P. J. and RONCHETTI, E. M. (2009). *Robust Statistics*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2488795 https://doi.org/10.1002/9780470434697

[16] KEARNS, M. and LI, M. (1993). Learning in the presence of malicious errors. *SIAM J. Comput.* **22** 807–837. MR1227763 https://doi.org/10.1137/0222052

MSC2020 subject classifications. Primary 62J02, 62G08; secondary 60G25.
Key words and phrases. Mean estimation, robust estimation.

[17] LECUÉ, G. and LERASLE, M. (2020). Robust machine learning by median-of-means: Theory and practice. *Ann. Statist.* **48** 906–931.

[18] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes. Ergebnisse der Mathematik und Ihrer Grenzgebiete* (3) [*Results in Mathematics and Related Areas* (3)] **23**. Springer, Berlin. MR1102015 https://doi.org/10.1007/978-3-642-20212-4

[19] LEPSKIĬ, O. V. (1991). Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.* **36** 682–697.

[20] 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

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

[22] MINSKER, S. (2018). Uniform bounds for robust mean estimators. Preprint. Available at arXiv:1812.03523.

[23] OLIVEIRA, R. I. and ORENSTEIN, P. (2019). The sub-gaussian property of trimmed means estimators. Technical report, IMPA.

[24] RODRIGUEZ, D. and VALDORA, M. (2019). The breakdown point of the median of means tournament. *Statist. Probab. Lett.* **153** 108–112. MR3962785 https://doi.org/10.1016/j.spl.2019.05.012

[25] 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, 21. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR3761781

[26] STIGLER, S. M. (1973). The asymptotic distribution of the trimmed mean. *Ann. Statist.* **1** 472–477. MR0359134

[27] TALAGRAND, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126** 505–563. MR1419006 https://doi.org/10.1007/s002220050108

[28] TUKEY, J. W. and MCLAUGHLIN, D. H. (1963). Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/Winsorization. I. *Sankhyā Ser. A* **25** 331–352. MR0169354

[29] VALIANT, L. G. (1985). Learning disjunction of conjunctions. In *IJCAI* 560–566.

The Annals of Statistics
2021, Vol. 49, No. 1, 411–434
https://doi.org/10.1214/20-AOS1962

CLASSIFICATION ACCURACY AS A PROXY FOR TWO-SAMPLE TESTING

BY ILMUN KIM[1,*], AADITYA RAMDAS[1,†], AARTI SINGH[1,‡] AND LARRY WASSERMAN[1,§]

[1]*Department of Statistics & Data Science, Carnegie Mellon University,* [*]*ilmunk@stat.cmu.edu;* [†]*aramdas@stat.cmu.edu;* [‡]*aarti@cs.cmu.edu;* [§]*larry@stat.cmu.edu*

When data analysts train a classifier and check if its accuracy is significantly different from chance, they are implicitly performing a two-sample test. We investigate the statistical properties of this flexible approach in the high-dimensional setting. We prove two results that hold for all classifiers in any dimensions: if its true error remains ϵ-better than chance for some $\epsilon > 0$ as $d, n \to \infty$, then (a) the permutation-based test is consistent (has power approaching to one), (b) a computationally efficient test based on a Gaussian approximation of the null distribution is also consistent. To get a finer understanding of the rates of consistency, we study a specialized setting of distinguishing Gaussians with mean-difference δ and common (known or unknown) covariance Σ, when $d/n \to c \in (0, \infty)$. We study variants of Fisher's linear discriminant analysis (LDA) such as "naive Bayes" in a nontrivial regime when $\epsilon \to 0$ (the Bayes classifier has true accuracy approaching 1/2), and contrast their power with corresponding variants of Hotelling's test. Surprisingly, the expressions for their power match exactly in terms of n, d, δ, Σ, and the LDA approach is only worse by a constant factor, achieving an asymptotic relative efficiency (ARE) of $1/\sqrt{\pi}$ for balanced samples. We also extend our results to high-dimensional elliptical distributions with finite kurtosis. Other results of independent interest include minimax lower bounds, and the optimality of Hotelling's test when $d = o(n)$. Simulation results validate our theory, and we present practical takeaway messages along with natural open problems.

REFERENCES

ANDERSON, T. W. (1951). Classification by multivariate analysis. *Psychometrika* **16** 31–50. MR0041403 https://doi.org/10.1007/BF02313425

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

ARIAS-CASTRO, E., PELLETIER, B. and SALIGRAMA, V. (2018). Remember the curse of dimensionality: The case of goodness-of-fit testing in arbitrary dimension. *J. Nonparametr. Stat.* **30** 448–471. MR3794401 https://doi.org/10.1080/10485252.2018.1435875

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

BEN-DAVID, S., BLITZER, J., CRAMMER, K. and PEREIRA, F. (2007). Analysis of representations for domain adaptation. In *Advances in Neural Information Processing Systems* 137–144.

BHATTACHARYA, B. B. (2020). Asymptotic distribution and detection thresholds for two-sample tests based on geometric graphs. *Ann. Statist.* **40** 2879–2903. MR4152627 https://doi.org/10.1214/19-AOS1913

BICKEL, P. J. and LEVINA, E. (2004). Some theory of Fisher's linear discriminant function, 'naive Bayes', and some alternatives when there are many more variables than observations. *Bernoulli* **10** 989–1010. MR2108040 https://doi.org/10.3150/bj/1106314847

BLANCHARD, G., LEE, G. and SCOTT, C. (2010). Semi-supervised novelty detection. *J. Mach. Learn. Res.* **11** 2973–3009. MR2746544

BORJI, A. (2019). Pros and cons of GAN evaluation measures. *Comput. Vis. Image Underst.* **179** 41–65.

MSC2020 subject classifications. Primary 62H15; secondary 62E20.
Key words and phrases. Classification accuracy, two sample testing, high-dimensional asymptotics, Hotelling's T^2 test, linear discriminant analysis, permutation test.

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

CHEN, N. F., SHEN, W., CAMPBELL, J. and SCHWARTZ, R. (2009). Large-scale analysis of formant frequency estimation variability in conversational telephone speech. In *Tenth Annual Conference of the International Speech Communication Association*.

ETZEL, J. A., GAZZOLA, V. and KEYSERS, C. (2009). An introduction to anatomical ROI-based fMRI classification analysis. *Brain Res.* **1282** 114–125.

FANG, K. T., KOTZ, S. and NG, K. W. (2018). *Symmetric Multivariate and Related Distributions*. Chapman and Hall/CRC.

FISHER, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annu. Eugen.* **7** 179–188.

FISHER, R. A. (1940). The precision of discriminant functions. *Annu. Eugen.* **10** 422–429. MR0003543

FRAHM, G. (2004). Generalized elliptical distributions: Theory and applications. Ph.D. thesis, Universität zu Köln.

FRIEDMAN, J. (2004). On multivariate goodness-of-fit and two-sample testing. Technical report, Stanford Linear Accelerator Center, Menlo Park, CA (US).

FRIEDMAN, J. H. and RAFSKY, L. C. (1979). Multivariate generalizations of the Wald–Wolfowitz and Smirnov two-sample tests. *Ann. Statist.* **7** 697–717. MR0532236

GAGNON-BARTSCH, J. and SHEM-TOV, Y. (2019). The classification permutation test: A flexible approach to testing for covariate imbalance in observational studies. *Ann. Appl. Stat.* **13** 1464–1483. MR4019146 https://doi.org/10.1214/19-AOAS1241

GIRI, N. and KIEFER, J. (1964). Local and asymptotic minimax properties of multivariate tests. *Ann. Math. Stat.* **35** 21–35. MR0159388 https://doi.org/10.1214/aoms/1177703730

GIRI, N., KIEFER, J. and STEIN, C. (1963). Minimax character of Hotelling's T^2 test in the simplest case. *Ann. Math. Stat.* **34** 1524–1535. MR0156408 https://doi.org/10.1214/aoms/1177703884

GOLLAND, P. and FISCHL, B. (2003). Permutation tests for classification: Towards statistical significance in image-based studies. In *Biennial International Conference on Information Processing in Medical Imaging* 330–341. Springer, New York.

GÓMEZ, E., GÓMEZ-VILLEGAS, M. A. and MARÍN, J. M. (2003). A survey on continuous elliptical vector distributions. *Rev. Mat. Complut.* **16** 345–361. MR2031887 https://doi.org/10.5209/rev_REMA.2003.v16.n1.16889

GRETTON, A., BORGWARDT, K. M., RASCH, M. J., SCHÖLKOPF, B. and SMOLA, A. (2012). A kernel two-sample test. *J. Mach. Learn. Res.* **13** 723–773. MR2913716

HEDIGER, S., MICHEL, L. and NÄF, J. (2019). On the use of random forest for two-sample testing. arXiv preprint, arXiv:1903.06287.

HEMERIK, J. and GOEMAN, J. J. (2018). 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

HENZE, N. (1988). A multivariate two-sample test based on the number of nearest neighbor type coincidences. *Ann. Statist.* **16** 772–783. MR0947577 https://doi.org/10.1214/aos/1176350835

HOTELLING, H. (1931). The generalization of student's ratio. *Ann. Math. Stat.* **2** 360–378.

HU, J. and BAI, Z. (2016). A review of 20 years of naive tests of significance for high-dimensional mean vectors and covariance matrices. *Sci. China Math.* **59** 2281–2300. MR3578957 https://doi.org/10.1007/s11425-016-0131-0

KARIYA, T. (1981). A robustness property of Hotelling's T^2-test. *Ann. Statist.* **9** 211–214. MR0600550

KIM, I., RAMDAS, A., SINGH, A. and WASSERMAN, L. (2021). Supplement to "Classification accuracy as a proxy for two-sample testing." https://doi.org/10.1214/20-AOS1962SUPP

LIU, Y., LI, C.-L. and PÓCZOS, B. (2018). Classifier two-sample test for video anomaly detections. In *British Machine Vision Conference* 2018, *BMVC* 2018 71. Northumbria Univ., Newcastle, UK.

LOPEZ-PAZ, D. and OQUAB, M. (2016). Revisiting classifier two-sample tests. arXiv preprint, arXiv:1610.06545.

LUSCHGY, H. (1982). Minimax character of the two-sample χ^2-test. *Stat. Neerl.* **36** 129–134. MR0673752 https://doi.org/10.1111/j.1467-9574.1982.tb00784.x

OLIVETTI, E., GREINER, S. and AVESANI, P. (2012). Induction in neuroscience with classification: Issues and solutions. In *Machine Learning and Interpretation in Neuroimaging* 42–50. Springer, New York.

PEREIRA, F., MITCHELL, T. and BOTVINICK, M. (2009). Machine learning classifiers and fMRI: A tutorial overview. *NeuroImage* **45** S199–S209.

RAUDYS, Š. and YOUNG, D. M. (2004). Results in statistical discriminant analysis: A review of the former Soviet Union literature. *J. Multivariate Anal.* **89** 1–35. MR2041207 https://doi.org/10.1016/S0047-259X(02)00021-0

ROSENBAUM, P. R. (2005). An exact distribution-free test comparing two multivariate distributions based on adjacency. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 515–530. MR2168202 https://doi.org/10.1111/j.1467-9868.2005.00513.x

ROSENBLATT, J. D., BENJAMINI, Y., GILRON, R., MUKAMEL, R. and GOEMAN, J. J. (2019). Better-than-chance classification for signal detection. *Biostatistics*. https://doi.org/10.1093/biostatistics/kxz035

SALAEVSKII, O. (1969). Minimax character of Hotelling's T^2 test. I. In *Investigations in Classical Problems of Probability Theory and Mathematical Statistics* 74–101. Springer, New York.

SCHILLING, M. F. (1986). Multivariate two-sample tests based on nearest neighbors. *J. Amer. Statist. Assoc.* **81** 799–806. MR0860514

SCOTT, C. and NOWAK, R. (2005). A Neyman–Pearson approach to statistical learning. *IEEE Trans. Inf. Theory* **51** 3806–3819. MR2239000 https://doi.org/10.1109/TIT.2005.856955

SIMAIKA, J. B. (1941). On an optimum property of two important statistical tests. *Biometrika* **32** 70–80. MR0003547 https://doi.org/10.1093/biomet/32.1.70

SRIPERUMBUDUR, B. K., FUKUMIZU, K., GRETTON, A., LANCKRIET, G. R. and SCHÖLKOPF, B. (2009). Kernel choice and classifiability for RKHS embeddings of probability distributions. In *Advances in Neural Information Processing Systems* 1750–1758.

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

SRIVASTAVA, M. S., KATAYAMA, S. and KANO, Y. (2013). A two sample test in high dimensional data. *J. Multivariate Anal.* **114** 349–358. MR2993891 https://doi.org/10.1016/j.jmva.2012.08.014

STELZER, J., CHEN, Y. and TURNER, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): Random permutations and cluster size control. *NeuroImage* **65** 69–82. https://doi.org/10.1016/j.neuroimage.2012.09.063

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

WALD, A. (1944). On a statistical problem arising in the classification of an individual into one of two groups. *Ann. Math. Stat.* **15** 145–162. MR0010371 https://doi.org/10.1214/aoms/1177731280

XIAO, J., WANG, R., TENG, G. and HU, Y. (2014). A transfer learning based classifier ensemble model for customer credit scoring. In 2014 *Seventh International Joint Conference on Computational Sciences and Optimization* 64–68. IEEE.

XIAO, J., XIAO, Y., HUANG, A., LIU, D. and WANG, S. (2015). Feature-selection-based dynamic transfer ensemble model for customer churn prediction. *Knowl. Inf. Syst.* **43** 29–51.

YU, K., MARTIN, R., ROTHMAN, N., ZHENG, T. and LAN, Q. (2007). Two-sample comparison based on prediction error, with applications to candidate gene association studies. *Ann. Hum. Genet.* **71** 107–118.

ZHU, C.-Z., ZANG, Y.-F., CAO, Q.-J., YAN, C.-G., HE, Y., JIANG, T.-Z., SUI, M.-Q. and WANG, Y.-F. (2008). Fisher discriminative analysis of resting-state brain function for attention-deficit/hyperactivity disorder. *NeuroImage* **40** 110–120.

ZOGRAFOS, K. (2008). On Mardia's and Song's measures of kurtosis in elliptical distributions. *J. Multivariate Anal.* **99** 858–879. MR2405095 https://doi.org/10.1016/j.jmva.2007.05.001

ZOLLANVARI, A., BRAGA-NETO, U. M. and DOUGHERTY, E. R. (2011). Analytic study of performance of error estimators for linear discriminant analysis. *IEEE Trans. Signal Process.* **59** 4238–4255. MR2865981 https://doi.org/10.1109/TSP.2011.2159210

The Annals of Statistics
2021, Vol. 49, No. 1, 435–458
https://doi.org/10.1214/20-AOS1963

ASYMMETRY HELPS: EIGENVALUE AND EIGENVECTOR ANALYSES OF ASYMMETRICALLY PERTURBED LOW-RANK MATRICES[1]

BY YUXIN CHEN[1] CHEN CHENG[2] AND JIANQING FAN[3]

[1]*Department of Electrical Engineering, Princeton University, yuxin.chen@princeton.edu*
[2]*Department of Statistics, Stanford University, chencheng@stanford.edu*
[3]*Department of Operations Research & Financial Engineering, Princeton University, jqfan@princeton.edu*

This paper is concerned with the interplay between statistical asymmetry and spectral methods. Suppose we are interested in estimating a rank-1 and symmetric matrix $\boldsymbol{M}^\star \in \mathbb{R}^{n\times n}$, yet only a randomly perturbed version $\boldsymbol{M}$ is observed. The noise matrix $\boldsymbol{M} - \boldsymbol{M}^\star$ is composed of independent (but not necessarily homoscedastic) entries and is, therefore, not symmetric in general. This might arise if, for example, when we have two independent samples for each entry of $\boldsymbol{M}^\star$ and arrange them in an *asymmetric* fashion. The aim is to estimate the leading eigenvalue and the leading eigenvector of $\boldsymbol{M}^\star$.

We demonstrate that the leading eigenvalue of the data matrix $\boldsymbol{M}$ can be $O(\sqrt{n})$ times more accurate (up to some log factor) than its (unadjusted) leading singular value of $\boldsymbol{M}$ in eigenvalue estimation. Moreover, the eigendecomposition approach is fully adaptive to heteroscedasticity of noise, without the need of any prior knowledge about the noise distributions. In a nutshell, this curious phenomenon arises since the statistical asymmetry automatically mitigates the bias of the eigenvalue approach, thus eliminating the need of careful bias correction. Additionally, we develop appealing nonasymptotic eigenvector perturbation bounds; in particular, we are able to bound the perturbation of any linear function of the leading eigenvector of $\boldsymbol{M}$ (e.g., entrywise eigenvector perturbation). We also provide partial theory for the more general rank-r case. The takeaway message is this: arranging the data samples in an asymmetric manner and performing eigendecomposition could sometimes be quite beneficial.

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

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

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

BAO, Z., DING, X. and WANG, K. (2018). Singular vector and singular subspace distribution for the matrix denoising model. Preprint. Available at arXiv:1809.10476.

BAUER, F. L. and FIKE, C. T. (1960). Norms and exclusion theorems. *Numer. Math.* **2** 137–141. MR0118729 https://doi.org/10.1007/BF01386217

BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** 494–521. MR2782201 https://doi.org/10.1016/j.aim.2011.02.007

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 62H12.
Key words and phrases. Spectral methods, eigenvalue perturbation, entrywise eigenvector perturbation, linear form of eigenvectors, heteroscedasticity.

BENAYCH-GEORGES, F. and ROCHET, J. (2016). Outliers in the single ring theorem. *Probab. Theory Related Fields* **165** 313–363. MR3500273 https://doi.org/10.1007/s00440-015-0632-x

BORDENAVE, C. and CAPITAINE, M. (2016). Outlier eigenvalues for deformed i.i.d. random matrices. *Comm. Pure Appl. Math.* **69** 2131–2194. MR3552011 https://doi.org/10.1002/cpa.21629

BRÉZIN, E. and ZEE, A. (1998). Non-Hermitean delocalization: Multiple scattering and bounds. *Nuclear Phys. B* **509** 599–614. MR1487025 https://doi.org/10.1016/S0550-3213(97)00652-4

BRYC, W. and SILVERSTEIN, J. W. (2018). Singular values of large non-central random matrices. Preprint. Available at arXiv:1802.02960.

CAI, T., HAN, X. and PAN, G. (2017). Limiting laws for divergent spiked eigenvalues and largest non-spiked eigenvalue of sample covariance matrices. Preprint. Available at arXiv:1711.00217.

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. (2019). Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees. Preprint. Available at arXiv:1910.04267.

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). Signal-plus-noise matrix models: Eigenvector deviations and fluctuations. *Biometrika* **106** 243–250. MR3912394 https://doi.org/10.1093/biomet/asy070

CAPITAINE, M., DONATI-MARTIN, C. and FÉRAL, D. (2009). The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations. *Ann. Probab.* **37** 1–47. MR2489158 https://doi.org/10.1214/08-AOP394

CHEN, Y., CHENG, C. and FAN, J. (2020). Supplement to "Asymmetry helps: Eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices." https://doi.org/10.1214/20-AOS1963SUPP

CHEN, Y., CHI, Y., FAN, J., MA, C. and YAN, Y. (2019a). Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. Available at arXiv:1902.07698.

CHEN, Y., FAN, J., MA, C. and WANG, K. (2019b). 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

CHEN, Y., FAN, J., MA, C. and YAN, Y. (2019c). Inference and uncertainty quantification for noisy matrix completion. *Proc. Natl. Acad. Sci. USA* **116** 22931–22937. MR4036123 https://doi.org/10.1073/pnas.1910053116

CHENG, C., WEI, Y. and CHEN, Y. (2020). Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise. Preprint. Available at arXiv:2001.04620.

CHI, Y., LU, Y. M. and CHEN, Y. (2019). Nonconvex optimization meets low-rank matrix factorization: An overview. *IEEE Trans. Signal Process.* **67** 5239–5269. MR4016283 https://doi.org/10.1109/TSP.2019.2937282

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

ELDRIDGE, J., BELKIN, M. and WANG, Y. (2018). Unperturbed: Spectral analysis beyond Davis–Kahan. In *Algorithmic Learning Theory* 2018. *Proc. Mach. Learn. Res.* (*PMLR*) **83** 38. Proceedings of Machine Learning Research PMLR. MR3857310

ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2013). Spectral statistics of Erdős–Rényi graphs I: Local semicircle law. *Ann. Probab.* **41** 2279–2375. MR3098073 https://doi.org/10.1214/11-AOP734

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

FEINBERG, J. and ZEE, A. (1997). Non-Hermitian random matrix theory: Method of Hermitian reduction. *Nuclear Phys. B* **504** 579–608. MR1488584 https://doi.org/10.1016/S0550-3213(97)00502-6

FÉRAL, D. and PÉCHÉ, S. (2007). The largest eigenvalue of rank one deformation of large Wigner matrices. *Comm. Math. Phys.* **272** 185–228. MR2291807 https://doi.org/10.1007/s00220-007-0209-3

FÜREDI, Z. and KOMLÓS, J. (1981). The eigenvalues of random symmetric matrices. *Combinatorica* **1** 233–241. MR0637828 https://doi.org/10.1007/BF02579329

JAIN, P. and NETRAPALLI, P. (2015). Fast exact matrix completion with finite samples. In *Conference on Learning Theory* 1007–1034.

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

KESHAVAN, R. H., MONTANARI, A. and OH, S. (2010). Matrix completion from a few entries. *IEEE Trans. Inf. Theory* **56** 2980–2998. MR2683452 https://doi.org/10.1109/TIT.2010.2046205

KHORUZHENKO, B. (1996). Large-N eigenvalue distribution of randomly perturbed asymmetric matrices. *J. Phys. A* **29** L165–L169. MR1395506 https://doi.org/10.1088/0305-4470/29/7/003

KNOWLES, A. and YIN, J. (2013). The isotropic semicircle law and deformation of Wigner matrices. *Comm. Pure Appl. Math.* **66** 1663–1750. MR3103909 https://doi.org/10.1002/cpa.21450

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

LI, X., LING, S., STROHMER, T. and WEI, K. (2019). Rapid, robust, and reliable blind deconvolution via nonconvex optimization. *Appl. Comput. Harmon. Anal.* **47** 893–934. MR3994997 https://doi.org/10.1016/j.acha.2018.01.001

LYTOVA, A. and TIKHOMIROV, K. (2018). On delocalization of eigenvectors of random non-Hermitian matrices. Preprint. Available at arXiv:1810.01590.

MA, C., WANG, K., CHI, Y. and CHEN, Y. (2020). Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion and blind deconvolution. *Found. Comput. Math.* **20** 451–632.

MEHLIG, B. and CHALKER, J. T. (1998). Eigenvector correlations in non-Hermitian random matrix ensembles. *Ann. Phys.* **7** 427–436. MR1666250 https://doi.org/10.1002/(SICI)1521-3889(199811)7:5/6<427::AID-ANDP427>3.0.CO;2-1

O'ROURKE, S., VU, V. and WANG, K. (2016). Eigenvectors of random matrices: A survey. *J. Combin. Theory Ser. A* **144** 361–442. MR3534074 https://doi.org/10.1016/j.jcta.2016.06.008

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

PÉCHÉ, S. (2006). The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probab. Theory Related Fields* **134** 127–173. MR2221787 https://doi.org/10.1007/s00440-005-0466-z

RAJAGOPALAN, A. B. (2015). Outlier eigenvalue fluctuations of perturbed iid matrices. Preprint. Available at arXiv:1507.01441.

RENFREW, D. and SOSHNIKOV, A. (2013). On finite rank deformations of Wigner matrices II: Delocalized perturbations. *Random Matrices Theory Appl.* **2** 1250015, 36. MR3039820 https://doi.org/10.1142/S2010326312500153

SILVERSTEIN, J. W. (1994). The spectral radii and norms of large-dimensional non-central random matrices. *Comm. Statist. Stochastic Models* **10** 525–532. MR1284550 https://doi.org/10.1080/15326349408807308

SOMMERS, H.-J., CRISANTI, A., SOMPOLINSKY, H. and STEIN, Y. (1988). Spectrum of large random asymmetric matrices. *Phys. Rev. Lett.* **60** 1895–1898. MR0948613 https://doi.org/10.1103/PhysRevLett.60.1895

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

TAO, T. (2013). Outliers in the spectrum of iid matrices with bounded rank perturbations. *Probab. Theory Related Fields* **155** 231–263. MR3010398 https://doi.org/10.1007/s00440-011-0397-9

TROPP, J. A. (2015). An introduction to matrix concentration inequalities. *Found. Trends Mach. Learn.* **8** 1–230.

VU, V. (2011). Singular vectors under random perturbation. *Random Structures Algorithms* **39** 526–538. MR2846302 https://doi.org/10.1002/rsa.20367

VU, V. and WANG, K. (2015). Random weighted projections, random quadratic forms and random eigenvectors. *Random Structures Algorithms* **47** 792–821. MR3418916 https://doi.org/10.1002/rsa.20561

WANG, R. (2015). Singular vector perturbation under Gaussian noise. *SIAM J. Matrix Anal. Appl.* **36** 158–177. MR3310977 https://doi.org/10.1137/130938177

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

XIA, D. (2016). Statistical inference for large matrices. Ph.D. thesis, Georgia Institute of Technology.

XIA, D. (2019). Confidence region of singular subspaces for low-rank matrix regression. *IEEE Trans. Inf. Theory* **65** 7437–7459. MR4030894 https://doi.org/10.1109/TIT.2019.2924900

YIN, Y. Q., BAI, Z. D. and KRISHNAIAH, P. R. (1988). On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Related Fields* **78** 509–521. MR0950344 https://doi.org/10.1007/BF00353874

ZHANG, A. and XIA, D. (2018). Tensor SVD: Statistical and computational limits. *IEEE Trans. Inf. Theory* **64** 7311–7338. MR3876445 https://doi.org/10.1109/TIT.2018.2841377

The Annals of Statistics
2021, Vol. 49, No. 1, 459–485
https://doi.org/10.1214/20-AOS1964

COMPLEX SAMPLING DESIGNS: UNIFORM LIMIT THEOREMS AND APPLICATIONS

BY QIYANG HAN[1] AND JON A. WELLNER[2]

[1]*Department of Statistics, Rutgers University, qh85@stat.rutgers.edu*
[2]*Department of Statistics, University of Washington, jaw@stat.washington.edu*

In this paper, we develop a general approach to proving global and local uniform limit theorems for the Horvitz–Thompson empirical process arising from complex sampling designs. Global theorems such as Glivenko–Cantelli and Donsker theorems, and local theorems such as local asymptotic modulus and related ratio-type limit theorems are proved for both the Horvitz–Thompson empirical process, and its calibrated version. Limit theorems of other variants and their conditional versions are also established. Our approach reveals an interesting feature: the problem of deriving uniform limit theorems for the Horvitz–Thompson empirical process is essentially no harder than the problem of establishing the corresponding finite-dimensional limit theorems, once the usual complexity conditions on the function class are satisfied. These global and local uniform limit theorems are then applied to important statistical problems including (i) M-estimation, (ii) Z-estimation and (iii) frequentist theory of pseudo-Bayes procedures, all with weighted likelihood, to illustrate their wide applicability.

REFERENCES

[1] ALEXANDER, K. S. (1985). Rates of growth for weighted empirical processes. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, *Vol. II* (*Berkeley*, *Calif.*, 1983). *Wadsworth Statist./Probab. Ser.* 475–493. Wadsworth, Belmont, CA. MR0822047

[2] ALEXANDER, K. S. (1987). The central limit theorem for weighted empirical processes indexed by sets. *J. Multivariate Anal.* **22** 313–339. MR0899666 https://doi.org/10.1016/0047-259X(87)90093-5

[3] ALEXANDER, K. S. (1987). Rates of growth and sample moduli for weighted empirical processes indexed by sets. *Probab. Theory Related Fields* **75** 379–423. MR0890285 https://doi.org/10.1007/BF00318708

[4] BARRETT, G. F. and DONALD, S. G. (2009). Statistical inference with generalized Gini indices of inequality, poverty, and welfare. *J. Bus. Econom. Statist.* **27** 1–17. MR2484980 https://doi.org/10.1198/jbes.2009.0001

[5] BERGER, Y. G. (1998). Rate of convergence for asymptotic variance of the Horvitz–Thompson estimator. *J. Statist. Plann. Inference* **74** 149–168. MR1665125 https://doi.org/10.1016/S0378-3758(98)00107-4

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

[7] BERTAIL, P., CHAUTRU, E. and CLÉMENÇON, S. (2017). Empirical processes in survey sampling with (conditional) Poisson designs. *Scand. J. Stat.* **44** 97–111. MR3619696 https://doi.org/10.1111/sjos.12243

[8] BHATTACHARYA, D. (2007). Inference on inequality from household survey data. *J. Econometrics* **137** 674–707. MR2354960 https://doi.org/10.1016/j.jeconom.2005.09.003

[9] BHATTACHARYA, D. and MAZUMDER, B. (2011). A nonparametric analysis of black–white differences in intergenerational income mobility in the United States. *Quant. Econ.* **2** 335–379.

[10] BOISTARD, H., LOPUHAÄ, H. P. and RUIZ-GAZEN, A. (2012). Approximation of rejective sampling inclusion probabilities and application to high order correlations. *Electron. J. Stat.* **6** 1967–1983. MR3020253 https://doi.org/10.1214/12-EJS736

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

MSC2020 subject classifications. Primary 60E15; secondary 62G05.
Key words and phrases. Complex sampling design, empirical process, uniform limit theorems.

[12] BREIDT, F. J. and OPSOMER, J. D. (2000). Local polynomial regresssion estimators in survey sampling. *Ann*. *Statist*. **28** 1026–1053. MR1810918 https://doi.org/10.1214/aos/1015956706

[13] BRESLOW, N., MCNENEY, B. and WELLNER, J. A. (2003). Large sample theory for semiparametric regression models with two-phase, outcome dependent sampling. *Ann*. *Statist*. **31** 1110–1139. MR2001644 https://doi.org/10.1214/aos/1059655907

[14] BRESLOW, N. E., LUMLEY, T., BALLANTYNE, C. M., CHAMBLESS, L. E. and KULICH, M. (2009). Improved Horvitz–Thompson estimation of model parameters from two-phase stratified samples: Applications in epidemiology. *Stat*. *Biosci*. **1** 32–49.

[15] BRESLOW, N. E., LUMLEY, T., BALLANTYNE, C. M., CHAMBLESS, L. E. and KULICH, M. (2009). Using the whole cohort in the analysis of case-cohort data. *Am*. *J*. *Epidemiol*. **169** 1398–1405.

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

[17] BRESLOW, N. E. and WELLNER, J. A. (2008). A Z-theorem with estimated nuisance parameters and correction note for: "Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression" [Scand. J. Statist. **34** (2007), no. 1, 86–102; MR2325244]. *Scand*. *J*. *Stat*. **35** 186–192. MR2391566 https://doi.org/10.1111/j.1467-9469.2007.00574.x

[18] CARDOT, H., CHAOUCH, M., GOGA, C. and LABRUÈRE, C. (2010). Properties of design-based functional principal components analysis. *J*. *Statist*. *Plann*. *Inference* **140** 75–91. MR2568123 https://doi.org/10.1016/j.jspi.2009.06.012

[19] CASTILLO, I. (2012). A semiparametric Bernstein–von Mises theorem for Gaussian process priors. *Probab*. *Theory Related Fields* **152** 53–99. MR2875753 https://doi.org/10.1007/s00440-010-0316-5

[20] CASTILLO, I. and NICKL, R. (2013). Nonparametric Bernstein–von Mises theorems in Gaussian white noise. *Ann*. *Statist*. **41** 1999–2028. MR3127856 https://doi.org/10.1214/13-AOS1133

[21] CASTILLO, I. and NICKL, R. (2014). On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. *Ann*. *Statist*. **42** 1941–1969. MR3262473 https://doi.org/10.1214/14-AOS1246

[22] CASTILLO, I. and ROUSSEAU, J. (2015). A Bernstein–von Mises theorem for smooth functionals in semiparametric models. *Ann*. *Statist*. **43** 2353–2383. MR3405597 https://doi.org/10.1214/15-AOS1336

[23] CHAUVET, G. (2015). Coupling methods for multistage sampling. *Ann*. *Statist*. **43** 2484–2506. MR3405601 https://doi.org/10.1214/15-AOS1348

[24] CHENG, G. and HUANG, J. Z. (2010). Bootstrap consistency for general semiparametric M-estimation. *Ann*. *Statist*. **38** 2884–2915. MR2722459 https://doi.org/10.1214/10-AOS809

[25] CLÉMENÇON, S., BERTAIL, P. and PAPA, G. (2016). Learning from survey training samples: Rate bounds for Horvitz–Thompson risk minimizers. In *Asian Conference on Machine Learning* 142–157.

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

[27] DAVIDSON, R. (2009). Reliable inference for the Gini index. *J*. *Econometrics* **150** 30–40. MR2525992 https://doi.org/10.1016/j.jeconom.2008.11.004

[28] DEVILLE, J.-C. and SÄRNDAL, C.-E. (1992). Calibration estimators in survey sampling. *J*. *Amer*. *Statist*. *Assoc*. **87** 376–382. MR1173804

[29] DEVROYE, L., GYÖRFI, L. and LUGOSI, G. (1996). *A Probabilistic Theory of Pattern Recognition*. *Applications of Mathematics* (*New York*) **31**. Springer, New York. MR1383093 https://doi.org/10.1007/978-1-4612-0711-5

[30] FULLER, W. A. (2011). *Sampling Statistics* **560**. Wiley.

[31] GHOSAL, S., GHOSH, J. K. and VAN DER VAART, A. W. (2000). Convergence rates of posterior distributions. *Ann*. *Statist*. **28** 500–531. MR1790007 https://doi.org/10.1214/aos/1016218228

[32] GHOSAL, S. and VAN DER VAART, A. (2007). Convergence rates of posterior distributions for non-i.i.d. observations. *Ann*. *Statist*. **35** 192–223. MR2332274 https://doi.org/10.1214/009053606000001172

[33] GINÉ, E. and KOLTCHINSKII, V. (2006). Concentration inequalities and asymptotic results for ratio type empirical processes. *Ann*. *Probab*. **34** 1143–1216. MR2243881 https://doi.org/10.1214/009117906000000070

[34] GINÉ, E., KOLTCHINSKII, V. and WELLNER, J. A. (2003). Ratio limit theorems for empirical processes. In *Stochastic Inequalities and Applications*. *Progress in Probability* **56** 249–278. Birkhäuser, Basel. MR2073436

[35] 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

[36] HÁJEK, J. (1961). Some extensions of the Wald–Wolfowitz–Noether theorem. *Ann*. *Math*. *Stat*. **32** 506–523. MR0130707 https://doi.org/10.1214/aoms/1177705057

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

[38] HÁJEK, J. (1981). *Sampling from a Finite Population*. *Statistics*: *Textbooks and Monographs* **37**. Dekker, New York. Edited by Václav Dupač, With a foreword by P. K. Sen. MR0627744

[39] HAN, Q. (2017). Bayes model selection. arXiv preprint, arXiv:1704.07513.

[40] HAN, Q. and WELLNER, J. A. (2019). Convergence rates of least squares regression estimators with heavy-tailed errors. *Ann. Statist.* **47** 2286–2319. MR3953452 https://doi.org/10.1214/18-AOS1748

[41] HAN, Q. and WELLNER, J. A. (2021). Supplement to "Complex sampling designs: Uniform limit theorems and applications." https://doi.org/10.1214/20-AOS1964SUPP

[42] HARTLEY, H. O. (1962). Multiple frame surveys. In *Proceedings of the Social Statistics Section* **19** 203–206. American Statistical Association, Washington, DC.

[43] HARTLEY, H. O. (1974). Multiple frame methodology and selected applications. *Sankhyā* **36** 118.

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

[45] HUANG, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Ann. Statist.* **24** 540–568. MR1394975 https://doi.org/10.1214/aos/1032894452

[46] KOLTCHINSKII, V. (2006). Local Rademacher complexities and oracle inequalities in risk minimization. *Ann. Statist.* **34** 2593–2656. MR2329442 https://doi.org/10.1214/009053606000001019

[47] KOLTCHINSKII, V. (2011). *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems*. *Lecture Notes in Math.* **2033**. Springer, Heidelberg. Lectures from the 38th Probability Summer School held in Saint-Flour, 2008, École d'Été de Probabilités de Saint-Flour. [Saint-Flour Probability Summer School]. MR2829871 https://doi.org/10.1007/978-3-642-22147-7

[48] KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. *Springer Series in Statistics*. Springer, New York. MR2724368 https://doi.org/10.1007/978-0-387-74978-5

[49] LEÓN-NOVELO, L. G. and SAVITSKY, T. D. (2019). Fully Bayesian estimation under informative sampling. *Electron. J. Stat.* **13** 1608–1645. MR3939589 https://doi.org/10.1214/19-ejs1538

[50] LIN, D. Y. (2000). On fitting Cox's proportional hazards models to survey data. *Biometrika* **87** 37–47. MR1766826 https://doi.org/10.1093/biomet/87.1.37

[51] LOHR, S. and RAO, J. N. K. (2006). Estimation in multiple-frame surveys. *J. Amer. Statist. Assoc.* **101** 1019–1030. MR2324141 https://doi.org/10.1198/016214506000000195

[52] LUMLEY, T., SHAW, P. A. and DAI, J. Y. (2011). Connections between survey calibration estimators and semiparametric models for incomplete data. *Int. Stat. Rev.* **79** 200–220. https://doi.org/10.1111/j.1751-5823.2011.00138.x

[53] MA, S. and KOSOROK, M. R. (2005). Robust semiparametric M-estimation and the weighted bootstrap. *J. Multivariate Anal.* **96** 190–217. MR2202406 https://doi.org/10.1016/j.jmva.2004.09.008

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

[55] MASON, D. M., SHORACK, G. R. and WELLNER, J. A. (1983). Strong limit theorems for oscillation moduli of the uniform empirical process. *Z. Wahrsch. Verw. Gebiete* **65** 83–97. MR0717935 https://doi.org/10.1007/BF00534996

[56] NAN, B., KALBFLEISCH, J. D. and YU, M. (2009). Asymptotic theory for the semiparametric accelerated failure time model with missing data. *Ann. Statist.* **37** 2351–2376. MR2543695 https://doi.org/10.1214/08-AOS657

[57] NAN, B. and WELLNER, J. A. (2013). A general semiparametric Z-estimation approach for case-cohort studies. *Statist. Sinica* **23** 1155–1180. MR3114709

[58] NICKL, R. (2017). Bernstein–von Mises theorems for statistical inverse problems I: Schrödinger equation, arXiv preprint arXiv:1707.01764.

[59] NICKL, R. and SÖHL, J. (2019). Bernstein–von Mises theorems for statistical inverse problems II: Compound Poisson processes. *Electron. J. Stat.* **13** 3513–3571. MR4013745 https://doi.org/10.1214/19-ejs1609

[60] PRÆSTGAARD, J. and WELLNER, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.* **21** 2053–2086. MR1245301

[61] ROSÉN, B. (1965). Limit theorems for sampling from finite populations. *Ark. Mat.* **5** 383–424. MR0177437 https://doi.org/10.1007/BF02591138

[62] ROSÉN, B. (1967). On the central limit theorem for a class of sampling procedures. *Z. Wahrsch. Verw. Gebiete* **7** 103–115. MR0210181 https://doi.org/10.1007/BF00536324

[63] ROSÉN, B. (1972). Asymptotic theory for successive sampling with varying probabilities without replacement. I, II. *Ann. Math. Stat.* **43** 373–397; ibid. 43 (1972), 748–776. MR0321223 https://doi.org/10.1214/aoms/1177692620

[64] ROSÉN, B. (1974). Asymptotic theory for Des Raj's estimator. I, II. *Scand. J. Stat.* **1** 71–83; ibid. 1 (1974), no. 3, 135–144. MR0375560

[65] RUBIN-BLEUER, S. and SCHIOPU KRATINA, I. (2005). On the two-phase framework for joint model and design-based inference. *Ann. Statist.* **33** 2789–2810. MR2253102 https://doi.org/10.1214/009053605000000651

[66] SAEGUSA, T. (2014). Bootstrapping two-phase sampling. arXiv preprint, arXiv:1406.5580.

[67] SAEGUSA, T. (2015). Variance estimation under two-phase sampling. *Scand. J. Stat.* **42** 1078–1091. MR3426311 https://doi.org/10.1111/sjos.12152

[68] SAEGUSA, T. (2019). Large sample theory for merged data from multiple sources. *Ann. Statist.* **47** 1585–1615. MR3911123 https://doi.org/10.1214/18-AOS1727

[69] SAEGUSA, T. and WELLNER, J. A. (2013). Weighted likelihood estimation under two-phase sampling. *Ann. Statist.* **41** 269–295. MR3059418 https://doi.org/10.1214/12-AOS1073

[70] SÄRNDAL, C.-E., SWENSSON, B. and WRETMAN, J. (1992). *Model Assisted Survey Sampling*. *Springer Series in Statistics*. Springer, New York. MR1140409 https://doi.org/10.1007/978-1-4612-4378-6

[71] SAVITSKY, T. D. and TOTH, D. (2016). Bayesian estimation under informative sampling. *Electron. J. Stat.* **10** 1677–1708. MR3522657 https://doi.org/10.1214/16-EJS1153

[72] SHORACK, G. R. (1973). Convergence of reduced empirical and quantile processes with application to functions of order statistics in the non-I.I.D. case. *Ann. Statist.* **1** 146–152. MR0336776

[73] SHORACK, G. R. and WELLNER, J. A. (1982). Limit theorems and inequalities for the uniform empirical process indexed by intervals. *Ann. Probab.* **10** 639–652. MR0659534

[74] STUTE, W. (1982). The oscillation behavior of empirical processes. *Ann. Probab.* **10** 86–107. MR0637378

[75] STUTE, W. (1984). The oscillation behavior of empirical processes: The multivariate case. *Ann. Probab.* **12** 361–379. MR0735843

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

[77] VAN DE GEER, S. A. (2000). *Applications of Empirical Process Theory*. *Cambridge Series in Statistical and Probabilistic Mathematics* **6**. Cambridge Univ. Press, Cambridge. MR1739079

[78] VAN DER VAART, A. (2002). Semiparametric statistics. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1999). *Lecture Notes in Math.* **1781** 331–457. Springer, Berlin. MR1915446

[79] VAN DER VAART, A. W. (1995). Efficiency of infinite-dimensional M-estimators. *Stat. Neerl.* **49** 9–30. MR1333176 https://doi.org/10.1111/j.1467-9574.1995.tb01452.x

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

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

[82] VÍŠEK, J. Á. (1979). Asymptotic distribution of simple estimate for rejective, Sampford and successive sampling. In *Contributions to Statistics* 263–275. Reidel, Dordrecht. MR0561274

[83] WELLNER, J. A. (1978). Limit theorems for the ratio of the empirical distribution function to the true distribution function. *Z. Wahrsch. Verw. Gebiete* **45** 73–88. MR0651392 https://doi.org/10.1007/BF00635964

[84] WELLNER, J. A. and ZHAN, Y. (1996). Bootstrapping Z-estimators. Technical Report 308, Univ. Washington Dept. Statistics.

[85] WELLNER, J. A. and ZHANG, Y. (2007). Two likelihood-based semiparametric estimation methods for panel count data with covariates. *Ann. Statist.* **35** 2106–2142. MR2363965 https://doi.org/10.1214/009053607000000181

The Annals of Statistics
2021, Vol. 49, No. 1, 486–507
https://doi.org/10.1214/20-AOS1965

PREDICTIVE INFERENCE WITH THE JACKKNIFE+

BY RINA FOYGEL BARBER[1], EMMANUEL J. CANDÈS[2], AADITYA RAMDAS[3,*] AND RYAN J. TIBSHIRANI[3,†]

[1]*Department of Statistics, University of Chicago, rina@uchicago.edu*

[2]*Departments of Statistics and Mathematics, Stanford University, candes@stanford.edu*

[3]*Department of Statistics and Data Science and Machine Learning Department, Carnegie Mellon University,* [*]*aramdas@cmu.edu;* [†]*ryantibs@cmu.edu*

This paper introduces the *jackknife+*, which is a novel method for constructing predictive confidence intervals. Whereas the jackknife outputs an interval centered at the predicted response of a test point, with the width of the interval determined by the quantiles of leave-one-out residuals, the jackknife+ also uses the leave-one-out predictions at the test point to account for the variability in the fitted regression function. Assuming exchangeable training samples, we prove that this crucial modification permits rigorous coverage guarantees regardless of the distribution of the data points, for any algorithm that treats the training points symmetrically. Such guarantees are not possible for the original jackknife and we demonstrate examples where the coverage rate may actually vanish. Our theoretical and empirical analysis reveals that the jackknife and the jackknife+ intervals achieve nearly exact coverage and have similar lengths whenever the fitting algorithm obeys some form of stability. Further, we extend the jackknife+ to K-fold cross validation and similarly establish rigorous coverage properties. Our methods are related to *cross-conformal prediction* proposed by Vovk (*Ann. Math. Artif. Intell.* **74** (2015) 9–28) and we discuss connections.

REFERENCES

[1] BARBER, R. F., CANDÈS, E. J., RAMDAS, A. and TIBSHIRANI, R. J. (2021). Supplement to "Predictive inference with the jackknife+." https://doi.org/10.1214/20-AOS1965SUPP

[2] BOUSQUET, O. and ELISSEEFF, A. (2002). Stability and generalization. *J. Mach. Learn. Res.* **2** 499–526. MR1929416 https://doi.org/10.1162/153244302760200704

[3] BURNAEV, E. and VOVK, V. (2014). Efficiency of conformalized ridge regression. In *Conference on Learning Theory* 605–622.

[4] BUTLER, R. and ROTHMAN, E. D. (1980). Predictive intervals based on reuse of the sample. *J. Amer. Statist. Assoc.* **75** 881–889. MR0600971

[5] BUZA, K. (2014). Feedback prediction for blogs. In *Data Analysis*, *Machine Learning and Knowledge Discovery* 145–152. Springer, Berlin.

[6] CHEN, W., CHUN, K.-J. and BARBER, R. F. (2018). Discretized conformal prediction for efficient distribution-free inference. *Stat* **7** e173. MR3769053 https://doi.org/10.1002/sta4.173

[7] DEVROYE, L. P. and WAGNER, T. J. (1979). Distribution-free inequalities for the deleted and holdout error estimates. *IEEE Trans. Inf. Theory* **25** 202–207. MR0521311 https://doi.org/10.1109/TIT.1979.1056032

[8] EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26. MR0515681

[9] EFRON, B. and GONG, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *Amer. Statist.* **37** 36–48. MR0694281 https://doi.org/10.2307/2685844

[10] GEISSER, S. (1975). The predictive sample reuse method with applications. *J. Amer. Statist. Assoc.* **70** 320–328.

[11] HASTIE, T., MONTANARI, A., ROSSET, S. and TIBSHIRANI, R. J. (2019). Surprises in high-dimensional ridgeless least squares interpolation. Preprint. Available at arXiv:1903.08560.

MSC2020 subject classifications. 62F40, 62G08, 62G09.

Key words and phrases. Distribution-free, jackknife, cross-validation, conformal inference, stability, leave-one-out.

[12] LANDAU, H. G. (1953). On dominance relations and the structure of animal societies. III. The condition for a score structure. *Bull. Math. Biophys.* **15** 143–148. MR0054933 https://doi.org/10.1007/bf02476378

[13] LEI, J. (2019). Fast exact conformalization of the lasso using piecewise linear homotopy. *Biometrika* **106** 749–764. MR4031197 https://doi.org/10.1093/biomet/asz046

[14] 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

[15] MILLER, R. G. (1974). The jackknife—a review. *Biometrika* **61** 1–15. MR0391366 https://doi.org/10.1093/biomet/61.1.1

[16] PAPADOPOULOS, H. (2008). Inductive conformal prediction: Theory and application to neural networks. In *Tools in Artificial Intelligence* InTech, Vienna.

[17] PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A. et al. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12** 2825–2830. MR2854348

[18] QUENOUILLE, M. H. (1949). Approximate tests of correlation in time-series. *J. Roy. Statist. Soc. Ser. B* **11** 68–84. MR0032176

[19] QUENOUILLE, M. H. (1956). Notes on bias in estimation. *Biometrika* **43** 353–360. MR0081040 https://doi.org/10.1093/biomet/43.3-4.353

[20] REDMOND, M. and BAVEJA, A. (2002). A data-driven software tool for enabling cooperative information sharing among police departments. *European J. Oper. Res.* **141** 660–678.

[21] STEINBERGER, L. and LEEB, H. (2016). Leave-one-out prediction intervals in linear regression models with many variables. Preprint. Available at arXiv:1602.05801.

[22] STEINBERGER, L. and LEEB, H. (2018). Conditional predictive inference for high-dimensional stable algorithms. Preprint. Available at arXiv:1809.01412.

[23] STINE, R. A. (1985). Bootstrap prediction intervals for regression. *J. Amer. Statist. Assoc.* **80** 1026–1031. MR0819610

[24] STONE, M. (1974). Cross-validatory choice and assessment of statistical predictions. *J. Roy. Statist. Soc. Ser. B* **36** 111–147. MR0356377

[25] TRENA, M., EZZATI-RICE, ROHDE, F. and GREENBLATT, J. (2008). Sample design of the medical expenditure panel survey household component, 1998–2007. MEPS Methodology Report No. 22, Agency for Healthcare Research and Quality.

[26] TUKEY, J. (1958). Bias and confidence in not quite large samples. *Ann. Math. Stat.* **29** 614.

[27] VOVK, V. (2012). Conditional validity of inductive conformal predictors. In *Asian Conference on Machine Learning* 475–490.

[28] VOVK, V. (2015). Cross-conformal predictors. *Ann. Math. Artif. Intell.* **74** 9–28. MR3353894 https://doi.org/10.1007/s10472-013-9368-4

[29] VOVK, V., GAMMERMAN, A. and SHAFER, G. (2005). *Algorithmic Learning in a Random World*. Springer, New York. MR2161220

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

[31] VOVK, V. and WANG, R. (2012). Combining p-values via averaging. Preprint. Available at arXiv:1212.4966.

[32] XU, H., CARAMANIS, C. and MANNOR, S. (2012). Sparse algorithms are not stable: A no-free-lunch theorem. *IEEE Trans. Pattern Anal. Mach. Intell.* **34** 187–193.

The Annals of Statistics
2021, Vol. 49, No. 1, 508–530
https://doi.org/10.1214/20-AOS1966

ROBUST ESTIMATION OF SUPERHEDGING PRICES

BY JAN OBŁÓJ[*] AND JOHANNES WIESEL[†]

Mathematical Institute and St. John's College, University of Oxford, [*]*jan.obloj@maths.ox.ac.uk;*
[†]*johannes.wiesel@maths.ox.ac.uk*

We consider statistical estimation of superhedging prices using historical stock returns in a frictionless market with d traded assets. We introduce a plug-in estimator based on empirical measures and show it is consistent but lacks suitable robustness. To address this, we propose novel estimators which use a larger set of martingale measures defined through a tradeoff between the radius of Wasserstein balls around the empirical measure and the allowed norm of martingale densities. We then extend our study, in part, to estimation of risk measures, to the case of markets with traded options, to a multi-period setting and to settings with model uncertainty. We also study convergence rates of estimators and convergence of super-hedging strategies.

REFERENCES

[1] BURZONI, M., FRITTELLI, M., HOU, Z., MAGGIS, M. and OBŁÓJ, J. (2019). Pointwise arbitrage pricing theory in discrete time. *Math. Oper. Res.* **44** 1034–1057. MR3996657 https://doi.org/10.1287/moor.2018.0956

[2] CHERIDITO, P., KUPPER, M. and TANGPI, L. (2017). Duality formulas for robust pricing and hedging in discrete time. *SIAM J. Financial Math.* **8** 738–765. MR3705784 https://doi.org/10.1137/16M1064088

[3] CHEVALIER, J. (1976). Estimation du support et du contour du support d'une loi de probabilité. *Ann. Inst. Henri Poincaré B, Calc. Probab. Stat.* **12** 339–364. MR0451491

[4] COHN, D. L. (1980). *Measure Theory*. Birkhäuser, Boston, MA. MR0578344

[5] CONT, R., DEGUEST, R. and SCANDOLO, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quant. Finance* **10** 593–606. MR2676786 https://doi.org/10.1080/14697681003685597

[6] CUEVAS, A. (1990). On pattern analysis in the nonconvex case. *Kybernetes* **19** 26–33. MR1084947 https://doi.org/10.1108/eb005866

[7] CUEVAS, A. and RODRÍGUEZ-CASAL, A. (2004). On boundary estimation. *Adv. in Appl. Probab.* **36** 340–354. MR2058139 https://doi.org/10.1239/aap/1086957575

[8] DEVROYE, L. and WISE, G. L. (1980). Detection of abnormal behavior via nonparametric estimation of the support. *SIAM J. Appl. Math.* **38** 480–488. MR0579432 https://doi.org/10.1137/0138038

[9] DUDLEY, R. M. (2002). *Real Analysis and Probability*. *Cambridge Studies in Advanced Mathematics* **74**. Cambridge Univ. Press, Cambridge. MR1932358 https://doi.org/10.1017/CBO9780511755347

[10] ECKSTEIN, S., KUPPER, M. and POHL, M. (2020). Robust risk aggregation with neural networks. *Math. Finance* **30** 1229–1272. MR4154769 https://doi.org/10.1111/mafi.12280

[11] EMBRECHTS, P., SCHIED, A. and WANG, R. (2018). Robustness in the optimization of risk measures. Preprint. Available at arXiv:1809.09268.

[12] EMBRECHTS, P., WANG, B. and WANG, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance Stoch.* **19** 763–790. MR3413935 https://doi.org/10.1007/s00780-015-0273-z

[13] FÖLLMER, H. and SCHIED, A. (2002). *Stochastic Finance: An Introduction in Discrete Time*. *De Gruyter Studies in Mathematics* **27**. de Gruyter, Berlin. MR1925197 https://doi.org/10.1515/9783110198065

[14] FOURNIER, N. and GUILLIN, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* **162** 707–738. MR3383341 https://doi.org/10.1007/s00440-014-0583-7

MSC2020 subject classifications. 91G70, 91G20, 62G20, 62G35.
Key words and phrases. Superhedging price, risk measures, statistical estimation, consistency, robustness, stock returns, Wasserstein metric, pricing-hedging duality, empirical measure.

[15] García Trillos, N. and Slepčev, D. (2015). On the rate of convergence of empirical measures in ∞-transportation distance. *Canad. J. Math.* **67** 1358–1383. MR3415656 https://doi.org/10.4153/CJM-2014-044-6

[16] Gatheral, J., Jaisson, T. and Rosenbaum, M. (2018). Volatility is rough. *Quant. Finance* **18** 933–949. MR3805308 https://doi.org/10.1080/14697688.2017.1393551

[17] Geffroy, J. (1964). Sur un problème d'estimation géométrique. *Publ. Inst. Stat. Univ. Paris* **13** 191–210. MR0202237

[18] Grenander, U. (1981). *Abstract Inference*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0599175

[19] Hampel, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Stat.* **42** 1887–1896. MR0301858 https://doi.org/10.1214/aoms/1177693054

[20] Härdle, W., Park, B. U. and Tsybakov, A. B. (1995). Estimation of non-sharp support boundaries. *J. Multivariate Anal.* **55** 205–218. MR1370400 https://doi.org/10.1006/jmva.1995.1075

[21] Hou, Z. and Obłój, J. (2018). Robust pricing-hedging dualities in continuous time. *Finance Stoch.* **22** 511–567. MR3816548 https://doi.org/10.1007/s00780-018-0363-9

[22] Huber, P. J. and Ronchetti, E. M. (2009). *Robust Statistics*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2488795 https://doi.org/10.1002/9780470434697

[23] Jouini, E., Schachermayer, W. and Touzi, N. (2006). Law invariant risk measures have the Fatou property. In *Advances in Mathematical Economics*, *Vol.* 9 49–71. Springer, Tokyo. MR2277714 https://doi.org/10.1007/4-431-34342-3_4

[24] 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

[25] Kellerer, H. G. (1985). Duality theorems and probability metrics. In *Proceedings of the Seventh Conference on Probability Theory* (*Braşov*, 1982) 211–220. VNU Sci. Press, Utrecht. MR0867434

[26] Korostelëv, A. P., Simar, L. and Tsybakov, A. B. (1995). Efficient estimation of monotone boundaries. *Ann. Statist.* **23** 476–489. MR1332577 https://doi.org/10.1214/aos/1176324531

[27] Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. *Springer Series in Statistics*. Springer, New York. MR2724368 https://doi.org/10.1007/978-0-387-74978-5

[28] Kou, S., Peng, X. and Heyde, C. C. (2013). External risk measures and Basel accords. *Math. Oper. Res.* **38** 393–417. MR3092538 https://doi.org/10.1287/moor.1120.0577

[29] Krätschmer, V., Schied, A. and Zähle, H. (2012). Qualitative and infinitesimal robustness of tail-dependent statistical functionals. *J. Multivariate Anal.* **103** 35–47. MR2823706 https://doi.org/10.1016/j.jmva.2011.06.005

[30] Krätschmer, V., Schied, A. and Zähle, H. (2014). Comparative and qualitative robustness for law-invariant risk measures. *Finance Stoch.* **18** 271–295. MR3177407 https://doi.org/10.1007/s00780-013-0225-4

[31] Kusuoka, S. (2001). On law invariant coherent risk measures. In *Advances in Mathematical Economics*, *Vol.* 3 83–95. Springer, Tokyo. MR1886557 https://doi.org/10.1007/978-4-431-67891-5_4

[32] Li, X. and Lin, Y. (2017). Generalized Wasserstein distance and weak convergence of sublinear expectations. *J. Theoret. Probab.* **30** 581–593. MR3647070 https://doi.org/10.1007/s10959-015-0651-7

[33] Mammen, E. and Tsybakov, A. B. (1995). Asymptotical minimax recovery of sets with smooth boundaries. *Ann. Statist.* **23** 502–524. MR1332579 https://doi.org/10.1214/aos/1176324533

[34] Mohajerin Esfahani, P. and Kuhn, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Math. Program.* **171** 115–166. MR3844536 https://doi.org/10.1007/s10107-017-1172-1

[35] Mykland, P. A. (2000). Conservative delta hedging. *Ann. Appl. Probab.* **10** 664–683. MR1768218 https://doi.org/10.1214/aoap/1019487360

[36] Mykland, P. A. (2003). Financial options and statistical prediction intervals. *Ann. Statist.* **31** 1413–1438. MR2012820 https://doi.org/10.1214/aos/1065705113

[37] Mykland, P. A. (2003). The interpolation of options. *Finance Stoch.* **7** 417–432. MR2014243 https://doi.org/10.1007/s007800200093

[38] Obłój, J. and Wiesel, J. (2021). Supplement to "Robust estimation of superhedging prices." https://doi.org/10.1214/20-AOS1966SUPP

[39] Basel Committee on Banking Supervision (2013). Fundamental review of the trading book: A revised market risk framework. Technical report.

[40] Basel Committee on Banking Supervision (2019). Minimum capital requirements for market risk. Bank for International Settlements, Basel, Switzerland.

[41] Pichler, A. (2013). Evaluations of risk measures for different probability measures. *SIAM J. Optim.* **23** 530–551. MR3033118 https://doi.org/10.1137/110857088

[42] POLONIK, W. (1995). Measuring mass concentrations and estimating density contour clusters—An excess mass approach. *Ann*. *Statist*. **23** 855–881. MR1345204 https://doi.org/10.1214/aos/1176324626

[43] RÁSONYI, M. (2002). A note on martingale measures with bounded densities. *Tr. Mat*. *Inst*. *Steklova* **237** 212–216. MR1976516

[44] RODRÍGUEZ CASAL, A. (2007). Set estimation under convexity type assumptions. *Ann*. *Inst*. *Henri Poincaré Probab*. *Stat*. **43** 763–774. MR3252430 https://doi.org/10.1016/j.anihpb.2006.11.001

[45] SARTIPIZADEH, H. and VINCENT, T. (2016). Computing the approximate convex hull in high dimensions. Preprint. Available at arXiv:1603.04422.

[46] TSYBAKOV, A. B. (1997). On nonparametric estimation of density level sets. *Ann*. *Statist*. **25** 948–969. MR1447735 https://doi.org/10.1214/aos/1069362732

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

The Annals of Statistics
2021, Vol. 49, No. 1, 531–556
https://doi.org/10.1214/20-AOS1967

CONCENTRATION OF KERNEL MATRICES WITH APPLICATION TO KERNEL SPECTRAL CLUSTERING

BY ARASH A. AMINI[1] AND ZAHRA S. RAZAEE[2]

[1]*Department of Statistics, University of California, Los Angeles, aaamini@ucla.edu*
[2]*Biostatistics and Bioinformatics Research Center, Cedars-Sinai Medical Center, zahra.razaee@cshs.org*

We study the concentration of random kernel matrices around their mean. We derive nonasymptotic exponential concentration inequalities for Lipschitz kernels assuming that the data points are independent draws from a class of multivariate distributions on $\mathbb{R}^d$, including the strongly log-concave distributions under affine transformations. A feature of our result is that the data points need not have identical distributions or zero mean, which is key in certain applications such as clustering. Our bound for the Lipschitz kernels is dimension-free and sharp up to constants. For comparison, we also derive the companion result for the Euclidean (inner product) kernel for a class of sub-Gaussian distributions. A notable difference between the two cases is that, in contrast to the Euclidean kernel, in the Lipschitz case, the concentration inequality does not depend on the mean of the underlying vectors. As an application of these inequalities, we derive a bound on the misclassification rate of a kernel spectral clustering (KSC) algorithm, under a perturbed nonparametric mixture model. We show an example where this bound establishes the high-dimensional consistency (as $d \to \infty$) of the KSC, when applied with a Gaussian kernel, to a noisy model of nested nonlinear manifolds.

REFERENCES

[1] ACKERMANN, M. R., BLÖMER, J. and SOHLER, C. (2010). Clustering for metric and nonmetric distance measures. *ACM Trans. Algorithms* **6** Art. 59, 26. MR2760422 https://doi.org/10.1145/1824777.1824779

[2] ADAMCZAK, R. (2015). A note on the Hanson–Wright inequality for random vectors with dependencies. *Electron. Commun. Probab.* **20** no. 72, 13. MR3407216 https://doi.org/10.1214/ECP.v20-3829

[3] AMINI, A. A. and RAZAEE, Z. S. (2021). Supplement to "Concentration of kernel matrices with application to kernel spectral clustering." https://doi.org/10.1214/20-AOS1967SUPP

[4] BELKIN, M. and NIYOGI, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **15** 1373–1396.

[5] BLANCHARD, G., BOUSQUET, O. and ZWALD, L. (2007). Statistical properties of kernel principal component analysis. *Mach. Learn.* **66** 259–294.

[6] BRAUN, M. L. (2006). Accurate error bounds for the eigenvalues of the kernel matrix. *J. Mach. Learn. Res.* **7** 2303–2328. MR2274441 https://doi.org/10.1080/14685240600860923

[7] CHEN, K. (2009). On coresets for k-median and k-means clustering in metric and Euclidean spaces and their applications. *SIAM J. Comput.* **39** 923–947. MR2538844 https://doi.org/10.1137/070699007

[8] CHENG, X. and SINGER, A. (2013). The spectrum of random inner-product kernel matrices. *Random Matrices Theory Appl.* **2** 1350010, 47. MR3149440 https://doi.org/10.1142/S201032631350010X

[9] DO, Y. and VU, V. (2013). The spectrum of random kernel matrices: Universality results for rough and varying kernels. *Random Matrices Theory Appl.* **2** 1350005, 29. MR3109422 https://doi.org/10.1142/S2010326313500056

[10] EL KAROUI, N. (2010). The spectrum of kernel random matrices. *Ann. Statist.* **38** 1–50. MR2589315 https://doi.org/10.1214/08-AOS648

[11] EL KAROUI, N. (2010). On information plus noise kernel random matrices. *Ann. Statist.* **38** 3191–3216. MR2722468 https://doi.org/10.1214/10-AOS801

MSC2020 subject classifications. 62H20, 60E15, 62G99.
Key words and phrases. Concentration inequalities, kernel matrices, nonasymptotic bounds, kernel spectral clustering.

[12] FAN, Z. and MONTANARI, A. (2019). The spectral norm of random inner-product kernel matrices. *Probab. Theory Related Fields* **173** 27–85. MR3916104 https://doi.org/10.1007/s00440-018-0830-4

[13] GINÉ, E. and KOLTCHINSKII, V. (2006). Empirical graph Laplacian approximation of Laplace–Beltrami operators: Large sample results. In *High Dimensional Probability*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **51** 238–259. IMS, Beachwood, OH. MR2387773 https://doi.org/10.1214/074921706000000888

[14] HEIN, M. (2006). Uniform convergence of adaptive graph-based regularization. In *Learning Theory*. *Lecture Notes in Computer Science* **4005** 50–64. Springer, Berlin. MR2277918 https://doi.org/10.1007/11776420_7

[15] HEIN, M., AUDIBERT, J.-Y. and VON LUXBURG, U. (2005). From graphs to manifolds—Weak and strong pointwise consistency of graph Laplacians. In *Learning Theory*. *Lecture Notes in Computer Science* **3559** 470–485. Springer, Berlin. MR2203281 https://doi.org/10.1007/11503415_32

[16] KASIVISWANATHAN, S. P. and RUDELSON, M. (2015). Spectral norm of random kernel matrices with applications to privacy. In *Approximation*, *Randomization*, *and Combinatorial Optimization*. *Algorithms and Techniques*. *LIPIcs*. *Leibniz Int*. *Proc*. *Inform*. **40** 898–914. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR3442004

[17] 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

[18] KUMAR, A., SABHARWAL, Y. and SEN, S. (2004). A simple linear time $(1+\varepsilon)$-approximation algorithm for k-means clustering in any dimensions. In *Annual Symposium on Foundations of Computer Science* **45** 454–462. IEEE Computer Society Press.

[19] LEDOUX, M. (2001). *The Concentration of Measure Phenomenon*. *Mathematical Surveys and Monographs* **89**. Amer. Math. Soc., Providence, RI. MR1849347

[20] ROSASCO, L., BELKIN, M. and DE VITO, E. (2010). On learning with integral operators. *J. Mach. Learn. Res.* **11** 905–934. MR2600634

[21] SCHIEBINGER, G., WAINWRIGHT, M. J. and YU, B. (2015). The geometry of kernelized spectral clustering. *Ann. Statist.* **43** 819–846. MR3325711 https://doi.org/10.1214/14-AOS1283

[22] SHAWE-TAYLOR, J., CRISTIANINI, N. and KANDOLA, J. S. (2002). On the concentration of spectral properties. In *Advances in Neural Information Processing Systems* 511–517.

[23] SHAWE-TAYLOR, J., WILLIAMS, C. K. I., CRISTIANINI, N. and KANDOLA, J. (2005). On the eigenspectrum of the Gram matrix and the generalization error of kernel-PCA. *IEEE Trans. Inf. Theory* **51** 2510–2522. MR2246374 https://doi.org/10.1109/TIT.2005.850052

[24] SINGER, A. (2006). From graph to manifold Laplacian: The convergence rate. *Appl. Comput. Harmon. Anal.* **21** 128–134. MR2238670 https://doi.org/10.1016/j.acha.2006.03.004

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

[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. MR3837109 https://doi.org/10.1017/9781108231596

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

[28] 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

[29] YAN, B. and SARKAR, P. (2016). On robustness of kernel clustering. In *Advances in Neural Information Processing Systems* 3098–3106.

[30] YANG, Y., PILANCI, M. and WAINWRIGHT, M. J. (2017). Randomized sketches for kernels: Fast and optimal nonparametric regression. *Ann. Statist.* **45** 991–1023. MR3662446 https://doi.org/10.1214/16-AOS1472

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

The Annals of Statistics
2021, Vol. 49, No. 1, 557–567
https://doi.org/10.1214/20-AOS1968

A RULE OF THUMB: RUN LENGTHS TO FALSE ALARM OF MANY TYPES OF CONTROL CHARTS RUN IN PARALLEL ON DEPENDENT STREAMS ARE ASYMPTOTICALLY INDEPENDENT

BY MOSHE POLLAK

Statistics Department, The Hebrew University of Jerusalem, moshe.pollak@mail.huji.ac.il

Consider a process that produces a series of independent identically distributed vectors. A change in an underlying state may become manifest in a modification of one or more of the marginal distributions. Often, the dependence structure between coordinates is unknown, impeding surveillance based on the joint distribution. A popular approach is to construct control charts for each coordinate separately and raise an alarm the first time any (or some) of the control charts signals. The difficulty is obtaining an expression for the overall average run length to false alarm (ARL2FA).

We argue that despite the dependence structure, when the process is in control, for large ARLs to false alarm, run lengths of many types of control charts run in parallel are asymptotically independent. Furthermore, often, in-control run lengths are asymptotically exponentially distributed, enabling uncomplicated asymptotic expressions for the ARL2FA.

We prove this assertion for certain Cusum and Shiryaev–Roberts-type control charts and illustrate it by simulations.

REFERENCES

EPPRECHT, E. K. (2015). Statistical control of multiple-stream processes: A literature review. In *Frontiers in Statistical Quality Control* 11 (S. Knoth and W. Schmid, eds.) 29–47. Springer-Verlag, New York.

LUZIA, N. (2018). A simple proof of the strong law of large numbers with rates. *Bull. Aust. Math. Soc.* **97** 513–517. MR3802466 https://doi.org/10.1017/S0004972718000059

MEI, Y. (2010). Efficient scalable schemes for monitoring a large number of data streams. *Biometrika* **97** 419–433. MR2650748 https://doi.org/10.1093/biomet/asq010

MENECES, N. S., OLIVERA, S. A., SACCONE, C. D. and TESSORE, J. (2008). Statistical control of multiple stream processes: A Shewhart control chart for each stream. *Qual. Engin.* **20** 185–194.

POLLAK, M. (1987). Average run lengths of an optimal method of detecting a change in distribution. *Ann. Statist.* **15** 749–779. MR0888438 https://doi.org/10.1214/aos/1176350373

POLLAK, M. (2020). Supplement to “A rule of thumb: Run lengths to false alarm of many types of control charts run in parallel on dependent streams are asymptotically independent.” https://doi.org/10.1214/20-AOS1968SUPP

POLLAK, M. and SIEGMUND, D. (1991). Sequential detection of a change in a normal mean when the initial value is unknown. *Ann. Statist.* **19** 394–416. MR1091859 https://doi.org/10.1214/aos/1176347990

POLLAK, M. and TARTAKOVSKY, A. G. (2008). Asymptotic exponentiality of the distribution of first exit times for a class of Markov processes with applications to quickest change detection. *Teor. Veroyatn. Primen.* **53** 500–515. MR2759707 https://doi.org/10.1137/S0040585X97983742

RÉNYI, A. (1957). On the asymptotic distribution of the sum of a random number of independent random variables. *Acta Math. Acad. Sci. Hung.* **8** 193–199. MR0088093 https://doi.org/10.1007/BF02025242

SIEGMUND, D. (1985). *Sequential Analysis: Tests and Confidence Intervals. Springer Series in Statistics*. Springer, New York. MR0799155 https://doi.org/10.1007/978-1-4757-1862-1

STONE, C. (1965). On moment generating functions and renewal theory. *Ann. Math. Stat.* **36** 1298–1301. MR0179857 https://doi.org/10.1214/aoms/1177700003

YAKIR, B. (1995). A note on the run length to false alarm of a change-point detection policy. *Ann. Statist.* **23** 272–281. MR1331668 https://doi.org/10.1214/aos/1176324467

YAKIR, B. (1998). On the average run length to false alarm in surveillance problems which possess an invariance structure. *Ann. Statist.* **26** 1198–1214. MR1635389 https://doi.org/10.1214/aos/1024691095

MSC2020 subject classifications. Primary 62L10; secondary 94A13.
Key words and phrases. Average run length, Cusum, Shiryaev–Roberts, exponential distribution, p-value.

The Annals of Statistics
2021, Vol. 49, No. 1, 568–589
https://doi.org/10.1214/20-AOS1970

SHARP MINIMAX DISTRIBUTION ESTIMATION FOR CURRENT STATUS CENSORING WITH OR WITHOUT MISSING

BY SAM EFROMOVICH

Department of Mathematical Sciences, University of Texas at Dallas, efrom@utdallas.edu

Nonparametric estimation of the cumulative distribution function and the probability density of a lifetime X modified by a current status censoring (CSC), including cases of right and left missing data, is a classical ill-posed problem with biased data. The biased nature of CSC data may preclude us from consistent estimation unless the biasing function is known or may be estimated, and its ill-posed nature slows down rates of convergence. Under a traditionally studied CSC, we observe a sample from (Z, Δ) where a continuous monitoring time Z is independent of X, $\Delta := I(X \le Z)$ is the status, and the bias of observations is created by the density of Z which is estimable. In presence of right or left missing, we observe corresponding samples from $(\Delta Z, \Delta)$ or $((1 - \Delta)Z, \Delta)$; the data are again biased but now the density of Z cannot be estimated from the data. As a result, to solve the estimation problem, either the density of Z must be known (like in a controlled study) or an extra cross-sectional sampling of Z, which is typically simpler than an underlying CSC study, be conducted. The main aim of the paper is to develop for this biased and ill-posed problem the theory of efficient (sharp-minimax) estimation which is inspired by known results for the case of directly observed X. Among interesting aspects of the developed theory: (i) While sharp-minimax analysis of missing CSC may follow the classical Pinsker's methodology, analysis of CSC requires a more complicated estimation procedure based on a special smoothing in both frequency and time domains; (ii) Efficient estimation requires solving an old-standing problem of approximating aperiodic Sobolev functions; (iii) If smoothness of the cdf of X is known, then its rate-minimax estimation is possible even if the density of Z is rougher. Real and simulated examples, as well as extensions of the core models to dependent X and Z and case-control CSC, are presented.

REFERENCES

BECKER, D. G., BRAUN, W. J. and WHITE, B. J. G. (2017). Interval-censored unimodal kernel density estimation via data sharpening. *J. Stat. Comput. Simul.* **87** 2023–2037. MR3647585 https://doi.org/10.1080/00949655.2017.1308510

CAI, T. T. (2012). Minimax and adaptive inference in nonparametric function estimation. *Statist. Sci.* **27** 31–50. MR2953494 https://doi.org/10.1214/11-STS355

CHEN, D.-G., SUN, J. and PEACE, K. E., eds. (2012). *Interval-Censored Time-to-Event Data: Methods and Applications. Chapman & Hall/CRC Biostatistics Series*. CRC Press, Boca Raton, FL. MR3053014

CHICKEN, E. and CAI, T. T. (2005). Block thresholding for density estimation: Local and global adaptivity. *J. Multivariate Anal.* **95** 76–106. MR2164124 https://doi.org/10.1016/j.jmva.2004.07.003

DIAO, G. and YUAN, A. (2019). A class of semiparametric cure models with current status data. *Lifetime Data Anal.* **25** 26–51. MR3896658 https://doi.org/10.1007/s10985-018-9420-0

DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24** 508–539. MR1394974 https://doi.org/10.1214/aos/1032894451

EFROMOVICH, S. Y. (1985). Nonparametric estimation of a density of unknown smoothness. *Theory Probab. Appl.* **30** 557–568.

EFROMOVICH, S. (1999). *Nonparametric Curve Estimation: Methods, Theory, and Applications. Springer Series in Statistics*. Springer, New York. MR1705298

MSC2020 subject classifications. Primary 62G07, 62G07; secondary 60J10.

Key words and phrases. Adaptation, aperiodic functions, case-control study, cumulative distribution function, density, missing, sharp minimax.

EFROMOVICH, S. (2018). *Missing and Modified Data in Nonparametric Estimation*: *With R Examples*. *Monographs on Statistics and Applied Probability* **156**. CRC Press, Boca Raton, FL. MR3752670

EFROMOVICH, S. (2021). Supplement to "Sharp minimax distribution estimation for current status censoring with or without missing." https://doi.org/10.1214/20-AOS1970SUPP

GINÉ, E. (1975). Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev norms. *Ann*. *Statist*. **3** 1243–1266. MR0388663

GROENEBOOM, P. and HENDRICKX, K. (2018). Current status linear regression. *Ann*. *Statist*. **46** 1415–1444. MR3819105 https://doi.org/10.1214/17-AOS1589

GROENEBOOM, P. and JONGBLOED, G. (2014). *Nonparametric Estimation Under Shape Constraints*: *Estimators*, *Algorithms and Asymptotics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **38**. Cambridge Univ. Press, New York. MR3445293 https://doi.org/10.1017/CBO9781139020893

GROENEBOOM, P., JONGBLOED, G. and WITTE, B. I. (2010). Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model. *Ann*. *Statist*. **38** 352–387. MR2589325 https://doi.org/10.1214/09-AOS721

HALL, P., KERKYACHARIAN, G. and PICARD, D. (1998). Block threshold rules for curve estimation using kernel and wavelet methods. *Ann*. *Statist*. **26** 922–942. MR1635418 https://doi.org/10.1214/aos/1024691082

HSU, L., GORFINE, M. and ZUCKER, D. (2018). On estimation of the hazard function from population-based case-control studies. *J*. *Amer*. *Statist*. *Assoc*. **113** 560–570. MR3832208 https://doi.org/10.1080/01621459.2017.1356315

JEWELL, N. P. and VAN DER LAAN, M. (2004a). Current status data: Review, recent developments and open problems. In *Advances in Survival Analysis*. *Handbook of Statist*. **23** 625–642. Elsevier, Amsterdam. MR2065792 https://doi.org/10.1016/S0169-7161(03)23035-2

JEWELL, N. P. and VAN DER LAAN, M. (2004b). Case-control current status data. *Biometrika* **91** 529–541. MR2090620 https://doi.org/10.1093/biomet/91.3.529

KEOGH, R. H. and COX, D. R. (2014). *Case-Control Studies*. *Institute of Mathematical Statistics* (*IMS*) *Monographs* **4**. Cambridge Univ. Press, Cambridge. MR3443808 https://doi.org/10.1017/CBO9781139094757

KLEIN, J. P., VAN HOUWELINGEN, H. C., IBRAHIM, J. G. and SCHEIKE, T. H., eds. (2014). *Handbook of Survival Analysis*. *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. CRC Press, Boca Raton, FL. MR3287588

LI, S., HU, T., WANG, P. and SUN, J. (2017). Regression analysis of current status data in the presence of dependent censoring with applications to tumorigenicity experiments. *Comput*. *Statist*. *Data Anal*. **110** 75–86. MR3612609 https://doi.org/10.1016/j.csda.2016.12.011

LI, S., HU, T., ZHAO, X. and SUN, J. (2019). A class of semiparametric transformation cure models for interval-censored failure time data. *Comput*. *Statist*. *Data Anal*. **133** 153–165. MR3926472 https://doi.org/10.1016/j.csda.2018.09.008

LI, S., SUN, J., TIAN, T. and CUI, X. (2020). Semiparametric regression analysis of doubly censored failure time data from cohort studies. *Lifetime Data Anal*. **26** 315–338. MR4079669 https://doi.org/10.1007/s10985-019-09477-x

MA, L., HU, T. and SUN, J. (2015). Sieve maximum likelihood regression analysis of dependent current status data. *Biometrika* **102** 731–738. MR3394289 https://doi.org/10.1093/biomet/asv020

MALOV, S. V. (2019). Nonparametric estimation for a current status right-censored data model. *Stat*. *Neerl*. **73** 475–495. MR4023706 https://doi.org/10.1111/stan.12180

PINSKER, M. S. (1980). Optimal filtration of square-integrable signals in Gaussian noise. *Probl*. *Inf*. *Transm*. **16** 52–68. MR0624591

SUN, J. (2007). *The Statistical Analysis of Interval-Censored Failure Time Data*. *Statistics for Biology and Health*. Springer, New York. MR2287318

SUN, J. and ZHAO, X. (2013). *Statistical Analysis of Panel Count Data*. *Statistics for Biology and Health*. Springer, New York. MR3136574 https://doi.org/10.1007/978-1-4614-8715-9

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

VAN ES, B. and GRAAFLAND, C. (2017). Nonparametric kernel density estimation for univariate current status data. Preprint. Available at arXiv:1707.00544v1.

VANDENBROUCKE, J. and PEARCE, N. (2014). Case-control studies: Basic concepts. *Int*. *J*. *Epidemiol*. **41** 1480–1489.

WANG, C., SUN, J., SUN, L., ZHOU, J. and WANG, D. (2012). Nonparametric estimation of current status data with dependent censoring. *Lifetime Data Anal*. **18** 434–445. MR2973774 https://doi.org/10.1007/s10985-012-9223-7

ZHANG, C.-H. (2005). General empirical Bayes wavelet methods and exactly adaptive minimax estimation. *Ann*. *Statist*. **33** 54–100. MR2157796 https://doi.org/10.1214/009053604000000995

The Annals of Statistics
2021, Vol. 49, No. 1, 590–611
https://doi.org/10.1214/20-AOS1971

WASSERSTEIN F-TESTS AND CONFIDENCE BANDS FOR THE FRÉCHET REGRESSION OF DENSITY RESPONSE CURVES

BY ALEXANDER PETERSEN[1,*], XI LIU[1,†] AND AFSHIN A. DIVANI[2]

[1]*Department of Statistics and Applied Probability, University of California, Santa Barbara,* [*]*petersen@pstat.ucsb.edu;* [†]*xliu@pstat.ucsb.edu*

[2]*Department of Neurology, University of New Mexico, adivani@salud.unm.edu*

Data consisting of samples of probability density functions are increasingly prevalent, necessitating the development of methodologies for their analysis that respect the inherent nonlinearities associated with densities. In many applications, density curves appear as functional response objects in a regression model with vector predictors. For such models, inference is key to understand the importance of density-predictor relationships, and the uncertainty associated with the estimated conditional mean densities, defined as conditional Fréchet means under a suitable metric. Using the Wasserstein geometry of optimal transport, we consider the Fréchet regression of density curve responses and develop tests for global and partial effects, as well as simultaneous confidence bands for estimated conditional mean densities. The asymptotic behavior of these objects is based on underlying functional central limit theorems within Wasserstein space, and we demonstrate that they are asymptotically of the correct size and coverage, with uniformly strong consistency of the proposed tests under sequences of contiguous alternatives. The accuracy of these methods, including nominal size, power and coverage, is assessed through simulations, and their utility is illustrated through a regression analysis of post-intracerebral hemorrhage hematoma densities and their associations with a set of clinical and radiological covariates.

REFERENCES

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

AITCHISON, J. (1986). *The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability*. CRC Press, London. MR0865647 https://doi.org/10.1007/978-94-009-4109-0

AL-MUFTI, F., THABET, A. M., SINGH, T., EL-GHANEM, M., AMULURU, K. and GANDHI, C. D. (2018). Clinical and radiographic predictors of intracerebral hemorrhage outcome. *Interv. Neurol.* **7** 118–136. https://doi.org/10.1159/000484571

AMBROSIO, L., GIGLI, N. and SAVARÉ, G. (2008). *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed. *Lectures in Mathematics ETH Zürich*. Birkhäuser, Basel. MR2401600

BARBER, R. F., REIMHERR, M. and SCHILL, T. (2017). The function-on-scalar LASSO with applications to longitudinal GWAS. *Electron. J. Stat.* **11** 1351–1389. MR3635916 https://doi.org/10.1214/17-EJS1260

BARDEN, D., LE, H. and OWEN, M. (2013). Central limit theorems for Fréchet means in the space of phylogenetic trees. *Electron. J. Probab.* **18** 1–25. MR3035753 https://doi.org/10.1214/EJP.v18-2201

BARRAS, C. D., TRESS, B. M., CHRISTENSEN, S., MACGREGOR, L., COLLINS, M., DESMOND, P. M., SKOLNICK, B. E., MAYER, S. A., BRODERICK, J. P. et al. (2009). Density and shape as CT predictors of intracerebral hemorrhage growth. *Stroke* **40** 1325–1331.

BIGOT, J., GOUET, R., KLEIN, T. and LÓPEZ, A. (2017). Geodesic PCA in the Wasserstein space by convex PCA. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 1–26. MR3606732 https://doi.org/10.1214/15-AIHP706

BOLSTAD, B. M., IRIZARRY, R. A., ÅSTRAND, M. and SPEED, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19** 185–193.

MSC2020 subject classifications. Primary 62J99, 62F03, 62F25; secondary 62F05, 62F12.

Key words and phrases. Random densities, least squares regression, Wasserstein metric, tests for regression effects, simultaneous confidence band.

BOULOUIS, G., MOROTTI, A., BROUWERS, H. B., CHARIDIMOU, A., JESSEL, M. J., AURIEL, E., PONTES-NETO, O., AYRES, A., VASHKEVICH, A. et al. (2016). Noncontrast computed tomography hypodensities predict poor outcome in intracerebral hemorrhage patients. *Stroke* **47** 2511–2516.

BROADHURST, R. E., STOUGH, J., PIZER, S. M. and CHANEY, E. L. (2006). A statistical appearance model based on intensity quantile histograms. In *3rd IEEE International Symposium on Biomedical Imaging*: *Nano to Macro*, 2006 422–425. IEEE Press, New York.

CAO, G., YANG, L. and TODEM, D. (2012). Simultaneous inference for the mean function based on dense functional data. *J. Nonparametr. Stat.* **24** 359–377. MR2921141 https://doi.org/10.1080/10485252.2011.638071

CHIOU, J.-M. and MÜLLER, H.-G. (2007). Diagnostics for functional regression via residual processes. *Comput. Statist. Data Anal.* **51** 4849–4863. MR2364544 https://doi.org/10.1016/j.csda.2006.07.042

CLAESKENS, G. and VAN KEILEGOM, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.* **31** 1852–1884. MR2036392 https://doi.org/10.1214/aos/1074290329

CORNEA, E., ZHU, H., KIM, P. and IBRAHIM, J. G. (2017). Regression models on Riemannian symmetric spaces. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 463–482. MR3611755 https://doi.org/10.1111/rssb.12169

DEGRAS, D. A. (2011). Simultaneous confidence bands for nonparametric regression with functional data. *Statist. Sinica* **21** 1735–1765. MR2895997 https://doi.org/10.5705/ss.2009.207

DELCOURT, C., ZHANG, S., ARIMA, H., SATO, S., SALMAN, R. A.-S., WANG, X., DAVIES, L., STAPF, C., ROBINSON, T. et al. (2016). Significance of hematoma shape and density in intracerebral hemorrhage: The intensive blood pressure reduction in acute intracerebral hemorrhage trial study. *Stroke* **47** 1227–1232.

DUBEY, P. and MÜLLER, H.-G. (2019). Fréchet analysis of variance for random objects. *Biometrika* **106** 803–821. MR4031200 https://doi.org/10.1093/biomet/asz052

EGOZCUE, J. J., DÍAZ-BARRERO, J. L. and PAWLOWSKY-GLAHN, V. (2006). Hilbert space of probability density functions based on Aitchison geometry. *Acta Math. Sin.* (*Engl. Ser.*) **22** 1175–1182. MR2245249 https://doi.org/10.1007/s10114-005-0678-2

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

FAN, J. and ZHANG, W. (2008). Statistical methods with varying coefficient models. *Stat. Interface* **1** 179–195. MR2425354 https://doi.org/10.4310/SII.2008.v1.n1.a15

FARAWAY, J. J. (1997). Regression analysis for a functional response. *Technometrics* **39** 254–261. MR1462586 https://doi.org/10.2307/1271130

FLETCHER, P. T. (2013). Geodesic regression and the theory of least squares on Riemannian manifolds. *Int. J. Comput. Vis.* **105** 171–185. MR3104017 https://doi.org/10.1007/s11263-012-0591-y

FLETCHER, P. T., LU, C., PIZER, S. M. and JOSHI, S. (2004). Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Trans. Med. Imag.* **23** 995–1005.

FRÉCHET, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. Henri Poincaré* **10** 215–310. MR0027464

HEVESI, M., BERSHAD, E. M., JAFARI, M., MAYER, S. A., SELIM, M., SUAREZ, J. I. and DIVANI, A. A. (2018). Untreated hypertension as predictor of in-hospital mortality in intracerebral hemorrhage: A multi-center study. *J. Crit. Care* **43** 235–239.

HRON, K., MENAFOGLIO, A., TEMPL, M., HRUZOVA, K. and FILZMOSER, P. (2016). Simplicial principal component analysis for density functions in Bayes spaces. *MOX-report* **25** 2014.

HSING, T. and EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis*, *with an Introduction to Linear Operators*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR3379106 https://doi.org/10.1002/9781118762547

KNEIP, A. and UTIKAL, K. J. (2001). Inference for density families using functional principal component analysis. *J. Amer. Statist. Assoc.* **96** 519–542. MR1946423 https://doi.org/10.1198/016214501753168235

LEVINA, E. and BICKEL, P. (2001). The Earth mover's distance is the Mallows distance: Some insights from statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision*. *ICCV* 2001 **2** 251–256. IEEE Press, New York.

MARRON, J. S. and ALONSO, A. M. (2014). Overview of object oriented data analysis. *Biom. J.* **56** 732–753. MR3258083 https://doi.org/10.1002/bimj.201300072

MORGENSTERN, L. B., HEMPHILL, J. C., III, ANDERSON, C., BECKER, K., BRODERICK, J. P., CONNOLLY, E. S., JR., GREENBERG, S. M., HUANG, J. N., MACDONALD, R. L., MESSÉ, S. R. et al. (2010). Guidelines for the management of spontaneous intracerebral hemorrhage: A guideline for healthcare professionals from the American Heart Association/American Stroke Association *Stroke* **41** 2108–2129.

NERINI, D. and GHATTAS, B. (2007). Classifying densities using functional regression trees: Applications in oceanology. *Comput. Statist. Data Anal.* **51** 4984–4993. MR2364554 https://doi.org/10.1016/j.csda.2006.09.028

NIETHAMMER, M., HUANG, Y. and VIALARD, F.-X. (2011). Geodesic regression for image time-series. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI* 2011 655–662. Springer, Berlin.

PANARETOS, V. M., PHAM, T. and YAO, Z. (2014). Principal flows. *J. Amer. Statist. Assoc.* **109** 424–436. MR3180574 https://doi.org/10.1080/01621459.2013.849199

PANARETOS, V. M. and ZEMEL, Y. (2016). Amplitude and phase variation of point processes. *Ann. Statist.* **44** 771–812. MR3476617 https://doi.org/10.1214/15-AOS1387

PANARETOS, V. M. and ZEMEL, Y. (2019). Statistical aspects of Wasserstein distances. *Annu. Rev. Stat. Appl.* **6** 405–431. MR3939527 https://doi.org/10.1146/annurev-statistics-030718-104938

PARZEN, E. (1979). Nonparametric statistical data modeling. *J. Amer. Statist. Assoc.* **74** 105–121. MR0529528

PASS, B. (2013). Optimal transportation with infinitely many marginals. *J. Funct. Anal.* **264** 947–963. MR3004954 https://doi.org/10.1016/j.jfa.2012.12.002

PATRANGENARU, V. and ELLINGSON, L. (2016). *Nonparametric Statistics on Manifolds and Their Applications to Object Data Analysis*. CRC Press, Boca Raton, FL. MR3444169

PETERSEN, A., LIU, X. and DIVANI, A. A (2021). Supplement to "Wasserstein F-tests and confidence bands for the Fréchet regression of density response curves." https://doi.org/10.1214/20-AOS1971SUPP.

PETERSEN, A. and MÜLLER, H.-G. (2016). Functional data analysis for density functions by transformation to a Hilbert space. *Ann. Statist.* **44** 183–218. MR3449766 https://doi.org/10.1214/15-AOS1363

PETERSEN, A. and MÜLLER, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *Ann. Statist.* **47** 691–719. MR3909947 https://doi.org/10.1214/17-AOS1624

SALAZAR, P., DI NAPOLI, M., JAFARI, M., JAFARLI, A., ZIAI, W., PETERSEN, A., MAYER, S. A., BERSHAD, E. M., DAMANI, R. and DIVANI, A. A. (2020). Exploration of multiparameter hematoma 3D image analysis for predicting outcome after intracerebral hemorrhage. *Neurocritical Care* **32** 539–549. https://doi.org/10.1007/s12028-019-00783-8

SALMAN, R. A.-S., LABOVITZ, D. L. and STAPF, C. (2009). Spontaneous intracerebral haemorrhage. *BMJ* **339** b2586.

SATTERTHWAITE, F. E. (1941). Synthesis of variance. *Psychometrika* **6** 309–316. MR0005564 https://doi.org/10.1007/BF02288586

SHEN, Q. and FARAWAY, J. (2004). An F test for linear models with functional responses. *Statist. Sinica* **14** 1239–1257. MR2126351

SRIVASTAVA, A., JERMYN, I. and JOSHI, S. (2007). Riemannian analysis of probability density functions with applications in vision. *Proceedings from IEEE Conference on Computer Vision and Pattern Recognition* **25** 1–8.

TALSKÁ, R., MENAFOGLIO, A., MACHALOVÁ, J., HRON, K. and FIŠEROVÁ, E. (2018). Compositional regression with functional response. *Comput. Statist. Data Anal.* **123** 66–85. MR3777086 https://doi.org/10.1016/j.csda.2018.01.018

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

VILLANI, C. (2003). *Topics in Optimal Transportation*. *Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI. MR1964483 https://doi.org/10.1090/gsm/058

WANG, J. and YANG, L. (2009). Polynomial spline confidence bands for regression curves. *Statist. Sinica* **19** 325–342. MR2487893

ZHANG, Z. and MÜLLER, H.-G. (2011). Functional density synchronization. *Comput. Statist. Data Anal.* **55** 2234–2249. MR2786984 https://doi.org/10.1016/j.csda.2011.01.007

ZHANG, X. and WANG, J.-L. (2016). From sparse to dense functional data and beyond. *Ann. Statist.* **44** 2281–2321. MR3546451 https://doi.org/10.1214/16-AOS1446

The Annals of Statistics
2021, Vol. 49, No. 1, 612–613
https://doi.org/10.1214/20-AOS1993
Main article: https://doi.org/10.1214/12-AOS1053

CORRECTION NOTE: "OPTIMAL TWO-STAGE PROCEDURES FOR ESTIMATING LOCATION AND SIZE OF THE MAXIMUM OF A MULTIVARIATE REGRESSION FUNCTION" *Ann. Statist.* 40 (2012) 2850–2876

BY EDUARD BELITSER[1,*] SUBHASHIS GHOSAL[2] AND HARRY VAN ZANTEN[1,†]

[1]*Department of Mathematics, VU Amsterdam,* [*]*e.n.belitser@vu.nl;* [†]*j.h.van.zanten@vu.nl*
[2]*Department of Statistics, North Carolina State University, sghosal@stat.ncsu.edu*

We rectify a wrongly stated fact in the paper of Belitser, Ghosal and van Zanten (*Ann. Statist.* **40** (2012) 2850–2876).

REFERENCES

[1] IBRAGIMOV, I. A. and HASMINSKI, R. Z. (1982). Estimation of the maximal signal value in Gaussian white noise. *Mat*. *Zametki* **32** 529–536. Translation: *Math. Notes* **32** (1982) 746–750. MR0679245
[2] LEPSKI, O. V. (2020). Personal communication.
[3] LEPSKI, O. V. (1993). Estimation of the maximum of a nonparametric signal up to a constant. *Teor. Veroyatn*. *Primen*. **38** 187–194. Translation: *Theory Probab. Appl.* **38** (1993) 152–158. MR1317793 https://doi.org/10.1137/1138013

MSC2020 subject classifications. Primary 62L12; secondary 62G05, 62H12, 62L05.
Key words and phrases. Two-stage procedure, optimal rate, sequential design, multistage procedure, adaptive estimation.

The Annals of Statistics
2021, Vol. 49, No. 1, 614
https://doi.org/10.1214/20-AOS2014
Main article: https://doi.org/10.1214/16-AOS1439

CORRECTION NOTE: HIGHER ORDER ELICITABILITY AND OSBAND'S PRINCIPLE

BY TOBIAS FISSLER[1] AND JOHANNA F. ZIEGEL[2]

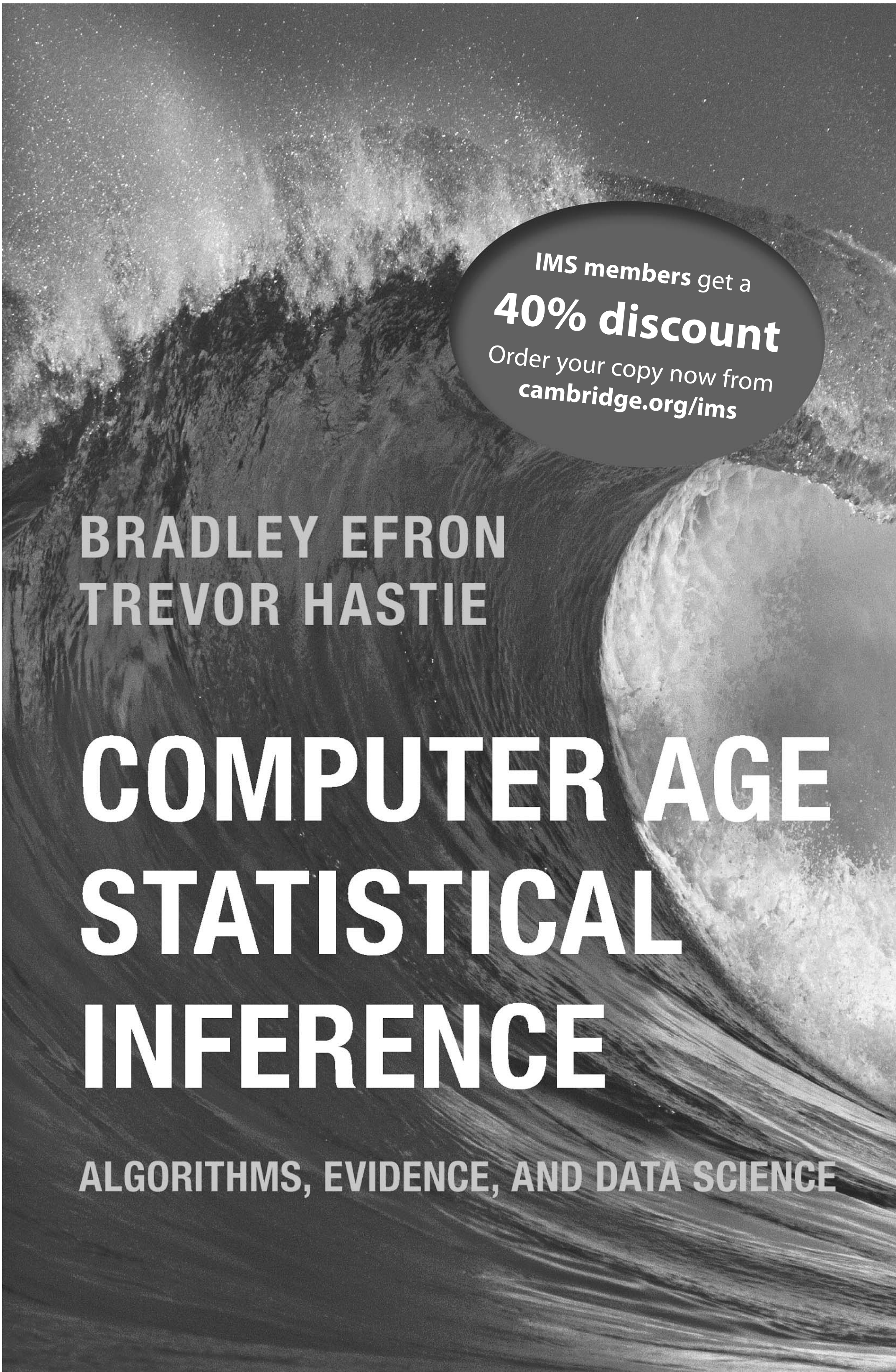
[1]*Institute for Statistics and Mathematics, WU Vienna University of Economics and Business, tobias.fissler@wu.ac.at*
[2]*Institute of Mathematical Statistics and Actuarial Science, University of Bern, johanna.ziegel@stat.unibe.ch*

REFERENCES

FISSLER, T. and ZIEGEL, J. F. (2016). Higher order elicitability and Osband's principle. *Ann. Statist.* **44** 1680–1707. MR3519937 https://doi.org/10.1214/16-AOS1439

FISSLER, T. and ZIEGEL, J. F. (2021). Supplement to "Correction note: Higher order elicitability and Osband's principle." https://doi.org/10.1214/20-AOS2014SUPP

MSC2020 subject classifications. Primary 62C99; secondary 91B06.
Key words and phrases. Consistency, decision theory, elicitability, expected shortfall, point forecasts, propriety, scoring functions, scoring rules, spectral risk measures, value at risk.


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