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. 2—April 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*

Vol. 49, No. 2—April 2021

THE ANNALS OF STATISTICS Vol. 49, No. 2, pp. 615–1238 April 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 2, April 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, PO Box 729, Middletown, Maryland 21769, USA.

The Annals of Statistics
2021, Vol. 49, No. 2, 615–649
https://doi.org/10.1214/20-AOS1969

DENSITY DECONVOLUTION UNDER GENERAL ASSUMPTIONS ON THE DISTRIBUTION OF MEASUREMENT ERRORS

BY DENIS BELOMESTNY[1,3] AND ALEXANDER GOLDENSHLUGER[2,3]

[1]*Faculty of Mathematics, Duisburg-Essen University, denis.belomestny@uni-due.de*
[2]*Department of Statistics, University of Haifa, goldensh@stat.haifa.ac.il*
[3]*Department of Computer Sciences, National Research University Higher School of Economics*

In this paper, we study the problem of density deconvolution under general assumptions on the measurement error distribution. Typically, deconvolution estimators are constructed using Fourier transform techniques, and it is assumed that the characteristic function of the measurement errors does not have zeros on the real line. This assumption is rather strong and is not fulfilled in many cases of interest. In this paper, we develop a methodology for constructing optimal density deconvolution estimators in the general setting that covers vanishing and nonvanishing characteristic functions of the measurement errors. We derive upper bounds on the risk of the proposed estimators and provide sufficient conditions under which zeros of the corresponding characteristic function have no effect on estimation accuracy. Moreover, we show that the derived conditions are also necessary in some specific problem instances.

REFERENCES

ANDERSSEN, R. S. (1980). On the use of linear functionals for Abel-type integral equations in applications. In *Application and Numerical Solution of Integral Equations* (*Proc*. *Sem*., *Australian Nat*. *Univ*., *Canberra*, 1978). *Monographs Textbooks Mech*. *Solids Fluids*: *Mech*. *Anal*. **6** 195–221. Nijhoff, The Hague. MR0582991

AUBIN, J.-P. (2000). *Applied Functional Analysis*, 2nd ed. Wiley, New York.

BUTUCEA, C. and TSYBAKOV, A. B. (2008a). Sharp optimality in density deconvolution with dominating bias. I. *Theory Probab*. *Appl*. **52** 24–39.

BUTUCEA, C. and TSYBAKOV, A. B. (2008b). Sharp optimality in density deconvolution with dominating bias. II. *Theory Probab*. *Appl*. **52** 237–249.

CARROLL, R. J. and HALL, P. (1988). Optimal rates of convergence for deconvolving a density. *J*. *Amer*. *Statist*. *Assoc*. **83** 1184–1186. MR0997599

COMTE, F. and LACOUR, C. (2013). Anisotropic adaptive kernel deconvolution. *Ann*. *Inst*. *Henri Poincaré Probab*. *Stat*. **49** 569–609. MR3088382 https://doi.org/10.1214/11-aihp470

DELAIGLE, A. and MEISTER, A. (2011). Nonparametric function estimation under Fourier-oscillating noise. *Statist*. *Sinica* **21** 1065–1092. MR2827515 https://doi.org/10.5705/ss.2009.082

DEVROYE, L. (1989). Consistent deconvolution in density estimation. *Canad*. *J*. *Statist*. **17** 235–239. MR1033106 https://doi.org/10.2307/3314852

FAN, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann*. *Statist*. **19** 1257–1272. MR1126324 https://doi.org/10.1214/aos/1176348248

FEUERVERGER, A., KIM, P. T. and SUN, J. (2008). On optimal uniform deconvolution. *J*. *Stat*. *Theory Pract*. **2** 433–451. MR2528791 https://doi.org/10.1080/15598608.2008.10411885

GOLBERG, M. A. (1979). A method of adjoints for solving some ill-posed equations of the first kind. *Appl*. *Math*. *Comput*. **5** 123–129. MR0522852 https://doi.org/10.1016/0096-3003(79)90002-X

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

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

MSC2020 subject classifications. 62G07, 62G20.
Key words and phrases. Density deconvolution, minimax risk, characteristic function, Laplace transform, lower bounds, density estimation.

GROENEBOOM, P. and JONGBLOED, G. (2003). Density estimation in the uniform deconvolution model. *Stat. Neerl.* **57** 136–157. MR2035863 https://doi.org/10.1111/1467-9574.00225

HALL, P. and MEISTER, A. (2007). A ridge-parameter approach to deconvolution. *Ann. Statist.* **35** 1535–1558. MR2351096 https://doi.org/10.1214/009053607000000028

HALL, P., RUYMGAART, F., VAN GAANS, O. and VAN ROOIJ, A. (2001). Inverting noisy integral equations using wavelet expansions: A class of irregular convolutions. In *State of the Art in Probability and Statistics* (*Leiden*, 1999). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **36** 533–546. IMS, Beachwood, OH. MR1836579 https://doi.org/10.1214/lnms/1215090087

JOHNSTONE, I. M. and RAIMONDO, M. (2004). Periodic boxcar deconvolution and Diophantine approximation. *Ann. Statist.* **32** 1781–1804. MR2102493 https://doi.org/10.1214/009053604000000391

LEPSKI, O. V. and WILLER, T. (2019). Oracle inequalities and adaptive estimation in the convolution structure density model. *Ann. Statist.* **47** 233–287. MR3909933 https://doi.org/10.1214/18-AOS1687

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

LOUNICI, K. and NICKL, R. (2011). Global uniform risk bounds for wavelet deconvolution estimators. *Ann. Statist.* **39** 201–231. MR2797844 https://doi.org/10.1214/10-AOS836

MEISTER, A. (2008). Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions. *Inverse Probl.* **24** 015003. MR2384762 https://doi.org/10.1088/0266-5611/24/1/015003

MEISTER, A. (2009). *Deconvolution Problems in Nonparametric Statistics*. *Lecture Notes in Statistics* **193**. Springer, Berlin. MR2768576 https://doi.org/10.1007/978-3-540-87557-4

MEISTER, A. and NEUMANN, M. H. (2010). Deconvolution from non-standard error densities under replicated measurements. *Statist. Sinica* **20** 1609–1636. MR2777338

STANLEY, R. P. (1997). *Enumerative Combinatorics*. *Vol.* 1. *Cambridge Studies in Advanced Mathematics* **49**. Cambridge Univ. Press, Cambridge. With a foreword by Gian-Carlo Rota. Corrected reprint of the 1986 original. MR1442260 https://doi.org/10.1017/CBO9780511805967

TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. Revised and extended from the 2004 French original. Translated by Vladimir Zaiats. MR2724359 https://doi.org/10.1007/b13794

WIDDER, D. V. (1946). *The Laplace Transform*. Princeton Univ. Press, Princeton.

WILF, H. S. (2006). *generatingfunctionology*, 3rd ed. AK Peters, Wellesley, MA. MR2172781

ZHANG, C.-H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18** 806–831. MR1056338 https://doi.org/10.1214/aos/1176347627

The Annals of Statistics
2021, Vol. 49, No. 2, 650–672
https://doi.org/10.1214/20-AOS1972

HIGH-DIMENSIONAL NONPARAMETRIC DENSITY ESTIMATION VIA SYMMETRY AND SHAPE CONSTRAINTS

BY MIN XU[1] AND RICHARD J. SAMWORTH[2]

[1]*Department of Statistics, Rutgers University, mx76@stat.rutgers.edu*
[2]*Statistical Laboratory, University of Cambridge r.samworth@statslab.cam.ac.uk*

We tackle the problem of high-dimensional nonparametric density estimation by taking the class of log-concave densities on $\mathbb{R}^p$ and incorporating within it symmetry assumptions, which facilitate scalable estimation algorithms and can mitigate the curse of dimensionality. Our main symmetry assumption is that the super-level sets of the density are K-homothetic (i.e., scalar multiples of a convex body $K \subseteq \mathbb{R}^p$). When K is known, we prove that the K-homothetic log-concave maximum likelihood estimator based on n independent observations from such a density achieves the minimax optimal rate of convergence with respect to, for example, squared Hellinger loss, of order $n^{-4/5}$, independent of p. Moreover, we show that the estimator is adaptive in the sense that if the data generating density admits a special form, then a nearly parametric rate may be attained. We also provide worst case and adaptive risk bounds in cases where K is only known up to a positive definite transformation, and where it is completely unknown and must be estimated nonparametrically. Our estimation algorithms are fast even when n and p are on the order of hundreds of thousands, and we illustrate the strong finite-sample performance of our methods on simulated data.

REFERENCES

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

BALABDAOUI, F. and DOSS, C. R. (2018). Inference for a two-component mixture of symmetric distributions under log-concavity. *Bernoulli* **24** 1053–1071. MR3706787 https://doi.org/10.3150/16-BEJ864

BALL, K. (1997). An elementary introduction to modern convex geometry. In *Flavors of Geometry*. *Math. Sci. Res. Inst. Publ.* **31** 1–58. Cambridge Univ. Press, Cambridge. MR1491097 https://doi.org/10.2977/prims/1195164788

BHATTACHARYA, S. and BICKEL, P. J. (2012). Adaptive estimation in elliptical distributions with extensions to high-dimensions. Preprint. Available at http://www.science.oregonstate.edu/~bhattash/Site/Research.html.

BIAU, G. and DEVROYE, L. (2003). On the risk of estimates for block decreasing densities. *J. Multivariate Anal.* **86** 143–165. MR1994726 https://doi.org/10.1016/S0047-259X(02)00028-3

BRONSTEIN, E. M. (2008). Approximation of convex sets by polytopes. *J. Math. Sci.* **153** 727–762.

BRUNEL, V.-E. (2018). Methods for estimation of convex sets. *Statist. Sci.* **33** 615–632. MR3881211 https://doi.org/10.1214/18-STS669

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.

CULE, M. and SAMWORTH, R. (2010). Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electron. J. Stat.* **4** 254–270. MR2645484 https://doi.org/10.1214/09-EJS505

CULE, M., SAMWORTH, R. and STEWART, M. (2010). Maximum likelihood estimation of a multi-dimensional log-concave density. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 545–607. MR2758237 https://doi.org/10.1111/j.1467-9868.2010.00753.x

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

MSC2020 subject classifications. 62G07.
Key words and phrases. Density estimation, nonparametric estimation, shape-constrained estimation, high-dimensional statistics.

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

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

DÜMBGEN, L., HÜSLER, A. and RUFIBACH, K. (2007). Active set and EM algorithms for log-concave densities based on complete and censored data. Preprint. Available at arXiv:0707.4643.

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

DYER, M., FRIEZE, A. and KANNAN, R. (1991). A random polynomial-time algorithm for approximating the volume of convex bodies. *J. Assoc. Comput. Mach.* **38** 1–17. MR1095916 https://doi.org/10.1145/102782.102783

FENG, O. Y., GUNTUBOYINA, A., KIM, A. K. H. and SAMWORTH, R. J. (2018). Adaptation in multivariate log-concave density estimation. Preprint. Available at arxiv:1812.11634.

FERNÁNDEZ, C., OSIEWALSKI, J. and STEEL, M. F. J. (1995). Modeling and inference with υ-spherical distributions. *J. Amer. Statist. Assoc.* **90** 1331–1340. MR1379475

FISHER, R. A. (1922). On the interpretation of χ^2 from contingency tables, and the calculation of p. *J. R. Stat. Soc.* **85** 87–94.

FISHER, R. A. (1925). Theory of statistical estimation. *Math. Proc. Cambridge Philos. Soc.* **22** 700–725.

GAO, F. and WELLNER, J. A. (2007). Entropy estimate for high-dimensional monotonic functions. *J. Multivariate Anal.* **98** 1751–1764. MR2392431 https://doi.org/10.1016/j.jmva.2006.09.003

GRENANDER, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39** 125–153. MR0093415 https://doi.org/10.1080/03461238.1956.10414944

GU, C. and QIU, C. (1993). Smoothing spline density estimation: Theory. *Ann. Statist.* **21** 217–234. MR1212174 https://doi.org/10.1214/aos/1176349023

HAN, Q. and WELLNER, J. A. (2016). Approximation and estimation of s-concave densities via Rényi divergences. *Ann. Statist.* **44** 1332–1359. MR3485962 https://doi.org/10.1214/15-AOS1408

IBRAGIMOV, I. A. and KHASMINSKII, R. (1983). Estimation of distribution density. *J. Math. Sci.* **21** 40–57.

JOHN, F. (1948). Extremum problems with inequalities as subsidiary conditions. In *Studies and Essays Presented to R. Courant on His* 60*th Birthday*, *January* 8, 1948 187–204. Interscience Publishers, New York, NY. MR0030135

KAMIYA, H., TAKEMURA, A. and KURIKI, S. (2008). Star-shaped distributions and their generalizations. *J. Statist. Plann. Inference* **138** 3429–3447. MR2450085 https://doi.org/10.1016/j.jspi.2006.03.016

KANNAN, R., LOVÁSZ, L. and SIMONOVITS, M. (1997). Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. *Random Structures Algorithms* **11** 1–50. MR1608200 https://doi.org/10.1002/(SICI)1098-2418(199708)11:1<1::AID-RSA1>3.0.CO;2-X

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

KUR, G., DAGAN, Y. and RAKHLIN, A. (2019). The log-concave maximum likelihood estimator is optimal in high dimensions. Preprint. Available at arXiv:1903.05315.

LIEBSCHER, E. and RICHTER, W.-D. (2016). Estimation of star-shaped distributions. *Risks* **4** 44.

LIU, H., XU, M., GU, H., GUPTA, A., LAFFERTY, J. and WASSERMAN, L. (2011). Forest density estimation. *J. Mach. Learn. Res.* **12** 907–951. MR2786914 https://doi.org/10.1016/j.micres.2009.11.010

POLONIK, W. (1995). Density estimation under qualitative assumptions in higher dimensions. *J. Multivariate Anal.* **55** 61–81. MR1365636 https://doi.org/10.1006/jmva.1995.1067

POLONIK, W. (1998). The silhouette, concentration functions and ML-density estimation under order restrictions. *Ann. Statist.* **26** 1857–1877. MR1673281 https://doi.org/10.1214/aos/1024691360

RICHTER, W.-D. (2014). Geometric disintegration and star-shaped distributions. *J. Stat. Distrib. Appl* **1**.

RICHTER, W.-D. (2015). Convex and radially concave contoured distributions. *J. Probab. Stat.* **12** Art. ID 165468. MR3431523 https://doi.org/10.1155/2015/165468

RICHTER, W.-D. (2019). High-dimensional star-shaped distributions. *J. Stat. Distrib. App* **6**.

ROBEVA, E., STURMFELS, B., TRAN, N. and UHLER, C. (2018). Maximum likelihood estimation for totally positive log-concave densities. Preprint. Available at arxiv:1806.10120.

ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Stat.* **27** 832–837. MR0079873 https://doi.org/10.1214/aoms/1177728190

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

SIMONOVITS, M. (2003). How to compute the volume in high dimension? *Math. Program.* **97** 337–374. MR2004402 https://doi.org/10.1007/s10107-003-0447-x

WAND, M. P. and JONES, M. C. (1995). *Kernel Smoothing. Monographs on Statistics and Applied Probability* **60**. CRC Press, London. MR1319818 https://doi.org/10.1007/978-1-4899-4493-1

XU, M. and SAMWORTH, R. J (2021). Supplement to "High-dimensional nonparametric density estimation via symmetry and shape constraints." https://doi.org/10.1214/20-AOS1972SUPP

The Annals of Statistics
2021, Vol. 49, No. 2, 673–701
https://doi.org/10.1214/20-AOS1973

AVERAGE TREATMENT EFFECTS IN THE PRESENCE OF UNKNOWN INTERFERENCE

BY FREDRIK SÄVJE[1], PETER M. ARONOW[2] AND MICHAEL G. HUDGENS[3]

[1]*Department of Political Science, Department of Statistics & Data Science, Yale University, fredrik.savje@yale.edu*

[2]*Department of Political Science, Department of Public Health (Biostatistics), Department of Statistics & Data Science, Yale University, peter.aronow@yale.edu*

[3]*Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, mhudgens@bios.unc.edu*

We investigate large-sample properties of treatment effect estimators under unknown interference in randomized experiments. The inferential target is a generalization of the average treatment effect estimand that marginalizes over potential spillover effects. We show that estimators commonly used to estimate treatment effects under no interference are consistent for the generalized estimand for several common experimental designs under limited but otherwise arbitrary and unknown interference. The rates of convergence depend on the growth rate of the unit-average amount of interference and the degree to which the interference aligns with dependencies in treatment assignment. Importantly for practitioners, the results imply that even if one erroneously assumes that units do not interfere in a setting with moderate interference, standard estimators are nevertheless likely to be close to an average treatment effect if the sample is sufficiently large. Conventional confidence statements may, however, not be accurate.

REFERENCES

ANGRIST, J. D. (2014). The perils of peer effects. *Labour Econ*. **30** 98–108. https://doi.org/10.1016/j.labeco.2014.05.008

ARONOW, P. M. (2012). A general method for detecting interference between units in randomized experiments. *Sociol. Methods Res*. **41** 3–16. MR3190698 https://doi.org/10.1177/0049124112437535

ARONOW, P. M., CRAWFORD, F. W. and ZUBIZARRETA, J. R. (2018). Confidence intervals for linear unbiased estimators under constrained dependence. *Electron. J. Stat*. **12** 2238–2252. MR3830833 https://doi.org/10.1214/18-EJS1448

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

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

BASSE, G. W. and AIROLDI, E. M. (2018a). Model-assisted design of experiments in the presence of network-correlated outcomes. *Biometrika* **105** 849–858. MR3877869 https://doi.org/10.1093/biomet/asy036

BASSE, G. W. and AIROLDI, E. M. (2018b). Limitations of design-based causal inference and A/B testing under arbitrary and network interference. *Sociol. Method*. **48** 136–151. https://doi.org/10.1177/0081175018782569

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

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

BLUM, J. R., HANSON, D. L. and KOOPMANS, L. H. (1963). On the strong law of large numbers for a class of stochastic processes. *Z. Wahrsch. Verw. Gebiete* **2** 1–11. MR0161369 https://doi.org/10.1007/BF00535293

BOWERS, J., FREDRICKSON, M. M. and PANAGOPOULOS, C. (2013). Reasoning about interference between units: A general framework. *Polit. Anal*. **21** 97–124. https://doi.org/10.1093/pan/mps038

BRAMOULLÉ, Y., DJEBBARI, H. and FORTIN, B. (2009). Identification of peer effects through social networks. *J. Econometrics* **150** 41–55. MR2525993 https://doi.org/10.1016/j.jeconom.2008.12.021

MSC2020 subject classifications. Primary 62G20; secondary 62D99, 62K99.
Key words and phrases. Causal effects, causal inference, experiments, SUTVA.

CHEN, L. H. Y. and SHAO, Q.-M. (2004). Normal approximation under local dependence. *Ann. Probab.* **32** 1985–2028. MR2073183 https://doi.org/10.1214/009117904000000450

CHETTY, R. (2009). Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Ann. Rev. Econ.* **1** 451–488. https://doi.org/10.1146/annurev.economics.050708.142910

CHOI, D. (2017). Estimation of monotone treatment effects in network experiments. *J. Amer. Statist. Assoc.* **112** 1147–1155. MR3735366 https://doi.org/10.1080/01621459.2016.1194845

COX, D. R. (1958). *Planning of Experiments*. Wiley, New York; CRC Press, London. MR0095561

ECK, D. J., MOROZOVA, O. and CRAWFORD, F. W. (2018). Randomization for the direct effect of an infectious disease intervention in a clustered study population. Available at arXiv:1808.05593.

ECKLES, D., KARRER, B. and UGANDER, J. (2016). Design and analysis of experiments in networks: Reducing bias from interference. *J. Causal Inference* **5**. https://doi.org/10.1515/jci-2015-0021

EDWARDS, D. A. (2011). On the Kantorovich–Rubinstein theorem. *Expo. Math.* **29** 387–398. MR2861765 https://doi.org/10.1016/j.exmath.2011.06.005

EGAMI, N. (2017). Unbiased estimation and sensitivity analysis for network-specific spillover effects: Application to an online network experiment. Available at arXiv:1708.08171.

FISHER, R. A. (1935). *The Design of Experiments*. Oliver & Boyd, London.

FOGARTY, C. B. (2018). On mitigating the analytical limitations of finely stratified experiments. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 1035–1056. MR3874309 https://doi.org/10.1111/rssb.12290

FORASTIERE, L., AIROLDI, E. M. and MEALLI, F. (2017). Identification and estimation of treatment and interference effects in observational studies on networks. Available at arXiv:1609.06245.

GOLDSMITH-PINKHAM, P. and IMBENS, G. W. (2013). Social networks and the identification of peer effects. *J. Bus. Econom. Statist.* **31** 253–264. MR3173674 https://doi.org/10.1080/07350015.2013.801251

GRAHAM, B. S. (2008). Identifying social interactions through conditional variance restrictions. *Econometrica* **76** 643–660. MR2406869 https://doi.org/10.1111/j.1468-0262.2008.00850.x

GREEN, D. P. and GERBER, A. S. (2004). *Get Out the Vote*: *How to Increase Voter Turnout*. Brookings Institution Press, Washington, DC.

HAHN, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica* **66** 315–331. MR1612242 https://doi.org/10.2307/2998560

HÁJEK, J. (1971). Comment: An essay on the logical foundations of survey sampling, part one. In *Foundations of Statistical Inference* (V. P. Godambe and D. A. Sprott, eds.) Holt, Rinehart and Winston, Toronto.

HALLORAN, M. E. and HUDGENS, M. G. (2016). Dependent happenings: A recent methodological review. *Curr. Epidemiol. Rep.* **3** 297–305. https://doi.org/10.1007/s40471-016-0086-4

HALLORAN, M. E. and STRUCHINER, C. J. (1995). Causal inference in infectious diseases. *Epidemiology* **6** 142–151. https://doi.org/10.2307/3702315

HERNÁN, M. A. and ROBINS, J. M. (2006). Estimating causal effects from epidemiological data. *J. Epidemiol. Community Health* **60** 578–586. https://doi.org/10.1136/jech.2004.029496

HIRANO, K., IMBENS, G. W. and RIDDER, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71** 1161–1189. MR1995826 https://doi.org/10.1111/1468-0262.00442

HOLLAND, P. W. (1986). Statistics and causal inference. *J. Amer. Statist. Assoc.* **81** 945–970. MR0867618

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

HUDGENS, M. G. and HALLORAN, M. E. (2008). Toward causal inference with interference. *J. Amer. Statist. Assoc.* **103** 832–842. MR2435472 https://doi.org/10.1198/016214508000000292

IMBENS, G. W. and ANGRIST, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica* **62** 467–475. https://doi.org/10.2307/2951620

ISAKI, C. T. and FULLER, W. A. (1982). Survey design under the regression superpopulation model. *J. Amer. Statist. Assoc.* **77** 89–96. MR0648029

JAGADEESAN, R., PILLAI, N. and VOLFOVSKY, A. (2017). Designs for estimating the treatment effect in networks with interference. Available at arXiv:1705.08524.

KANG, H. and IMBENS, G. W. (2016). Peer encouragement designs in causal inference with partial interference and identification of local average network effects. Available at arXiv:1609.04464.

LEE, L. (2007). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *J. Econometrics* **140** 333–374. MR2408910 https://doi.org/10.1016/j.jeconom.2006.07.001

LIN, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *Ann. Appl. Stat.* **7** 295–318. MR3086420 https://doi.org/10.1214/12-AOAS583

LIU, L. and HUDGENS, M. G. (2014). Large sample randomization inference of causal effects in the presence of interference. *J. Amer. Statist. Assoc.* **109** 288–301. MR3180564 https://doi.org/10.1080/01621459.2013.844698

LIU, L., HUDGENS, M. G. and BECKER-DREPS, S. (2016). On inverse probability-weighted estimators in the presence of interference. *Biometrika* **103** 829–842. MR3620442 https://doi.org/10.1093/biomet/asw047

LUO, X., SMALL, D. S., LI, C.-S. R. and ROSENBAUM, P. R. (2012). Inference with interference between units in an fMRI experiment of motor inhibition. *J. Amer. Statist. Assoc.* **107** 530–541. MR2980065 https://doi.org/10.1080/01621459.2012.655954

MANSKI, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Rev. Econ. Stud.* **60** 531–542. MR1236836 https://doi.org/10.2307/2298123

MANSKI, C. F. (2013). Identification of treatment response with social interactions. *Econom. J.* **16** S1–S23. MR3030060 https://doi.org/10.1111/j.1368-423X.2012.00368.x

NICKERSON, D. W. (2008). Is voting contagious? Evidence from two field experiments. *Am. Polit. Sci. Rev.* **102** 49–57. https://doi.org/10.1017/S0003055408080039

OGBURN, E. L. and VANDERWEELE, T. J. (2017). Vaccines, contagion, and social networks. *Ann. Appl. Stat.* **11** 919–948. MR3693552 https://doi.org/10.1214/17-AOAS1023

RIGDON, J. and HUDGENS, M. G. (2015). Exact confidence intervals in the presence of interference. *Statist. Probab. Lett.* **105** 130–135. MR3371989 https://doi.org/10.1016/j.spl.2015.06.011

ROBINSON, P. M. (1982). On the convergence of the Horvitz–Thompson estimator. *Aust. J. Stat.* **24** 234–238. MR0678263 https://doi.org/10.1111/j.1467-842x.1982.tb00829.x

ROSENBAUM, P. R. (2007). Interference between units in randomized experiments. *J. Amer. Statist. Assoc.* **102** 191–200. MR2345537 https://doi.org/10.1198/016214506000001112

ROSENBLATT, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA* **42** 43–47. MR0074711 https://doi.org/10.1073/pnas.42.1.43

RUBIN, D. B. (1980). Comment: Randomization analysis of experimental data: The Fisher randomization test. *J. Amer. Statist. Assoc.* **75** 591. https://doi.org/10.2307/2287653

SÄVJE, F., ARONOW, P. M and HUDGENS, M. G (2021). Supplements to "Average treatment effects in the presence of unknown interference." https://doi.org/10.1214/20-AOS1973SUPPA, https://doi.org/10.1214/20-AOS1973SUPPB, https://doi.org/10.1214/20-AOS1973SUPPC

SINCLAIR, B., MCCONNELL, M. and GREEN, D. P. (2012). Detecting spillover effects: Design and analysis of multilevel experiments. *Amer. J. Polit. Sci.* **56** 1055–1069. https://doi.org/10.1111/j.1540-5907.2012.00592.x

SOBEL, M. E. (2006). What do randomized studies of housing mobility demonstrate?: Causal inference in the face of interference. *J. Amer. Statist. Assoc.* **101** 1398–1407. MR2307573 https://doi.org/10.1198/016214506000000636

SPLAWA-NEYMAN, J. (1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statist. Sci.* **5** 465–472. MR1092986

SUSSMAN, D. L. and AIROLDI, E. M. (2017). Elements of estimation theory for causal effects in the presence of network interference. Available at arXiv:1702.03578.

TCHETGEN TCHETGEN, E. J., FULCHER, I. and SHPITSER, I. (2019). Auto-g-computation of causal effects on a network. Available at arXiv:1709.01577.

TCHETGEN TCHETGEN, E. J. and VANDERWEELE, T. J. (2012). On causal inference in the presence of interference. *Stat. Methods Med. Res.* **21** 55–75. MR2867538 https://doi.org/10.1177/0962280210386779

TOULIS, P. and KAO, E. (2013). Estimation of causal peer influence effects. In *Proceedings of the* 30*th International Conference on Machine Learning* (S. Dasgupta and D. McAllester, eds.). *Proceedings of Machine Learning Research* **28** 1489–1497. PMLR, Atlanta, GA.

UGANDER, J., KARRER, B., BACKSTROM, L. and KLEINBERG, J. (2013). Graph cluster randomization: Network exposure to multiple universes. In *Proceedings of the* 19*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *KDD* '13 329–337. ACM, New York. https://doi.org/10.1145/2487575.2487695

VANDERWEELE, T. J. and TCHETGEN TCHETGEN, E. J. (2011). Effect partitioning under interference in two-stage randomized vaccine trials. *Statist. Probab. Lett.* **81** 861–869. MR2793754 https://doi.org/10.1016/j.spl.2011.02.019

The Annals of Statistics
2021, Vol. 49, No. 2, 702–722
https://doi.org/10.1214/20-AOS1974

THE ADAPTIVE WYNN ALGORITHM IN GENERALIZED LINEAR MODELS WITH UNIVARIATE RESPONSE

BY FRITJOF FREISE[1], NORBERT GAFFKE[2,*] AND RAINER SCHWABE[2,†]

[1]*Department of Biometry, Epidemiology and Information Processing, University of Veterinary Medicine Hannover, fritjof.freise@tiho-hannover.de*

[2]*Faculty of Mathematics, University of Magdeburg, *norbert.gaffke@ovgu.de; †rainer.schwabe@ovgu.de*

For a nonlinear regression model, the information matrices of designs depend on the parameter of the model. The adaptive Wynn algorithm for D-optimal design estimates the parameter at each step on the basis of the observed responses and employed design points so far, and selects the next design point as in the classical Wynn algorithm for D-optimal design. The name "Wynn algorithm" is in honor of Henry P. Wynn who established the latter "classical" algorithm in his 1970 paper (*Ann. Math. Stat.* **41** (1970) 1655–1664). The asymptotics of the sequences of designs and maximum likelihood estimates generated by the adaptive algorithm is studied for an important class of nonlinear regression models: generalized linear models whose (univariate) response variables follow a distribution from a one-parameter exponential family. Under the assumptions of compactness of the experimental region and of the parameter space together with some natural continuity assumptions, it is shown that the adaptive ML-estimators are strongly consistent and the design sequence is asymptotically locally D-optimal at the true parameter point. If the true parameter point is an interior point of the parameter space, then under some smoothness assumptions the asymptotic normality of the adaptive ML-estimators is obtained.

REFERENCES

[1] ATKINSON, A. C., FEDOROV, V. V., HERZBERG, A. M. and ZHANG, R. (2014). Elemental information matrices and optimal experimental design for generalized regression models. *J. Statist. Plann. Inference* **144** 81–91. MR3131364 https://doi.org/10.1016/j.jspi.2012.09.012

[2] ATKINSON, A. C. and WOODS, D. C. (2015). Designs for generalized linear models. In *Handbook of Design and Analysis of Experiments*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 471–514. CRC Press, Boca Raton, FL. MR3699359

[3] BREIMAN, L. (1992). *Probability*. *Classics in Applied Mathematics* **7**. SIAM, Philadelphia, PA. MR1163370 https://doi.org/10.1137/1.9781611971286

[4] CHEN, K., HU, I. and YING, Z. (1999). Strong consistency of maximum quasi-likelihood estimators in generalized linear models with fixed and adaptive designs. *Ann. Statist.* **27** 1155–1163. MR1740117 https://doi.org/10.1214/aos/1017938919

[5] FAHRMEIR, L. and KAUFMANN, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann. Statist.* **13** 342–368. MR0773172 https://doi.org/10.1214/aos/1176346597

[6] FEDOROV, V. V. (1972). *Theory of Optimal Experiments*. Academic Press, New York. MR0403103

[7] FREISE, F. (2016). On Convergence of the Maximum Likelihood Estimator in Adaptive Designs. Dissertation, University of Magdeburg.

[8] FREISE, F., GAFFKE, N. and SCHWABE, R. (2019). Convergence of least squares estimators in the adaptive Wynn algorithm for a class of nonlinear regression models. Preprint. Available at arXiv:1909.03763.

[9] FREISE, F., GAFFKE, N. and SCHWABE, R. (2021). Supplement to "The adaptive Wynn algorithm in generalized linear models with univariate response." https://doi.org/10.1214/20-AOS1974SUPP.

[10] HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*: *Probability and Mathematical Statistics*. Academic Press, New York-London. MR0624435

MSC2020 subject classifications. Primary 62L05; secondary 62F12, 62J12.

Key words and phrases. Nonlinear regression, approximate design, D-optimality, adaptive estimation, strong consistency, asymptotic normality.

[11] HU, I. (1998). On sequential designs in nonlinear problems. *Biometrika* **85** 496–503. MR1649131 https://doi.org/10.1093/biomet/85.2.496

[12] LAI, T. L. (1994). Asymptotic properties of nonlinear least squares estimates in stochastic regression models. *Ann. Statist.* **22** 1917–1930. MR1329175 https://doi.org/10.1214/aos/1176325764

[13] LAI, T. L. and WEI, C. Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Ann. Statist.* **10** 154–166. MR0642726

[14] PRONZATO, L. (2009). Asymptotic properties of nonlinear estimates in stochastic models with finite design space. *Statist. Probab. Lett.* **79** 2307–2313. MR2556381 https://doi.org/10.1016/j.spl.2009.07.025

[15] PRONZATO, L. (2010). One-step ahead adaptive D-optimal design on a finite design space is asymptotically optimal. *Metrika* **71** 219–238. MR2602189 https://doi.org/10.1007/s00184-008-0227-y

[16] PUKELSHEIM, F. (1993). *Optimal Design of Experiments*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York. MR1211416

[17] 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

[18] WU, C.-F. (1981). Asymptotic theory of nonlinear least squares estimation. *Ann. Statist.* **9** 501–513. MR0615427

[19] WU, C. F. and WYNN, H. P. (1978). The convergence of general step-length algorithms for regular optimum design criteria. *Ann. Statist.* **6** 1273–1285. MR0523762

[20] WYNN, H. P. (1970). The sequential generation of D-optimum experimental designs. *Ann. Math. Stat.* **41** 1655–1664. MR0267704 https://doi.org/10.1214/aoms/1177696809

The Annals of Statistics
2021, Vol. 49, No. 2, 723–744
https://doi.org/10.1214/20-AOS1975

OPTIMAL DISCLOSURE RISK ASSESSMENT

BY FEDERICO CAMERLENGHI[1], STEFANO FAVARO[2], ZACHARIE NAULET[3] AND FRANCESCA PANERO[4]

Dedicated to the memory of Chris Skinner

[1]*Department of Economics, Management and Statistics, University of Milano–Bicocca*, federico.camerlenghi@unimib.it
[2]*Department of Economics and Statistics, Univeristy of Torino*, stefano.favaro@unito.it
[3]*Université Paris-Saclay, Laboratoire de mathématiques d'Orsay*, zacharie.naulet@universite-paris-saclay.fr
[4]*Department of Statistics, University of Oxford*, francesca.panero@stats.ox.ac.uk

Protection against disclosure is a legal and ethical obligation for agencies releasing microdata files for public use. Consider a microdata sample of size n from a finite population of size $\bar{n} = n + \lambda n$, with $\lambda > 0$, such that each sample record contains two disjoint types of information: identifying categorical information and sensitive information. Any decision about releasing data is supported by the estimation of measures of disclosure risk, which are defined as discrete functionals of the number of sample records with a unique combination of values of identifying variables. The most common measure is arguably the number τ_1 of sample unique records that are population uniques. In this paper, we first study nonparametric estimation of τ_1 under the Poisson abundance model for sample records. We introduce a class of linear estimators of τ_1 that are simple, computationally efficient and scalable to massive datasets, and we give uniform theoretical guarantees for them. In particular, we show that they provably estimate τ_1 all of the way up to the sampling fraction $(\lambda + 1)^{-1} \propto (\log n)^{-1}$, with vanishing normalized mean-square error (NMSE) for large n. We then establish a lower bound for the minimax NMSE for the estimation of τ_1, which allows us to show that: (i) $(\lambda + 1)^{-1} \propto (\log n)^{-1}$ is the smallest possible sampling fraction for consistently estimating τ_1; (ii) estimators' NMSE is near optimal, in the sense of matching the minimax lower bound, for large n. This is the main result of our paper, and it provides a rigorous answer to an open question about the feasibility of nonparametric estimation of τ_1 under the Poisson abundance model and for a sampling fraction $(\lambda + 1)^{-1} < 1/2$.

REFERENCES

[1] BEN-HAMOU, A., BOUCHERON, S. and OHANNESSIAN, M. I. (2017). Concentration inequalities in the infinite urn scheme for occupancy counts and the missing mass, with applications. *Bernoulli* **23** 249–287. MR3556773 https://doi.org/10.3150/15-BEJ743

[2] BETHLEHEM, J. G., KELLER, W. J. and PANNEKOEK, J. (1990). Disclosure control of microdata. *J. Amer. Statist. Assoc.* **85** 38–45.

[3] CAMERLENGHI, F., FAVARO, S., NAULET, Z. and PANERO, F. (2021). Supplement to "Optimal disclosure risk assessment." https://doi.org/10.1214/20-AOS1975SUPP

[4] CAROTA, C., FILIPPONE, M., LEOMBRUNI, R. and POLETTINI, S. (2015). Bayesian nonparametric disclosure risk estimation via mixed effects log-linear models. *Ann. Appl. Stat.* **9** 525–546. MR3341126 https://doi.org/10.1214/15-AOAS807

[5] CAROTA, C., FILIPPONE, M. and POLETTINI, S. (2018). Assessing Bayesian nonparametric log-linear models: An application to disclosure risk estimation. Preprint. Available at arXiv:1801.05244.

[6] DAVIDSON, K. R. and DONSIG, A. P. (2010). *Real Analysis and Applications*: *Theory in Practice*. *Undergraduate Texts in Mathematics*. Springer, New York. MR2568574 https://doi.org/10.1007/978-0-387-98098-0

MSC2020 subject classifications. Primary 62G05; secondary 62C20.
Key words and phrases. Disclosure risk assessment, microdata sample, nonparametric inference, optimal minimax procedure, Poisson abundance model, polynomial approximation.

[7] EFRON, B. and MORRIS, C. (1973). Stein's estimation rule and its competitors—An empirical Bayes approach. *J. Amer. Statist. Assoc.* **68** 117–130. MR0388597

[8] EFRON, B. and THISTED, R. (1976). Estimating the number of unseen species: How many words did Shakespeare know? *Biometrika* **63** 435–447.

[9] FIENBERG, S. E. and MAKOV, U. E. (1998). Condentiality, uniqueness, and disclosure limitation for categorical data. *J. Off. Stat.* **14** 385–397.

[10] GOOD, I. J. and TOULMIN, G. H. (1956). The number of new species, and the increase in population coverage, when a sample is increased. *Biometrika* **43** 45–63. MR0077039 https://doi.org/10.1093/biomet/43.1-2.45

[11] JIAO, J., VENKAT, K., HAN, Y. and WEISSMAN, T. (2015). Minimax estimation of functionals of discrete distributions. *IEEE Trans. Inf. Theory* **61** 2835–2885. MR3342309 https://doi.org/10.1109/TIT.2015.2412945

[12] LUKE, Y. L. (1972). Inequalities for generalized hypergeometric functions. *J. Approx. Theory* **5** 41–65. MR0350082 https://doi.org/10.1016/0021-9045(72)90028-7

[13] MANRIQUE-VALLIER, D. and REITER, J. P. (2012). Estimating identification disclosure risk using mixed membership models. *J. Amer. Statist. Assoc.* **107** 1385–1394. MR3036402 https://doi.org/10.1080/01621459.2012.710508

[14] MANRIQUE-VALLIER, D. and REITER, J. P. (2014). Bayesian estimation of discrete multivariate latent structure models with structural zeros. *J. Comput. Graph. Statist.* **23** 1061–1079. MR3270711 https://doi.org/10.1080/10618600.2013.844700

[15] MOSSEL, E. and OHANNESSIAN, M. I. (2019). On the impossibility of learning the missing mass. *Entropy* **21** Art. ID 28.

[16] NEWMAN, D. J. and RIVLIN, T. J. (1976). Approximation of monomials by lower degree polynomials. *Aequationes Math.* **14** 451–455. MR0410181 https://doi.org/10.1007/BF01835995

[17] OHANNESSIAN, M. I. and DAHLEH, M. A. (2012). Rare probability estimation under regularly varying heavy tails. *J. Mach. Learn. Res.* **23** 1–24.

[18] OLVER, F. W. J., LOZIER, D. W., BOISVERT, R. F. and CLARK, C. W. (2010). *NIST Handbook of Mathematical Functions*. Cambridge Univ. Press, Cambridge.

[19] ORLITSKY, A., SURESH, A. T. and WU, Y. (2016). Optimal prediction of the number of unseen species. *Proc. Natl. Acad. Sci. USA* **113** 13283–13288. MR3582444 https://doi.org/10.1073/pnas.1607774113

[20] REITER, J. P. (2005). Estimating risks of identification disclosure in microdata. *J. Amer. Statist. Assoc.* **100** 1103–1112. MR2236926 https://doi.org/10.1198/016214505000000619

[21] RINOTT, Y. and SHLOMO, N. (2006). A generalized negative binomial smoothing model for sample disclosure risk estimation. In *Privacy in Statistical Databases*. *Lecture Notes in Computer Science*. Springer, Berlin.

[22] ROBBINS, H. (1956). An empirical Bayes approach to statistics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, *Vol. I* 157–163. Univ. California Press, Berkeley, CA. MR0084919

[23] SAMUELS, S. M. (1998). A Bayesian, species-sampling-inspired approach to the uniques problem in microdata disclosure risk assessment. *J. Off. Stat.* **14** 373–383.

[24] SKINNER, C., MARSH, C., OPENSHAW, S. and WYMER, C. (1994). Disclosure control for census microdata. *J. Off. Stat.* **10** 31–51.

[25] SKINNER, C. and SHLOMO, N. (2008). Assessing identification risk in survey microdata using log-linear models. *J. Amer. Statist. Assoc.* **103** 989–1001. MR2462887 https://doi.org/10.1198/016214507000001328

[26] SKINNER, C. J. and ELLIOT, M. J. (2002). A measure of disclosure risk for microdata. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 855–867. MR1979391 https://doi.org/10.1111/1467-9868.00365

[27] SONI, R. P. (1965). On an inequality for modified Bessel functions. *J. Math. Phys.* **44** 406–407. MR0185164

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

[29] VALIANT, P. and VALIANT, G. (2013). Estimating the unseen: Improved estimators for entropy and other properties. *Adv. Neural Inf. Process. Syst.* **27** 2157–2165.

[30] WILLENBORG, L. and DE WAAL, T. (2001). *Elements of Statistical Disclosure Control*. *Lecture Notes in Statistics* **155**. Springer, New York. MR1866909 https://doi.org/10.1007/978-1-4613-0121-9

[31] WU, Y. and YANG, P. (2016). Minimax rates of entropy estimation on large alphabets via best polynomial approximation. *IEEE Trans. Inf. Theory* **62** 3702–3720. MR3506758 https://doi.org/10.1109/TIT.2016.2548468

[32] 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. 2, 745–768
https://doi.org/10.1214/20-AOS1976

NETWORK REPRESENTATION USING GRAPH ROOT DISTRIBUTIONS

BY JING LEI

Department of Statistics and Data Science, Carnegie Mellon University, jinglei@andrew.cmu.edu

Exchangeable random graphs serve as an important probabilistic framework for the statistical analysis of network data. In this work, we develop an alternative parameterization for a large class of exchangeable random graphs, where the nodes are independent random vectors in a linear space equipped with an indefinite inner product, and the edge probability between two nodes equals the inner product of the corresponding node vectors. Therefore, the distribution of exchangeable random graphs in this subclass can be represented by a node sampling distribution on this linear space, which we call the *graph root distribution*. We study existence and identifiability of such representations, the topological relationship between the graph root distribution and the exchangeable random graph sampling distribution and estimation of graph root distributions.

REFERENCES

[1] ADAMIC, L. A. and GLANCE, N. (2005). The political blogosphere and the 2004 US election: Divided they blog. In *Proceedings of the 3rd International Workshop on Link Discovery* 36–43. ACM.

[2] AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P. (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.

[3] AIROLDI, E. M., COSTA, T. B. and CHAN, S. H. (2013). Stochastic blockmodel approximation of a graphon: Theory and consistent estimation. In *Advances in Neural Information Processing Systems* 692–700.

[4] ALDOUS, D. J. (1981). Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* **11** 581–598. MR0637937 https://doi.org/10.1016/0047-259X(81)90099-3

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

[6] BHATTACHARYYA, S. and CHATTERJEE, S. (2018). Spectral clustering for multiple sparse networks: I. arXiv preprint arXiv:1805.10594.

[7] BICKEL, P. J. and CHEN, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA* **106** 21068–21073.

[8] BICKEL, P. J., CHEN, A. and LEVINA, E. (2011). The method of moments and degree distributions for network models. *Ann. Statist.* **39** 2280–2301. MR2906868 https://doi.org/10.1214/11-AOS904

[9] BOLLOBÁS, B., JANSON, S. and RIORDAN, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures Algorithms* **31** 3–122. MR2337396 https://doi.org/10.1002/rsa.20168

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

[11] CHAN, S. and AIROLDI, E. (2014). A consistent histogram estimator for exchangeable graph models. In *International Conference on Machine Learning* 208–216.

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

[13] 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

[14] CHIN, P., RAO, A. and VU, V. (2015). Stochastic block model and community detection in sparse graphs: A spectral algorithm with optimal rate of recovery. In *Conference on Learning Theory* 391–423.

[15] COJA-OGHLAN, A. (2010). Graph partitioning via adaptive spectral techniques. *Combin. Probab. Comput.* **19** 227–284. MR2593622 https://doi.org/10.1017/S0963548309990514

MSC2020 subject classifications. Primary 62E10; secondary 62G05, 62M15.
Key words and phrases. Network data, exchangeable random graph, Kreĭn space, spectral embedding.

[16] CRANE, H. and DEMPSEY, W. (2018). Edge exchangeable models for interaction networks. *J. Amer. Statist. Assoc.* **113** 1311–1326. MR3862359 https://doi.org/10.1080/01621459.2017.1341413
[17] 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
[18] GAO, C., MA, Z., ZHANG, A. Y. and ZHOU, H. H. (2018). Community detection in degree-corrected block models. *Ann. Statist.* **46** 2153–2185. MR3845014 https://doi.org/10.1214/17-AOS1615
[19] GOHBERG, I. C. and KREĬN, M. G. (1988). *Introduction to the Theory of Linear Nonselfadjoint Operators*. **18**. Amer. Math. Soc., Providence, R.I. MR0246142
[20] GOLDENBERG, A., ZHENG, A. X., FIENBERG, S. E. and AIROLDI, E. M. (2010). A survey of statistical network models. *Found. Trends Mach. Learn.* **2** 129–233.
[21] HALL, P. and HOROWITZ, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35** 70–91. MR2332269 https://doi.org/10.1214/009053606000000957
[22] HOFF, P. (2008). Modeling homophily and stochastic equivalence in symmetric relational data. In *Advances in Neural Information Processing Systems* 657–664.
[23] HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. MR0718088 https://doi.org/10.1016/0378-8733(83)90021-7
[24] HOOVER, D. N. (1982). Row-column exchangeability and a generalized model for probability. In *Exchangeability in Probability and Statistics* (*Rome*, 1981) 281–291. North-Holland, Amsterdam. MR0675982
[25] JIN, J. (2015). Fast community detection by SCORE. *Ann. Statist.* **43** 57–89. MR3285600 https://doi.org/10.1214/14-AOS1265
[26] KALLENBERG, O. (1989). On the representation theorem for exchangeable arrays. *J. Multivariate Anal.* **30** 137–154. MR1003713 https://doi.org/10.1016/0047-259X(89)90092-4
[27] KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* (3) **83** 016107, 10. MR2788206 https://doi.org/10.1103/PhysRevE.83.016107
[28] KLOPP, O., TSYBAKOV, A. B. and VERZELEN, N. (2017). Oracle inequalities for network models and sparse graphon estimation. *Ann. Statist.* **45** 316–354. MR3611494 https://doi.org/10.1214/16-AOS1454
[29] KLOPP, O. and VERZELEN, N. (2019). Optimal graphon estimation in cut distance. *Probab. Theory Related Fields* **174** 1033–1090. MR3980311 https://doi.org/10.1007/s00440-018-0878-1
[30] KOLACZYK, E. D. (2009). *Statistical Analysis of Network Data*: *Methods and Models*. *Springer Series in Statistics*. Springer, New York. MR2724362 https://doi.org/10.1007/978-0-387-88146-1
[31] KREBS, V. (2004). Social network analysis software & services for organizations, communities, and their consultants. http://www.orgnet.com/.
[32] LAX, P. D. (2002). *Functional Analysis*. *Pure and Applied Mathematics* (*New York*). Wiley Interscience, New York. MR1892228
[33] LEI, J. (2014). Adaptive global testing for functional linear models. *J. Amer. Statist. Assoc.* **109** 624–634. MR3223738 https://doi.org/10.1080/01621459.2013.856794
[34] LEI, J. (2016). A goodness-of-fit test for stochastic block models. *Ann. Statist.* **44** 401–424. MR3449773 https://doi.org/10.1214/15-AOS1370
[35] LEI, J. (2020). Convergence and concentration of empirical measures under Wasserstein distance in unbounded functional spaces. *Bernoulli* **26** 767–798. MR4036051 https://doi.org/10.3150/19-BEJ1151
[36] LEI, J. (2021). Supplement to "Network representation using graph root distributions." https://doi.org/10.1214/20-AOS1976SUPP
[37] LOVÁSZ, L. (2012). *Large Networks and Graph Limits*. *American Mathematical Society Colloquium Publications* **60**. Amer. Math. Soc., Providence, RI. MR3012035 https://doi.org/10.1090/coll/060
[38] MCSHERRY, F. (2001). Spectral partitioning of random graphs. In 42*nd IEEE Symposium on Foundations of Computer Science* (*Las Vegas*, *NV*, 2001) 529–537. IEEE Computer Soc., Los Alamitos, CA. MR1948742
[39] MEISTER, A. (2011). Asymptotic equivalence of functional linear regression and a white noise inverse problem. *Ann. Statist.* **39** 1471–1495. MR2850209 https://doi.org/10.1214/10-AOS872
[40] NEWMAN, M. E. J. (2010). *Networks*: *An Introduction*. Oxford Univ. Press, Oxford. MR2676073 https://doi.org/10.1093/acprof:oso/9780199206650.001.0001
[41] NICKEL, C. L. M. (2007). *Random Dot Product Graphs*: *A Model for Social Networks*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)–The Johns Hopkins University. MR2710154
[42] ONG, C. S., MARY, X., CANU, S. and SMOLA, A. J. (2004). Learning with non-positive kernels. In *Proceedings of the Twenty-First International Conference on Machine Learning* 81. ACM.
[43] ORBANZ, P. and ROY, D. M. (2015). Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Trans. Pattern Anal. Mach. Intell.* **37** 437–461.
[44] PENROSE, M. (2003). *Random Geometric Graphs*. *Oxford Studies in Probability* **5**. Oxford Univ. Press, Oxford. MR1986198 https://doi.org/10.1093/acprof:oso/9780198506263.001.0001

[45] ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. MR2893856 https://doi.org/10.1214/11-AOS887

[46] RUBIN-DELANCHY, P., PRIEBE, C. E. and TANG, M. (2017). The generalised random dot product graph. arXiv preprint arXiv:1709.05506.

[47] 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

[48] SUSSMAN, D. L., TANG, M., FISHKIND, D. E. and PRIEBE, C. E. (2012). A consistent adjacency spectral embedding for stochastic blockmodel graphs. *J. Amer. Statist. Assoc.* **107** 1119–1128. MR3010899 https://doi.org/10.1080/01621459.2012.699795

[49] TANG, M., ATHREYA, A., SUSSMAN, D. L., LYZINSKI, V. and PRIEBE, C. E. (2017). A nonparametric two-sample hypothesis testing problem for random graphs. *Bernoulli* **23** 1599–1630. MR3624872 https://doi.org/10.3150/15-BEJ789

[50] WOLFE, P. J. and OLHEDE, S. C. (2013). Nonparametric graphon estimation. arXiv preprint arXiv:1309.5936.

[51] XU, J. (2018). Rates of convergence of spectral methods for graphon estimation. In *Proceedings of the* 35*th International Conference on Machine Learning* (J. Dy and A. Krause, eds.). *Proceedings of Machine Learning Research* **80** 5433–5442. PMLR, Stockholmsmässan, Stockholm Sweden.

[52] XU, J., MASSOULIÉ, L. and LELARGE, M. (2014). Edge label inference in generalized stochastic block models: From spectral theory to impossibility results. In *Conference on Learning Theory* 903–920.

[53] ZHANG, Y., LEVINA, E. and ZHU, J. (2017). Estimating network edge probabilities by neighbourhood smoothing. *Biometrika* **104** 771–783. MR3737303 https://doi.org/10.1093/biomet/asx042

[54] ZHAO, Y., LEVINA, E. and ZHU, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *Ann. Statist.* **40** 2266–2292. MR3059083 https://doi.org/10.1214/12-AOS1036

The Annals of Statistics
2021, Vol. 49, No. 2, 769–792
https://doi.org/10.1214/20-AOS1977

MULTIVARIATE EXTENSIONS OF ISOTONIC REGRESSION AND TOTAL VARIATION DENOISING VIA ENTIRE MONOTONICITY AND HARDY–KRAUSE VARIATION

BY BILLY FANG[1,*], ADITYANAND GUNTUBOYINA[1,†] AND BODHISATTVA SEN[2]

[1]*Department of Statistics, University of California,* [*]*blfang@berkeley.edu;* [†]*aditya@stat.berkeley.edu*
[2]*Department of Statistics, Columbia University, bodhi@stat.columbia.edu*

We consider the problem of nonparametric regression when the covariate is d dimensional, where $d \geq 1$. In this paper, we introduce and study two nonparametric least squares estimators (LSEs) in this setting—the entirely monotonic LSE and the constrained Hardy–Krause variation LSE. We show that these two LSEs are natural generalizations of univariate isotonic regression and univariate total variation denoising, respectively, to multiple dimensions. We discuss the characterization and computation of these two LSEs obtained from n data points. We provide a detailed study of their risk properties under the squared error loss and fixed uniform lattice design. We show that the finite sample risk of these LSEs is always bounded from above by $n^{-2/3}$ modulo logarithmic factors depending on d; thus these nonparametric LSEs avoid the curse of dimensionality to some extent. We also prove nearly matching minimax lower bounds. Further, we illustrate that these LSEs are particularly useful in fitting rectangular piecewise constant functions. Specifically, we show that the risk of the entirely monotonic LSE is almost parametric (at most $1/n$ up to logarithmic factors) when the true function is well approximable by a rectangular piecewise constant entirely monotone function with not too many constant pieces. A similar result is also shown to hold for the constrained Hardy–Krause variation LSE for a simple subclass of rectangular piecewise constant functions. We believe that the proposed LSEs yield a novel approach to estimating multivariate functions using convex optimization that avoid the curse of dimensionality to some extent.

REFERENCES

[1] AISTLEITNER, C. and DICK, J. (2015). Functions of bounded variation, signed measures, and a general Koksma–Hlawka inequality. *Acta Arith.* **167** 143–171. MR3312093 https://doi.org/10.4064/aa167-2-4

[2] AYER, M., BRUNK, H. D., EWING, G. M., REID, W. T. and SILVERMAN, E. (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Stat.* **26** 641–647. MR0073895 https://doi.org/10.1214/aoms/1177728423

[3] BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions. The Theory and Application of Isotonic Regression*. Wiley, London-Sydney. Wiley Series in Probability and Mathematical Statistics. MR0326887

[4] BARRON, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans. Inf. Theory* **39** 930–945. MR1237720 https://doi.org/10.1109/18.256500

[5] 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

[6] BENKESER, D. and VAN DER LAAN, M. (2016). The highly adaptive lasso estimator. In 2016 *IEEE International Conference on Data Science and Advanced Analytics* (*DSAA*) 689–696. IEEE.

[7] BLEI, R., GAO, F. and LI, W. V. (2007). Metric entropy of high dimensional distributions. *Proc. Amer. Math. Soc.* **135** 4009–4018. MR2341952 https://doi.org/10.1090/S0002-9939-07-08935-6

MSC2020 subject classifications. 62G08.
Key words and phrases. Almost parametric risk, bounded mixed derivative, (constrained) least squares estimation, curse of dimensionality, dimension independent risk, multivariate shape constrained regression, nonparametric regression, risk under the squared error loss.

[8] BRUNK, H. D. (1955). Maximum likelihood estimates of monotone parameters. *Ann. Math. Stat.* **26** 607–616. MR0073894 https://doi.org/10.1214/aoms/1177728420
[9] BUNGARTZ, H.-J. and GRIEBEL, M. (2004). Sparse grids. *Acta Numer.* **13** 147–269. MR2249147 https://doi.org/10.1017/S0962492904000182
[10] CHAMBOLLE, A., CASELLES, V., CREMERS, D., NOVAGA, M. and POCK, T. (2010). An introduction to total variation for image analysis. In *Theoretical Foundations and Numerical Methods for Sparse Recovery*. *Radon Ser. Comput. Appl. Math.* **9** 263–340. de Gruyter, Berlin. MR2731599 https://doi.org/10.1515/9783110226157.263
[11] CHATTERJEE, S. and GOSWAMI, S. (2019). New risk bounds for 2d total variation denoising. arXiv preprint. Available at arXiv:1902.01215.
[12] 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
[13] CHKIFA, A., DEXTER, N., TRAN, H. and WEBSTER, C. G. (2018). Polynomial approximation via compressed sensing of high-dimensional functions on lower sets. *Math. Comp.* **87** 1415–1450. MR3766392 https://doi.org/10.1090/mcom/3272
[14] CONDAT, L. (2013). A direct algorithm for 1-d total variation denoising. *IEEE Signal Process. Lett.* **20** 1054–1057.
[15] CONDAT, L. (2017). Discrete total variation: New definition and minimization. *SIAM J. Imaging Sci.* **10** 1258–1290. MR3684410 https://doi.org/10.1137/16M1075247
[16] DALALYAN, A. S., HEBIRI, M. and LEDERER, J. (2017). On the prediction performance of the Lasso. *Bernoulli* **23** 552–581. MR3556784 https://doi.org/10.3150/15-BEJ756
[17] DENG, H. and ZHANG, C.-H. (2018). Isotonic regression in multi-dimensional spaces and graphs. arXiv preprint. Available at arXiv:1812.08944.
[18] DONOHO, D. L. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. *AMS Math. Challenges Lect.* **1** 375.
[19] 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
[20] FAN, Z. and GUAN, L. (2018). Approximate ℓ_0-penalized estimation of piecewise-constant signals on graphs. *Ann. Statist.* **46** 3217–3245. MR3852650 https://doi.org/10.1214/17-AOS1656
[21] FANG, B., GUNTUBOYINA, A. and SEN, B. (2021). Supplement to "Multivariate extensions of isotonic regression and total variation denoising via entire monotonicity and Hardy–Krause variation." https://doi.org/10.1214/20-AOS1977SUPP
[22] FELLER, W. (2015). Completely monotone functions and sequences. In *Selected Papers I* 497–510. Springer, Berlin.
[23] GAO, F. (2013). Bracketing entropy of high dimensional distributions. In *High Dimensional Probability VI*. *Progress in Probability* **66** 3–17. Birkhäuser/Springer, Basel. MR3443489
[24] GAO, F., LI, W. V. and WELLNER, J. A. (2010). How many Laplace transforms of probability measures are there? *Proc. Amer. Math. Soc.* **138** 4331–4344. MR2680059 https://doi.org/10.1090/S0002-9939-2010-10448-3
[25] GILL, R. D., VAN DER LAAN, M. J. and WELLNER, J. A. (1995). Inefficient estimators of the bivariate survival function for three models. *Ann. Inst. Henri Poincaré Probab. Stat.* **31** 545–597. MR1338452
[26] GROENEBOOM, P. and JONGBLOED, G. (2014). *Nonparametric Estimation Under Shape Constraints*. *Cambridge Series in Statistical and Probabilistic Mathematics* **38**. Cambridge Univ. Press, New York. Estimators, algorithms and asymptotics. MR3445293 https://doi.org/10.1017/CBO9781139020893
[27] GUNTUBOYINA, A. and SEN, B. (2018). Nonparametric shape-restricted regression. *Statist. Sci.* **33** 568–594. MR3881209 https://doi.org/10.1214/18-STS665
[28] GUNTUBOYINA, A., LIEU, D., CHATTERJEE, S. and SEN, B. (2020). Adaptive risk bounds in univariate total variation denoising and trend filtering. *Ann. Statist.* **48** 205–229. MR4065159 https://doi.org/10.1214/18-AOS1799
[29] GUO, D. and WANG, X. (2006). Quasi-Monte Carlo filtering in nonlinear dynamic systems. *IEEE Trans. Signal Process.* **54** 2087–2098.
[30] 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
[31] HOBSON, E. W. (1958). *The Theory of Functions of a Real Variable and the Theory of Fourier's Series*. *Vol. I*. Dover, New York, NY. MR0092828
[32] HÜTTER, J.-C. and RIGOLLET, P. (2016). Optimal rates for total variation denoising. In *Conference on Learning Theory* 1115–1146.
[33] KIM, S.-J., KOH, K., BOYD, S. and GORINEVSKY, D. (2009). l_1 trend filtering. *SIAM Rev.* **51** 339–360. MR2505584 https://doi.org/10.1137/070690274

[34] LEONOV, A. S. (1996). On the total variation for functions of several variables and a multidimensional analog of Helly's selection principle. *Math. Notes* **63** 61–71.

[35] LIN, Y. (2000). Tensor product space ANOVA models. *Ann. Statist.* **28** 734–755. MR1792785 https://doi.org/10.1214/aos/1015951996

[36] LIN, K., SHARPNACK, J. L., RINALDO, A. and TIBSHIRANI, R. J. (2017). A sharp error analysis for the fused lasso, with application to approximate changepoint screening. In *Advances in Neural Information Processing Systems* 6884–6893.

[37] MAMMEN, E. and VAN DE GEER, S. (1997). Locally adaptive regression splines. *Ann. Statist.* **25** 387–413. MR1429931 https://doi.org/10.1214/aos/1034276635

[38] MEYER, M. and WOODROOFE, M. (2000). On the degrees of freedom in shape-restricted regression. *Ann. Statist.* **28** 1083–1104. MR1810920 https://doi.org/10.1214/aos/1015956708

[39] NEMIROVSKI, A. (2000). Topics in non-parametric statistics. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1998). *Lecture Notes in Math.* **1738** 85–277. Springer, Berlin. MR1775640

[40] NIYOGI, P. and GIROSI, F. (1999). Generalization bounds for function approximation from scattered noisy data. *Adv. Comput. Math.* **10** 51–80. MR1671871 https://doi.org/10.1023/A:1018966213079

[41] ORTELLI, F. and VAN DE GEER, S. (2018). On the total variation regularized estimator over a class of tree graphs. *Electron. J. Stat.* **12** 4517–4570. MR3892703 https://doi.org/10.1214/18-ejs1519

[42] ORTELLI, F. and VAN DE GEER, S. (2019). Oracle inequalities for image denoising with total variation regularization. arXiv preprint. Available at arXiv:1911.07231.

[43] ORTELLI, F. and VAN DE GEER, S. (2019). Synthesis and analysis in total variation regularization. arXiv preprint. Available at arXiv:1901.06418.

[44] OWEN, A. B. (2005). Multidimensional variation for quasi-Monte Carlo. In *Contemporary Multivariate Analysis and Design of Experiments*. *Ser. Biostat.* **2** 49–74. World Sci. Publ., Hackensack, NJ. MR2271076

[45] PRAUSE, A. and STELAND, A. (2017). Sequential detection of three-dimensional signals under dependent noise. *Sequential Anal.* **36** 151–178. MR3665833 https://doi.org/10.1080/07474946.2017.1319674

[46] ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, Chichester. MR0961262

[47] RUDIN, L. I., OSHER, S. and FATEMI, E. (1992). Nonlinear total variation based noise removal algorithms. Experimental mathematics: Computational issues in nonlinear science (Los Alamos, NM, 1991). *Phys. D* **60** 259–268. MR3363401 https://doi.org/10.1016/0167-2789(92)90242-F

[48] RUIZ, D Á, M., LI, H. and MUNK, A. (2018). Frame-constrained total variation regularization for white noise regression. arXiv preprint. Available at arXiv:1807.02038.

[49] SADHANALA, V., WANG, Y.-X. and TIBSHIRANI, R. J. (2016). Total variation classes beyond 1d: Minimax rates, and the limitations of linear smoothers. In *Advances in Neural Information Processing Systems* 3513–3521.

[50] SADHANALA, V., WANG, Y.-X., SHARPNACK, J. L. and TIBSHIRANI, R. J. (2017). Higher-order total variation classes on grids: Minimax theory and trend filtering methods. In *Advances in Neural Information Processing Systems* 5800–5810.

[51] STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053. MR0673642

[52] TEMLYAKOV, V. (2018). *Multivariate Approximation*. *Cambridge Monographs on Applied and Computational Mathematics* **32**. Cambridge Univ. Press, Cambridge. MR3837133 https://doi.org/10.1017/9781108689687

[53] TIBSHIRANI, R. J. (2014). Adaptive piecewise polynomial estimation via trend filtering. *Ann. Statist.* **42** 285–323. MR3189487 https://doi.org/10.1214/13-AOS1189

[54] VAN DER LAAN, M. (2017). Finite sample inference for targeted learning. arXiv preprint. Available at arXiv:1708.09502.

[55] VAN DER LAAN, M. (2017). A generally efficient targeted minimum loss based estimator based on the highly adaptive Lasso. *Int. J. Biostat.* **13** 20150097. MR3724476 https://doi.org/10.1515/ijb-2015-0097

[56] VAN DER LAAN, M. J., BENKESER, D. and CAI, W. (2019). Efficient estimation of pathwise differentiable target parameters with the undersmoothed highly adaptive lasso. arXiv preprint. Available at arXiv:1908.05607.

[57] VAN DER LAAN, M. J. and BIBAUT, A. F. (2017). Uniform consistency of the highly adaptive lasso estimator of infinite dimensional parameters. arXiv preprint. Available at arXiv:1709.06256.

[58] VAPNIK, V. N. and CHERVONENKIS, A. Y. (2015). On the uniform convergence of relative frequencies of events to their probabilities. In *Measures of Complexity* 11–30. Springer, Cham. Reprint of Theor. Probability Appl. **1**(6) (1971), 264–280. MR3408730

[59] WAHBA, G., WANG, Y., GU, C., KLEIN, R. and KLEIN, B. (1995). Smoothing spline ANOVA for exponential families, with application to the Wisconsin Epidemiological Study of Diabetic Retinopathy. *Ann. Statist.* **23** 1865–1895. MR1389856 https://doi.org/10.1214/aos/1034713638

[60] WIDDER, D. V. (1941). *The Laplace Transform. Princeton Mathematical Series, V.* 6. Princeton Univ. Press, Princeton, NJ. MR0005923

[61] YANG, Y. and BARRON, A. (1999). Information-theoretic determination of minimax rates of convergence. *Ann. Statist.* **27** 1564–1599. MR1742500 https://doi.org/10.1214/aos/1017939142

[62] YOUNG, W. H. and YOUNG, G. C. (1924). On the Discontinuties of Monotone Functions of Several Variables. *Proc. Lond. Math. Soc.* (2) **22** 124–142. MR1575698 https://doi.org/10.1112/plms/s2-22.1.124

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

[64] ZHANG, T. (2019). Element-wise estimation error of a total variation regularized estimator for change point detection. arXiv preprint. Available at arXiv:1901.00914.

[65] ZIEMER, W. P. (2012). *Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation* **120**. Springer, Berlin.

The Annals of Statistics
2021, Vol. 49, No. 2, 793–819
https://doi.org/10.1214/20-AOS1978

ANALYSIS OF "LEARN-AS-YOU-GO" (LAGO) STUDIES

BY DANIEL NEVO[1], JUDITH J. LOK[2] AND DONNA SPIEGELMAN[3]

[1]*Department of Statistics and Operations Research, Tel Aviv University, danielnevo@gmail.com*
[2]*Department of Mathematics and Statistics, Boston University, jjlok@bu.edu*
[3]*Department of Biostatistics and Center for Methods on Implementation and Prevention Science (CMIPS), Yale School of Public Health, donna.spiegelman@yale.edu*

In Learn-As-you-GO (LAGO) adaptive studies, the intervention is a complex multicomponent package, and is adapted in stages during the study based on past outcome data. This design formalizes standard practice in public health intervention studies. An effective intervention package is sought, while minimizing intervention package cost. In LAGO study data, the interventions in later stages depend upon the outcomes in the previous stages, violating standard statistical theory. We develop an estimator for the intervention effects, and prove consistency and asymptotic normality using a novel coupling argument, ensuring the validity of the test for the hypothesis of no overall intervention effect. We develop a confidence set for the optimal intervention package and confidence bands for the success probabilities under alternative package compositions. We illustrate our methods in the BetterBirth Study, which aimed to improve maternal and neonatal outcomes among 157,689 births in Uttar Pradesh, India through a multicomponent intervention package.

REFERENCES

ALBERT, A. and ANDERSON, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* **71** 1–10. MR0738319 https://doi.org/10.1093/biomet/71.1.1

BAUER, P. and KOHNE, K. (1994). Evaluation of experiments with adaptive interim analyses. *Biometrics* 1029–1041.

BAUER, P., BRETZ, F., DRAGALIN, V., KÖNIG, F. and WASSMER, G. (2016). Twenty-five years of confirmatory adaptive designs: Opportunities and pitfalls. *Stat. Med.* **35** 325–347. MR3455501 https://doi.org/10.1002/sim.6472

BRANNATH, W., POSCH, M. and BAUER, P. (2002). Recursive combination tests. *J. Amer. Statist. Assoc.* **97** 236–244. MR1947283 https://doi.org/10.1198/016214502753479374

COLLINS, L. M., MURPHY, S. A. and STRECHER, V. (2007). The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): New methods for more potent eHealth interventions. *Am. J. Prev. Med.* **32** S112–S118.

COLLINS, L. M., NAHUM-SHANI, I. and ALMIRALL, D. (2014). Optimization of behavioral dynamic treatment regimens based on the sequential, multiple assignment, randomized trial (SMART). *Clin. Trials* **11** 426–434.

COX, D. R. (1958). *Planning of Experiments. A Wiley Publication in Applied Statistics*. Wiley, New York; CRC Press, London. MR0095561

ESCOFFERY, C., LEBOW-SKELLEY, E., UDELSON, H., BÖING, E. A., WOOD, R., FERNANDEZ, M. E. and MULLEN, P. D. (2018). A scoping study of frameworks for adapting public health evidence-based interventions. *Translational Behavioral Medicine*.

FDA (2016). Adaptive designs for medical device clinical studies: Guidance for industry and foodand drug administration staff.

GAO, P., LIU, L. and MEHTA, C. (2013). Exact inference for adaptive group sequential designs. *Stat. Med.* **32** 3991–4005. MR3102430 https://doi.org/10.1002/sim.5847

GAWANDE, A. (2014). *Being Mortal: Medicine and What Matters in the End*. Metropolitan Books.

HERNÁN, M. A. and ROBINS, J. M. (2020). *Causal Inference: What If*. Chapman & Hall/CRC Boca Raton.

MSC2020 subject classifications. 62K99, 62L99, 62F12, 62F10, 62F05, 62J12.
Key words and phrases. Adaptive designs, dependent sample, coupling, public health.

HIRSCHHORN, L. R., SEMRAU, K., KODKANY, B., CHURCHILL, R., KAPOOR, A., SPECTOR, J., RINGER, S., FIRESTONE, R., KUMAR, V. et al. (2015). Learning before leaping: Integration of an adaptive study design process prior to initiation of BetterBirth, a large-scale randomized controlled trial in Uttar Pradesh, India. *Implementation Science* **10** 1.

HU, F. and ROSENBERGER, W. F. (2003). Optimality, variability, power: Evaluating response-adapative randomization procedures for treatment comparisons. *J. Amer. Statist. Assoc.* **98** 671–678. MR2011680 https://doi.org/10.1198/016214503000000576

IWUJI, C. C., ORNE-GLIEMANN, J., LARMARANGE, J., BALESTRE, E., THIEBAUT, R., TANSER, F., OKESOLA, N., MAKOWA, T., DREYER, J. et al. (2017). Universal test and treat and the HIV epidemic in rural South Africa: A phase 4, open-label, community cluster randomised trial. *The Lancet HIV*.

KAIRALLA, J. A., COFFEY, C. S., THOMANN, M. A. and MULLER, K. E. (2012). Adaptive trial designs: A review of barriers and opportunities. *Trials* **13** 145. https://doi.org/10.1186/1745-6215-13-145

LINDVALL, T. (2002). *Lectures on the Coupling Method*. Dover, Mineola, NY. Corrected reprint of the 1992 original. MR1924231

MÜLLER, H.-H. and SCHÄFER, H. (2001). Adaptive group sequential designs for clinical trials: Combining the advantages of adaptive and of classical group sequential approaches. *Biometrics* **57** 886–891. MR1859823 https://doi.org/10.1111/j.0006-341X.2001.00886.x

MÜLLER, H.-H. and SCHÄFER, H. (2004). A general statistical principle for changing a design any time during the course of a trial. *Stat. Med.* **23** 2497–2508.

MURPHY, S. A. (2005). An experimental design for the development of adaptive treatment strategies. *Stat. Med.* **24** 1455–1481. MR2137651 https://doi.org/10.1002/sim.2022

MURPHY, S. A., LYNCH, K. G., OSLIN, D., MCKAY, J. R. and TENHAVE, T. (2007). Developing adaptive treatment strategies in substance abuse research. *Drug Alcohol Depend.* **88** S24–S30.

NEVO, D., LOK, J. J. and SPIEGELMAN, D. (2021). Supplement to "Analysis of "learn-as-you-go" (LAGO) studies." https://doi.org/10.1214/20-AOS1978SUPP

O'QUIGLEY, J., PEPE, M. and FISHER, L. (1990). Continual reassessment method: A practical design for phase 1 clinical trials in cancer. *Biometrics* **46** 33–48. MR1059105 https://doi.org/10.2307/2531628

O'QUIGLEY, J. and SHEN, L. Z. (1996). Continual reassessment method: A likelihood approach. *Biometrics* 673–684.

PROSCHAN, M. A. and HUNSBERGER, S. A. (1995). Designed extension of studies based on conditional power. *Biometrics* 1315–1324.

ROSENBERGER, W. F., FLOURNOY, N. and DURHAM, S. D. (1997). Asymptotic normality of maximum likelihood estimators from multiparameter response-driven designs. *J. Statist. Plann. Inference* **60** 69–76. MR1453033 https://doi.org/10.1016/S0378-3758(96)00120-6

ROSENBERGER, W. F. and HAINES, L. M. (2002). Competing designs for phase I clinical trials: A review. *Stat. Med.* **21** 2757–2770.

ROSENBLUM, M. and VAN DER LAAN, M. J. (2011). Optimizing randomized trial designs to distinguish which subpopulations benefit from treatment. *Biometrika* **98** 845–860. MR2860328 https://doi.org/10.1093/biomet/asr055

SCHEFFÉ, H. (1959). *The Analysis of Variance*. Wiley, New York; CRC Press, London. MR0116429

SEMRAU, K. E., HIRSCHHORN, L. R., MARX DELANEY, M., SINGH, V. P., SAURASTRI, R., SHARMA, N., TULLER, D. E., FIRESTONE, R., LIPSITZ, S. et al. (2017). Outcomes of a coaching-based WHO safe childbirth checklist program in India. *N. Engl. J. Med.* **377** 2313–2324.

SIMON, R., RUBINSTEIN, L., ARBUCK, S. G., CHRISTIAN, M. C., FREIDLIN, B. and COLLINS, J. (1997). Accelerated titration designs for phase I clinical trials in oncology. *J. Natl. Cancer Inst.* **89** 1138–1147.

SNAPINN, S., CHEN, M.-G., JIANG, Q. and KOUTSOUKOS, T. (2006). Assessment of futility in clinical trials. *Pharm. Stat.* **5** 273–281.

SPIEGELMAN, D. and ZHOU, X. (2018). Evaluating public health interventions: 8. Causal inference for time-invariant interventions. *Am. J. Publ. Health* **108** 1187–1190.

THALL, P. F., MILLIKAN, R. E., MUELLER, P. and LEE, S.-J. (2003). Dose-finding with two agents in Phase I oncology trials. *Biometrics* **59** 487–496. MR2004253 https://doi.org/10.1111/1541-0420.00058

VAN DER LAAN, M. J. (2008). The construction and analysis of adaptive group sequential designs. Technical Report, Technical Report 232, Division of Biostatistics, UC Berkeley.

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

WANG, K. and IVANOVA, A. (2005). Two-dimensional dose finding in discrete dose space. *Biometrics* **61** 217–222. MR2135863 https://doi.org/10.1111/j.0006-341X.2005.030540.x

WEDDERBURN, R. W. M. (1976). On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models. *Biometrika* **63** 27–32. MR0408092 https://doi.org/10.1093/biomet/63.1.27

WHO (2012). *WHO Recommendations for the Prevention and Treatment of Postpartum Haemorrhage*. World Health Organization.

WONG, K. M., CAPASSO, A. and ECKHARDT, S. G. (2016). The changing landscape of phase I trials in oncology. *Nat. Rev. Clin. Oncol.* **13** 106–117.

The Annals of Statistics
2021, Vol. 49, No. 2, 820–844
https://doi.org/10.1214/20-AOS1979

NECESSARY AND SUFFICIENT CONDITIONS FOR VARIABLE SELECTION CONSISTENCY OF THE LASSO IN HIGH DIMENSIONS

BY SOUMENDRA N. LAHIRI

Department of Mathematics and Statistics, Washington University in St. Louis, s.lahiri@wustl.edu

This paper investigates conditions for variable selection consistency of the LASSO in high dimensional regression models and gives necessary and sufficient conditions for the same, potentially allowing the model dimension p to grow *arbitrarily* fast as a function of the sample size n. These conditions require both upper and lower bounds on the growth rate of the penalty parameter. It turns out that a variant of the irrepresentable Condition (IRC) of (*J. Mach. Learn. Res.* **7** (2006) 2541–2563), herein called the *lower irrepresentable Condition* (or LIRC), is determined by the lower bound considerations while the upper bound considerations lead to a new condition, called the *upper irrepresentable Condition* (or UIRC) in this paper. It is shown that the LIRC together with the UIRC is necessary and sufficient for the variable selection consistency of the LASSO, thereby settling a conjecture of (*J. Mach. Learn. Res.* **7** (2006) 2541–2563). Further, it is shown that under some mild regularity conditions, the penalty parameter must necessarily tend to infinity at a certain minimal rate to ensure variable selection consistency of the LASSO and that the corresponding LASSO estimators of the nonzero regression parameters cannot be $\sqrt{n}$-consistent (even for individual parameters). Thus, under fairly general conditions, the LASSO with a single choice of the penalty parameter cannot achieve both variable selection consistency and $\sqrt{n}$-consistency simultaneously.

REFERENCES

ADLER, R. J. (1990). *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **12**. IMS, Hayward, CA. MR1088478

ATHREYA, K. B. and LAHIRI, S. N. (2006). *Measure Theory and Probability Theory. Springer Texts in Statistics*. Springer, New York. MR2247694

BELLONI, A. and CHERNOZHUKOV, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli* **19** 521–547. MR3037163 https://doi.org/10.3150/11-BEJ410

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

BROCKWELL, P. J. and DAVIS, R. A. (1996). *Time Series: Theory and Methods. Springer Series in Statistics*. Springer, New York. MR2839251

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

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564–1597. MR3262461 https://doi.org/10.1214/14-AOS1230

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

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

FRANK, I. E. and FRIEDMAN, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics* **35** 109–148.

FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33** 1–22.

MSC2020 subject classifications. Primary 62E20; secondary 62J05.
Key words and phrases. Asymptotic normality, irrepresentable condition, oracle property, regularization.

FUK, D. H. and NAGAEV, S. V. (1971). Probabilistic inequalities for sums of independent random variables. *Teor. Veroyatn. Primen.* **16** 660–675. MR0293695

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

KNIGHT, K. and FU, W. (2000). Asymptotics for lasso-type estimators. *Ann. Statist.* **28** 1356–1378. MR1805787 https://doi.org/10.1214/aos/1015957397

MAMMEN, E. (1989). Asymptotics with increasing dimension for robust regression with applications to the bootstrap. *Ann. Statist.* **17** 382–400. MR0981457 https://doi.org/10.1214/aos/1176347023

MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. MR2278363 https://doi.org/10.1214/009053606000000281

PORTNOY, S. (1984). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I. Consistency. *Ann. Statist.* **12** 1298–1309. MR0760690 https://doi.org/10.1214/aos/1176346793

PORTNOY, S. (1985). Asymptotic behavior of M estimators of p regression parameters when p^2/n is large. II. Normal approximation. *Ann. Statist.* **13** 1403–1417. MR0811499 https://doi.org/10.1214/aos/1176349744

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

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

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

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

ZOU, H. and LI, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.* **36** 1509–1533. MR2435443 https://doi.org/10.1214/009053607000000802

The Annals of Statistics
2021, Vol. 49, No. 2, 845–862
https://doi.org/10.1214/20-AOS1981

EMPIRICAL PROCESS RESULTS FOR EXCHANGEABLE ARRAYS

BY LAURENT DAVEZIES[*], XAVIER D'HAULTFŒUILLE[†] AND YANNICK GUYONVARCH[‡]

CREST (UMR CNRS 9194), [*]*laurent.davezies@ensae.fr;* [†]*xavier.dhaultfoeuille@ensae.fr;* [‡]*yannick.guyonvarch@ensae.fr*

Exchangeable arrays are natural tools to model common forms of dependence between units of a sample. Jointly exchangeable arrays are well suited to dyadic data, where observed random variables are indexed by two units from the same population. Examples include trade flows between countries or relationships in a network. Separately exchangeable arrays are well suited to multiway clustering, where units sharing the same cluster (e.g., geographical areas or sectors of activity when considering individual wages) may be dependent in an unrestricted way. We prove uniform laws of large numbers and central limit theorems for such exchangeable arrays. We obtain these results under the same moment restrictions and conditions on the class of functions as those typically assumed with i.i.d. data. We also show the convergence of bootstrap processes adapted to such arrays.

REFERENCES

ALDOUS, D. J. (1981). Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* **11** 581–598. MR0637937 https://doi.org/10.1016/0047-259X(81)90099-3

ARCONES, M. A. and GINÉ, E. (1992). On the bootstrap of U and V statistics. *Ann. Statist.* **20** 655–674. MR1165586 https://doi.org/10.1214/aos/1176348650

ARCONES, M. A. and GINÉ, E. (1993). Limit theorems for U-processes. *Ann. Probab.* **21** 1494–1542. MR1235426

ARCONES, M. A. and GINÉ, E. (1994). U-processes indexed by Vapnik–Červonenkis classes of functions with applications to asymptotics and bootstrap of U-statistics with estimated parameters. *Stochastic Process. Appl.* **52** 17–38. MR1289166 https://doi.org/10.1016/0304-4149(94)90098-1

BARBE, P. and BERTAIL, P. (1995). *The Weighted Bootstrap*. *Lecture Notes in Statistics* **98**. Springer, New York. MR2195545 https://doi.org/10.1007/978-1-4612-2532-4

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

BERTRAND, M., DUFLO, E. and MULLAINATHAN, S. (2004). How much should we trust differences-in-differences estimates? *Q. J. Econ.* **119** 249–275.

BICKEL, P. J. and CHEN, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA* **106** 21068–21073.

BICKEL, P. J., CHEN, A. and LEVINA, E. (2011). The method of moments and degree distributions for network models. *Ann. Statist.* **39** 2280–2301. MR2906868 https://doi.org/10.1214/11-AOS904

BRETAGNOLLE, J. (1983). Lois limites du bootstrap de certaines fonctionnelles. *Ann. Inst. Henri Poincaré B, Probab. Stat.* **19** 281–296. MR0725561

CAMERON, A. C., GELBACH, J. B. and MILLER, D. L. (2011). Robust inference with multiway clustering. *J. Bus. Econom. Statist.* **29** 238–249. MR2807878 https://doi.org/10.1198/jbes.2010.07136

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

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564–1597. MR3262461 https://doi.org/10.1214/14-AOS1230

DAVEZIES, L., D'HAULTFŒUILLE, X. and GUYONVARCH, Y. (2021). Supplement to "Empirical process results for exchangeable arrays." https://doi.org/10.1214/20-AOS1981SUPP

DEHLING, H. and PHILIPP, W. (2002). Empirical process techniques for dependent data. Springer, Berlin.

DE LA PEÑA, V. H. (1992). Decoupling and Khintchine's inequalities for U-statistics. *Ann. Probab.* **20** 1877–1892. MR1188046

MSC2020 subject classifications. Primary 60F17; secondary 60G09, 62F40.
Key words and phrases. Exchangeable arrays, empirical processes, bootstrap.

DE LA PEÑA, V. H. and GINÉ, E. (1999). *Decoupling*. *Probability and Its Applications* (*New York*). Springer, New York. MR1666908 https://doi.org/10.1007/978-1-4612-0537-1

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

FAFCHAMPS, M. and GUBERT, F. (2007). The formation of risk sharing networks. *J. Dev. Econ.* **83** 326–350.

GINÉ, E. and NICKL, R. (2015). *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

GINÉ, E. and ZINN, J. (1990). Bootstrapping general empirical measures. *Ann. Probab.* **18** 851–869. MR1055437

GOURIÉROUX, C., MONFORT, A. and TROGNON, A. (1984). Pseudomaximum likelihood methods: Applications to Poisson models. *Econometrica* **52** 701–720. MR0740306 https://doi.org/10.2307/1913472

GRAHAM, B. S. (2019). Dyadic regression. arXiv e-prints. Available at arXiv:1908.09029.

GRAHAM, B. S., NIU, F. and POWELL, J. L. (2019). Kernel density estimation for undirected dyadic data. arXiv e-prints. Available at arXiv:1907.13630.

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

HANSEN, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50** 1029–1054. MR0666123 https://doi.org/10.2307/1912775

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. MR0144363

HOLLAND, P. W. and LEINHARDT, S. (1976). Local structure in social networks. *Sociol. Method.* **7** 1–45.

HOOVER, D. N. (1979). Relations on probability spaces and arrays of random variables. Working paper.

KALLENBERG, O. (1989). On the representation theorem for exchangeable arrays. *J. Multivariate Anal.* **30** 137–154. MR1003713 https://doi.org/10.1016/0047-259X(89)90092-4

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

KATO, K. (2019). Lecture notes on empirical process theory. Technical report, Cornell Univ., Ithaca, NY.

KOSOROK, M. R. (2003). Bootstraps of sums of independent but not identically distributed stochastic processes. *J. Multivariate Anal.* **84** 299–318. MR1965224 https://doi.org/10.1016/S0047-259X(02)00040-4

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

MACKINNON, J. G., NIELSEN, M. Ø., WEBB and MATTHEW, D. (2019). Wild bootstrap and asymptotic inference with multiway clustering. *J. Bus. Econom. Statist.* https://doi.org/10.1080/07350015.2019.1677473

MCCULLAGH, P. (2000). Resampling and exchangeable arrays. *Bernoulli* **6** 285–301. MR1748722 https://doi.org/10.2307/3318577

MENZEL, K. (2019). Bootstrap with cluster-dependence in two or more dimensions. Working paper.

OWEN, A. B. (2007). The pigeonhole bootstrap. *Ann. Appl. Stat.* **1** 386–411. MR2415741 https://doi.org/10.1214/07-AOAS122

SANTOS SILVA, J. and TENREYRO, S. (2006). The log of gravity. *Rev. Econ. Stat.* **88** 641–658.

SILVERMAN, B. W. (1976). Limit theorems for dissociated random variables. *Adv. in Appl. Probab.* **8** 806–819. MR0431348 https://doi.org/10.2307/1425935

TABORD-MEEHAN, M. (2019). Inference with dyadic data: Asymptotic behavior of the dyadic-robust t-statistic. *J. Bus. Econom. Statist.* **37** 671–680. MR4016162 https://doi.org/10.1080/07350015.2017.1409630

VAN DER VAART, A. W. (2000). *Asymptotics Statistics*. Cambridge Univ. Press, Cambridge.

VAN DER VAART, A. and WELLNER, J. (1996). *Weak Convergence of Empirical Processes*: *With Applications to Statistics*. Springer, New York.

WASSERMAN, S. and FAUST, K. (1994). *Social Network Analysis*: *Methods and Applications* 8. Cambridge Univ. Press, Cambridge.

WELLNER, J. A. and ZHAN, Y. (1996). Bootstrapping Z-estimators. Technical Report 308, Univ. Washington, Dept. Statistics.

The Annals of Statistics
2021, Vol. 49, No. 2, 863–884
https://doi.org/10.1214/20-AOS1982

SURVIVAL ANALYSIS VIA HIERARCHICALLY DEPENDENT MIXTURE HAZARDS

BY FEDERICO CAMERLENGHI[1], ANTONIO LIJOI[2,*] AND IGOR PRÜNSTER[2,†]

[1]*Department of Economics, Management, and Statistics, University of Milano-Bicocca, federico.camerlenghi@unimib.it*
[2]*Department of Decision Sciences and BIDSA, Bocconi University, *antonio.lijoi@unibocconi.it; †igor@unibocconi.it*

Hierarchical nonparametric processes are popular tools for defining priors on collections of probability distributions, which induce dependence across multiple samples. In survival analysis problems, one is typically interested in modeling the hazard rates, rather than the probability distributions themselves, and the currently available methodologies are not applicable. Here, we fill this gap by introducing a novel, and analytically tractable, class of multivariate mixtures whose distribution acts as a prior for the vector of sample-specific baseline hazard rates. The dependence is induced through a hierarchical specification of the mixing random measures that ultimately corresponds to a composition of random discrete combinatorial structures. Our theoretical results allow to develop a full Bayesian analysis for this class of models, which can also account for right-censored survival data and covariates, and we also show posterior consistency. In particular, we emphasize that the posterior characterization we achieve is the key for devising both marginal and conditional algorithms for evaluating Bayesian inferences of interest. The effectiveness of our proposal is illustrated through some synthetic and real data examples.

REFERENCES

[1] BARRIENTOS, A. F., JARA, A. and QUINTANA, F. A. (2012). On the support of MacEachern's dependent Dirichlet processes and extensions. *Bayesian Anal*. **7** 277–309. MR2934952 https://doi.org/10.1214/12-BA709

[2] BERTOIN, J. (1999). Subordinators: Examples and applications. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1997). *Lecture Notes in Math*. **1717** 1–91. Springer, Berlin. MR1746300 https://doi.org/10.1007/978-3-540-48115-7_1

[3] BLEI, D. M., NG, A. Y. and JORDAN, M. I. (2003). Latent Dirichlet allocation. *J. Mach. Learn. Res*. **3** 993–1022.

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

[5] CAMERLENGHI, F., LIJOI, A. and PRÜNSTER, I. (2017). Bayesian prediction with multiple-samples information. *J. Multivariate Anal*. **156** 18–28. MR3624682 https://doi.org/10.1016/j.jmva.2017.01.010

[6] CAMERLENGHI, F., LIJOI, A. and PRÜNSTER, I. (2021). Supplement to "Survival analysis via hierarchically dependent mixture hazards." https://doi.org/10.1214/20-AOS1982SUPP

[7] CLARK, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* **41** 135–155. MR0415944 https://doi.org/10.2307/1913889

[8] CONT, R. and TANKOV, P. (2004). *Financial Modelling with Jump Processes*. *Chapman & Hall/CRC Financial Mathematics Series*. CRC Press/CRC, Boca Raton, FL. MR2042661

[9] DE BLASI, P., PECCATI, G. and PRÜNSTER, I. (2009). Asymptotics for posterior hazards. *Ann. Statist*. **37** 1906–1945. MR2533475 https://doi.org/10.1214/08-AOS631

[10] DE IORIO, M., JOHNSON, W. O., MÜLLER, P. and ROSNER, G. L. (2009). Bayesian nonparametric nonproportional hazards survival modeling. *Biometrics* **65** 762–771. MR2649849 https://doi.org/10.1111/j.1541-0420.2008.01166.x

MSC2020 subject classifications. 62F15, 62N02, 60G57.
Key words and phrases. Bayesian Nonparametrics, completely random measures, generalized gamma processes, hazard rate mixtures, hierarchical processes, meta-analysis, partial exchangeability.

[11] De Iorio, M., Müller, P., Rosner, G. L. and MacEachern, S. N. (2004). An ANOVA model for dependent random measures. *J. Amer. Statist. Assoc.* **99** 205–215. MR2054299 https://doi.org/10.1198/016214504000000205
[12] Doksum, K. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Probab.* **2** 183–201. MR0373081 https://doi.org/10.1214/aop/1176996703
[13] Donnet, S., Rivoirard, V., Rousseau, J. and Scricciolo, C. (2017). Posterior concentration rates for counting processes with Aalen multiplicative intensities. *Bayesian Anal.* **12** 53–87. MR3597567 https://doi.org/10.1214/15-BA986
[14] Drăghici, L. and Ramamoorthi, R. V. (2003). Consistency of Dykstra–Laud priors. *Sankhyā* **65** 464–481. MR2028910
[15] Dykstra, R. L. and Laud, P. (1981). A Bayesian nonparametric approach to reliability. *Ann. Statist.* **9** 356–367. MR0606619
[16] Epifani, I. and Lijoi, A. (2010). Nonparametric priors for vectors of survival functions. *Statist. Sinica* **20** 1455–1484. MR2777332
[17] Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2** 615–629. MR0438568
[18] Ferguson, T. S. and Klass, M. J. (1972). A representation of independent increment processes without Gaussian components. *Ann. Math. Stat.* **43** 1634–1643. MR0373022 https://doi.org/10.1214/aoms/1177692395
[19] Ferguson, T. S. and Phadia, E. G. (1979). Bayesian nonparametric estimation based on censored data. *Ann. Statist.* **7** 163–186. MR0515691
[20] de Finetti, B. (1938). Sur la condition d'equivalence partielle. *Actualités Scientiques et Industrielles* **739** 5–18.
[21] 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
[22] Hjort, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Ann. Statist.* **18** 1259–1294. MR1062708 https://doi.org/10.1214/aos/1176347749
[23] Ishwaran, H. and James, L. F. (2004). Computational methods for multiplicative intensity models using weighted gamma processes: Proportional hazards, marked point processes, and panel count data. *J. Amer. Statist. Assoc.* **99** 175–190. MR2054297 https://doi.org/10.1198/016214504000000179
[24] Jacod, J. (1974/75). Multivariate point processes: Predictable projection, Radon–Nikodým derivatives, representation of martingales. *Z. Wahrsch. Verw. Gebiete* **31** 235–253. MR0380978 https://doi.org/10.1007/BF00536010
[25] James, L. F. (2005). Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *Ann. Statist.* **33** 1771–1799. MR2166562 https://doi.org/10.1214/009053605000000336
[26] Kingman, J. F. C. (1993). *Poisson Processes*. *Oxford Studies in Probability* **3**. Oxford University Press, Oxford. MR1207584
[27] Lijoi, A. and Nipoti, B. (2014). A class of hazard rate mixtures for combining survival data from different experiments. *J. Amer. Statist. Assoc.* **109** 802–814. MR3223751 https://doi.org/10.1080/01621459.2013.869499
[28] Lijoi, A., Nipoti, B. and Prünster, I. (2014). Bayesian inference with dependent normalized completely random measures. *Bernoulli* **20** 1260–1291. MR3217444 https://doi.org/10.3150/13-BEJ521
[29] Lijoi, A. and Prünster, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics*. *Camb. Ser. Stat. Probab. Math.* **28** 80–136. Cambridge Univ. Press, Cambridge. MR2730661
[30] Lo, A. Y. and Weng, C.-S. (1989). On a class of Bayesian nonparametric estimates. II. Hazard rate estimates. *Ann. Inst. Statist. Math.* **41** 227–245. MR1006487 https://doi.org/10.1007/BF00049393
[31] Mandelbrot, B. and Taylor, H. M. (1967). On the distribution of stock price differences. *Oper. Res.* **15** 1057–1062.
[32] Müller, P., Quintana, F. A. and Page, G. (2018). Nonparametric Bayesian inference in applications. *Stat. Methods Appl.* **27** 175–206. MR3807363 https://doi.org/10.1007/s10260-017-0405-z
[33] Müller, P., Quintana, F. A., Jara, A. and Hanson, T. (2015). *Bayesian Nonparametric Data Analysis*. *Springer Series in Statistics*. Springer, Cham. MR3309338 https://doi.org/10.1007/978-3-319-18968-0
[34] Müller, P., Quintana, F. A., Rosner, G. L. and Maitland, M. L. (2013). Bayesian inference for longitudinal data with non-parametric treatment effects. *Biostatistics* **15** 341–352.
[35] Nieto-Barajas, L. E. (2014). Bayesian semiparametric analysis of short- and long-term hazard ratios with covariates. *Comput. Statist. Data Anal.* **71** 477–490. MR3131984 https://doi.org/10.1016/j.csda.2013.03.012

[36] NIETO-BARAJAS, L. E. and WALKER, S. G. (2004). Bayesian nonparametric survival analysis via Lévy driven Markov processes. *Statist*. *Sinica* **14** 1127–1146. MR2126344

[37] NIETO-BARAJAS, L. E. and WALKER, S. G. (2005). A semi-parametric Bayesian analysis of survival data based on Lévy-driven processes. *Lifetime Data Anal*. **11** 529–543. MR2213503 https://doi.org/10.1007/s10985-005-5238-7

[38] PECCATI, G. and PRÜNSTER, I. (2008). Linear and quadratic functionals of random hazard rates: An asymptotic analysis. *Ann*. *Appl*. *Probab*. **18** 1910–1943. MR2462554 https://doi.org/10.1214/07-AAP509

[39] SATO, K. (2013). *Lévy Processes and Infinitely Divisible Distributions*. *Cambridge Studies in Advanced Mathematics* **68**. Cambridge Univ. Press, Cambridge. Translated from the 1990 Japanese original, Revised edition of the 1999 English translation. MR3185174

[40] TEH, Y. W. and JORDAN, M. I. (2010). Hierarchical Bayesian nonparametric models with applications. In *Bayesian Nonparametrics*. *Camb*. *Ser*. *Stat*. *Probab*. *Math*. **28** 158–207. Cambridge Univ. Press, Cambridge. MR2730663

[41] TEH, Y. W., JORDAN, M. I., BEAL, M. J. and BLEI, D. M. (2006). Hierarchical Dirichlet processes. *J*. *Amer*. *Statist*. *Assoc*. **101** 1566–1581. MR2279480 https://doi.org/10.1198/016214506000000302

[42] WOLPERT, R. L. and ICKSTADT, K. (1998). Simulation of Lévy random fields. In *Practical Nonparametric and Semiparametric Bayesian Statistics*. *Lect*. *Notes Stat*. **133** 227–242. Springer, New York. MR1630084 https://doi.org/10.1007/978-1-4612-1732-9_12

The Annals of Statistics
2021, Vol. 49, No. 2, 885–917
https://doi.org/10.1214/20-AOS1983

SAFE ADAPTIVE IMPORTANCE SAMPLING: A MIXTURE APPROACH

BY BERNARD DELYON[1] AND FRANÇOIS PORTIER[2]

[1] *Université de Rennes 1, bernard.delyon@univ-rennes1.fr*
[2] *Institut Polytechnique de Paris, francois.portier@gmail.com*

This paper investigates *adaptive importance sampling* algorithms for which the *policy*, the sequence of distributions used to generate the particles, is a mixture distribution between a flexible *kernel density estimate* (based on the previous particles), and a "safe" heavy-tailed density. When the share of samples generated according to the safe density goes to zero but not too quickly, two results are established: (i) uniform convergence rates are derived for the policy toward the target density; (ii) a central limit theorem is obtained for the resulting integral estimates. The fact that the asymptotic variance is the same as the variance of an "oracle" procedure with variance-optimal policy, illustrates the benefits of the approach. In addition, a subsampling step (among the particles) can be conducted before constructing the kernel estimate in order to decrease the computational effort without altering the performance of the method. The practical behavior of the algorithms is illustrated in a simulation study.

REFERENCES

AZAÏS, R., DELYON, B. and PORTIER, F. (2018). Integral estimation based on Markovian design. *Adv. in Appl. Probab.* **50** 833–857. MR3877256 https://doi.org/10.1017/apr.2018.38

BERTAIL, P. and PORTIER, F. (2019). Rademacher complexity for Markov chains: Applications to kernel smoothing and Metropolis–Hastings. *Bernoulli* **25** 3912–3938. MR4010977 https://doi.org/10.3150/19-bej1115

BORNN, L., JACOB, P. E., DEL MORAL, P. and DOUCET, A. (2013). An adaptive interacting Wang–Landau algorithm for automatic density exploration. *J. Comput. Graph. Statist.* **22** 749–773. MR3173740 https://doi.org/10.1080/10618600.2012.723569

BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford. With a foreword by Michel Ledoux. MR3185193 https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

CAPPÉ, O., GUILLIN, A., MARIN, J. M. and ROBERT, C. P. (2004). Population Monte Carlo. *J. Comput. Graph. Statist.* **13** 907–929. MR2109057 https://doi.org/10.1198/106186004X12803

CAPPÉ, O., DOUC, R., GUILLIN, A., MARIN, J.-M. and ROBERT, C. P. (2008). Adaptive importance sampling in general mixture classes. *Stat. Comput.* **18** 447–459. MR2461888 https://doi.org/10.1007/s11222-008-9059-x

CHOPIN, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.* **32** 2385–2411. MR2153989 https://doi.org/10.1214/009053604000000698

DAI, B., HE, N., DAI, H. and SONG, L. (2016). Provable Bayesian inference via particle mirror descent. In *Artificial Intelligence and Statistics* 985–994.

DEL MORAL, P. (2013). *Mean Field Simulation for Monte Carlo Integration*. *Monographs on Statistics and Applied Probability* **126**. CRC Press, Boca Raton, FL. MR3060209

DELYON, B. and PORTIER, F. (2018). Asymptotic optimality of adaptive importance sampling. In *Proceedings of the* 32*nd International Conference on Neural Information Processing Systems* 3138–3148.

DOUC, R., GUILLIN, A., MARIN, J.-M. and ROBERT, C. P. (2007a). Convergence of adaptive mixtures of importance sampling schemes. *Ann. Statist.* **35** 420–448. MR2332281 https://doi.org/10.1214/009053606000001154

DOUC, R., GUILLIN, A., MARIN, J.-M. and ROBERT, C. P. (2007b). Minimum variance importance sampling via population Monte Carlo. *ESAIM Probab. Stat.* **11** 427–447. MR2339302 https://doi.org/10.1051/ps:2007028

MSC2020 subject classifications. Primary 65C05, 62G07; secondary 60F05.
Key words and phrases. Monte Carlo methods, adaptive importance sampling, kernel density estimation, martingale methods.

ELVIRA, V., MARTINO, L., LUENGO, D. and BUGALLO, M. F. (2019). Generalized multiple importance sampling. *Statist. Sci.* **34** 129–155. MR3938972 https://doi.org/10.1214/18-STS668

EVANS, M. and SWARTZ, T. (2000). *Approximating Integrals via Monte Carlo and Deterministic Methods*. *Oxford Statistical Science Series*. Oxford Univ. Press, Oxford. MR1859163

FENG, M. B., MAGGIAR, A., STAUM, J. and WÄCHTER, A. (2018). Uniform convergence of sample average approximation with adaptive multiple importance sampling. In 2018 *Winter Simulation Conference* (*WSC*) 1646–1657. IEEE.

FOLLAND, G. B. (2013). *Real Analysis*: *Modern Techniques and Their Applications*. Wiley, New York.

FORT, G., JOURDAIN, B., KUHN, E., LELIÈVRE, T. and STOLTZ, G. (2015). Convergence of the Wang–Landau algorithm. *Math. Comp.* **84** 2297–2327. MR3356027 https://doi.org/10.1090/S0025-5718-2015-02952-4

FREEDMAN, D. A. (1975). On tail probabilities for martingales. *Ann. Probab.* **3** 100–118. MR0380971 https://doi.org/10.1214/aop/1176996452

GEWEKE, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* **57** 1317–1339. MR1035115 https://doi.org/10.2307/1913710

GINÉ, E. and GUILLOU, A. (2001). On consistency of kernel density estimators for randomly censored data: Rates holding uniformly over adaptive intervals. *Ann. Inst. Henri Poincaré Probab. Stat.* **37** 503–522. MR1876841 https://doi.org/10.1016/S0246-0203(01)01081-0

GINÉ, E. and GUILLOU, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. Henri Poincaré Probab. Stat.* **38** 907–921. MR1955344 https://doi.org/10.1016/S0246-0203(02)01128-7

GIVENS, G. H. and RAFTERY, A. E. (1996). Local adaptive importance sampling for multivariate densities with strong nonlinear relationships. *J. Amer. Statist. Assoc.* **91** 132–141. MR1394067 https://doi.org/10.2307/2291389

GORDON, N. J., SALMOND, D. J. and SMITH, A. F. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc., F, Radar Signal Process.* **140** 107–113.

HAARIO, H., SAKSMAN, E. and TAMMINEN, J. (2001). An adaptive Metropolis algorithm. *Bernoulli* **7** 223–242. MR1828504 https://doi.org/10.2307/3318737

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. *Probability and Mathematical Statistics*. Academic Press, New York. MR0624435

HANSEN, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* **24** 726–748. MR2409261 https://doi.org/10.1017/S0266466608080304

HESTERBERG, T. (1995). Weighted average importance sampling and defensive mixture distributions. *Technometrics* **37** 185–194.

KLOEK, T. and VAN DIJK, H. K. (1978). Bayesian estimates of equation system parameters: An application of integration by Monte Carlo. *Econometrica* 1–19.

MARIN, J.-M., PUDLO, P. and SEDKI, M. (2019). Consistency of adaptive importance sampling and recycling schemes. *Bernoulli* **25** 1977–1998. MR3961237 https://doi.org/10.3150/18-BEJ1042

NEDDERMEYER, J. C. (2009). Computationally efficient nonparametric importance sampling. *J. Amer. Statist. Assoc.* **104** 788–802. MR2541595 https://doi.org/10.1198/jasa.2009.0122

OH, M.-S. and BERGER, J. O. (1992). Adaptive importance sampling in Monte Carlo integration. *J. Stat. Comput. Simul.* **41** 143–168. MR1276184 https://doi.org/10.1080/00949659208810398

OWEN, A. B. (2013). *Monte Carlo Theory, Methods and Examples*.

OWEN, A. and ZHOU, Y. (2000). Safe and effective importance sampling. *J. Amer. Statist. Assoc.* **95** 135–143. MR1803146 https://doi.org/10.2307/2669533

POLLARD, D. (1984). *Convergence of Stochastic Processes*. *Springer Series in Statistics*. Springer, New York. MR0762984 https://doi.org/10.1007/978-1-4612-5254-2

ROBERT, C. P. and CASELLA, G. (2004). *Monte Carlo Statistical Methods*, 2nd ed. *Springer Texts in Statistics*. Springer, New York. MR2080278 https://doi.org/10.1007/978-1-4757-4145-2

SILVERMAN, B. W. (2018). *Density Estimation for Statistics and Data Analysis*. Routledge, London.

STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360. MR0594650

TEMLYAKOV, V. N. (1998). The best m-term approximation and greedy algorithms. *Adv. Comput. Math.* **8** 249–265. MR1628182 https://doi.org/10.1023/A:1018900431309

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

WANG, F. and LANDAU, D. P. (2001). Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram. *Phys. Rev. E* **64** 056101.

WEST, M. (1993). Approximating posterior distributions by mixtures. *J. Roy. Statist. Soc. Ser. B* **55** 409–422. MR1224405

ZHANG, P. (1996). Nonparametric importance sampling. *J. Amer. Statist. Assoc.* **91** 1245–1253. MR1424622 https://doi.org/10.2307/2291743

The Annals of Statistics
2021, Vol. 49, No. 2, 918–943
https://doi.org/10.1214/20-AOS1984

DISTRIBUTED LINEAR REGRESSION BY AVERAGING

BY EDGAR DOBRIBAN[1] AND YUE SHENG[2]

[1]*Department of Statistics, University of Pennsylvania, dobriban@wharton.upenn.edu*
[2]*Graduate Group in Applied Mathematics and Computational Science, Department of Mathematics, University of Pennsylvania, yuesheng@sas.upenn.edu*

Distributed statistical learning problems arise commonly when dealing with large datasets. In this setup, datasets are partitioned over machines, which compute locally, and communicate short messages. Communication is often the bottleneck. In this paper, we study one-step and iterative *weighted parameter averaging* in statistical *linear models* under *data parallelism*. We do linear regression on each machine, send the results to a central server and take a weighted average of the parameters. Optionally, we iterate, sending back the weighted average and doing local ridge regressions centered at it. How does this work compared to doing linear regression on the full data? Here, we study the performance loss in *estimation* and *test error*, and *confidence interval length* in high dimensions, where the number of parameters is comparable to the training data size.

We find the performance loss in one-step weighted averaging, and also give results for iterative averaging. We also find that different problems are affected differently by the distributed framework. Estimation error and confidence interval length increases a lot, while prediction error increases much less. We rely on recent results from random matrix theory, where we develop a new calculus of deterministic equivalents as a tool of broader interest.

REFERENCES

AGARWAL, A., CHAPELLE, O., DUDÍK, M. and LANGFORD, J. (2014). A reliable effective terascale linear learning system. *J. Mach. Learn. Res.* **15** 1111–1133. MR3195340

ALI, A., KOLTER, J. Z. and TIBSHIRANI, R. J. (2019). A continuous-time view of early stopping for least squares regression. In *Proceedings of Machine Learning Research* **89** 1370–1378.

BAI, Z. and SILVERSTEIN, J. W. (2009). *Spectral Analysis of Large Dimensional Random Matrices*. *Springer Series in Statistics*. Springer, New York. MR2567175 https://doi.org/10.1007/978-1-4419-0661-8

BANERJEE, M. and DUROT, C. (2018). Removing the curse of superefficiency: An effective strategy for distributed computing in isotonic regression. Preprint. Available at arXiv:1806.08542.

BANERJEE, M., DUROT, C. and SEN, B. (2019). Divide and conquer in nonstandard problems and the superefficiency phenomenon. *Ann. Statist.* **47** 720–757. MR3909948 https://doi.org/10.1214/17-AOS1633

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

BEKKERMAN, R., BILENKO, M. and LANGFORD, J. (2011). *Scaling up Machine Learning: Parallel and Distributed Approaches*. Cambridge Univ. Press, Cambridge.

BERTSEKAS, D. P. and TSITSIKLIS, J. N. (1989). *Parallel and Distributed Computation: Numerical Methods* **23**. Prentice Hall, Englewood Cliffs, NJ.

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

BOYD, S., PARIKH, N., CHU, E., PELEATO, B. and ECKSTEIN, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3** 1–122.

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

MSC2020 subject classifications. Primary 62J05; secondary 65Y05, 68W10, 68W15.
Key words and phrases. Linear regression, distributed learning, parallel computation, random matrix theory, high dimensional.

CHEN, X., LIU, W. and ZHANG, Y. First-order newton-type estimator for distributed estimation and inference. Preprint. Available at arXiv:1811.11368.

CHEN, X., LIU, W. and ZHANG, Y. (2019). Quantile regression under memory constraint. *Ann. Statist.* **47** 3244–3273. MR4025741 https://doi.org/10.1214/18-AOS1777

CHEN, X. and XIE, M. (2014). A split-and-conquer approach for analysis of extraordinarily large data. *Statist. Sinica* **24** 1655–1684. MR3308656

CHU, C.-T., KIM, S. K., LIN, Y.-A., YU, Y., BRADSKI, G., OLUKOTUN, K. and NG, A. Y. (2007). Map-reduce for machine learning on multicore. In *Advances in Neural Information Processing Systems* 281–288.

COUILLET, R. and DEBBAH, M. (2011). *Random Matrix Methods for Wireless Communications*. Cambridge Univ. Press, Cambridge. MR2884783 https://doi.org/10.1017/CBO9780511994746

COUILLET, R., DEBBAH, M. and SILVERSTEIN, J. W. (2011). A deterministic equivalent for the analysis of correlated MIMO multiple access channels. *IEEE Trans. Inf. Theory* **57** 3493–3514. MR2817033 https://doi.org/10.1109/TIT.2011.2133151

DAVIS, C. (1957). All convex invariant functions of Hermitian matrices. *Arch. Math.* **8** 276–278. MR0090572 https://doi.org/10.1007/BF01898787

DEAN, J. and GHEMAWAT, S. (2008). Mapreduce: Simplified data processing on large clusters. *Commun. ACM* **51** 107–113.

DOBRIBAN, E. and SHENG, Y. (2019). One-shot distributed ridge regression in high dimensions. Preprint. Available at arXiv:1903.09321.

DOBRIBAN, E. and SHENG, Y. (2021). Supplement to “Distributed linear regression by averaging.” https://doi.org/10.1214/20-AOS1984SUPP

DONOHO, D. and MONTANARI, A. (2016). High dimensional robust M-estimation: Asymptotic variance via approximate message passing. *Probab. Theory Related Fields* **166** 935–969. MR3568043 https://doi.org/10.1007/s00440-015-0675-z

DUCHI, J. C., AGARWAL, A. and WAINWRIGHT, M. J. (2012). Dual averaging for distributed optimization: Convergence analysis and network scaling. *IEEE Trans. Automat. Control* **57** 592–606. MR2932818 https://doi.org/10.1109/TAC.2011.2161027

DUCHI, J. C., JORDAN, M. I., WAINWRIGHT, M. J. and ZHANG, Y. (2014). Optimality guarantees for distributed statistical estimation. Preprint. Available at arXiv:1405.0782.

EL KAROUI, N., BEAN, D., BICKEL, P. J., LIM, C. and YU, B. (2013). On robust regression with high-dimensional predictors. *Proc. Natl. Acad. Sci. USA* **110** 14557–14562.

FAN, J., GUO, Y. and WANG, K. (2019). Communication-efficient accurate statistical estimation. Preprint. Available at arXiv:1906.04870.

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

HACHEM, W., LOUBATON, P. and NAJIM, J. (2007). Deterministic equivalents for certain functionals of large random matrices. *Ann. Appl. Probab.* **17** 875–930. MR2326235 https://doi.org/10.1214/105051606000000925

HUO, X. and CAO, S. (2019). Aggregated inference. *Wiley Interdiscip. Rev.: Comput. Stat.* **11** e1451, 13. MR3897175 https://doi.org/10.1002/wics.1451

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

LEE, J. D., LIU, Q., SUN, Y. and TAYLOR, J. E. (2017). Communication-efficient sparse regression. *J. Mach. Learn. Res.* **18** Paper No. 5, 30. MR3625709

LEWIS, A. S. (1996). Convex analysis on the Hermitian matrices. *SIAM J. Optim.* **6** 164–177. MR1377729 https://doi.org/10.1137/0806009

LIN, S.-B., GUO, X. and ZHOU, D.-X. (2017). Distributed learning with regularized least squares. *J. Mach. Learn. Res.* **18** Paper No. 92, 31. MR3714255 https://doi.org/10.1016/j.physletb.2016.11.035

LIU, Q. and IHLER, A. T. (2014). Distributed estimation, information loss and exponential families. In *Advances in Neural Information Processing Systems* 1098–1106.

MARCHENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb.* (*N.S.*) **72 (114)** 507–536. MR0208649

MARDIA, K. V., KENT, J. T. and BIBBY, J. M. (1979). *Multivariate Analysis*. *Probability and Mathematical Statistics*: *A Series of Monographs and Textbooks*. Academic Press, London. MR0560319

MCDONALD, R., MOHRI, M., SILBERMAN, N., WALKER, D. and MANN, G. S. (2009). Efficient large-scale distributed training of conditional maximum entropy models. In *Advances in Neural Information Processing Systems* 1231–1239.

MÜLLER, A. and DEBBAH, M. (2016). Random matrix theory tutorial—Introduction to deterministic equivalents. *Traitement Signal* **33** 223–248.

NEDIĆ, A. and OZDAGLAR, A. (2009). Distributed subgradient methods for multi-agent optimization. *IEEE Trans. Automat. Control* **54** 48–61. MR2478070 https://doi.org/10.1109/TAC.2008.2009515

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

PEACOCK, M. J. M., COLLINGS, I. B. and HONIG, M. L. (2008). Eigenvalue distributions of sums and products of large random matrices via incremental matrix expansions. *IEEE Trans*. *Inf. Theory* **54** 2123–2138. MR2450853 https://doi.org/10.1109/TIT.2008.920221

ROSENBLATT, J. D. and NADLER, B. (2016). On the optimality of averaging in distributed statistical learning. *Inf. Inference* **5** 379–404. MR3609865 https://doi.org/10.1093/imaiai/iaw013

RUBIO, F. and MESTRE, X. (2011). Spectral convergence for a general class of random matrices. *Statist*. *Probab*. *Lett*. **81** 592–602. MR2772917 https://doi.org/10.1016/j.spl.2011.01.004

SERDOBOLSKII, V. I. (1983). On minimum error probability in discriminant analysis. *Dokl*. *Akad*. *Nauk SSSR* **27** 720–725.

SERDOBOLSKII, V. I. (2007). *Multiparametric Statistics*. Elsevier, Amsterdam. MR2531357

SHAMIR, O., SREBRO, N. and ZHANG, T. (2014). Communication-efficient distributed optimization using an approximate Newton-type method. In *Proceedings of the* 31*st International Conference on Machine Learning* **32** 1000–1008.

SHI, C., LU, W. and SONG, R. (2018). A massive data framework for M-estimators with cubic-rate. *J. Amer*. *Statist*. *Assoc*. **113** 1698–1709. MR3902239 https://doi.org/10.1080/01621459.2017.1360779

SMITH, V., FORTE, S., MA, C., TAKÁČ, M., JORDAN, M. I. and JAGGI, M. (2017). CoCoA: A general framework for communication-efficient distributed optimization. *J. Mach*. *Learn*. *Res*. **18** Paper No. 230, 49. MR3845529

SZABO, B. and VAN ZANTEN, H. (2018). Adaptive distributed methods under communication constraints. Preprint. Available at arXiv:1804.00864.

TANDON, R., LEI, Q., DIMAKIS, A. G. and KARAMPATZIAKIS, N. (2017). Gradient coding: Avoiding stragglers in distributed learning. In *International Conference on Machine Learning* 3368–3376.

TSITSIKLIS, J. N., BERTSEKAS, D. P. and ATHANS, M. (1986). Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Trans*. *Automat*. *Control* **31** 803–812. MR0853431 https://doi.org/10.1109/TAC.1986.1104412

TULINO, A. M. and VERDÚ, S. (2004). Random matrix theory and wireless communications. *Commun*. *Inf*. *Theory* **1** 1–182.

VOLGUSHEV, S., CHAO, S.-K. and CHENG, G. (2019). Distributed inference for quantile regression processes. *Ann*. *Statist*. **47** 1634–1662. MR3911125 https://doi.org/10.1214/18-AOS1730

WANG, J., KOLAR, M., SREBRO, N. and ZHANG, T. (2017). Efficient distributed learning with sparsity. In *Proceedings of the* 34*th International Conference on Machine Learning* **70** 3636–3645. JMLR.org.

YAO, Y., ROSASCO, L. and CAPONNETTO, A. (2007). On early stopping in gradient descent learning. *Constr*. *Approx*. **26** 289–315. MR2327601 https://doi.org/10.1007/s00365-006-0663-2

ZAHARIA, M., CHOWDHURY, M., FRANKLIN, M. J., SHENKER, S. and STOICA, I. (2010). Spark: Cluster computing with working sets. *HotCloud* **10** 95.

ZHANG, Y., DUCHI, J. and WAINWRIGHT, M. (2013a) Divide and conquer kernel ridge regression. In *Conference on Learning Theory* 592–617.

ZHANG, Y., DUCHI, J. C. and WAINWRIGHT, M. J. (2013b). Communication-efficient algorithms for statistical optimization. *J. Mach*. *Learn*. *Res*. **14** 3321–3363. MR3144464

ZHANG, Y., DUCHI, J. and WAINWRIGHT, M. (2015). Divide and conquer kernel ridge regression: A distributed algorithm with minimax optimal rates. *J. Mach*. *Learn*. *Res*. **16** 3299–3340. MR3450540

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, T., CHENG, G. and LIU, H. (2016). A partially linear framework for massive heterogeneous data. *Ann*. *Statist*. **44** 1400–1437. MR3519928 https://doi.org/10.1214/15-AOS1410

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

ZINKEVICH, M., LANGFORD, J. and SMOLA, A. J. (2009). Slow learners are fast. In *Advances in Neural Information Processing Systems* 2331–2339.

ZINKEVICH, M., WEIMER, M., LI, L. and SMOLA, A. J. (2010). Parallelized stochastic gradient descent. In *Advances in Neural Information Processing Systems* 2595–2603.

The Annals of Statistics
2021, Vol. 49, No. 2, 944–967
https://doi.org/10.1214/20-AOS1986

SUBSPACE ESTIMATION FROM UNBALANCED AND INCOMPLETE DATA MATRICES: $\ell_{2,\infty}$ STATISTICAL GUARANTEES

BY CHANGXIAO CAI[1,*], GEN LI[2], YUEJIE CHI[3], H. VINCENT POOR[1,†] AND YUXIN CHEN[1,‡]

[1]*Department of Electrical Engineering, Princeton University, *ccai@princeton.edu; †poor@princeton.edu; ‡yuxin.chen@princeton.edu*

[2]*Department of Electronic Engineering, Tsinghua University, g-li16@mails.tsinghua.edu.cn*

[3]*Department of Electrical and Computer Engineering, Carnegie Mellon University, yuejiechi@cmu.edu*

This paper is concerned with estimating the column space of an unknown low-rank matrix $\boldsymbol{A}^\star \in \mathbb{R}^{d_1 \times d_2}$, given noisy and partial observations of its entries. There is no shortage of scenarios where the observations—while being too noisy to support faithful recovery of the entire matrix—still convey sufficient information to enable reliable estimation of the column space of interest. This is particularly evident and crucial for the highly unbalanced case where the column dimension d_2 far exceeds the row dimension d_1, which is the focal point of the current paper.

We investigate an efficient spectral method, which operates upon the sample Gram matrix with diagonal deletion. While this algorithmic idea has been studied before, we establish new statistical guarantees for this method in terms of both ℓ_2 and $\ell_{2,\infty}$ estimation accuracy, which improve upon prior results if d_2 is substantially larger than d_1. To illustrate the effectiveness of our findings, we derive matching minimax lower bounds with respect to the noise levels, and develop consequences of our general theory for three applications of practical importance: (1) tensor completion from noisy data, (2) covariance estimation/principal component analysis with missing data and (3) community recovery in bipartite graphs. Our theory leads to improved performance guarantees for all three cases.

REFERENCES

[1] ABBE, E. (2017). Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** 6446–6531. MR3827065

[2] ABBE, E., BANDEIRA, A. S. and HALL, G. (2016). Exact recovery in the stochastic block model. *IEEE Trans. Inf. Theory* **62** 471–487. MR3447993 https://doi.org/10.1109/TIT.2015.2490670

[3] ABBE, E., FAN, J. and WANG, K. (2020). An ℓ_p theory of PCA and spectral clustering. Preprint. Available at arXiv:2006.14062.

[4] ABBE, E., FAN, J., WANG, K. and ZHONG, Y. (2017). Entrywise eigenvector analysis of random matrices with low expected rank. Preprint. Available at arXiv:1709.09565.

[5] AGARWAL, A., SHAH, D., SHEN, D. and SONG, D. (2019). On robustness of principal component regression. In *Advances in Neural Information Processing Systems* 9893–9903.

[6] AGARWAL, N., BANDEIRA, A. S., KOILIARIS, K. and KOLLA, A. (2017). Multisection in the stochastic block model using semidefinite programming. In *Compressed Sensing and Its Applications*. *Appl. Numer. Harmon. Anal.* 125–162. Birkhäuser/Springer, Cham. MR3751736

[7] ALZAHRANI, T. and HORADAM, K. J. (2016). Community detection in bipartite networks: Algorithms and case studies. In *Complex Systems and Networks*. *Underst. Complex Syst.* 25–50. Springer, Heidelberg. MR3586347

[8] AMINI, A. A. and LEVINA, E. (2018). On semidefinite relaxations for the block model. *Ann. Statist.* **46** 149–179. MR3766949 https://doi.org/10.1214/17-AOS1545

MSC2020 subject classifications. Primary 62F10; secondary 62H25.

Key words and phrases. Spectral method, principal component analysis, tensor completion, covariance estimation, missing data, spectral clustering, leave-one-out analysis.

[9] 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
[10] BANDEIRA, A. S. (2018). Random Laplacian matrices and convex relaxations. *Found. Comput. Math*. **18** 345–379. MR3777782 https://doi.org/10.1007/s10208-016-9341-9
[11] BARAK, B. and MOITRA, A. (2016). Noisy tensor completion via the sum-of-squares hierarchy. In *Proceedings of the Conference on Learning Theory* 417–445.
[12] BICKEL, P. J. and LEVINA, E. (2008). Covariance regularization by thresholding. *Ann. Statist*. **36** 2577–2604. MR2485008 https://doi.org/10.1214/08-AOS600
[13] CAI, C., LI, G., CHI, Y., POOR, H. V. and CHEN, Y. (2021). Supplement to "Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees." https://doi.org/10.1214/20-AOS1986SUPP
[14] CAI, C., LI, G., POOR, H. V. and CHEN, Y. (2019). Nonconvex low-rank tensor completion from noisy data. In *Advances in Neural Information Processing Systems* 1861–1872.
[15] CAI, C., POOR, H. V. and CHEN, Y. (2020). Uncertainty quantification for nonconvex tensor completion: Confidence intervals, heteroscedasticity and optimality. In *Proceedings of the International Conference on Machine Learning*. To appear. Available at arXiv:2006.08580.
[16] CAI, T., MA, Z. and WU, Y. (2015). Optimal estimation and rank detection for sparse spiked covariance matrices. *Probab. Theory Related Fields* **161** 781–815. MR3334281 https://doi.org/10.1007/s00440-014-0562-z
[17] CAI, T. T. and LI, X. (2015). Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. *Ann. Statist*. **43** 1027–1059. MR3346696 https://doi.org/10.1214/14-AOS1290
[18] 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
[19] CAI, T. T. and YUAN, M. (2012). Adaptive covariance matrix estimation through block thresholding. *Ann. Statist*. **40** 2014–2042. MR3059075 https://doi.org/10.1214/12-AOS999
[20] CAI, T. T. and ZHANG, A. (2016). Minimax rate-optimal estimation of high-dimensional covariance matrices with incomplete data. *J. Multivariate Anal*. **150** 55–74. MR3534902 https://doi.org/10.1016/j.jmva.2016.05.002
[21] 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
[22] CANDÈS, E. J. and TAO, T. (2010). The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inf. Theory* **56** 2053–2080. MR2723472 https://doi.org/10.1109/TIT.2010.2044061
[23] 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
[24] 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
[25] CHATTERJEE, S. (2015). Matrix estimation by universal singular value thresholding. *Ann. Statist*. **43** 177–214. MR3285604 https://doi.org/10.1214/14-AOS1272
[26] 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
[27] CHEN, Y., CHENG, C. and FAN, J. (2020). Asymmetry helps: Eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices. *Ann. Statist.* To appear. Available at arXiv:1811.12804.
[28] CHEN, Y. and CHI, Y. (2018). Harnessing structures in big data via guaranteed low-rank matrix estimation: Recent theory and fast algorithms via convex and nonconvex optimization. *IEEE Signal Process. Mag*. **35** 14–31.
[29] CHEN, Y., CHI, Y., FAN, J., MA, C. and YAN, Y. (2019). Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. Preprint. Available at arXiv:1902.07698.
[30] CHEN, Y., FAN, J., MA, C. and WANG, K. (2019). Spectral method and regularized MLE are both optimal for top-K ranking. *Ann. Statist*. **47** 2204–2235. MR3953449 https://doi.org/10.1214/18-AOS1745
[31] CHEN, Y., KAMATH, G., SUH, C. and TSE, D. (2016). Community recovery in graphs with locality. In *Proceedings of the International Conference on Machine Learning* 689–698.
[32] CHEN, Y., LI, X. and XU, J. (2018). Convexified modularity maximization for degree-corrected stochastic block models. *Ann. Statist*. **46** 1573–1602. MR3819110 https://doi.org/10.1214/17-AOS1595
[33] CHEN, Y., SUH, C. and GOLDSMITH, A. J. (2016). Information recovery from pairwise measurements. *IEEE Trans. Inf. Theory* **62** 5881–5905. MR3552429 https://doi.org/10.1109/TIT.2016.2600566
[34] CHEN, Y. and WAINWRIGHT, M. J. (2015). Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees. Available at arXiv:1509.03025.

[35] 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.
[36] 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
[37] CHIN, P., RAO, A. and VU, V. (2015). Stochastic block model and community detection in sparse graphs: A spectral algorithm with optimal rate of recovery. In *Proceedings of the Conference on Learning Theory* 391–423.
[38] CHO, J., KIM, D. and ROHE, K. (2017). Asymptotic theory for estimating the singular vectors and values of a partially-observed low rank matrix with noise. *Statist. Sinica* **27** 1921–1948. MR3726772
[39] COJA-OGHLAN, A. (2006). A spectral heuristic for bisecting random graphs. *Random Structures Algorithms* **29** 351–398. MR2254496 https://doi.org/10.1002/rsa.20116
[40] COJA-OGHLAN, A. (2010). Graph partitioning via adaptive spectral techniques. *Combin. Probab. Comput.* **19** 227–284. MR2593622 https://doi.org/10.1017/S0963548309990514
[41] DHILLON, I. S. (2001). Co-clustering documents and words using bipartite spectral graph partitioning. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 269–274. ACM, New York.
[42] ELDRIDGE, J., BELKIN, M. and WANG, Y. (2018). Unperturbed: Spectral analysis beyond Davis–Kahan. In *Proceedigs of the* 29*th Conference on Algorithmic Learning Theory* 321–358.
[43] ELSENER, A. and VAN DE GEER, S. (2019). Sparse spectral estimation with missing and corrupted measurements. *Stat* **8** e229, 11. MR3978409 https://doi.org/10.1002/sta4.229
[44] FAN, J., WANG, K., ZHONG, Y. and ZHU, Z. (2018). Robust high dimensional factor models with applications to statistical machine learning. Preprint. Available at arXiv:1808.03889.
[45] FAN, J., WANG, W. and ZHONG, Y. (2018). An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. *J. Mach. Learn. Res.* **18** 1-42. MR3827095
[46] FELDMAN, V., PERKINS, W. and VEMPALA, S. (2015). Subsampled power iteration: A unified algorithm for block models and planted csp's. In *Advances in Neural Information Processing Systems* 2836–2844.
[47] FLORESCU, L. and PERKINS, W. (2016). Spectral thresholds in the bipartite stochastic block model. In *Proceedings of the Conference on Learning Theory* 943–959.
[48] GANDY, S., RECHT, B. and YAMADA, I. (2011). Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Probl.* **27** 025010, 19. MR2765628 https://doi.org/10.1088/0266-5611/27/2/025010
[49] GAO, C., LU, Y., MA, Z. and ZHOU, H. H. (2016). Optimal estimation and completion of matrices with biclustering structures. *J. Mach. Learn. Res.* **17** 5602–5630. MR3569248
[50] GAO, C., MA, Z., ZHANG, A. Y. and ZHOU, H. H. (2017). Achieving optimal misclassification proportion in stochastic block models. *J. Mach. Learn. Res.* **18** 1980–2024. MR3687603
[51] GHASSEMI, M., SHAKERI, Z., SARWATE, A. D. and BAJWA, W. U. (2017). STARK: Structured dictionary learning through rank-one tensor recovery. In *Proceedings of the* 2017 *IEEE* 7*th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing* 1–5. IEEE, New York.
[52] GONEN, A., ROSENBAUM, D., ELDAR, Y. and SHALEV-SHWARTZ, S. (2016). Subspace learning with partial information. *J. Mach. Learn. Res.* **17** 1821–1841. MR3504612
[53] GROSS, D. (2011). Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inf. Theory* **57** 1548–1566. MR2815834 https://doi.org/10.1109/TIT.2011.2104999
[54] GUÉDON, O. and VERSHYNIN, R. (2016). Community detection in sparse networks via Grothendieck's inequality. *Probab. Theory Related Fields* **165** 1025–1049. MR3520025 https://doi.org/10.1007/s00440-015-0659-z
[55] HAJEK, B., WU, Y. and XU, J. (2016). Achieving exact cluster recovery threshold via semidefinite programming: Extensions. *IEEE Trans. Inf. Theory* **62** 5918–5937. MR3552431 https://doi.org/10.1109/TIT.2016.2594812
[56] HAO, B., ZHANG, A. and CHENG, G. (2018). Sparse and low-rank tensor estimation via cubic sketchings. Preprint. Available at arXiv:1801.09326.
[57] HOPKINS, S. B., SHI, J., SCHRAMM, T. and STEURER, D. (2016). Fast spectral algorithms from sum-of-squares proofs: Tensor decomposition and planted sparse vectors. In *Proceedings of the* 48*th ACM Symposium on Theory of Computing* 178–191. ACM, New York. MR3536564 https://doi.org/10.1145/2897518.2897529
[58] JAIN, P., NETRAPALLI, P. and SANGHAVI, S. (2013). Low-rank matrix completion using alternating minimization (extended abstract). In *Proceedings of the* 45*th ACM Symposium on Theory of Computing* 665–674. ACM, New York. MR3210828 https://doi.org/10.1145/2488608.2488693
[59] JAIN, P. and OH, S. (2014). Provable tensor factorization with missing data. In *Advances in Neural Information Processing Systems* 1431–1439.

[60] JALALI, A., CHEN, Y., SANGHAVI, S. and XU, H. (2011). Clustering partially observed graphs via convex optimization. In *Proceedings of the International Conference on Machine Learning* **11** 1001–1008.

[61] 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

[62] 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

[63] 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

[64] JOSSE, J. and HUSSON, F. (2012). Handling missing values in exploratory multivariate data analysis methods. *J. SFdS* **153** 79–99. MR3008600

[65] 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

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

[67] KIERS, H. A. (1997). Weighted least squares fitting using ordinary least squares algorithms. *Psychometrika* **62** 251–266.

[68] KIM, H.-J., OLLILA, E., KOIVUNEN, V. and CROUX, C. (2013). Robust and sparse estimation of tensor decompositions. In *Proceedings of the* 2013 *IEEE Global Conference on Signal and Information Processing* 965–968. IEEE.

[69] 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

[70] 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

[71] LARREMORE, D. B., CLAUSET, A. and JACOBS, A. Z. (2014). Efficiently inferring community structure in bipartite networks. *Phys. Rev. E* **90** 012805.

[72] LAWLEY, D. N. and MAXWELL, A. E. (1962). Factor analysis as a statistical method. *Journal of the Royal Statistical Society. Series D* (*The Statistician*) **12** 209–229.

[73] LEI, J. and RINALDO, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann. Statist.* **43** 215–237. MR3285605 https://doi.org/10.1214/14-AOS1274

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

[75] LELARGE, M., MASSOULIÉ, L. and XU, J. (2015). Reconstruction in the labelled stochastic block model. *IEEE Trans. Netw. Sci. Eng.* **2** 152–163. MR3453283 https://doi.org/10.1109/TNSE.2015.2490580

[76] 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.

[77] LOH, P.-L. and WAINWRIGHT, M. J. (2012). High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *Ann. Statist.* **40** 1637–1664. MR3015038 https://doi.org/10.1214/12-AOS1018

[78] LOUNICI, K. (2013). Sparse principal component analysis with missing observations. In *High Dimensional Probability VI. Progress in Probability* **66** 327–356. Birkhäuser/Springer, Basel. MR3443508

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

[80] 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. MR4099988 https://doi.org/10.1007/s10208-019-09429-9

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

[82] MAO, X., SARKAR, P. and CHAKRABARTI, D. (2017). Estimating mixed memberships with sharp eigenvector deviations. Preprint. Available at arXiv:1709.00407.

[83] MASSOULIÉ, L. (2014). Community detection thresholds and the weak Ramanujan property. In *Proceedings of the* 46*th ACM Symposium on Theory of Computing* 694–703. ACM, New York. MR3238997

[84] 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

[85] MOSSEL, E., NEEMAN, J. and SLY, A. (2014). Consistency thresholds for binary symmetric block models. Preprint. Available at arXiv:1407.1591.

[86] MOSSEL, E., NEEMAN, J. and SLY, A. (2015). Reconstruction and estimation in the planted partition model. *Probab. Theory Related Fields* **162** 431–461. MR3383334 https://doi.org/10.1007/s00440-014-0576-6

[87] MU, C., HUANG, B., WRIGHT, J. and GOLDFARB, D. (2014). Square deal: Lower bounds and improved relaxations for tensor recovery. In *Proceedings of the International Conference on Machine Learning* 73–81.

[88] NADLER, B. (2008). Finite sample approximation results for principal component analysis: A matrix perturbation approach. *Ann. Statist.* **36** 2791–2817. MR2485013 https://doi.org/10.1214/08-AOS618

[89] 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

[90] PANANJADY, A. and WAINWRIGHT, M. J. (2019). Value function estimation in Markov reward processes: Instance-dependent ℓ_∞-bounds for policy evaluation. Preprint. Available at arXiv:1909.08749.

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

[92] PAVEZ, E. and ORTEGA, A. (2019). Covariance matrix estimation with non uniform and data dependent missing observations. Preprint. Available at arXiv:1910.00667.

[93] POTECHIN, A. and STEURER, D. (2017). Exact tensor completion with sum-of-squares. In *Proceedings of the Conference on Learning Theory* 1619–1673.

[94] RICHARD, E. and MONTANARI, A. (2014). A statistical model for tensor PCA. In *Advances in Neural Information Processing Systems* 2897–2905.

[95] ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. MR2893856 https://doi.org/10.1214/11-AOS887

[96] ROMERA-PAREDES, B. and PONTIL, M. (2013). A new convex relaxation for tensor completion. In *Advances in Neural Information Processing Systems* 2967–2975.

[97] RUDELSON, M. and VERSHYNIN, R. (2015). Delocalization of eigenvectors of random matrices with independent entries. *Duke Math. J.* **164** 2507–2538. MR3405592 https://doi.org/10.1215/00127094-3129809

[98] SALMI, J., RICHTER, A. and KOIVUNEN, V. (2009). Sequential unfolding SVD for tensors with applications in array signal processing. *IEEE Trans. Signal Process.* **57** 4719–4733. MR2751759 https://doi.org/10.1109/TSP.2009.2027740

[99] 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

[100] SHEN, Y., HUANG, Q., SREBRO, N. and SANGHAVI, S. (2016). Normalized spectral map synchronization. In *Advances in Neural Information Processing Systems* 4925–4933.

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

[102] SUN, R. and LUO, Z.-Q. (2016). Guaranteed matrix completion via non-convex factorization. *IEEE Trans. Inf. Theory* **62** 6535–6579. MR3565131 https://doi.org/10.1109/TIT.2016.2598574

[103] SUSSMAN, D. L., TANG, M., FISHKIND, D. E. and PRIEBE, C. E. (2012). A consistent adjacency spectral embedding for stochastic blockmodel graphs. *J. Amer. Statist. Assoc.* **107** 1119–1128. MR3010899 https://doi.org/10.1080/01621459.2012.699795

[104] VU, V. (2018). A simple SVD algorithm for finding hidden partitions. *Combin. Probab. Comput.* **27** 124–140. MR3734334 https://doi.org/10.1017/S0963548317000463

[105] WEDIN, P. (1973). Perturbation theory for pseudo-inverses. *Nordisk Tidskr. Informationsbehandling* (*BIT*) **13** 217–232. MR0336982 https://doi.org/10.1007/bf01933494

[106] 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

[107] XIA, D., YUAN, M. and ZHANG, C.-H. (2017). Statistically optimal and computationally efficient low rank tensor completion from noisy entries. Preprint. Available at arXiv:1711.04934.

[108] XIA, D. and ZHOU, F. (2019). The sup-norm perturbation of HOSVD and low rank tensor denoising. *J. Mach. Learn. Res.* **20** 61–1. MR3960915

[109] 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

[110] 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

[111] YUN, S.-Y. and PROUTIERE, A. (2014). Accurate community detection in the stochastic block model via spectral algorithms. Preprint. Available at arXiv:1412.7335.

[112] YUN, S.-Y. and PROUTIERE, A. (2016). Optimal cluster recovery in the labeled stochastic block model. In *Advances in Neural Information Processing Systems* 965–973.

[113] ZHANG, A., CAI, T. T. and WU, Y. (2018). Heteroskedastic PCA: Algorithm, optimality, and applications. Preprint. Available at arXiv:1810.08316v2.

[114] 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

[115] 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

[116] ZHOU, Z. and AMINI, A. A. (2018). Optimal bipartite network clustering. Preprint. Available at arXiv:1803.06031.

[117] ZHU, Z., WANG, T. and SAMWORTH, R. J. (2019). High-dimensional principal component analysis with heterogeneous missingness. Preprint. Available at arXiv:1906.12125.

The Annals of Statistics
2021, Vol. 49, No. 2, 968–987
https://doi.org/10.1214/20-AOS1987

ASYMPTOTIC DISTRIBUTION AND CONVERGENCE RATES OF STOCHASTIC ALGORITHMS FOR ENTROPIC OPTIMAL TRANSPORTATION BETWEEN PROBABILITY MEASURES

BY BERNARD BERCU[*] AND JÉRÉMIE BIGOT[†]

Institut de Mathématiques de Bordeaux et CNRS (UMR 5251), Université de Bordeaux, [*]*bernard.bercu@u-bordeaux.fr;* [†]*jeremie.bigot@u-bordeaux.fr*

This paper is devoted to the stochastic approximation of entropically regularized Wasserstein distances between two probability measures, also known as Sinkhorn divergences. The semi-dual formulation of such regularized optimal transportation problems can be rewritten as a nonstrongly concave optimisation problem. It allows to implement a Robbins–Monro stochastic algorithm to estimate the Sinkhorn divergence using a sequence of data sampled from one of the two distributions. Our main contribution is to establish the almost sure convergence and the asymptotic normality of a new recursive estimator of the Sinkhorn divergence between two probability measures in the discrete and semi-discrete settings. We also study the rate of convergence of the expected excess risk of this estimator in the absence of strong concavity of the objective function. Numerical experiments on synthetic and real datasets are also provided to illustrate the usefulness of our approach for data analysis.

REFERENCES

[1] ABID, B. K. and GOWER, R. M. (2018). Greedy stochastic algorithms for entropy-regularized optimal transport problems. In *Proceedings of the* 21*th International Conference on Artificial Intelligence and Statistics*, *Lanzarote*, *Spain*.

[2] ALTSCHULER, J., WEED, J. and RIGOLLET, P. (2017). Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration. In *Proceedings of the* 31*st International Conference on Neural Information Processing Systems* (*USA*, 2017), *NIPS* '17 1961–1971. Curran Associates, Red Hook.

[3] ALTSCHULER, J., WEED, J. and RIGOLLET, P. (2017). Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration. In *Advances in Neural Information Processing Systems* 30: *Annual Conference on Neural Information Processing Systems* 2017, 4–9 *December* 2017, *Long Beach*, *CA*, *USA* 1961–1971.

[4] ARJOVSKY, M., CHINTALA, S. and BOTTOU, L. (2017). Wasserstein generative adversarial networks. In *Proceedings of the* 34*th International Conference on Machine Learning*, *ICML* 2017, *Sydney*, *NSW*, *Australia*, 6–11 *August* 2017 214–223.

[5] BACH, F. (2014). Adaptivity of averaged stochastic gradient descent to local strong convexity for logistic regression. *J*. *Mach*. *Learn*. *Res*. **15** 595–627. MR3190851

[6] BACH, F. R. and MOULINES, E. (2011). Non-asymptotic analysis of stochastic approximation algorithms for machine learning. In *Advances in Neural Information Processing Systems* 24: 25*th Annual Conference on Neural Information Processing Systems* 2011. *Proceedings of a Meeting Held* 12–14 *December* 2011, *Granada*, *Spain* 451–459.

[7] BERCU, B. and BIGOT, J. (2021). Supplement to "Asymptotic distribution and convergence rates of stochastic algorithms for entropic optimal transportation between probability measures." https://doi.org/10.1214/20-AOS1987SUPP

[8] BIGOT, J., CAZELLES, E. and PAPADAKIS, N. (2017). Central limit theorems for Sinkhorn divergence between probability distributions on finite spaces and statistical applications. Preprint. Available at arXiv:1711.08947.

[9] 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

MSC2020 subject classifications. Primary 62G05; secondary 62G20.
Key words and phrases. Stochastic optimisation, convergence of random variables, optimal transport, entropic regularization, Sinkhorn divergence, Wasserstein distance.

[10] BRENIER, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.* **44** 375–417. MR1100809 https://doi.org/10.1002/cpa.3160440402

[11] CATTIAUX, P. and GAMBOA, F. (1999). Large deviations and variational theorems for marginal problems. *Bernoulli* **5** 81–108. MR1673564 https://doi.org/10.2307/3318614

[12] CHIZAT, L., PEYRÉ, G., SCHMITZER, B. and VIALARD, F.-X. (2018). Scaling algorithms for unbalanced optimal transport problems. *Math. Comp.* **87** 2563–2609. MR3834678 https://doi.org/10.1090/mcom/3303

[13] CSISZÁR, I. (1975). I-divergence geometry of probability distributions and minimization problems. *Ann. Probab.* **3** 146–158. MR0365798 https://doi.org/10.1214/aop/1176996454

[14] CUESTA, J. A. and MATRÁN, C. (1989). Notes on the Wasserstein metric in Hilbert spaces. *Ann. Probab.* **17** 1264–1276. MR1009457

[15] CUTURI, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems* 26 (C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani and K. Q. Weinberger, eds.) 2292–2300. Curran Associates, Red Hook.

[16] CUTURI, M. and PEYRÉ, G. (2016). A smoothed dual approach for variational Wasserstein problems. *SIAM J. Imaging Sci.* **9** 320–343. MR3466197 https://doi.org/10.1137/15M1032600

[17] CUTURI, M. and PEYRÉ, G. (2017). *Computational Optimal Transport*. Available at https://optimaltransport.github.io/book/.

[18] DEL BARRIO, E., CUESTA-ALBERTOS, J. A., MATRÁN, C. and RODRÍGUEZ-RODRÍGUEZ, J. M. (1999). Tests of goodness of fit based on the L_2-Wasserstein distance. *Ann. Statist.* **27** 1230–1239. MR1740113 https://doi.org/10.1214/aos/1017938923

[19] DEL BARRIO, E., GINÉ, E. and UTZET, F. (2005). Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli* **11** 131–189. MR2121458 https://doi.org/10.3150/bj/1110228245

[20] DEL BARRIO, E. and LOUBES, J.-M. (2019). Central limit theorems for empirical transportation cost in general dimension. *Ann. Probab.* **47** 926–951. MR3916938 https://doi.org/10.1214/18-AOP1275

[21] EVANS, S. N. and MATSEN, F. A. (2012). The phylogenetic Kantorovich–Rubinstein metric for environmental sequence samples. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 569–592. MR2925374 https://doi.org/10.1111/j.1467-9868.2011.01018.x

[22] FREITAG, G. and MUNK, A. (2005). On Hadamard differentiability in k-sample semiparametric models—With applications to the assessment of structural relationships. *J. Multivariate Anal.* **94** 123–158. MR2161214 https://doi.org/10.1016/j.jmva.2004.03.006

[23] FROGNER, C., ZHANG, C., MOBAHI, H., ARAYA, M. and POGGIO, T. A. (2015). Learning with a Wasserstein loss. In *Advances in Neural Information Processing Systems* 2053–2061.

[24] GENEVAY, A. (2019). Entropy-regularized optimal transport for machine learning. Ph.D. thesis, Univ. Paris-Dauphine.

[25] GENEVAY, A., CUTURI, M., PEYRÉ, G. and BACH, F. (2016). Stochastic optimization for large-scale optimal transport. In *Advances in Neural Information Processing Systems* 29 (D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon and R. Garnett, eds.) 3440–3448. Curran Associates, Red Hook.

[26] GENEVAY, A., PEYRE, G. and CUTURI, M. (2018). Learning generative models with Sinkhorn divergences. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics* (A. Storkey and F. Perez-Cruz, eds.). *Proceedings of Machine Learning Research*, *PMLR* **84** 1608–1617.

[27] GODICHON-BAGGIONI, A. (2019). L^p and almost sure rates of convergence of averaged stochastic gradient algorithms: Locally strongly convex objective. *ESAIM Probab. Stat.* **23** 841–873. MR4045544 https://doi.org/10.1051/ps/2019011

[28] GRAMFORT, A., PEYRÉ, G. and CUTURI, M. (2015). Fast optimal transport averaging of neuroimaging data. In *International Conference on Information Processing in Medical Imaging* 261–272. Springer, Berlin.

[29] HARTMANN, V. and SCHUHMACHER, D. (2018). Semi-discrete optimal transport—The case $p = 1$. Preprint. Available at arXiv:1706.07650.

[30] KITAGAWA, J., MÉRIGOT, Q. and THIBERT, B. (2019). Convergence of a Newton algorithm for semi-discrete optimal transport. *J. Eur. Math. Soc.* (*JEMS*) **21** 2603–2651. MR3985609 https://doi.org/10.4171/JEMS/889

[31] KLATT, M., TAMELING, C. and MUNK, A. (2020). Empirical regularized optimal transport: Statistical theory and applications. *SIAM J. Math. Data Sci.* **2** 419–443. MR4105566 https://doi.org/10.1137/19M1278788

[32] LÉONARD, C. (2014). A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.* **34** 1533–1574. MR3121631 https://doi.org/10.3934/dcds.2014.34.1533

[33] MÉRIGOT, Q., MEYRON, J. and THIBERT, B. (2018). An algorithm for optimal transport between a simplex soup and a point cloud. *SIAM J. Imaging Sci.* **11** 1363–1389. MR3805844 https://doi.org/10.1137/17M1137486

[34] MUNK, A. and CZADO, C. (1998). Nonparametric validation of similar distributions and assessment of goodness of fit. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 223–241. MR1625620 https://doi.org/10.1111/1467-9868.00121

[35] PELLETIER, M. (1998). On the almost sure asymptotic behaviour of stochastic algorithms. *Stochastic Process. Appl.* **78** 217–244. MR1654569 https://doi.org/10.1016/S0304-4149(98)00029-5

[36] PELLETIER, M. (1998). Weak convergence rates for stochastic approximation with application to multiple targets and simulated annealing. *Ann. Appl. Probab.* **8** 10–44. MR1620405 https://doi.org/10.1214/aoap/1027961032

[37] 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

[38] R CORE TEAM (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[39] RABIN, J. and PAPADAKIS, N. (2015). Convex color image segmentation with optimal transport distances. In *Scale Space and Variational Methods in Computer Vision*. *Lecture Notes in Computer Science* **9087** 256–269. Springer, Cham. MR3394935 https://doi.org/10.1007/978-3-319-18461-6_21

[40] RIPPL, T., MUNK, A. and STURM, A. (2016). Limit laws of the empirical Wasserstein distance: Gaussian distributions. *J. Multivariate Anal.* **151** 90–109. MR3545279 https://doi.org/10.1016/j.jmva.2016.06.005

[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] ROLET, A., CUTURI, M. and PEYRÉ, G. (2016). Fast dictionary learning with a smoothed Wasserstein loss. In *Proc. International Conference on Artificial Intelligence and Statistics* (*AISTATS*).

[43] SEGUY, V. and CUTURI, M. (2015). Principal geodesic analysis for probability measures under the optimal transport metric. In *Advances in Neural Information Processing Systems* 28 (C. Cortes, N. Lawrence, D. Lee, M. Sugiyama and R. Garnett, eds.) 3294–3302. Curran Associates, Red Hook.

[44] SOMMERFELD, M. and MUNK, A. (2018). Inference for empirical Wasserstein distances on finite spaces. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 219–238. MR3744719 https://doi.org/10.1111/rssb.12236

[45] 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

[46] 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

[47] YE, J., WU, P., WANG, J. Z. and LI, J. (2017). Fast discrete distribution clustering using Wasserstein barycenter with sparse support. *IEEE Trans. Signal Process.* **65** 2317–2332. MR3620352 https://doi.org/10.1109/TSP.2017.2659647

The Annals of Statistics
2021, Vol. 49, No. 2, 988–1010
https://doi.org/10.1214/20-AOS1988

MULTISCALE GEOMETRIC FEATURE EXTRACTION FOR HIGH-DIMENSIONAL AND NON-EUCLIDEAN DATA WITH APPLICATIONS

BY GABRIEL CHANDLER[1] AND WOLFGANG POLONIK[2]

[1]*Department of Mathematics and Statistics, Pomona College, gabriel.chandler@pomona.edu*
[2]*Department of Statistics, University of California, Davis, wpolonik@ucdavis.edu*

A method for extracting multiscale geometric features from a data cloud is proposed and analyzed. Based on geometric considerations, we map each pair of data points into a real-valued feature function defined on the unit interval. Further statistical analysis is then based on the collection of feature functions. The potential of the method is illustrated by different applications, including classification and anomaly detection. Connections to other concepts, such as random set theory, localized depth measures and nonlinear dimension reduction, are also explored.

REFERENCES

AEBERHARD, S., COOMANS, D. and DE VEL, O. (1994). CComparative analysis of statistical pattern recognition methods in high dimensional settings. *Pattern Recognition* **27** 1065–1077.

AGOSTINELLI, C. (2018). Local half-region depth for functional data. *J. Multivariate Anal*. **163** 67–79. MR3732341 https://doi.org/10.1016/j.jmva.2017.10.004

AGOSTINELLI, C. and ROMANAZZI, M. (2011). Local depth. *J. Statist. Plann. Inference* **141** 817–830. MR2732952 https://doi.org/10.1016/j.jspi.2010.08.001

ALON, U., BARKAI, N., NOTTERMAN, D. A., GISH, K., YBARRA, S., MACK, D. and LEVINE, A. J. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl. Acad. Sci. USA* **96** 6745–6750.

BUBENIK, P. (2015). Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res*. **16** 77–102. MR3317230

BUBENIK, P. and KIM, P. T. (2007). A statistical approach to persistent homology. *Homology, Homotopy Appl*. **9** 337–362. MR2366953

CAMPOS, G. O., ZIMEK, A., SANDER, J., CAMPELLO, R. J. G. B., MICENKOVÁ, B., SCHUBERT, E., ASSENT, I. and HOULE, M. E. (2016). On the evaluation of unsupervised outlier detection: Measures, datasets, and an empirical study. *Data Min. Knowl. Discov*. **30** 891–927. MR3513246 https://doi.org/10.1007/s10618-015-0444-8

CHANDLER, G. and POLONIK, W. (2021). Supplement to "Multiscale geometric feature extraction for high-dimensional and non-Euclidean data with applications." https://doi.org/10.1214/20-AOS1988SUPP

CHAUDHURI, P. and MARRON, J. S. (1999). SiZer for exploration of structures in curves. *J. Amer. Statist. Assoc*. **94** 807–823. MR1723347 https://doi.org/10.2307/2669996

CHEN, Y., DANG, X., PENG, H. and BART, H. L. (2008). Outlier detection with the kernelized spatial depth function. *IEEE Trans. Pattern Anal. Mach. Intell*. **31** 288–305.

CUTURI, M. (2010). Positive definite kernels in machine learning Technical report.

DUTTA, S., SARKAR, S. and GHOSH, A. K. (2016). Multi-scale classification using localized spatial depth. *J. Mach. Learn. Res*. **17** Art. ID 218. MR3595152

EINMAHL, J. H. J. and MASON, D. M. (1992). Generalized quantile processes. *Ann. Statist*. **20** 1062–1078. MR1165606 https://doi.org/10.1214/aos/1176348670

ELMORE, R. T., HETTMANSPERGER, T. P. and XUAN, F. (2006). Spherical data depth and a multivariate median. In *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications. DIMACS Ser. Discrete Math. Theoret. Comput. Sci*. **72** 87–101. Amer. Math. Soc., Providence, RI. MR2343115

FRAIMAN, R., GAMBOA, F. and MORENO, L. (2019). Connecting pairwise geodesic spheres by depth: DCOPS. *J. Multivariate Anal*. **169** 81–94. MR3875588 https://doi.org/10.1016/j.jmva.2018.08.008

MSC2020 subject classifications. 62G07, 62GHXX.
Key words and phrases. Kernel-trick, local depth, kernelized depth, visualization, VC-classes, curse of dimensionality, nonlinear dimension reduction, random set.

GENTON, M. G. (2002). Classes of kernels for machine learning: A statistics perspective. *J. Mach. Learn. Res.* **2** 293–312. MR1904760 https://doi.org/10.1162/15324430260185637

KOTÍK, L. and HLUBINKA, D. (2017). A weighted localization of halfspace depth and its properties. *J. Multivariate Anal.* **157** 53–69. MR3641736 https://doi.org/10.1016/j.jmva.2017.02.008

LENG, X. and MÜLLER, H.-G. (2006). Classification using functional data analysis for temporal gene expression data. *Bioinformatics* **22** 68–76. https://doi.org/10.1093/bioinformatics/bti742

LIU, R. Y. (1988). On a notion of simplicial depth. *Proc. Natl. Acad. Sci. USA* **85** 1732–1734. MR0930658 https://doi.org/10.1073/pnas.85.6.1732

LIU, R. Y. (1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18** 405–414. MR1041400 https://doi.org/10.1214/aos/1176347507

MASSÉ, J.-C. (2004). Asymptotics for the Tukey depth process, with an application to a multivariate trimmed mean. *Bernoulli* **10** 397–419. MR2061438 https://doi.org/10.3150/bj/1089206404

MINOTTE, M. and SCOTT, D. (1993). The mode tree: A tool for visualization of nonparametric density features. *J. Comput. Graph. Statist.* **2** 51–68.

PAINDAVEINE, D. and VAN BEVER, G. (2013). From depth to local depth: A focus on centrality. *J. Amer. Statist. Assoc.* **108** 1105–1119. MR3174687 https://doi.org/10.1080/01621459.2013.813390

PHAM, N. (2018). L1-depth revisited: A robust angle-based outlier factor in high-dimensional space. In *Machine Learning and Knowledge Discovery in Databases* 105–121. Springer, Berlin.

RÉNYI, A. and SULANKE, R. (1963). Über die konvexe Hülle von n zufällig gewählten Punkten. *Z. Wahrsch. Verw. Gebiete* **2** 75–84. MR0156262 https://doi.org/10.1007/BF00535300

SERFLING, R. (2019). Depth functions on general data spaces, I. Perspectives, with consideration of "density" and "local" depths. Available at https://personal.utdallas.edu/~serfling/papers/I.Perspectives.pdf.

SHAWE-TAYLOR, J. and CRISTIANINI, N. (2004). *Kernel Methods for Pattern Analysis*. Cambridge Univ. Press, Cambridge.

SPELLMAN, P. T., SHERLOCK, G., ZHANG, M. Q., IYER, V. R., ANDERS, K., EISEN, M. B., BROWN, P. O., BOTSTEIN, D. and FUTCHER, B. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast saccharomyces cerevisiae by microarray hybridization. *Mol. Biol. Cell* **9** 3273–3297.

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

TING, K. M., ZHOU, G.-T., LIU, F. T. and TAN, S. C. (2013). Mass estimation. *Mach. Learn.* **90** 127–160. MR3008690 https://doi.org/10.1007/s10994-012-5303-x

ZUO, Y. and SERFLING, R. (2000). General notions of statistical depth function. *Ann. Statist.* **28** 461–482. MR1790005 https://doi.org/10.1214/aos/1016218226

The Annals of Statistics
2021, Vol. 49, No. 2, 1011–1028
https://doi.org/10.1214/20-AOS1989

COVERAGE OF CREDIBLE INTERVALS IN NONPARAMETRIC MONOTONE REGRESSION

BY MOUMITA CHAKRABORTY* AND SUBHASHIS GHOSAL†

Department of Statistics, North Carolina State University, *mchakra@ncsu.edu; †sghosal@ncsu.edu

For nonparametric univariate regression under a monotonicity constraint on the regression function f, we study the coverage of a Bayesian credible interval for $f(x_0)$, where x_0 is an interior point. Analysis of the posterior becomes a lot more tractable by considering a "projection-posterior" distribution based on a finite random series of step functions with normal basis coefficients as a prior for f. A sample f from the resulting conjugate posterior distribution is projected on the space of monotone increasing functions to obtain a monotone function f^* closest to f, inducing the "projection-posterior." We use projection-posterior samples to obtain credible intervals for $f(x_0)$. We obtain the asymptotic coverage of the credible interval thus constructed and observe that it is free of nuisance parameters involving the true function. We observe a very interesting phenomenon that the coverage is typically higher than the nominal credibility level, the opposite of a phenomenon observed by Cox (*Ann. Statist.* **21** (1993) 903–923) in the Gaussian sequence model. We further show that a recalibration gives the right asymptotic coverage by starting from a lower credibility level that can be explicitly calculated.

REFERENCES

[1] ARMSTRONG, T. (2015). Adaptive testing on a regression function at a point. *Ann. Statist.* **43** 2086–2101. MR3375877 https://doi.org/10.1214/15-AOS1342

[2] AYER, M., BRUNK, H. D., EWING, G. M., REID, W. T. and SILVERMAN, E. (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Stat.* **26** 641–647. MR0073895 https://doi.org/10.1214/aoms/1177728423

[3] BANERJEE, M. (2000). Likelihood ratio inference in regular and non-regular problems. Ph. D. Dissertation—University of Washington. MR2701606

[4] BANERJEE, M. (2007). Likelihood based inference for monotone response models. *Ann. Statist.* **35** 931–956. MR2341693 https://doi.org/10.1214/009053606000001578

[5] BANERJEE, M. and WELLNER, J. A. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29** 1699–1731. MR1891743 https://doi.org/10.1214/aos/1015345959

[6] BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions. The Theory and Application of Isotonic Regression*. Wiley, London-Sydney. MR0326887

[7] BARLOW, R. E. and BRUNK, H. D. (1972). The isotonic regression problem and its dual. *J. Amer. Statist. Assoc.* **67** 140–147. MR0314205

[8] BELITSER, E. (2017). On coverage and local radial rates of credible sets. *Ann. Statist.* **45** 1124–1151. MR3662450 https://doi.org/10.1214/16-AOS1477

[9] BELITSER, E. and GHOSAL, S. (2020). Empirical Bayes oracle uncertainty quantification for regression. *Ann. Statist.* To appear.

[10] BELITSER, E. and NURUSHEV, N. (2019). Robust inference for general framework of projection structures. Preprint. Available at arXiv:1904.01003.

[11] BELITSER, E. and NURUSHEV, N. (2020). Needles and straw in a haystack: Robust confidence for possibly sparse sequences. *Bernoulli* **26** 191–225. MR4036032 https://doi.org/10.3150/19-BEJ1122

[12] BHAUMIK, P. and GHOSAL, S. (2015). Bayesian two-step estimation in differential equation models. *Electron. J. Stat.* **9** 3124–3154. MR3453972 https://doi.org/10.1214/15-EJS1099

MSC2020 subject classifications. Primary 62G08, 62F15, 62G05, 62G10, 62G15, 62G20.
Key words and phrases. Nonparametric regression, monotonicity, credible set, coverage, Chernoff's distribution, projection-posterior.

[13] BHAUMIK, P. and GHOSAL, S. (2017). Efficient Bayesian estimation and uncertainty quantification in ordinary differential equation models. *Bernoulli* **23** 3537–3570. MR3654815 https://doi.org/10.3150/16-BEJ856

[14] BOWMAN, A. W., JONES, M. C. and GIJBELS, I. (1998). Testing monotonicity of regression. *J. Comput. Graph. Statist.* **7** 489–500.

[15] BRUNK, H. D. (1970). Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference* (*Proc. Sympos., Indiana Univ., Bloomington, Ind.*, 1969) 177–197. Cambridge Univ. Press, London. MR0277070

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

[17] CAI, T. T., LOW, M. G. and XIA, Y. (2013). Adaptive confidence intervals for regression functions under shape constraints. *Ann. Statist.* **41** 722–750. MR3099119 https://doi.org/10.1214/12-AOS1068

[18] 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

[19] 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

[20] CASTILLO, I. and SZABÓ, B. (2020). Spike and slab empirical Bayes sparse credible sets. *Bernoulli* **26** 127–158. MR4036030 https://doi.org/10.3150/19-BEJ1119

[21] CHAKRABORTY, M. and GHOSAL, S. (2021). Supplement to "Coverage of credible intervals in nonparametric monotone regression." https://doi.org/10.1214/20-AOS1989SUPP

[22] CHERNOFF, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.* **16** 31–41. MR0172382 https://doi.org/10.1007/BF02868560

[23] 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

[24] DE LEEUW, J., KURT, H. and MAIR, P. (2009). Isotone optimization in R: Pool-adjacent-violators algorithm (PAVA) and active set methods. *J. Stat. Softw.* **32**.

[25] DÜMBGEN, L. (2003). Optimal confidence bands for shape-restricted curves. *Bernoulli* **9** 423–449. MR1997491 https://doi.org/10.3150/bj/1065444812

[26] DÜMBGEN, L. and JOHNS, R. B. (2004). Confidence bands for isotonic median curves using sign tests. *J. Comput. Graph. Statist.* **13** 519–533. MR2063998 https://doi.org/10.1198/1061860043506

[27] GHOSAL, S., SEN, A. and VAN DER VAART, A. W. (2000). Testing monotonicity of regression. *Ann. Statist.* **28** 1054–1082. MR1810919 https://doi.org/10.1214/aos/1015956707

[28] GIJBELS, I., HALL, P., JONES, M. C. and KOCH, I. (2000). Tests for monotonicity of a regression mean with guaranteed level. *Biometrika* **87** 663–673. MR1789816 https://doi.org/10.1093/biomet/87.3.663

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

[30] GRENANDER, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39** 125–153. MR0093415 https://doi.org/10.1080/03461238.1956.10414944

[31] GROENEBOOM, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probab. Theory Related Fields* **81** 79–109. MR0981568 https://doi.org/10.1007/BF00343738

[32] 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

[33] GROENEBOOM, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation. DMV Seminar* **19**. Birkhäuser, Basel. MR1180321 https://doi.org/10.1007/978-3-0348-8621-5

[34] GROENEBOOM, P. and WELLNER, J. A. (2001). Computing Chernoff's distribution. *J. Comput. Graph. Statist.* **10** 388–400. MR1939706 https://doi.org/10.1198/10618600152627997

[35] HALL, P. and HECKMAN, N. E. (2000). Testing for monotonicity of a regression mean by calibrating for linear functions. *Ann. Statist.* **28** 20–39. MR1762902 https://doi.org/10.1214/aos/1016120363

[36] HOFFMANN, M. and NICKL, R. (2011). On adaptive inference and confidence bands. *Ann. Statist.* **39** 2383–2409. MR2906872 https://doi.org/10.1214/11-AOS903

[37] HUANG, J. and WELLNER, J. A. (1995). Estimation of a monotone density or monotone hazard under random censoring. *Scand. J. Stat.* **22** 3–33. MR1334065

[38] HUANG, Y. and ZHANG, C.-H. (1994). Estimating a monotone density from censored observations. *Ann. Statist.* **22** 1256–1274. MR1311975 https://doi.org/10.1214/aos/1176325628

[39] KNAPIK, B. T., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2011). Bayesian inverse problems with Gaussian priors. *Ann. Statist.* **39** 2626–2657. MR2906881 https://doi.org/10.1214/11-AOS920

[40] KOSOROK, M. R. (2008). Bootstrapping in Grenander estimator. In *Beyond Parametrics in Interdisciplinary Research*: *Festschrift in Honor of Professor Pranab K. Sen*. *Inst. Math. Stat.* (*IMS*) *Collect.* **1** 282–292. IMS, Beachwood, OH. MR2462212 https://doi.org/10.1214/193940307000000202

[41] LEAHU, H. (2011). On the Bernstein–von Mises phenomenon in the Gaussian white noise model. *Electron. J. Stat.* **5** 373–404. MR2802048 https://doi.org/10.1214/11-EJS611

[42] LI, K.-C. (1989). Honest confidence regions for nonparametric regression. *Ann. Statist.* **17** 1001–1008. MR1015135 https://doi.org/10.1214/aos/1176347253

[43] LIN, L. and DUNSON, D. B. (2014). Bayesian monotone regression using Gaussian process projection. *Biometrika* **101** 303–317. MR3215349 https://doi.org/10.1093/biomet/ast063

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

[45] PATSCHKOWSKI, T. and ROHDE, A. (2019). Locally adaptive confidence bands. *Ann. Statist.* **47** 349–381. MR3909936 https://doi.org/10.1214/18-AOS1690

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

[47] PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36. MR0267677

[48] RAY, K. (2017). Adaptive Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.* **45** 2511–2536. MR3737900 https://doi.org/10.1214/16-AOS1533

[49] ROBINS, J. and VAN DER VAART, A. (2006). Adaptive nonparametric confidence sets. *Ann. Statist.* **34** 229–253. MR2275241 https://doi.org/10.1214/009053605000000877

[50] SALOMOND, J.-B. (2018). Testing un-separated hypotheses by estimating a distance. *Bayesian Anal.* **13** 461–484. MR3780431 https://doi.org/10.1214/17-BA1059

[51] SCHMIDT-HIEBER, J., MUNK, A. and DÜMBGEN, L. (2013). Multiscale methods for shape constraints in deconvolution: Confidence statements for qualitative features. *Ann. Statist.* **41** 1299–1328. MR3113812 https://doi.org/10.1214/13-AOS1089

[52] SEN, B., BANERJEE, M. and WOODROOFE, M. (2010). Inconsistency of bootstrap: The Grenander estimator. *Ann. Statist.* **38** 1953–1977. MR2676880 https://doi.org/10.1214/09-AOS777

[53] SNIEKERS, S. and VAN DER VAART, A. (2015). Credible sets in the fixed design model with Brownian motion prior. *J. Statist. Plann. Inference* **166** 78–86. MR3390135 https://doi.org/10.1016/j.jspi.2014.07.008

[54] 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

[55] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

[56] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Process with Applications to Statistics*. Springer, New York.

[57] YOO, W. W. and GHOSAL, S. (2016). Supremum norm posterior contraction and credible sets for nonparametric multivariate regression. *Ann. Statist.* **44** 1069–1102. MR3485954 https://doi.org/10.1214/15-AOS1398

The Annals of Statistics
2021, Vol. 49, No. 2, 1029–1054
https://doi.org/10.1214/20-AOS1990

LINEARIZED TWO-LAYERS NEURAL NETWORKS IN HIGH DIMENSION

BY BEHROOZ GHORBANI[1], SONG MEI[2], THEODOR MISIAKIEWICZ[3] AND ANDREA MONTANARI[4]

[1]*Department of Electrical Engineering, Stanford University, ghorbani@stanford.edu*
[2]*Institute for Computational and Mathematical Engineering, Stanford University, songmei@stanford.edu*
[3]*Department of Statistics, Stanford University, misiakie@stanford.edu*
[4]*Department of Electrical Engineering and Department of Statistics, Stanford University, montanari@stanford.edu*

We consider the problem of learning an unknown function $f_\star$ on the d-dimensional sphere with respect to the square loss, given i.i.d. samples $\{(y_i, \boldsymbol{x}_i)\}_{i\le n}$ where $\boldsymbol{x}_i$ is a feature vector uniformly distributed on the sphere and $y_i = f_\star(\boldsymbol{x}_i) + \varepsilon_i$. We study two popular classes of models that can be regarded as linearizations of two-layers neural networks around a random initialization: the random features model of Rahimi–Recht (RF); the neural tangent model of Jacot–Gabriel–Hongler (NT). Both these models can also be regarded as randomized approximations of kernel ridge regression (with respect to different kernels), and enjoy universal approximation properties when the number of neurons N diverges, for a fixed dimension d.

We consider two specific regimes: the infinite-sample finite-width regime, in which $n = \infty$ while d and N are large but finite, and the infinite-width finite-sample regime in which $N = \infty$ while d and n are large but finite. In the first regime, we prove that if $d^{\ell+\delta} \le N \le d^{\ell+1-\delta}$ for small $\delta > 0$, then RF effectively fits a degree-ℓ polynomial in the raw features, and NT fits a degree-$(\ell+1)$ polynomial. In the second regime, both RF and NT reduce to kernel methods with rotationally invariant kernels. We prove that, if the sample size satisfies $d^{\ell+\delta} \le n \le d^{\ell+1-\delta}$, then kernel methods can fit at most a degree-ℓ polynomial in the raw features. This lower bound is achieved by kernel ridge regression, and near-optimal prediction error is achieved for vanishing ridge regularization.

REFERENCES

[1] ALAOUI, A. E. and MAHONEY, M. W. (2015). Fast randomized kernel ridge regression with statistical guarantees. In *Advances in Neural Information Processing Systems* 775–783.

[2] ALLEN-ZHU, Z., LI, Y. and SONG, Z. (2019). A convergence theory for deep learning via over-parameterization. In *International Conference on Machine Learning* 242–252.

[3] ANTHONY, M. and BARTLETT, P. L. (1999). *Neural Network Learning*: *Theoretical Foundations*. Cambridge Univ. Press, Cambridge. MR1741038 https://doi.org/10.1017/CBO9780511624216

[4] ARORA, S., DU, S., HU, W., LI, Z. and WANG, R. (2019). Fine-grained analysis of optimization and generalization for overparameterized two-layer neural networks. In *International Conference on Machine Learning* 322–332.

[5] BACH, F. (2013). Sharp analysis of low-rank kernel matrix approximations. In *Conference on Learning Theory* 185–209.

[6] BACH, F. (2017). On the equivalence between kernel quadrature rules and random feature expansions. *J. Mach. Learn. Res.* **18** Paper No. 21, 38. MR3634888

[7] BACH, F. (2017). Breaking the curse of dimensionality with convex neutral networks. *J. Mach. Learn. Res.* **18** Paper No. 19, 53. MR3634886

[8] BARRON, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans. Inf. Theory* **39** 930–945. MR1237720 https://doi.org/10.1109/18.256500

MSC2020 subject classifications. Primary 62G08; secondary 62J07.
Key words and phrases. Two-layers neural networks, random features, neural tangent kernel, approximation bounds, Kernel ridge regression.

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

[10] BELKIN, M., HSU, D. and XU, J. (2019). Two models of double descent for weak features. Available at arXiv:1903.07571.

[11] BELKIN, M., HSU, D. J. and MITRA, P. (2018). Overfitting or perfect fitting? Risk bounds for classification and regression rules that interpolate. In *Advances in Neural Information Processing Systems* 2300–2311.

[12] BELKIN, M., RAKHLIN, A. and TSYBAKOV, A. B. (2019). Does data interpolation contradict statistical optimality? In *The 22nd International Conference on Artificial Intelligence and Statistics* 1611–1619.

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

[14] CAPONNETTO, A. and DE VITO, E. (2007). Optimal rates for the regularized least-squares algorithm. *Found. Comput. Math.* **7** 331–368. MR2335249 https://doi.org/10.1007/s10208-006-0196-8

[15] CHIHARA, T. S. (2011). *An Introduction to Orthogonal Polynomials*. Courier Corporation.

[16] CHIZAT, L. and BACH, F. (2018). On the global convergence of gradient descent for over-parameterized models using optimal transport. In *Advances in Neural Information Processing Systems* 3036–3046.

[17] CHIZAT, L., OYALLON, E. and BACH, F. (2019). On lazy training in differentiable programming. In *Advances in Neural Information Processing Systems* 2933–2943.

[18] CRISTIANINI, N., SHAWE-TAYLOR, J. et al. (2000). *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge Univ. Press, Cambridge.

[19] CYBENKO, G. (1989). Approximation by superpositions of a sigmoidal function. *Math. Control Signals Systems* **2** 303–314. MR1015670 https://doi.org/10.1007/BF02551274

[20] DEVORE, R. A., HOWARD, R. and MICCHELLI, C. (1989). Optimal nonlinear approximation. *Manuscripta Math.* **63** 469–478. MR0991266 https://doi.org/10.1007/BF01171759

[21] DONOHO, D. L. and JOHNSTONE, I. M. (1989). Projection-based approximation and a duality with kernel methods. *Ann. Statist.* **17** 58–106. MR0981438 https://doi.org/10.1214/aos/1176347004

[22] DU, S., LEE, J., LI, H., WANG, L. and ZHAI, X. (2019). Gradient descent finds global minima of deep neural networks. In *International Conference on Machine Learning* 1675–1685.

[23] DU, S. S., ZHAI, X., POCZOS, B. and SINGH, A. (2019). Gradient descent provably optimizes over-parameterized neural networks. In *International Conference on Learning Representations*.

[24] EFTHIMIOU, C. and FRYE, C. (2014). *Spherical Harmonics in p Dimensions*. World Scientific Co. Pte. Ltd., Hackensack, NJ. MR3290046 https://doi.org/10.1142/9134

[25] EL KAROUI, N. (2010). The spectrum of kernel random matrices. *Ann. Statist.* **38** 1–50. MR2589315 https://doi.org/10.1214/08-AOS648

[26] EL KAROUI, N. (2010). On information plus noise kernel random matrices. *Ann. Statist.* **38** 3191–3216. MR2722468 https://doi.org/10.1214/10-AOS801

[27] GEIGER, M., SPIGLER, S., JACOT, A. and WYART, M. (2019). Disentangling feature and lazy learning in deep neural networks: An empirical study. Available at arXiv:1906.08034.

[28] GHORBANI, B., MEI, S., MISIAKIEWICZ, T. and MONTANARI, A. (2019). Limitations of lazy training of two-layers neural network. In *Advances in Neural Information Processing Systems* 9108–9118.

[29] GHORBANI, B., MEI, S., MISIAKIEWICZ, T. and MONTANARI, A. (2020). When do neural networks outperform kernel methods? In *Advances in Neural Information Processing Systems* (H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan and H. Lin, eds.) 14820–14830. Curran Associates.

[30] GHORBANI, B., MEI, S., MISIAKIEWICZ, T. and MONTANARI, A. (2021). Supplement to "Linearized two-layers neural networks in high dimension." https://doi.org/10.1214/20-AOS1990SUPP

[31] GIROSI, F., JONES, M. and POGGIO, T. (1995). Regularization theory and neural networks architectures. *Neural Comput.* **7** 219–269.

[32] GYÖRFI, L., KOHLER, M., KRZYŻAK, A. and WALK, H. (2002). *A Distribution-Free Theory of Nonparametric Regression. Springer Series in Statistics*. Springer, New York. MR1920390 https://doi.org/10.1007/b97848

[33] HASTIE, T., MONTANARI, A., ROSSET, S. and TIBSHIRANI, R. J. (2019). Surprises in high-dimensional ridgeless least squares interpolation. Available at arXiv:1903.08560.

[34] HORNIK, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Netw.* **4** 251–257.

[35] JACOT, A., GABRIEL, F. and HONGLER, C. (2018). Neural tangent kernel: Convergence and generalization in neural networks. In *Advances in Neural Information Processing Systems* 8571–8580.

[36] LEE, J., XIAO, L., SCHOENHOLZ, S., BAHRI, Y., NOVAK, R., SOHL-DICKSTEIN, J. and PENNINGTON, J. (2019). Wide neural networks of any depth evolve as linear models under gradient descent. In *Advances in Neural Information Processing Systems* 8570–8581.

[37] LIANG, T. and RAKHLIN, A. (2020). Just interpolate: Kernel "Ridgeless" regression can generalize. *Ann. Statist.* **48** 1329–1347. MR4124325 https://doi.org/10.1214/19-AOS1849
[38] LIANG, T., RAKHLIN, A. and ZHAI, X. (2019). On the risk of minimum-norm interpolants and restricted lower isometry of kernels. Available at arXiv:1908.10292.
[39] MAIOROV, V. E. (1999). On best approximation by ridge functions. *J. Approx. Theory* **99** 68–94. MR1696577 https://doi.org/10.1006/jath.1998.3304
[40] MAIOROV, V. E. and MEIR, R. (2000). On the near optimality of the stochastic approximation of smooth functions by neural networks. *Adv. Comput. Math.* **13** 79–103. MR1759189 https://doi.org/10.1023/A:1018993908478
[41] MEI, S., BAI, Y. and MONTANARI, A. (2018). The landscape of empirical risk for nonconvex losses. *Ann. Statist.* **46** 2747–2774. MR3851754 https://doi.org/10.1214/17-AOS1637
[42] MEI, S., MISIAKIEWICZ, T. and MONTANARI, A. (2019). Mean-field theory of two-layers neural networks: Dimension-free bounds and kernel limit. In *Conference on Learning Theory* 2388–2464.
[43] MEI, S. and MONTANARI, A. (2019). The generalization error of random features regression: Precise asymptotics and double descent curve. Available at arXiv:1908.05355.
[44] MEI, S., MONTANARI, A. and NGUYEN, P.-M. (2018). A mean field view of the landscape of two-layer neural networks. *Proc. Natl. Acad. Sci. USA* **115** E7665–E7671. MR3845070 https://doi.org/10.1073/pnas.1806579115
[45] MHASKAR, H. N. (1996). Neural networks for optimal approximation of smooth and analytic functions. *Neural Comput.* **8** 164–177.
[46] MHASKAR, H. N. and MICCHELLI, C. A. (1994). Dimension-independent bounds on the degree of approximation by neural networks. *IBM J. Res. Develop.* **38** 277–284.
[47] NEAL, R. M. (1996). Priors for infinite networks. In *Bayesian Learning for Neural Networks* 29–53. Springer, Berlin.
[48] PETRUSHEV, P. P. (1999). Approximation by ridge functions and neural networks. *SIAM J. Math. Anal.* **30** 155–189. MR1646689 https://doi.org/10.1137/S0036141097322959
[49] PINKUS, A. (1999). Approximation theory of the MLP model in neural networks. In *Acta Numerica*, 1999. *Acta Numer.* **8** 143–195. Cambridge Univ. Press, Cambridge. MR1819645 https://doi.org/10.1017/S0962492900002919
[50] RAHIMI, A. and RECHT, B. (2008). Random features for large-scale kernel machines. In *Advances in Neural Information Processing Systems* 1177–1184.
[51] ROTSKOFF, G. M. and VANDEN-EIJNDEN, E. (2018). Neural networks as interacting particle systems: Asymptotic convexity of the loss landscape and universal scaling of the approximation error. Available at arXiv:1805.00915.
[52] RUDI, A. and ROSASCO, L. (2017). Generalization properties of learning with random features. In *Advances in Neural Information Processing Systems* 3215–3225.
[53] SIRIGNANO, J. and SPILIOPOULOS, K. (2020). Mean field analysis of neural networks: A central limit theorem. *Stochastic Process. Appl.* **130** 1820–1852. MR4058290 https://doi.org/10.1016/j.spa.2019.06.003
[54] SOUDRY, D., HOFFER, E., NACSON, M. S., GUNASEKAR, S. and SREBRO, N. (2018). The implicit bias of gradient descent on separable data. *J. Mach. Learn. Res.* **19** Paper No. 70, 57. MR3899772
[55] SZEGÖ, G. (1939). *Orthogonal Polynomials*. Amer. Math. Soc., New York. MR0000077
[56] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. MR2724359 https://doi.org/10.1007/b13794
[57] VEMPALA, S. and WILMES, J. (2018). Polynomial convergence of gradient descent for training one-hidden-layer neural networks. Available at arXiv:1805.02677.
[58] YEHUDAI, G. and SHAMIR, O. (2019). On the power and limitations of random features for understanding neural networks. In *Advances in Neural Information Processing Systems* 6594–6604.
[59] ZOU, D., CAO, Y., ZHOU, D. and GU, Q. (2020). Gradient descent optimizes over-parameterized deep ReLU networks. *Mach. Learn.* **109** 467–492. MR4075425 https://doi.org/10.1007/s10994-019-05839-6

The Annals of Statistics
2021, Vol. 49, No. 2, 1055–1080
https://doi.org/10.1214/20-AOS1991

TIME-UNIFORM, NONPARAMETRIC, NONASYMPTOTIC CONFIDENCE SEQUENCES

BY STEVEN R. HOWARD[1,*], AADITYA RAMDAS[2], JON MCAULIFFE[1,†] AND JASJEET SEKHON[3]

[1]*Department of Statistics, University of California, Berkeley,* [*]*stevehoward@berkeley.edu;* [†]*jonmcauliffe@berkeley.edu*
[2]*Departments of Statistics and Machine Learning, Carnegie Mellon University, aramdas@stat.cmu.edu*
[3]*Department of Statistics and Data Science, Yale University, sekhon@berkeley.edu*

A confidence sequence is a sequence of confidence intervals that is uniformly valid over an unbounded time horizon. Our work develops confidence sequences whose widths go to zero, with nonasymptotic coverage guarantees under nonparametric conditions. We draw connections between the Cramér–Chernoff method for exponential concentration, the law of the iterated logarithm (LIL) and the sequential probability ratio test—our confidence sequences are time-uniform extensions of the first; provide tight, nonasymptotic characterizations of the second; and generalize the third to nonparametric settings, including sub-Gaussian and Bernstein conditions, self-normalized processes and matrix martingales. We illustrate the generality of our proof techniques by deriving an empirical-Bernstein bound growing at a LIL rate, as well as a novel upper LIL for the maximum eigenvalue of a sum of random matrices. Finally, we apply our methods to covariance matrix estimation and to estimation of sample average treatment effect under the Neyman–Rubin potential outcomes model.

REFERENCES

[1] ARMITAGE, P., MCPHERSON, C. K. and ROWE, B. C. (1969). Repeated significance tests on accumulating data. *J. Roy. Statist. Soc. Ser. A* **132** 235–244. MR0250405 https://doi.org/10.2307/2343787

[2] ARONOW, P. M. and MIDDLETON, J. A. (2013). A class of unbiased estimators of the average treatment effect in randomized experiments. *J. Causal Inference* **1** 135–154.

[3] AUDIBERT, J.-Y., MUNOS, R. and SZEPESVÁRI, C. (2009). Exploration-exploitation tradeoff using variance estimates in multi-armed bandits. *Theoret. Comput. Sci.* **410** 1876–1902. MR2514714 https://doi.org/10.1016/j.tcs.2009.01.016

[4] AZUMA, K. (1967). Weighted sums of certain dependent random variables. *Tohoku Math. J.* (2) **19** 357–367. MR0221571 https://doi.org/10.2748/tmj/1178243286

[5] BALSUBRAMANI, A. (2014). Sharp finite-time iterated-logarithm martingale concentration. arXiv:1405.2639.

[6] BALSUBRAMANI, A. and RAMDAS, A. (2016). Sequential nonparametric testing with the law of the iterated logarithm. In *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence. UAI'*16 42–51. AUAI Press.

[7] BERCU, B., DELYON, B. and RIO, E. (2015). *Concentration Inequalities for Sums and Martingales. SpringerBriefs in Mathematics*. Springer, Cham. MR3363542 https://doi.org/10.1007/978-3-319-22099-4

[8] BERMAN, R., PEKELIS, L., SCOTT, A. and VAN DEN BULTE, C. (2018). p-hacking and false discovery in A/B testing. Technical Report No. 3204791. SSRN.

[9] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford. With a foreword by Michel Ledoux. MR3185193 https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

[10] CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Stat.* **23** 493–507. MR0057518 https://doi.org/10.1214/aoms/1177729330

MSC2020 subject classifications. Primary 62L12, 62G05; secondary 60G42, 60B20.

Key words and phrases. Confidence sequence, finite LIL bound, empirical-Bernstein bound, sequential probability ratio test, matrix concentration, potential outcomes.

[11] CRAMÉR, H. (1938). Sur un nouveau théorème-limite de la théorie des probabilités. *Actualités Scientifiques* **736**.

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

[13] DARLING, D. A. and ROBBINS, H. (1967). Iterated logarithm inequalities. *Proc. Natl. Acad. Sci. USA* **57** 1188–1192. MR0211441 https://doi.org/10.1073/pnas.57.5.1188

[14] DARLING, D. A. and ROBBINS, H. (1968). Some further remarks on inequalities for sample sums. *Proc. Natl. Acad. Sci. USA* **60** 1175–1182. MR0235604 https://doi.org/10.1073/pnas.60.4.1175

[15] DE LA PEÑA, V. H., KLASS, M. J. and LAI, T. L. (2004). Self-normalized processes: Exponential inequalities, moment bounds and iterated logarithm laws. *Ann. Probab.* **32** 1902–1933. MR2073181 https://doi.org/10.1214/009117904000000397

[16] DE LA PEÑA, V. H., KLASS, M. J. and LAI, T. L. (2007). Pseudo-maximization and self-normalized processes. *Probab. Surv.* **4** 172–192. MR2368950 https://doi.org/10.1214/07-PS119

[17] DE LA PEÑA, V. H., KLASS, M. J. and LAI, T. L. (2009). Theory and applications of multivariate self-normalized processes. *Stochastic Process. Appl.* **119** 4210–4227. MR2565565 https://doi.org/10.1016/j.spa.2009.10.003

[18] DE LA PEÑA, V. H., LAI, T. L. and SHAO, Q.-M. (2009). *Self-Normalized Processes: Limit Theory and Statistical Applications. Probability and Its Applications (New York)*. Springer, Berlin. MR2488094 https://doi.org/10.1007/978-3-540-85636-8

[19] EFRON, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* **58** 403–417. MR0312660 https://doi.org/10.1093/biomet/58.3.403

[20] GARIVIER, A. (2013). Informational confidence bounds for self-normalized averages and applications. In 2013 *IEEE Information Theory Workshop (ITW)* 1–5. IEEE.

[21] GARIVIER, A. and LEONARDI, F. (2011). Context tree selection: A unifying view. *Stochastic Process. Appl.* **121** 2488–2506. MR2832411 https://doi.org/10.1016/j.spa.2011.06.012

[22] GITTENS, A. and TROPP, J. A. (2011). Tail bounds for all eigenvalues of a sum of random matrices. ACM Report 2014-02, Caltech.

[23] GRÜNWALD, P., DE HEIDE, R. and KOOLEN, W. (2019). Safe testing. arXiv:1906.07801.

[24] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. MR0144363

[25] HOWARD, S. R., RAMDAS, A., MCAULIFFE, J. and SEKHON, J. (2020). Time-uniform Chernoff bounds via nonnegative supermartingales. *Probab. Surv.* **17** 257–317. MR4100718 https://doi.org/10.1214/18-PS321

[26] HOWARD, S. R., RAMDAS, A., MCAULIFFE, J. and SEKHON, J. (2021). Supplement to “Time-uniform, nonparametric, nonasymptotic confidence sequences.” https://doi.org/10.1214/20-AOS1991SUPP

[27] IMBENS, G. W. and RUBIN, D. B. (2015). *Causal Inference—for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge Univ. Press, New York. MR3309951 https://doi.org/10.1017/CBO9781139025751

[28] JAMIESON, K. and JAIN, L. (2018). A bandit approach to multiple testing with false discovery control. In *Proceedings of the* 32*nd International Conference on Neural Information Processing Systems* 3664–3674.

[29] JAMIESON, K., MALLOY, M., NOWAK, R. and BUBECK, S. (2014). lil’ UCB: An optimal exploration algorithm for multi-armed bandits. In *Proceedings of the* 27*th Conference on Learning Theory* **35** 423–439.

[30] JAMIESON, K. and NOWAK, R. (2014). Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting. In 48*th Annual Conference on Information Sciences and Systems (CISS)* 1–6.

[31] JENNISON, C. and TURNBULL, B. W. (1984). Repeated confidence intervals for group sequential clinical trials. *Control. Clin. Trials* **5** 33–45.

[32] JENNISON, C. and TURNBULL, B. W. (1989). Interim analyses: The repeated confidence interval approach. *J. Roy. Statist. Soc. Ser. B* **51** 305–361. With discussion and a reply by the authors. MR1017201

[33] JENNISON, C. and TURNBULL, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials*. CRC Press/CRC, Boca Raton, FL. MR1710781

[34] JOHARI, R., KOOMEN, P., PEKELIS, L. and WALSH, D. (2017). Peeking at A/B tests: Why it matters, and what to do about it. 1517–1525. ACM Press.

[35] JOHARI, R., PEKELIS, L. and WALSH, D. J. (2015). Always valid inference: Bringing sequential analysis to A/B testing. arXiv preprint arXiv:1512.04922.

[36] JØRGENSEN, B. (1997). *The Theory of Dispersion Models. Monographs on Statistics and Applied Probability* **76**. CRC Press, London. MR1462891

[37] KAUFMANN, E., CAPPÉ, O. and GARIVIER, A. (2016). On the complexity of best-arm identification in multi-armed bandit models. *J. Mach. Learn. Res.* **17** Paper No. 1, 42. MR3482921

[38] KAUFMANN, E. and KOOLEN, W. (2018). Mixture martingales revisited with applications to sequential tests and confidence intervals. arXiv:1811.11419.

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

[40] KULLDORFF, M., DAVIS, R. L., KOLCZAK, M., LEWIS, E., LIEU, T. and PLATT, R. (2011). A maximized sequential probability ratio test for drug and vaccine safety surveillance. *Sequential Anal*. **30** 58–78. MR2770706 https://doi.org/10.1080/07474946.2011.539924

[41] LAI, T. L. (1976). On confidence sequences. *Ann*. *Statist*. **4** 265–280. MR0395103

[42] LAI, T. L. (1976). Boundary crossing probabilities for sample sums and confidence sequences. *Ann*. *Probab*. **4** 299–312. MR0405578 https://doi.org/10.1214/aop/1176996135

[43] LAI, T. L. (1984). Incorporating scientific, ethical and economic considerations into the design of clinical trials in the pharmaceutical industry: A sequential approach. *Comm*. *Statist*. *Theory Methods* **13** 2355–2368.

[44] LAI, T. L. (1997). On optimal stopping problems in sequential hypothesis testing. *Statist*. *Sinica* **7** 33–51. MR1441143

[45] LAI, T. L. and SIEGMUND, D. (1977). A nonlinear renewal theory with applications to sequential analysis. I. *Ann*. *Statist*. **5** 946–954. MR0445599

[46] LAI, T. L. and SIEGMUND, D. (1979). A nonlinear renewal theory with applications to sequential analysis. II. *Ann*. *Statist*. **7** 60–76. MR0515684

[47] LAN, K. K. G. and DEMETS, D. L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika* **70** 659–663. MR0725380 https://doi.org/10.2307/2336502

[48] LORDEN, G. and POLLAK, M. (2005). Nonanticipating estimation applied to sequential analysis and changepoint detection. *Ann*. *Statist*. **33** 1422–1454. MR2195641 https://doi.org/10.1214/009053605000000183

[49] MALEK, A., KATARIYA, S., CHOW, Y. and GHAVAMZADEH, M. (2017). Sequential multiple hypothesis testing with type I error control. In *Artificial Intelligence and Statistics* 1468–1476.

[50] MAURER, A. and PONTIL, M. (2009). Empirical Bernstein bounds and sample variance penalization. In *Proceedings of the Conference on Learning Theory*.

[51] MCDIARMID, C. (1998). Concentration. In *Probabilistic Methods for Algorithmic Discrete Mathematics*. *Algorithms Combin*. **16** 195–248. Springer, Berlin. MR1678578 https://doi.org/10.1007/978-3-662-12788-9_6

[52] O'BRIEN, P. C. and FLEMING, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics* **35** 549–556.

[53] POCOCK, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika* **64** 191–199.

[54] RAGINSKY, M., SASON, I. et al. (2013). Concentration of measure inequalities in information theory, communications, and coding. *Found*. *Trends Commun*. *Inf. Theory* **10** 1–246.

[55] ROBBINS, H. (1970). Statistical methods related to the law of the iterated logarithm. *Ann*. *Math*. *Stat*. **41** 1397–1409. MR0277063 https://doi.org/10.1214/aoms/1177696786

[56] ROBBINS, H. and SIEGMUND, D. (1968). Iterated logarithm inequalities and related statistical procedures. In *Mathematics of the Decision Sciences*, *Part* 2 (*Seminar, Stanford Calif.*, 1967) 267–279. Amer. Math. Soc., Providence, RI. MR0251777

[57] ROBBINS, H. and SIEGMUND, D. (1969). Probability distributions related to the law of the iterated logarithm. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **62** 11–13. MR0242228 https://doi.org/10.1073/pnas.62.1.11

[58] ROBBINS, H. and SIEGMUND, D. (1970). Boundary crossing probabilities for the Wiener process and sample sums. *Ann*. *Math*. *Stat*. **41** 1410–1429. MR0277059 https://doi.org/10.1214/aoms/1177696787

[59] ROBBINS, H. and SIEGMUND, D. (1972). A class of stopping rules for testing parametric hypotheses. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (*Univ. California*, *Berkeley*, *Calif.*, 1970/1971), *Vol. IV*: *Biology and Health* 37–41. MR0403111

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

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

[62] RUDELSON, M. (1999). Random vectors in the isotropic position. *J*. *Funct*. *Anal*. **164** 60–72. MR1694526 https://doi.org/10.1006/jfan.1998.3384

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

[64] SIEGMUND, D. (1978). Estimation following sequential tests. *Biometrika* **65** 341–349. MR0513934 https://doi.org/10.2307/2335213

[65] 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

[66] SIEGMUND, D. and GREGORY, P. (1980). A sequential clinical trial for testing $p_1 = p_2$. *Ann*. *Statist*. **8** 1219–1228. MR0594639

[67] SPLAWA-NEYMAN, J. (1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statist*. *Sci*. **5** 465–472. Translated from the Polish and edited by D. M. Dabrowska and T. P. Speed. MR1092986

[68] STOUT, W. F. (1970). The Hartman–Wintner law of the iterated logarithm for martingales. *Ann*. *Math*. *Stat*. **41** 2158–2160.

[69] TROPP, J. A. (2011). Freedman's inequality for matrix martingales. *Electron*. *Commun*. *Probab*. **16** 262–270. MR2802042 https://doi.org/10.1214/ECP.v16-1624

[70] TROPP, J. A. (2015). An introduction to matrix concentration inequalities. *Found*. *Trends Mach*. *Learn*. **8** 1–230.

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

[72] VILLE, J. (1939). *Étude Critique de la Notion de Collectif*. NUMDAM. MR3533075

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

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

[75] WHITEHEAD, J. and STRATTON, I. (1983). Group sequential clinical trials with triangular continuation regions. *Biometrics* **39** 227–236. MR0712749 https://doi.org/10.2307/2530822

[76] YANG, F., RAMDAS, A., JAMIESON, K. G. and WAINWRIGHT, M. J. (2017). A framework for Multi-A(rmed)/B(andit) testing with online FDR control. In 31*st Conference on Neural Information Processing Systems*.

[77] ZHAO, S., ZHOU, E., SABHARWAL, A. and ERMON, S. (2016). Adaptive concentration inequalities for sequential decision problems. In 30*th Conference on Neural Information Processing Systems*.

The Annals of Statistics
2021, Vol. 49, No. 2, 1081–1112
https://doi.org/10.1214/20-AOS1994

MINIMAX RATES IN SPARSE, HIGH-DIMENSIONAL CHANGE POINT DETECTION

BY HAOYANG LIU[1,*], CHAO GAO[1,†] AND RICHARD J. SAMWORTH[2]

[1]*Department of Statistics, University of Chicago,* [*]*hyliu@uchicago.edu;* [†]*chaogao@uchicago.edu*
[2]*Statistical Laboratory, Centre for Mathematical Sciences, University of Cambridge, r.samworth@statslab.cam.ac.uk*

We study the detection of a sparse change in a high-dimensional mean vector as a minimax testing problem. Our first main contribution is to derive the exact minimax testing rate across all parameter regimes for n independent, p-variate Gaussian observations. This rate exhibits a phase transition when the sparsity level is of order $\sqrt{p \log\log(8n)}$ and has a very delicate dependence on the sample size: in a certain sparsity regime, it involves a triple iterated logarithmic factor in n. Further, in a dense asymptotic regime, we identify the sharp leading constant, while in the corresponding sparse asymptotic regime, this constant is determined to within a factor of $\sqrt{2}$. Extensions that cover spatial and temporal dependence, primarily in the dense case, are also provided.

REFERENCES

ARIAS-CASTRO, E., CANDÈS, E. J. and DURAND, A. (2011). Detection of an anomalous cluster in a network. *Ann. Statist.* **39** 278–304. MR2797847 https://doi.org/10.1214/10-AOS839

ARIAS-CASTRO, E., DONOHO, D. L. and HUO, X. (2005). Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Inf. Theory* **51** 2402–2425. MR2246369 https://doi.org/10.1109/TIT.2005.850056

ASTON, J. A. D. and KIRCH, C. (2012). Evaluating stationarity via change-point alternatives with applications to fMRI data. *Ann. Appl. Stat.* **6** 1906–1948. MR3058688 https://doi.org/10.1214/12-AOAS565

ASTON, J. A. D. and KIRCH, C. (2018). High dimensional efficiency with applications to change point tests. *Electron. J. Stat.* **12** 1901–1947. MR3815301 https://doi.org/10.1214/18-EJS1442

AUE, A., HÖRMANN, S., HORVÁTH, L. and REIMHERR, M. (2009). Break detection in the covariance structure of multivariate time series models. *Ann. Statist.* **37** 4046–4087. MR2572452 https://doi.org/10.1214/09-AOS707

BAI, J. (2010). Common breaks in means and variances for panel data. *J. Econometrics* **157** 78–92. MR2652280 https://doi.org/10.1016/j.jeconom.2009.10.020

BASSEVILLE, M. and NIKIFOROV, I. V. (1993). *Detection of Abrupt Changes: Theory and Application. Prentice Hall Information and System Sciences Series*. Prentice Hall, Inc., Englewood Cliffs, NJ. MR1210954

BERTHET, Q. and RIGOLLET, P. (2013). Optimal detection of sparse principal components in high dimension. *Ann. Statist.* **41** 1780–1815. MR3127849 https://doi.org/10.1214/13-AOS1127

CHAN, H. P.and WALTHER, G., (2015). Optimal detection of multi-sample aligned sparse signals. *Ann. Statist.* **43** 1865–1895. MR3375870 https://doi.org/10.1214/15-AOS1328

CHEN, Y., WANG, T. and SAMWORTH, R. J. (2020). High-dimensional, multiscale online changepoint detection. arXiv preprint. Available at arXiv:2003.03668.

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

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

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

MSC2020 subject classifications. 62C20.
Key words and phrases. Minimax detection boundary, iterated logarithm, time series.

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

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

ENIKEEVA, F. and HARCHAOUI, Z. (2019). High-dimensional change-point detection under sparse alternatives. *Ann. Statist.* **47** 2051–2079. MR3953444 https://doi.org/10.1214/18-AOS1740

FRICK, K., MUNK, A. and SIELING, H. (2014). Multiscale change point inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 495–580. MR3210728 https://doi.org/10.1111/rssb.12047

GAO, C., HAN, F. and ZHANG, C.-H. (2020). On estimation of isotonic piecewise constant signals. *Ann. Statist.* **48** 629–654. MR4102670 https://doi.org/10.1214/18-AOS1792

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

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. MR0144363

HORVÁTH, L. and HUŠKOVÁ, M. (2012). Change-point detection in panel data. *J. Time Series Anal.* **33** 631–648. MR2944843 https://doi.org/10.1111/j.1467-9892.2012.00796.x

JIRAK, M. (2015). Uniform change point tests in high dimension. *Ann. Statist.* **43** 2451–2483. MR3405600 https://doi.org/10.1214/15-AOS1347

KILLICK, R., FEARNHEAD, P. and ECKLEY, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *J. Amer. Statist. Assoc.* **107** 1590–1598. MR3036418 https://doi.org/10.1080/01621459.2012.737745

KIRCH, C., MUHSAL, B. and OMBAO, H. (2015). Detection of changes in multivariate time series with application to EEG data. *J. Amer. Statist. Assoc.* **110** 1197–1216. MR3420695 https://doi.org/10.1080/01621459.2014.957545

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

LIU, H., GAO, C. and SAMWORTH, R. J. (2021). Supplement to "Minimax rates in sparse, high-dimensional change point detection." https://doi.org/10.1214/20-AOS1994SUPP

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

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

PENG, T., LECKIE, C. and RAMAMOHANARAO, K. (2004). Proactively detecting distributed denial of service attacks using source IP address monitoring. In *Networking* 2004 (N. Mitrou, K. Kontovasilis, G. N. Rouskas, I. Iliadis and L. Merakos, eds.) 771–782. Springer, Berlin.

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 localization in high dimension. arXiv preprint. Available at arXiv:1712.09912.

WANG, D., YU, Y. and RINALDO, A. (2018). Optimal change point detection and localization in sparse dynamic networks. arXiv preprint. Available at arXiv:1809.09602.

WANG, D., YU, Y. and RINALDO, A. (2020). Univariate mean change point detection: Penalization, CUSUM and optimality. *Electron. J. Stat.* **14** 1917–1961. MR4091859 https://doi.org/10.1214/20-EJS1710

XIE, Y. and SIEGMUND, D. (2013). Sequential multi-sensor change-point detection. *Ann. Statist.* **41** 670–692. MR3099117 https://doi.org/10.1214/13-AOS1094

ZHANG, N. R., SIEGMUND, D. O., JI, H. and LI, J. Z. (2010). Detecting simultaneous changepoints in multiple sequences. *Biometrika* **97** 631–645. MR2672488 https://doi.org/10.1093/biomet/asq025

The Annals of Statistics
2021, Vol. 49, No. 2, 1113–1138
https://doi.org/10.1214/20-AOS1995

SPIKED SEPARABLE COVARIANCE MATRICES AND PRINCIPAL COMPONENTS

BY XIUCAI DING[1] AND FAN YANG[2]

[1]*Department of Mathematics, Duke University, xiuca.ding@duke.edu*
[2]*Department of Statistics, University of Pennsylvania, fyang75@wharton.upenn.edu*

We study a class of separable sample covariance matrices of the form $\widetilde{\mathcal{Q}}_1 := \widetilde{A}^{1/2} X \widetilde{B} X^* \widetilde{A}^{1/2}$. Here, $\widetilde{A}$ and $\widetilde{B}$ are positive definite matrices whose spectrums consist of bulk spectrums plus several spikes, that is, larger eigenvalues that are separated from the bulks. Conceptually, we call $\widetilde{\mathcal{Q}}_1$ a *spiked separable covariance matrix model*. On the one hand, this model includes the spiked covariance matrix as a special case with $\widetilde{B} = I$. On the other hand, it allows for more general correlations of datasets. In particular, for spatio-temporal dataset, $\widetilde{A}$ and $\widetilde{B}$ represent the spatial and temporal correlations, respectively.

In this paper, we study the outlier eigenvalues and eigenvectors, that is, the principal components, of the spiked separable covariance model $\widetilde{\mathcal{Q}}_1$. We prove the convergence of the outlier eigenvalues $\widetilde{\lambda}_i$ and the generalized components (i.e., $\langle \mathbf{v}, \widetilde{\boldsymbol{\xi}}_i \rangle$ for any deterministic vector $\mathbf{v}$) of the outlier eigenvectors $\widetilde{\boldsymbol{\xi}}_i$ with optimal convergence rates. Moreover, we also prove the delocalization of the nonoutlier eigenvectors. We state our results in full generality, in the sense that they also hold near the so-called BBP transition and for degenerate outliers. Our results highlight both the similarity and difference between the spiked separable covariance matrix model and the spiked covariance matrix model in (*Probab. Theory Related Fields* **164** (2016) 459–552). In particular, we show that the spikes of both $\widetilde{A}$ and $\widetilde{B}$ will cause outliers of the eigenvalue spectrum, and the eigenvectors can help to select the outliers that correspond to the spikes of $\widetilde{A}$ (or $\widetilde{B}$).

REFERENCES

[1] 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

[2] 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

[3] 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

[4] 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

[5] 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** 33. MR3183577 https://doi.org/10.1214/ejp.v19-3054

[6] 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

[7] BOURGADE, P., YAU, H.-T. and YIN, J. (2014). Local circular law for random matrices. *Probab. Theory Related Fields* **159** 545–595. MR3230002 https://doi.org/10.1007/s00440-013-0514-z

MSC2020 subject classifications. Primary 15B52, 62E20; secondary 62H99.
Key words and phrases. Spiked separable covariance matrices, principal components, BBP transition, local laws.

[8] COUILLET, R. and HACHEM, W. (2014). Analysis of the limiting spectral measure of large random matrices of the separable covariance type. *Random Matrices Theory Appl.* **3** 1450016. MR3279621 https://doi.org/10.1142/S2010326314500166
[9] DING, X. (2019). Singular vector distribution of sample covariance matrices. *Adv. Appl. Probab.* **51** 236–267. MR3984017 https://doi.org/10.1017/apr.2019.10
[10] 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
[11] DING, X. (2020). Spiked sample covariance matrices with possibly multiple bulk components. *Random Matrices Theory Appl.*. (In press).
[12] DING, X. and YANG, F. (2021). Supplement to "Spiked separable covariance matrices and principal components." https://doi.org/10.1214/20-AOS1995SUPP
[13] DING, X. and YANG, F. (2018). A necessary and sufficient condition for edge universality at the largest singular values of covariance matrices. *Ann. Appl. Probab.* **28** 1679–1738. MR3809475 https://doi.org/10.1214/17-AAP1341
[14] DOBRIBAN, E. and OWEN, A. B. (2019). Deterministic parallel analysis: An improved method for selecting factors and principal components. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 163–183. MR3904784
[15] DONOHO, D., GAVISH, M. and JOHNSTONE, I. (2018). Optimal shrinkage of eigenvalues in the spiked covariance model. *Ann. Statist.* **46** 1742–1778. MR3819116 https://doi.org/10.1214/17-AOS1601
[16] 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
[17] EL KAROUI, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Ann. Statist.* **36** 2757–2790. MR2485012 https://doi.org/10.1214/07-AOS581
[18] 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
[19] ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2013). Delocalization and diffusion profile for random band matrices. *Comm. Math. Phys.* **323** 367–416. MR3085669 https://doi.org/10.1007/s00220-013-1773-3
[20] ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2013). The local semicircle law for a general class of random matrices. *Electron. J. Probab.* **18** 59. MR3068390 https://doi.org/10.1214/EJP.v18-2473
[21] GENTON, M. G. (2007). Separable approximations of space-time covariance matrices. *Environmetrics* **18** 681–695. MR2408938 https://doi.org/10.1002/env.854
[22] GNEITING, T. (2002). Nonseparable, stationary covariance functions for space-time data. *J. Amer. Statist. Assoc.* **97** 590–600. MR1941475 https://doi.org/10.1198/016214502760047113
[23] HACHEM, W., LOUBATON, P. and NAJIM, J. (2007). Deterministic equivalents for certain functionals of large random matrices. *Ann. Appl. Probab.* **17** 875–930. MR2326235 https://doi.org/10.1214/105051606000000925
[24] 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
[25] KE, Z. T. (2016). Detecting rare and weak spikes in large covariance matrices. arXiv preprint. Available at arXiv:1609.00883.
[26] KERMOAL, J. P., SCHUMACHER, L., PEDERSEN, K. I., MOGENSEN, P. E. and FREDERIKSEN, F. (2002). A stochastic MIMO radio channel model with experimental validation. *IEEE J. Sel. Areas Commun.* **20** 1211–1226.
[27] 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
[28] 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
[29] KONG, W. and VALIANT, G. (2017). Spectrum estimation from samples. *Ann. Statist.* **45** 2218–2247. MR3718167 https://doi.org/10.1214/16-AOS1525
[30] KYRIAKIDIS, P. C. and JOURNEL, A. G. (1999). Geostatistical space-time models: A review. *Math. Geol.* **31** 651–684. MR1694654 https://doi.org/10.1023/A:1007528426688
[31] LEDOIT, O. and PÉCHÉ, S. (2011). Eigenvectors of some large sample covariance matrix ensembles. *Probab. Theory Related Fields* **151** 233–264. MR2834718 https://doi.org/10.1007/s00440-010-0298-3
[32] LEDOIT, O. and WOLF, M. (2015). Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. *J. Multivariate Anal.* **139** 360–384. MR3349498 https://doi.org/10.1016/j.jmva.2015.04.006
[33] LEDOIT, O. and WOLF, M. (2017). Numerical implementation of the QuEST function. *Comput. Statist. Data Anal.* **115** 199–223. MR3683138 https://doi.org/10.1016/j.csda.2017.06.004

[34] LEE, J. O. and SCHNELLI, K. (2016). Tracy–Widom distribution for the largest eigenvalue of real sample covariance matrices with general population. *Ann. Appl. Probab.* **26** 3786–3839. MR3582818 https://doi.org/10.1214/16-AAP1193

[35] LEEB, W. (2019). Matrix denoising for weighted loss functions and heterogeneous signals. arXiv preprint. Available at arXiv:1902.09474.

[36] LI, B., GENTON, M. G. and SHERMAN, M. (2008). Testing the covariance structure of multivariate random fields. *Biometrika* **95** 813–829. MR2461213 https://doi.org/10.1093/biomet/asn053

[37] LU, N. and ZIMMERMAN, D. L. (2005). The likelihood ratio test for a separable covariance matrix. *Statist. Probab. Lett.* **73** 449–457. MR2187860 https://doi.org/10.1016/j.spl.2005.04.020

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

[39] NADAKUDITI, R. R. (2014). OptShrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage. *IEEE Trans. Inf. Theory* **60** 3002–3018. MR3200641 https://doi.org/10.1109/TIT.2014.2311661

[40] NADLER, B. (2010). Nonparametric detection of signals by information theoretic criteria: Performance analysis and an improved estimator. *IEEE Trans. Signal Process.* **58** 2746–2756. MR2789420 https://doi.org/10.1109/TSP.2010.2042481

[41] 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

[42] ONATSKI, A. (2008). The Tracy–Widom limit for the largest eigenvalues of singular complex Wishart matrices. *Ann. Appl. Probab.* **18** 470–490. MR2398763 https://doi.org/10.1214/07-AAP454

[43] ONATSKI, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Rev. Econ. Stat.* **92** 1004–1016.

[44] ONATSKI, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *J. Econometrics* **168** 244–258. MR2923766 https://doi.org/10.1016/j.jeconom.2012.01.034

[45] PASSEMIER, D. and YAO, J. (2014). Estimation of the number of spikes, possibly equal, in the high-dimensional case. *J. Multivariate Anal.* **127** 173–183. MR3188885 https://doi.org/10.1016/j.jmva.2014.02.017

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

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

[48] PAUL, D. and SILVERSTEIN, J. W. (2009). No eigenvalues outside the support of the limiting empirical spectral distribution of a separable covariance matrix. *J. Multivariate Anal.* **100** 37–57. MR2460475 https://doi.org/10.1016/j.jmva.2008.03.010

[49] 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

[50] SKUP, M. (2010). Longitudinal fMRI analysis: A review of methods. *Stat. Interface* **3** 235–252. MR2659514 https://doi.org/10.4310/SII.2010.v3.n2.a10

[51] TRACY, C. A. and WIDOM, H. (1994). Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.* **159** 151–174. MR1257246

[52] TRACY, C. A. and WIDOM, H. (1996). On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.* **177** 727–754. MR1385083

[53] TULINO, A. and VERDÚ, S. (2004). *Random Matrix Theory and Wireless Communications. Foundations and Trends in Communications and Information Theory*. Publishers Inc..

[54] WANG, L. and PAUL, D. (2014). Limiting spectral distribution of renormalized separable sample covariance matrices when $p/n \to 0$. *J. Multivariate Anal.* **126** 25–52. MR3173080 https://doi.org/10.1016/j.jmva.2013.12.015

[55] WERNER, K., JANSSON, M. and STOICA, P. (2008). On estimation of covariance matrices with Kronecker product structure. *IEEE Trans. Signal Process.* **56** 478–491. MR2445531 https://doi.org/10.1109/TSP.2007.907834

[56] XI, H., YANG, F. and YIN, J. (2020). Convergence of eigenvector empirical spectral distribution of sample covariance matrices. *Ann. Statist.* **48** 953–982. MR4102683 https://doi.org/10.1214/19-AOS1832

[57] YANG, F. Local laws of random matrices and their applications Ph.D. thesis, UCLA.

[58] YANG, F. (2019). Edge universality of separable covariance matrices. *Electron. J. Probab.* **24** 123. MR4029426 https://doi.org/10.1214/19-ejp381

[59] YEO, J. and PAPANICOLAOU, G. (2016). Random matrix approach to estimation of high-dimensional factor models. arXiv preprint. Available at arXiv:1611.05571.

[60] ZHANG, L. Spectral Analysis of Large Dimensional Random Matrices Ph.D. thesis, National University of Singapore.

The Annals of Statistics
2021, Vol. 49, No. 2, 1139–1165
https://doi.org/10.1214/20-AOS1996

DISTRIBUTION AND QUANTILE FUNCTIONS, RANKS AND SIGNS IN DIMENSION d: A MEASURE TRANSPORTATION APPROACH

BY MARC HALLIN[1], EUSTASIO DEL BARRIO[2,*], JUAN CUESTA-ALBERTOS[3] AND CARLOS MATRÁN[2,†]

[1]*ECARES and Département de Mathématique, Université libre de Bruxelles, Belgium,* mhallin@ulb.ac.be
[2]*IMUVA and Departamento de Estadística e Investigación Operativa, Universidad de Valladolid, Spain,* [*]*tasio@eio.uva.es;* [†]*carlos.matran@uva.es*
[3]*Departamento de Matemáticas, Estadística y Computación, Facultad de Ciencias, Universidad de Cantabria, Spain,* *juan.cuesta@unican.es*

Unlike the real line, the real space $\mathbb{R}^d$, for $d \geq 2$, is not canonically ordered. As a consequence, such fundamental univariate concepts as quantile and distribution functions and their empirical counterparts, involving ranks and signs, do not canonically extend to the multivariate context. Palliating that lack of a canonical ordering has been an open problem for more than half a century, generating an abundant literature and motivating, among others, the development of statistical depth and copula-based methods. We show that, unlike the many definitions proposed in the literature, the measure transportation-based ranks and signs introduced in Chernozhukov, Galichon, Hallin and Henry (*Ann. Statist.* **45** (2017) 223–256) enjoy all the properties that make univariate ranks a successful tool for semiparametric inference. Related with those ranks, we propose a new *center-outward* definition of multivariate distribution and quantile functions, along with their empirical counterparts, for which we establish a Glivenko–Cantelli result. Our approach is based on McCann (*Duke Math. J.* **80** (1995) 309–323) and our results do not require any moment assumptions. The resulting ranks and signs are shown to be strictly distribution-free and essentially maximal ancillary in the sense of Basu (*Sankhyā* **21** (1959) 247–256) which, in semiparametric models involving noise with unspecified density, can be interpreted as a finite-sample form of semiparametric efficiency. Although constituting a sufficient summary of the sample, empirical center-outward distribution functions are defined at observed values only. A continuous extension to the entire d-dimensional space, yielding smooth empirical quantile contours and sign curves while preserving the essential monotonicity and Glivenko–Cantelli features of the concept, is provided. A numerical study of the resulting empirical quantile contours is conducted.

REFERENCES

AHUJA, R. K., MAGNANTI, T. L. and ORLIN, J. B. (1993). *Network Flows*: *Theory*, *Algorithms*, *and Applications*. Prentice Hall, Englewood Cliffs, NJ. MR1205775

BASU, D. (1955). On statistics independent of a complete sufficient statistic. *Sankhyā* **15** 377–380. MR0074745 https://doi.org/10.1007/978-1-4419-5825-9_14

BASU, D. (1959). The family of ancillary statistics. *Sankhyā* **21** 247–256. MR0110115 https://doi.org/10.1007/978-1-4419-5825-9_18

BEIRLANT, J., DEL BARRIO, E., BUITENDAG, S., HALLIN, M. and KAMPER, FR. (2020). Center-outward quantiles and the measurement of multivariate risk. *Insurance Math*. *Econom*. **95** 79–100.

BELLONI, A. and WINKLER, R. L. (2011). On multivariate quantiles under partial orders. *Ann*. *Statist*. **39** 1125–1179. MR2816350 https://doi.org/10.1214/10-AOS863

MSC2020 subject classifications. Primary 62G30; secondary 62B05.

Key words and phrases. Multivariate distribution function, multivariate quantiles, multivariate ranks, multivariate signs, Glivenko–Cantelli theorem, Basu theorem, distribution-freeness, ancillarity, cyclical monotonicity.

BERTSEKAS, D. P. (1991). *Linear Network Optimization*: *Algorithms and Codes*. MIT Press, Cambridge, MA. MR1201582

BICKEL, P. J. (1965). On some asymptotically nonparametric competitors of Hotelling's T^2. *Ann*. *Math*. *Stat*. **36** 160–173; correction, ibid. 1583. MR0181052

BISWAS, M., MUKHOPADHYAY, M. and GHOSH, A. K. (2014). A distribution-free two-sample run test applicable to high-dimensional data. *Biometrika* **101** 913–926. MR3286925 https://doi.org/10.1093/biomet/asu045

BOECKEL, M., SPOKOINY, V. and SUVORIKOVA, A. (2018). Multivariate Brenier cumulative distribution functions and their application to nonparametric testing. Available at arXiv:1809.04090.

BURKARD, R., DELL'AMICO, M. and MARTELLO, S. (2009). *Assignment Problems*. SIAM, Philadelphia, PA. MR2488749 https://doi.org/10.1137/1.9780898717754

CHAKRABORTY, B. and CHAUDHURI, P. (1996). On a transformation and re-transformation technique for constructing an affine equivariant multivariate median. *Proc*. *Amer*. *Math*. *Soc*. **124** 2539–2547. MR1363452 https://doi.org/10.1090/S0002-9939-96-03657-X

CHAKRABORTY, B. and CHAUDHURI, P. (1998). On an adaptive transformation-retransformation estimate of multivariate location. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **60** 145–157. MR1625644 https://doi.org/10.1111/1467-9868.00114

CHAKRABORTY, A. and CHAUDHURI, P. (2014). The spatial distribution in infinite dimensional spaces and related quantiles and depths. *Ann*. *Statist*. **42** 1203–1231. MR3224286 https://doi.org/10.1214/14-AOS1226

CARLIER, G., CHERNOZHUKOV, V. and GALICHON, A. (2016). Vector quantile regression: an optimal transport approach. *Ann*. *Statist*. **44** 1165–1192.

CHAKRABORTY, A. and CHAUDHURI, P. (2017). Tests for high-dimensional data based on means, spatial signs and spatial ranks. *Ann*. *Statist*. **45** 771–799. MR3650400 https://doi.org/10.1214/16-AOS1467

CHAUDHURI, P. (1996). On a geometric notion of quantiles for multivariate data. *J*. *Amer*. *Statist*. *Assoc*. **91** 862–872. MR1395753 https://doi.org/10.2307/2291681

CHAUDHURI, P. and SENGUPTA, D. (1993). Sign tests in multidimension: Inference based on the geometry of the data cloud. *J*. *Amer*. *Statist*. *Assoc*. **88** 1363–1370. MR1245371

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

CHOI, K. and MARDEN, J. (1997). An approach to multivariate rank tests in multivariate analysis of variance. *J*. *Amer*. *Statist*. *Assoc*. **92** 1581–1590. MR1615267 https://doi.org/10.2307/2965429

CUESTA-ALBERTOS, J. A., MATRÁN, C. and TUERO-DÍAZ, A. (1997). Optimal transportation plans and convergence in distribution. *J*. *Multivariate Anal*. **60** 72–83. MR1441460 https://doi.org/10.1006/jmva.1996.1627

DE VALK, C. and SEGERS, J. (2018). Stability and tail limits of transport-based quantile contours. Available at arXiv:1811.12061.

DEB, N. and SEN, B. (2019). Multivariate rank-based distribution-free nonparametric testing using measure transportation. Available at arXiv:1909.08733.

DEL BARRIO, E., GONZÁLEZ-SANZ, A. and HALLIN, M. (2020). A note on the regularity of center-outward distribution and quantile functions. *J*. *Multivariate Anal*. **180** S0047259X20302529.

DEL BARRIO, E. and LOUBES, J.-M. (2019). Central limit theorems for empirical transportation cost in general dimension. *Ann*. *Probab*. **47** 926–951. MR3916938 https://doi.org/10.1214/18-AOP1275

DICK, J. and PILLICHSHAMMER, F. (2014). Discrepancy theory and quasi-Monte Carlo integration. In *A Panorama of Discrepancy Theory* (W. Chen, A. Srivastava and G. Travaglini, eds.). *Lecture Notes in Math*. **2107** 539–619. Springer, Cham. MR3330331 https://doi.org/10.1007/978-3-319-04696-9_9

FIGALLI, A. (2017). *The Monge–Ampère Equation and Its Applications*. *Zurich Lectures in Advanced Mathematics*. European Mathematical Society (EMS), Zürich. MR3617963 https://doi.org/10.4171/170

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

GENEST, C. and RIVEST, L.-P. (2001). On the multivariate probability integral transformation. *Statist*. *Probab*. *Lett*. **53** 391–399. MR1856163 https://doi.org/10.1016/S0167-7152(01)00047-5

GHOSAL, P. and SEN, B. (2019). Multivariate ranks and quantiles using optimal transportation and applications to goodness-of-fit testing. Available at arXiv:1905.05340.

GUSHCHIN, A. A. and BORZYKH, D. A. (2017). Integrated quantile functions: Properties and applications. *Mod*. *Stoch*. *Theory Appl*. **4** 285–314. MR3739012 https://doi.org/10.15559/17-vmsta88

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

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

HALLIN, M., HLUBINKA, D. and HUDECOVÁ, S. (2020). Fully distribution-free center-outward rank tests for multiple-output regression and MANOVA. Available at arXiv:2007.15496.

HALLIN, M., INGENBLEEK, J.-F. and PURI, M. L. (1989). Asymptotically most powerful rank tests for multivariate randomness against serial dependence. *J. Multivariate Anal.* **30** 34–71. MR1003708 https://doi.org/10.1016/0047-259X(89)90087-0

HALLIN, M., MORDANT, G. and SEGERS, J. (2019). Multivariate goodness-of-fit tests based on Wasserstein distance. Available at arXiv:2003.06684.

HALLIN, M., LA VECCHIA, D. and LIU, H. (2020). Center-outward R-estimation for semiparametric VARMA models. *J. Amer. Statist. Assoc.* https://doi.org/10.1080/01621459.2020.1832501

HALLIN, M., LA VECCHIA, D. and LIU, H. (2020). Rank-based testing for semiparametric VAR models: a measure transportation approach. Available at arXiv:2011.06062.

HALLIN, M., LU, Z., PAINDAVEINE, D. and ŠIMAN, M. (2015). Local bilinear multiple-output quantile regression. *Bernoulli* **21** 1435–1466.

HALLIN, M. and MEHTA, C. (2015). R-estimation for asymmetric independent component analysis. *J. Amer. Statist. Assoc.* **110** 218–232. MR3338498 https://doi.org/10.1080/01621459.2014.909316

HALLIN, M., OJA, H. and PAINDAVEINE, D. (2006). Semiparametrically efficient rank-based inference for shape. II. Optimal R-estimation of shape. *Ann. Statist.* **34** 2757–2789. MR2329466 https://doi.org/10.1214/009053606000000948

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

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

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

HALLIN, M. and PAINDAVEINE, D. (2004a). Rank-based optimal tests of the adequacy of an elliptic VARMA model. *Ann. Statist.* **32** 2642–2678. MR2153998 https://doi.org/10.1214/009053604000000724

HALLIN, M. and PAINDAVEINE, D. (2004b). Multivariate signed-rank tests in vector autoregressive order identification. *Statist. Sci.* **19** 697–711. MR2185591 https://doi.org/10.1214/088342304000000602

HALLIN, M. and PAINDAVEINE, D. (2005). Affine-invariant aligned rank tests for the multivariate general linear model with VARMA errors. *J. Multivariate Anal.* **93** 122–163. MR2119768 https://doi.org/10.1016/j.jmva.2004.01.005

HALLIN, M. and PAINDAVEINE, D. (2006a). Semiparametrically efficient rank-based inference for shape. I. Optimal rank-based tests for sphericity. *Ann. Statist.* **34** 2707–2756. MR2329465 https://doi.org/10.1214/009053606000000731

HALLIN, M. and PAINDAVEINE, D. (2006b). Parametric and semiparametric inference for shape: The role of the scale functional. *Statist. Decisions* **24** 327–350. MR2305111 https://doi.org/10.1524/stnd.2006.24.3.327

HALLIN, M. and PAINDAVEINE, D. (2008). Optimal rank-based tests for homogeneity of scatter. *Ann. Statist.* **36** 1261–1298. MR2418657 https://doi.org/10.1214/07-AOS508

HALLIN, M., PAINDAVEINE, D. and ŠIMAN, M. (2010). Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth. *Ann. Statist.* **38** 635–669. MR2604670 https://doi.org/10.1214/09-AOS723

HALLIN, M., PAINDAVEINE, D. and VERDEBOUT, T. (2010). Optimal rank-based testing for principal components. *Ann. Statist.* **38** 3245–3299. MR2766852 https://doi.org/10.1214/10-AOS810

HALLIN, M., PAINDAVEINE, D. and VERDEBOUT, T. (2013). Optimal rank-based tests for common principal components. *Bernoulli* **19** 2524–2556. MR3160563 https://doi.org/10.3150/12-BEJ461

HALLIN, M., PAINDAVEINE, D. and VERDEBOUT, T. (2014). Efficient R-estimation of principal and common principal components. *J. Amer. Statist. Assoc.* **109** 1071–1083. MR3265681 https://doi.org/10.1080/01621459.2014.880057

HALLIN, M. and WERKER, B. J. M. (1999). Optimal testing for semi-parametric AR models—From Gaussian Lagrange multipliers to autoregression rank scores and adaptive tests. In *Asymptotics*, *Nonparametrics*, *and Time Series* (S. Ghosh, ed.). *Statist. Textbooks Monogr.* **158** 295–350. Dekker, New York. MR1724702

HALLIN, M. and WERKER, B. J. M. (2003). Semi-parametric efficiency, distribution-freeness and invariance. *Bernoulli* **9** 137–165. MR1963675 https://doi.org/10.3150/bj/1068129013

HALLIN, M., DEL BARRIO, E., CUESTA-ALBERTOS, J. A. and MATRÁN, C. (2021). Supplement to "Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach." https://doi.org/10.1214/20-AOS1996SUPP

HAMEL, A. H. and KOSTNER, D. (2018). Cone distribution functions and quantiles for multivariate random variables. *J. Multivariate Anal.* **167** 97–113. MR3830636 https://doi.org/10.1016/j.jmva.2018.04.004

HE, X. and WANG, G. (1997). Convergence of depth contours for multivariate datasets. *Ann. Statist.* **25** 495–504. MR1439311 https://doi.org/10.1214/aos/1031833661

HODGES, J. L. JR. (1955). A bivariate sign test. *Ann. Math. Stat.* **26** 523–527. MR0070921 https://doi.org/10.1214/aoms/1177728498

ILMONEN, P. and PAINDAVEINE, D. (2011). Semiparametrically efficient inference based on signed ranks in symmetric independent component models. *Ann. Statist.* **39** 2448–2476. MR2906874 https://doi.org/10.1214/11-AOS906

JUDD, K. L. (1998). *Numerical Methods in Economics*. MIT Press, Cambridge, MA. MR1646331

KARP, R. M. (1978). A characterization of the minimum cycle mean in a digraph. *Discrete Math.* **23** 309–311. MR0523080 https://doi.org/10.1016/0012-365X(78)90011-0

KOLTCHINSKII, V. I. (1997). M-estimation, convexity and quantiles. *Ann. Statist.* **25** 435–477. MR1439309 https://doi.org/10.1214/aos/1031833659

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

LEHMANN, E. L. and SCHOLZ, F.-W. (1992). Ancillarity. In *Current Issues in Statistical Inference: Essays in Honor of D. Basu. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **17** 32–51. IMS, Hayward, CA. MR1194408 https://doi.org/10.1214/lnms/1215458837

LIU, R. Y. (1992). Data depth and multivariate rank tests. In *L_1-Statistical Analysis and Related Methods* (*Neuchâtel*, 1992) 279–294. North-Holland, Amsterdam. MR1214839

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

LÓPEZ-PINTADO, S. and ROMO, J. (2009). On the concept of depth for functional data. *J. Amer. Statist. Assoc.* **104** 718–734. MR2541590 https://doi.org/10.1198/jasa.2009.0108

MARDEN, J. I. (1999). Multivariate rank tests. In *Multivariate Analysis, Design of Experiments, and Survey Sampling. Statist. Textbooks Monogr.* **159** 401–432. Dekker, New York. MR1719102

MCCANN, R. J. (1995). Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.* **80** 309–323. MR1369395 https://doi.org/10.1215/S0012-7094-95-08013-2

MCKEAGUE, I. W., LÓPEZ-PINTADO, S., HALLIN, M. and ŠIMAN, M. (2011). Analyzing growth trajectories. *J. Dev. Orig. Health Dis.* **2** 322–329.

MÖTTÖNEN, J. and OJA, H. (1995). Multivariate spatial sign and rank methods. *J. Nonparametr. Stat.* **5** 201–213. MR1346895 https://doi.org/10.1080/10485259508832643

MÖTTÖNEN, J., OJA, H. and TIENARI, J. (1997). On the efficiency of multivariate spatial sign and rank tests. *Ann. Statist.* **25** 542–552. MR1439313 https://doi.org/10.1214/aos/1031833663

NIEDERREITER, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods. CBMS–NSF Regional Conference Series in Applied Mathematics* **63**. SIAM, Philadelphia, PA. MR1172997 https://doi.org/10.1137/1.9781611970081

NORDHAUSEN, K., OJA, H. and PAINDAVEINE, D. (2009). Signed-rank tests for location in the symmetric independent component model. *J. Multivariate Anal.* **100** 821–834. MR2498716 https://doi.org/10.1016/j.jmva.2008.08.004

OJA, H. (1999). Affine invariant multivariate sign and rank tests and corresponding estimates: A review. *Scand. J. Stat.* **26** 319–343. MR1712063 https://doi.org/10.1111/1467-9469.00152

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

OJA, H. and PAINDAVEINE, D. (2005). Optimal signed-rank tests based on hyperplanes. *J. Statist. Plann. Inference* **135** 300–323. MR2200471 https://doi.org/10.1016/j.jspi.2004.04.022

OJA, H. and RANDLES, R. H. (2004). Multivariate nonparametric tests. *Statist. Sci.* **19** 598–605. MR2185581 https://doi.org/10.1214/088342304000000558

PATEIRO-LÓPEZ, B. and RODRÍGUEZ-CASAL, A. (2010). Generalizing the convex hull of a sample: The R package α-hull. *J. Stat. Softw.* **34**.

PEYRÉ, G. and CUTURI, M. (2019). Computational optimal transport. *Found. Trends Mach. Learn.* **11** 355–607.

PURI, M. L. and SEN, P. K. (1966). On a class of multivariate multisample rank-order tests. *Sankhyā Ser. A* **28** 353–376. MR0231488

PURI, M. L. and SEN, P. K. (1969). A class of rank order tests for a general linear hypothesis. *Ann. Math. Stat.* **40** 1325–1343. MR0245155 https://doi.org/10.1214/aoms/1177697505

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

RANDLES, R. H. (1989). A distribution-free multivariate sign test based on interdirections. *J. Amer. Statist. Assoc.* **84** 1045–1050. MR1134492

ROCKAFELLAR, R. T. (1966). Characterization of the subdifferentials of convex functions. *Pacific J. Math.* **17** 497–510. MR0193549

ROCKAFELLAR, R. T. and WETS, R. J.-B. (1998). *Variational Analysis*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **317**. Springer, Berlin. MR1491362 https://doi.org/10.1007/978-3-642-02431-3

SANTNER, T. J., WILLIAMS, B. J. and NOTZ, W. I. (2003). *The Design and Analysis of Computer Experiments*. *Springer Series in Statistics*. Springer, New York. MR2160708 https://doi.org/10.1007/978-1-4757-3799-8

SEGERS, J., VAN DEN AKKER, R. and WERKER, B. J. M. (2014). Semiparametric Gaussian copula models: Geometry and efficient rank-based estimation. *Ann*. *Statist*. **42** 1911–1940. MR3262472 https://doi.org/10.1214/14-AOS1244

SEN, P. K. and PURI, M. L. (1967). On the theory of rank order tests for location in the multivariate one sample problem. *Ann*. *Math*. *Stat*. **38** 1216–1228. MR0212954 https://doi.org/10.1214/aoms/1177698790

SERFLING, R. (2002). Quantile functions for multivariate analysis: Approaches and applications. *Stat*. *Neerl*. **56** 214–232. MR1916321 https://doi.org/10.1111/1467-9574.00195

SHI, H., DRTON, M. and HAN, F. (2019). Distribution-free consistent independence tests via Hallin's multivariate ranks. Available at arXiv:1909.10024.

UM, Y. and RANDLES, R. H. (1998). Nonparametric tests for the multivariate multi-sample location problem. *Statist*. *Sinica* **8** 801–812. MR1651509

YOSIDA, K. (1965). *Functional Analysis*. *Die Grundlehren der Mathematischen Wissenschaften* **123**. Academic Press, New York; Springer, Berlin. MR0180824

ZUO, Y. (2018). On general notions of depth for regression. Available at arXiv:1805.02046v1.

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

ZUO, Y. and SERFLING, R. (2000). General notions of statistical depth function. *Ann*. *Statist*. **28** 461–482. MR1790005 https://doi.org/10.1214/aos/1016218226

The Annals of Statistics
2021, Vol. 49, No. 2, 1166–1194
https://doi.org/10.1214/20-AOS1997

MINIMAX ESTIMATION OF SMOOTH OPTIMAL TRANSPORT MAPS

BY JAN-CHRISTIAN HÜTTER[1] AND PHILIPPE RIGOLLET[2]

[1]*Broad Institute of MIT, jhuetter@broadinstitute.org*
[2]*Department of Mathematics, Massachusetts Institute of Technology, rigollet@math.mit.edu*

Brenier's theorem is a cornerstone of optimal transport that guarantees the existence of an optimal transport map T between two probability distributions P and Q over $\mathbb{R}^d$ under certain regularity conditions. The main goal of this work is to establish the minimax estimation rates for such a transport map from data sampled from P and Q under additional smoothness assumptions on T. To achieve this goal, we develop an estimator based on the minimization of an empirical version of the semidual optimal transport problem, restricted to truncated wavelet expansions. This estimator is shown to achieve near minimax optimality using new stability arguments for the semidual and a complementary minimax lower bound. Furthermore, we provide numerical experiments on synthetic data supporting our theoretical findings and highlighting the practical benefits of smoothness regularization. These are the first minimax estimation rates for transport maps in general dimension.

REFERENCES

[1] AHIDAR-COUTRIX, A., LE GOUIC, T. and PARIS, Q. (2020). Convergence rates for empirical barycenters in metric spaces: Curvature, convexity and extendable geodesics. *Probab. Theory Related Fields* **177** 323–368. MR4095017 https://doi.org/10.1007/s00440-019-00950-0

[2] ALAUX, J., GRAVE, E., CUTURI, M. and JOULIN, A. (2018). Unsupervised hyperalignment for multilingual word embeddings. Preprint. Available at arXiv:1811.01124.

[3] ALTSCHULER, J., BACH, F., RUDI, A. and NILES-WEED, J. (2019). Massively scalable Sinkhorn distances via the Nyström method. In *Advances in Neural Information Processing Systems* 4429–4439.

[4] ALTSCHULER, J., WEED, J. and RIGOLLET, P. (2017). Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration. In *Advances in Neural Information Processing Systems* 1964–1974.

[5] ALVAREZ-MELIS, D., JAAKKOLA, T. S. and JEGELKA, S. (2018). Structured optimal transport. In *International Conference on Artificial Intelligence and Statistics*, *AISTATS* 2018, 9–11 *April* 2018, *Playa Blanca*, *Lanzarote*, *Canary Islands*, *Spain* 1771–1780.

[6] ANDERES, E., BORGWARDT, S. and MILLER, J. (2016). Discrete Wasserstein barycenters: Optimal transport for discrete data. *Math. Methods Oper. Res.* **84** 389–409. MR3566931 https://doi.org/10.1007/s00186-016-0549-x

[7] ARJOVSKY, M., CHINTALA, S. and BOTTOU, L. (2017). Wasserstein generative adversarial networks. In *International Conference on Machine Learning* 214–223.

[8] BIGOT, J., CAZELLES, E. and PAPADAKIS, N. (2019). Central limit theorems for entropy-regularized optimal transport on finite spaces and statistical applications. *Electron. J. Stat.* **13** 5120–5150. MR4041704 https://doi.org/10.1214/19-EJS1637

[9] 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

[10] BLONDEL, M., SEGUY, V. and ROLET, A. (2018). Smooth and sparse optimal transport. In *International Conference on Artificial Intelligence and Statistics* 880–889.

[11] BRENIER, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.* **44** 375–417. MR1100809 https://doi.org/10.1002/cpa.3160440402

[12] CAFFARELLI, L. A. (1992). The regularity of mappings with a convex potential. *J. Amer. Math. Soc.* **5** 99–104. MR1124980 https://doi.org/10.2307/2152752

MSC2020 subject classifications. 62G05.
Key words and phrases. Optimal transport, nonparametric estimation, minimax rates, wavelet estimator.

[13] CAFFARELLI, L. A. (1992). Boundary regularity of maps with convex potentials. *Comm. Pure Appl. Math.* **45** 1141–1151. MR1177479 https://doi.org/10.1002/cpa.3160450905

[14] CAFFARELLI, L. A. (1996). Boundary regularity of maps with convex potentials. II. *Ann. of Math.* (2) **144** 453–496. MR1426885 https://doi.org/10.2307/2118564

[15] CAILLERIE, C., CHAZAL, F., DEDECKER, J. and MICHEL, B. (2011). Deconvolution for the Wasserstein metric and geometric inference. *Electron. J. Stat.* **5** 1394–1423. MR2851684 https://doi.org/10.1214/11-EJS646

[16] CANAS, G. and ROSASCO, L. (2012). Learning probability measures with respect to optimal transport metrics. In *Advances in Neural Information Processing Systems* 2492–2500.

[17] CAZELLES, E., SEGUY, V., BIGOT, J., CUTURI, M. and PAPADAKIS, N. (2018). Geodesic PCA versus log-PCA of histograms in the Wasserstein space. *SIAM J. Sci. Comput.* **40** B429–B456. MR3780753 https://doi.org/10.1137/17M1143459

[18] CLAICI, S. and SOLOMON, J. (2018). Wasserstein coresets for Lipschitz costs. Preprint. Available at arXiv:1805.07412v1.

[19] COURTY, N., FLAMARY, R., HABRARD, A. and RAKOTOMAMONJY, A. (2017). Joint distribution optimal transportation for domain adaptation. In *Advances in Neural Information Processing Systems* 3730–3739.

[20] COURTY, N., FLAMARY, R. and TUIA, D. (2014). Domain adaptation with regularized optimal transport. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* 274–289. Springer, Berlin.

[21] COURTY, N., FLAMARY, R., TUIA, D. and RAKOTOMAMONJY, A. (2017). Optimal transport for domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39** 1853–1865.

[22] CUTURI, M. (2013). Sinkhorn distances: lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems* 2292–2300.

[23] DALALYAN, A. and REISS, M. (2007). Asymptotic statistical equivalence for ergodic diffusions: The multidimensional case. *Probab. Theory Related Fields* **137** 25–47. MR2278451 https://doi.org/10.1007/s00440-006-0502-7

[24] DAMODARAN, B. B., KELLENBERGER, B., FLAMARY, R., TUIA, D. and COURTY, N. (2018). DeepJDOT: Deep joint distribution optimal transport for unsupervised domain adaptation. In *Computer Vision—ECCV* 2018—15*th European Conference*, *Munich*, *Germany*, *September* 8–14, 2018, *Proceedings*, *Part IV* 467–483.

[25] DEDECKER, J., FISCHER, A. and MICHEL, B. (2015). Improved rates for Wasserstein deconvolution with ordinary smooth error in dimension one. *Electron. J. Stat.* **9** 234–265. MR3314482 https://doi.org/10.1214/15-EJS997

[26] DEDECKER, J. and MICHEL, B. (2013). Minimax rates of convergence for Wasserstein deconvolution with supersmooth errors in any dimension. *J. Multivariate Anal.* **122** 278–291. MR3189324 https://doi.org/10.1016/j.jmva.2013.08.009

[27] DEL BARRIO, E., GAMBOA, F., GORDALIZA, P. and LOUBES, J.-M. (2019). Obtaining fairness using optimal transport theory. In *International Conference on Machine Learning* 2357–2365.

[28] DEL BARRIO, E., GORDALIZA, P., LESCORNEL, H. and LOUBES, J.-M. (2019). Central limit theorem and bootstrap procedure for Wasserstein's variations with an application to structural relationships between distributions. *J. Multivariate Anal.* **169** 341–362. MR3875604 https://doi.org/10.1016/j.jmva.2018.09.014

[29] DVURECHENSKY, P., GASNIKOV, A. and KROSHNIN, A. (2018). Computational optimal transport: Complexity by accelerated gradient descent is better than by Sinkhorn's algorithm. In *Proceedings of the* 35*th International Conference on Machine Learning* (J. Dy and A. Krause, eds.). *Proceedings of Machine Learning Research* **80** 1367–1376. PMLR, Stockholmsmässan.

[30] ESPOSITO, L., NITSCH, C. and TROMBETTI, C. (2013). Best constants in Poincaré inequalities for convex domains. *J. Convex Anal.* **20** 253–264. MR3086452

[31] FEYDY, J., CHARLIER, B., VIALARD, F. and PEYRÉ, G. (2017). Optimal transport for diffeomorphic registration. In *Medical Image Computing and Computer Assisted Intervention—MICCAI* 2017—20*th International Conference*, *Quebec City*, *QC*, *Canada*, *September* 11–13, 2017, *Proceedings*, *Part I* 291–299.

[32] FLAMARY, R., CUTURI, M., COURTY, N. and RAKOTOMAMONJY, A. (2018). Wasserstein discriminant analysis. *Mach. Learn.* **107** 1923–1945. MR3854402 https://doi.org/10.1007/s10994-018-5717-1

[33] FLAMARY, R., LOUNICI, K. and FERRARI, A. (2019). Concentration bounds for linear Monge mapping estimation and optimal transport domain adaptation. Preprint. Available at arXiv:1905.10155.

[34] FORROW, A., HÜTTER, J.-C., NITZAN, M., SCHIEBINGER, G., RIGOLLET, P. and WEED, J. (2019). Statistical optimal transport via factored couplings. In *The* 22*nd International Conference on Artificial Intelligence and Statistics* 2454–2465.

[35] GENEVAY, A., CUTURI, M., PEYRÉ, G. and BACH, F. (2016). Stochastic optimization for large-scale optimal transport. In *Advances in Neural Information Processing Systems* 3440–3448.

[36] GENEVAY, A., PEYRÉ, G. and CUTURI, M. (2018). Learning generative models with Sinkhorn divergences. In *International Conference on Artificial Intelligence and Statistics*, *AISTATS* 2018, 9–11 *April* 2018, *Playa Blanca*, *Lanzarote*, *Canary Islands*, *Spain* 1608–1617.

[37] GIGLI, N. (2011). On Hölder continuity-in-time of the optimal transport map towards measures along a curve. *Proc*. *Edinb*. *Math*. *Soc*. (2) **54** 401–409. MR2794662 https://doi.org/10.1017/S001309150800117X

[38] 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

[39] GRAVE, E., JOULIN, A. and BERTHET, Q. (2019). Unsupervised alignment of embeddings with Wasserstein procrustes. In *Proceedings of Machine Learning Research* (K. Chaudhuri and M. Sugiyama, eds.). *Proceedings of Machine Learning Research* **89** 1880–1890. PMLR.

[40] GUNSILIUS, F. F. (2021). On the convergence rate of potentials of Brenier maps. *Econometric Theory* 1–37. https://doi.org/doi:10.1017/S0266466621000037

[41] HÄRDLE, W., KERKYACHARIAN, G., PICARD, D. and TSYBAKOV, A. (1998). *Wavelets*, *Approximation*, *and Statistical Applications*. *Lecture Notes in Statistics* **129**. Springer, New York. MR1618204 https://doi.org/10.1007/978-1-4612-2222-4

[42] HÜTTER, J.-C. (2019). Minimax estimation with structured data: Shape constraints, causal models, and optimal transport. Ph.D. thesis, Massachusetts Institute of Technology.

[43] HÜTTER, J.-C. and RIGOLLET, P. (2021). Supplement to "Minimax estimation of smooth optimal transport maps." https://doi.org/10.1214/20-AOS1997SUPP

[44] IBRAGIMOV, I. A. and HAS'MINSKIĬ, R. Z. (1981). *Statistical Estimation*: *Asymptotic Theory*. *Applications of Mathematics* **16**. Springer, New York. Translated from the Russian by Samuel Kotz. MR0620321

[45] JANATI, H., CUTURI, M. and GRAMFORT, A. (2019). Wasserstein regularization for sparse multi-task regression. In *The* 22*nd International Conference on Artificial Intelligence and Statistics* 1407–1416.

[46] JONES, P. W. (1981). Quasiconformal mappings and extendability of functions in Sobolev spaces. *Acta Math*. **147** 71–88. MR0631089 https://doi.org/10.1007/BF02392869

[47] KANTOROVICH, L. V. (1948). On a problem of Monge. *C*. *R*. (*Dokl*.) *Acad*. *Sci*. *URSS* **3** 225–226.

[48] KANTOROVITCH, L. (1942). On the translocation of masses. *C*. *R*. (*Dokl*.) *Acad*. *Sci*. *URSS* **37** 199–201. MR0009619

[49] KLATT, M., TAMELING, C. and MUNK, A. (2020). Empirical regularized optimal transport: Statistical theory and applications. *SIAM J*. *Math*. *Data Sci*. **2** 419–443. MR4105566 https://doi.org/10.1137/19M1278788

[50] KNOTT, M. and SMITH, C. S. (1984). On the optimal mapping of distributions. *J*. *Optim*. *Theory Appl*. **43** 39–49. MR0745785 https://doi.org/10.1007/BF00934745

[51] KOLTCHINSKII, V. (2011). *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems*. *Lecture Notes in Math*. **2033**. Springer, Heidelberg. MR2829871 https://doi.org/10.1007/978-3-642-22147-7

[52] KROSHNIN, A., SPOKOINY, V. and SUVORIKOVA, A. (2019). Statistical inference for Bures–Wasserstein barycenters. Preprint. Available at arXiv:1901.00226.

[53] LAVENANT, H., CLAICI, S., CHIEN, E. and SOLOMON, J. (2018). Dynamical optimal transport on discrete surfaces. *ACM Trans*. *Graph*. **37** 250:1–250:16.

[54] LEDOUX, M. (2019). On optimal matching of Gaussian samples. *J*. *Math*. *Sci*. **238** 495–522.

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

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

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

[58] LUCET, Y. (1997). Faster than the fast Legendre transform, the linear-time Legendre transform. *Numer*. *Algorithms* **16** 171–185. MR1619184 https://doi.org/10.1023/A:1019191114493

[59] MALLAT, S. (1999). *A Wavelet Tour of Signal Processing*. Elsevier, Amsterdam.

[60] MASAROTTO, V., PANARETOS, V. M. and ZEMEL, Y. (2019). Procrustes metrics on covariance operators and optimal transportation of Gaussian processes. *Sankhya A* **81** 172–213. MR3982195 https://doi.org/10.1007/s13171-018-0130-1

[61] MASSART, P. (2007). *Concentration Inequalities and Model Selection*: *Ecole D'Eté de Probabilités de Saint-Flour XXXIII—*2003. *École D'Été de Probabilités de Saint-Flour*. Springer, Berlin.

[62] MONGE, G. (1781). *Mémoire Sur La Théorie Des Déblais et Des Remblais*. Histoire de L'Académie Royale des Sciences de Paris.

[63] MONTAVON, G., MÜLLER, K. and CUTURI, M. (2016). Wasserstein training of restricted Boltzmann machines. In *Advances in Neural Information Processing Systems* 29: *Annual Conference on Neural Information Processing Systems* 2016, *December* 5–10, 2016, *Barcelona*, *Spain* 3711–3719.

[64] MÜLLER, H.-G. and GASSER, T. (1979). Optimal convergence properties of kernel estimates of derivatives of a density function. In *Smoothing Techniques for Curve Estimation* (*Proc*. *Workshop*, *Heidelberg*, 1979). *Lecture Notes in Math*. **757** 144–154. Springer, Berlin. MR0564256

[65] MUNK, A. and CZADO, C. (1998). Nonparametric validation of similar distributions and assessment of goodness of fit. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **60** 223–241. MR1625620 https://doi.org/10.1111/1467-9868.00121

[66] MURPHY, K. P. (2012). *Machine Learning*: *A Probabilistic Perspective*. MIT Press, Cambridge.

[67] NILES-WEED, J. and RIGOLLET, P. (2019). Estimation of Wasserstein distances in the spiked transport model. Preprint. Available at arXiv:1909.07513.

[68] 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

[69] PERROT, M., COURTY, N., FLAMARY, R. and HABRARD, A. (2016). Mapping estimation for discrete optimal transport. In *Advances in Neural Information Processing Systems* 4197–4205.

[70] PEYRÉ, G. and CUTURI, M. (2019). Computational optimal transport. *Found*. *Trends Mach*. *Learn*. **11** 355–607.

[71] RAMDAS, A., TRILLOS, N. G. and CUTURI, M. (2017). On Wasserstein two-sample testing and related families of nonparametric tests. *Entropy* **19** Paper No. 47. MR3608466 https://doi.org/10.3390/e19020047

[72] RIGOLLET, P. and WEED, J. (2018). Entropic optimal transport is maximum-likelihood deconvolution. *C*. *R*. *Math*. *Acad*. *Sci*. *Paris* **356** 1228–1235. MR3907589 https://doi.org/10.1016/j.crma.2018.10.010

[73] RIGOLLET, P. and WEED, J. (2019). Uncoupled isotonic regression via minimum Wasserstein deconvolution. *Inf. Inference* **8** 691–717. MR4045481 https://doi.org/10.1093/imaiai/iaz006

[74] ROLET, A., CUTURI, M. and PEYRÉ, G. (2016). Fast dictionary learning with a smoothed Wasserstein loss. In *Proceedings of the* 19*th International Conference on Artificial Intelligence and Statistics*, *AISTATS* 2016, *Cadiz*, *Spain*, *May* 9–11, 2016 630–638.

[75] RÜSCHENDORF, L. and RACHEV, S. T. (1990). A characterization of random variables with minimum L^2-distance. *J*. *Multivariate Anal*. **32** 48–54. MR1035606 https://doi.org/10.1016/0047-259X(90)90070-X

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

[77] SCHIEBINGER, G., SHU, J., TABAKA, M., CLEARY, B., SUBRAMANIAN, V., SOLOMON, A., GOULD, J., LIU, S., LIN, S. et al. (2019). Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell* **176** 928–943.e22. https://doi.org/10.1016/j.cell.2019.01.006

[78] SCHMITZ, M. A., HEITZ, M., BONNEEL, N., MBOULA, F. M. N., COEURJOLLY, D., CUTURI, M., PEYRÉ, G. and STARCK, J. (2018). Wasserstein dictionary learning: Optimal transport-based unsupervised non-linear dictionary learning. *SIAM J. Imaging Sci*. **11** 643–678.

[79] SCHMITZ, M. A., HEITZ, M., BONNEEL, N., NGOLÈ, F., COEURJOLLY, D., CUTURI, M., PEYRÉ, G. and STARCK, J.-L. (2018). Wasserstein dictionary learning: Optimal transport-based unsupervised nonlinear dictionary learning. *SIAM J. Imaging Sci*. **11** 643–678. MR3769700 https://doi.org/10.1137/17M1140431

[80] SEGUY, V. and CUTURI, M. (2015). Principal geodesic analysis for probability measures under the optimal transport metric. In *Advances in Neural Information Processing Systems* 28: *Annual Conference on Neural Information Processing Systems* 2015, *December* 7–12, 2015, *Montreal*, *Quebec*, *Canada* 3312–3320.

[81] SEGUY, V., DAMODARAN, B. B., FLAMARY, R., COURTY, N., ROLET, A. and BLONDEL, M. (2018). Large-scale optimal transport and mapping estimation. In *Proceedings of the International Conference in Learning Representations*.

[82] SINGH, S. and PÓCZOS, B. (2018). Minimax distribution estimation in Wasserstein distance. Preprint. Available at arXiv:1802.08855.

[83] SOLOMON, J., DE GOES, F., PEYRÉ, G., CUTURI, M., BUTSCHER, A., NGUYEN, A., DU, T. and GUIBAS, L. (2015). Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Trans*. *Graph*. **34** 66:1–66:11. https://doi.org/10.1145/2766963

[84] SOLOMON, J., PEYRÉ, G., KIM, V. G. and SRA, S. (2016). Entropic metric alignment for correspondence problems. *ACM Trans. Graph.* **35** 72:1–72:13. https://doi.org/10.1145/2897824.2925903

[85] STAIB, M., CLAICI, S., SOLOMON, J. M. and JEGELKA, S. (2017). Parallel streaming Wasserstein barycenters. In *Advances in Neural Information Processing Systems* 30: *Annual Conference on Neural Information Processing Systems* 2017, 4–9 *December* 2017, *Long Beach*, *CA*, *USA* 2644–2655.

[86] STRAUCH, C. (2018). Adaptive invariant density estimation for ergodic diffusions over anisotropic classes. *Ann. Statist.* **46** 3451–3480. MR3852658 https://doi.org/10.1214/17-AOS1664

[87] TALAGRAND, M. (2014). *Upper and Lower Bounds for Stochastic Processes*: *Modern Methods and Classical Problems*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete*. 3. *Folge*. *A Series of Modern Surveys in Mathematics* [*Results in Mathematics and Related Areas*. 3*rd Series*] **60**. Springer, Heidelberg. MR3184689 https://doi.org/10.1007/978-3-642-54075-2

[88] TAMELING, C. and MUNK, A. (2018). Computational strategies for statistical inference based on empirical optimal transport. In 2018 *IEEE Data Science Workshop*, *DSW* 2018, *Lausanne*, *Switzerland*, *June* 4–6, 2018 175–179.

[89] TANAKA, K., MIZUGUCHI, M., SEKINE, K., TAKAYASU, A. and OISHI, S. (2013). Estimation of an embedding constant on Lipschitz domains using extension operators. In *JSST* 2013 *International Conference on Simulation Technology*.

[90] TRIEBEL, H. (2006). *Theory of Function Spaces*. *III*. *Monographs in Mathematics* **100**. Birkhäuser, Basel. MR2250142

[91] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. Revised and extended from the 2004 French original. Translated by Vladimir Zaiats. MR2724359 https://doi.org/10.1007/b13794

[92] VAN DE GEER, S. (1987). A new approach to least-squares estimation, with applications. *Ann. Statist.* **15** 587–602. MR0888427 https://doi.org/10.1214/aos/1176350362

[93] VAN DE GEER, S. (2002). M-estimation using penalties or sieves. *J. Statist. Plann. Inference* **108** 55–69. MR1947391 https://doi.org/10.1016/S0378-3758(02)00270-7

[94] VAN DER VAART, A. W. (2000). *Asymptotic Statistics* **3**. Cambridge Univ. Press, Cambridge.

[95] 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

[96] 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

[97] WEED, J. and BACH, F. (2019). Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli* **25** 2620–2648. MR4003560 https://doi.org/10.3150/18-BEJ1065

[98] WEED, J. and BERTHET, Q. (2019). Estimation of smooth densities in Wasserstein distance. In *Conference on Learning Theory* 3118–3119.

[99] YANG, K. D., DAMODARAN, K., VENKATCHALAPATHY, S., SOYLEMEZOGLU, A. C., SHIVASHANKAR, G. V. and UHLER, C. (2018). Autoencoder and optimal transport to infer single-cell trajectories of biological processes. *BioRxiv*. Available at https://www.biorxiv.org/content/10.1101/455469v1.

[100] YUKICH, J. E. (2006). *Probability Theory of Classical Euclidean Optimization Problems*. Springer, Berlin.

[101] ZEMEL, Y. and PANARETOS, V. M. (2019). Fréchet means and Procrustes analysis in Wasserstein space. *Bernoulli* **25** 932–976. MR3920362 https://doi.org/10.3150/17-bej1009

The Annals of Statistics
2021, Vol. 49, No. 2, 1195–1217
https://doi.org/10.1214/20-AOS1998

ESTIMATION AND INFERENCE IN THE PRESENCE OF FRACTIONAL $d = 1/2$ AND WEAKLY NONSTATIONARY PROCESSES

BY JAMES A. DUFFY[1] AND IOANNIS KASPARIS[2]

[1]*Department of Economics, University of Oxford, james.duffy@economics.ox.ac.uk*
[2]*Department of Economics, University of Cyprus, kasparis@ucy.ac.cy*

We provide new limit theory for functionals of a general class of processes lying at the boundary between stationarity and nonstationarity—what we term weakly nonstationary processes (WNPs). This includes, as leading examples, fractional processes with $d = 1/2$, and arrays of autoregressive processes with roots drifting slowly towards unity. We first apply the theory to study inference in parametric and nonparametric regression models involving WNPs as covariates. We then use these results to develop a new specification test for parametric regression models. By construction, our specification test statistic has a χ^2 limiting distribution regardless of the form and extent of persistence of the regressor, implying that a practitioner can validly perform the test using a fixed critical value, while remaining agnostic about the mechanism generating the regressor. Simulation exercises confirm that the test controls size across a wide range of data generating processes, and outperforms a comparable test due to Wang and Phillips (*Ann. Statist.* **40** (2012) 727–758) against many alternatives.

REFERENCES

ABADIR, K. M., DISTASO, W., GIRAITIS, L. and KOUL, H. L. (2014). Asymptotic normality for weighted sums of linear processes. *Econometric Theory* **30** 252–284. MR3177798 https://doi.org/10.1017/S0266466613000182

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

ANDREWS, D. W. K., CHENG, X. and GUGGENBERGER, P. (2020). Generic results for establishing the asymptotic size of confidence sets and tests. *J. Econometrics* **218** 496–531. MR4149237 https://doi.org/10.1016/j.jeconom.2020.04.027

BAILLIE, R. T., CHUNG, C.-F. and TIESLAU, M. T. (1996). Analysing inflation by the fractionally integrated ARFIMA–GARCH model. *J. Appl. Econometrics* **11** 23–40.

BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1987). *Regular Variation. Encyclopedia of Mathematics and Its Applications* **27**. Cambridge Univ. Press, Cambridge. MR0898871 https://doi.org/10.1017/CBO9780511721434

BUCHMANN, B. and CHAN, N. H. (2007). Asymptotic theory of least squares estimators for nearly unstable processes under strong dependence. *Ann. Statist.* **35** 2001–2017. MR2363961 https://doi.org/10.1214/009053607000000136

CHAN, N. and WANG, Q. (2015). Nonlinear regressions with nonstationary time series. *J. Econometrics* **185** 182–195. MR3300342 https://doi.org/10.1016/j.jeconom.2014.04.025

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

CHAN, N. H. and WEI, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.* **16** 367–401. MR0924877 https://doi.org/10.1214/aos/1176350711

CHRISTENSEN, B. J. and NIELSEN, M. Ø. (2006). Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting. *J. Econometrics* **133** 343–371. MR2250183 https://doi.org/10.1016/j.jeconom.2005.03.018

MSC2020 subject classifications. 62M10, 62F12, 62G08.
Key words and phrases. Fractional process, half unit root, kernel regression, specification testing, mildly integrated process, weakly nonstationary process.

CHRISTOPEIT, N. (2009). Weak convergence of nonlinear transformations of integrated processes: The multivariate case. *Econometric Theory* **25** 1180–1207. MR2540497 https://doi.org/10.1017/S0266466608090476

DEMETRESCU, M., GEORGIEV, I., RODRIGUES, P. and TAYLOR, R. (2019). Testing for episodic predictability in stock returns. *J. Econometrics*. In press. https://doi.org/10.1016/j.jeconom.2020.01.001

DONG, C., GAO, J., TJØSTHEIM, D. and YIN, J. (2017). Specification testing for nonlinear multivariate cointegrating regressions. *J. Econometrics* **200** 104–117. MR3683664 https://doi.org/10.1016/j.jeconom.2017.05.016

DUFFY, J. A. (2020). Asymptotic theory for kernel estimators under moderate deviations from a unit root, with an application to the asymptotic size of nonparametric tests. *Econometric Theory* **36** 559–582. MR4125472 https://doi.org/10.1017/s0266466619000240

DUFFY, J. A. and KASPARIS, I. (2021). Supplement to "Estimation and Inference in the presence of fractional $d = 1/2$ and weakly nonstationary processes." https://doi.org/10.1214/20-AOS1998SUPP

GAO, J., KING, M., LU, Z. and TJØSTHEIM, D. (2009). Specification testing in nonlinear and nonstationary time series autoregression. *Ann. Statist.* **37** 3893–3928. MR2572447 https://doi.org/10.1214/09-AOS698

GIKHMAN, I. I. and SKOROKHOD, A. V. (1969). *Introduction to the Theory of Random Processes*. Saunders, Philadelphia. MR0247660

GIRAITIS, L. and PHILLIPS, P. C. B. (2006). Uniform limit theory for stationary autoregression. *J. Time Series Anal.* **27** 51–60. MR2235146 https://doi.org/10.1111/j.1467-9892.2005.00452.x

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York. MR0624435

HASSLER, U. and POHLE, J.-M. (2019). Forecasting under long memory and nonstationarity. Working paper, Goethe University Frankfurt.

HASSLER, U. and WOLTERS, J. (1995). Long memory in inflation rates: International evidence. *J. Bus. Econom. Statist.* **13** 37–45.

HOROWITZ, J. L. and SPOKOINY, V. G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* **69** 599–631. MR1828537 https://doi.org/10.1111/1468-0262.00207

HUALDE, J. and ROBINSON, P. M. (2010). Semiparametric inference in multivariate fractionally cointegrated systems. *J. Econometrics* **157** 492–511. MR2661617 https://doi.org/10.1016/j.jeconom.2010.04.002

HUALDE, J. and ROBINSON, P. M. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models. *Ann. Statist.* **39** 3152–3181. MR3012404 https://doi.org/10.1214/11-AOS931

JEGANATHAN, P. (2004). Convergence of functionals of sums of r.v.s to local times of fractional stable motions. *Ann. Probab.* **32** 1771–1795. MR2073177 https://doi.org/10.1214/009117904000000658

JEGANATHAN, P. (2008). Limit theorems for functionals of sums that converge to fractional Brownian and stable motions. Cowles Foundation Discussion Paper 1949.

JOHANSEN, S. and NIELSEN, M. Ø. (2012a). Likelihood inference for a fractionally cointegrated vector autoregressive model. *Econometrica* **80** 2667–2732. MR3001138 https://doi.org/10.3982/ECTA9299

JOHANSEN, S. and NIELSEN, M. Ø. (2012b). A necessary moment condition for the fractional functional central limit theorem. *Econometric Theory* **28** 671–679. MR2927923 https://doi.org/10.1017/S0266466611000697

KARLSEN, H. A., MYKLEBUST, T. and TJØSTHEIM, D. (2007). Nonparametric estimation in a nonlinear cointegration type model. *Ann. Statist.* **35** 252–299. MR2332276 https://doi.org/10.1214/009053606000001181

KARLSEN, H. A. and TJØSTHEIM, D. (2001). Nonparametric estimation in null recurrent time series. *Ann. Statist.* **29** 372–416. MR1863963 https://doi.org/10.1214/aos/1009210546

KASAHARA, Y. and MAEJIMA, M. (1988). Weighted sums of i.i.d. random variables attracted to integrals of stable processes. *Probab. Theory Related Fields* **78** 75–96. MR0940869 https://doi.org/10.1007/BF00718037

KASPARIS, I., ANDREOU, E. and PHILLIPS, P. C. B. (2015). Nonparametric predictive regression. *J. Econometrics* **185** 468–494. MR3311834 https://doi.org/10.1016/j.jeconom.2014.05.015

KOSTAKIS, A., STAMATOGIANNIS, M. P. and MAGDALINOS, T. (2015). Robust econometric inference for stock return predictability. *Rev. Financ. Stud.* **28** 1506–1553.

LOÈVE, M. (1978). *Probability Theory. II*, 4th ed. Springer, New York. MR0651018

MARINUCCI, D. and ROBINSON, P. M. (1998). Semiparametric frequency domain analysis of fractional cointegration. Lond. School Econ. Mimeo.

MARINUCCI, D. and ROBINSON, P. M. (1999). Alternative forms of fractional Brownian motion. *J. Statist. Plann. Inference* **80** 111–122. MR1713794 https://doi.org/10.1016/S0378-3758(98)00245-6

MIKUSHEVA, A. (2007). Uniform inference in autoregressive models. *Econometrica* **75** 1411–1452. MR2347350 https://doi.org/10.1111/j.1468-0262.2007.00798.x

PARK, J. Y. and PHILLIPS, P. C. B. (1999). Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory* **15** 269–298. MR1704225 https://doi.org/10.1017/S0266466699153015

PARK, J. Y. and PHILLIPS, P. C. B. (2001). Nonlinear regressions with integrated time series. *Econometrica* **69** 117–161. MR1806536 https://doi.org/10.1111/1468-0262.00180

PHILLIPS, P. C. B. (1987a). Time series regression with a unit root. *Econometrica* **55** 277–301. MR0882096 https://doi.org/10.2307/1913237

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

PHILLIPS, P. C. B. (1995). Fully modified least squares and vector autoregression. *Econometrica* **63** 1023–1078. MR1348513 https://doi.org/10.2307/2171721

PHILLIPS, P. C. B. and LEE, J. H. (2013). Predictive regression under various degrees of persistence and robust long-horizon regression. *J. Econometrics* **177** 250–264. MR3118559 https://doi.org/10.1016/j.jeconom.2013.04.011

PHILLIPS, P. C. B. and MAGDALINOS, T. (2007). Limit theory for moderate deviations from a unit root. *J. Econometrics* **136** 115–130. MR2328587 https://doi.org/10.1016/j.jeconom.2005.08.002

PHILLIPS, P. C. B. and MAGDALINOS, T. (2009). Econometric inference in the vicinity of unity. Singapore Management University. Mimeo.

PHILLIPS, P. C. B. and SHIMOTSU, K. (2004). Local Whittle estimation in nonstationary and unit root cases. *Ann. Statist.* **32** 656–692. MR2060173 https://doi.org/10.1214/009053604000000139

PÖTSCHER, B. M. (2004). Nonlinear functions and convergence to Brownian motion: Beyond the continuous mapping theorem. *Econometric Theory* **20** 1–22. MR2028350 https://doi.org/10.1017/S0266466604201013

ROBINSON, P. M. and HUALDE, J. (2003). Cointegration in fractional systems with unknown integration orders. *Econometrica* **71** 1727–1766. MR2015418 https://doi.org/10.1111/1468-0262.00468

SHIMOTSU, K. and PHILLIPS, P. C. B. (2005). Exact local Whittle estimation of fractional integration. *Ann. Statist.* **33** 1890–1933. MR2166565 https://doi.org/10.1214/009053605000000309

TANAKA, K. (1999). The nonstationary fractional unit root. *Econometric Theory* **15** 549–582. MR1717967 https://doi.org/10.1017/S0266466699154045

TAQQU, M. S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Probab. Theory Related Fields* **31** 287–302.

WANG, Q. (2014). Martingale limit theorem revisited and nonlinear cointegrating regression. *Econometric Theory* **30** 509–535. MR3205605 https://doi.org/10.1017/S026646661300039X

WANG, Q. and PHILLIPS, P. C. B. (2009a). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* **25** 710–738. MR2507529 https://doi.org/10.1017/S0266466608090269

WANG, Q. and PHILLIPS, P. C. B. (2009b). Structural nonparametric cointegrating regression. *Econometrica* **77** 1901–1948. MR2573873 https://doi.org/10.3982/ECTA7732

WANG, Q. and PHILLIPS, P. C. B. (2011). Asymptotic theory for zero energy functionals with nonparametric regression applications. *Econometric Theory* **27** 235–259. MR2782038 https://doi.org/10.1017/S0266466610000277

WANG, Q. and PHILLIPS, P. C. B. (2012). A specification test for nonlinear nonstationary models. *Ann. Statist.* **40** 727–758. MR2933664 https://doi.org/10.1214/12-AOS975

WU, W. B., HUANG, Y. and HUANG, Y. (2010). Kernel estimation for time series: An asymptotic theory. *Stochastic Process. Appl.* **120** 2412–2431. MR2728171 https://doi.org/10.1016/j.spa.2010.08.001

WU, W. B. and MIELNICZUK, J. (2002). Kernel density estimation for linear processes. *Ann. Statist.* **30** 1441–1459. MR1936325 https://doi.org/10.1214/aos/1035844982

YANG, B., LONG, W., PENG, L. and CAI, Z. (2019). Testing the predictability of U.S. housing price index returns based on an IVX-AR model. *J. Amer. Statist. Assoc.* In press. https://doi.org/10.1080/01621459.2019.1686392

The Annals of Statistics
2021, Vol. 49, No. 2, 1218–1238
https://doi.org/10.1214/20-AOS1999

ONLY CLOSED TESTING PROCEDURES ARE ADMISSIBLE FOR CONTROLLING FALSE DISCOVERY PROPORTIONS

BY JELLE J. GOEMAN[1], JESSE HEMERIK[2] AND ALDO SOLARI[3]

[1]*Department of Biomedical Data Sciences, Leiden University Medical Center, j.j.goeman@lumc.nl*

[2]*Oslo Centre for Biostatistics and Epidemiology, University of Oslo, and Biometris, Wageningen University & Research, jesse.hemerik@wur.nl*

[3]*Department of Economics, Management and Statistics, University of Milano-Bicocca, solari@stat.unipd.it*

We consider the class of all multiple testing methods controlling tail probabilities of the false discovery proportion, either for one random set or simultaneously for many such sets. This class encompasses methods controlling familywise error rate, generalized familywise error rate, false discovery exceedance, joint error rate, simultaneous control of all false discovery proportions, and others, as well as gene set testing in genomics and cluster inference in neuroimaging. We show that all such methods are either equivalent to a closed testing procedure, or are uniformly improved by one. Moreover, we show that a closed testing method is admissible if and only if all its local tests are admissible. This implies that, when designing methods, it is sufficient to restrict attention to closed testing. We demonstrate the practical usefulness of this design principle by obtaining more informative inferences from the method of higher criticism, and by constructing a uniform improvement of a recently proposed method.

REFERENCES

BARNETT, I. J. and LIN, X. (2014). Analytical p-value calculation for the higher criticism test in finite-d problems. *Biometrika* **101** 964–970. MR3286929 https://doi.org/10.1093/biomet/asu033

BENJAMINI, Y. (2010). Simultaneous and selective inference: Current successes and future challenges. *Biom. J.* **52** 708–721. MR2758547 https://doi.org/10.1002/bimj.200900299

BENJAMINI, Y. and HELLER, R. (2008). Screening for partial conjunction hypotheses. *Biometrics* **64** 1215–1222. MR2522270 https://doi.org/10.1111/j.1541-0420.2007.00984.x

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

BENJAMINI, Y. and LIU, W. (1999). A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence **82** 163–170. J. Statist. Plann. Inference, 1-2, Multiple comparisons (Tel Aviv, 1996). MR1736441 https://doi.org/10.1016/S0378-3758(99)00040-3

BERK, R., BROWN, L., BUJA, A., ZHANG, K. and ZHAO, L. (2013). Valid post-selection inference. *Ann. Statist.* **41** 802–837. MR3099122 https://doi.org/10.1214/12-AOS1077

BLANCHARD, G., NEUVIAL, P. and ROQUAIN, E. (2020). Post hoc confidence bounds on false positives using reference families. *Ann. Statist.* **48** 1281–1303. MR4124323 https://doi.org/10.1214/19-AOS1847

BLANCHARD, G. and ROQUAIN, É. (2009). Adaptive false discovery rate control under independence and dependence. *J. Mach. Learn. Res.* **10** 2837–2871. MR2579914

BRANNATH, W. and BRETZ, F. (2010). Shortcuts for locally consonant closed test procedures. *J. Amer. Statist. Assoc.* **105** 660–669. MR2724850 https://doi.org/10.1198/jasa.2010.tm08127

BRETZ, F., MAURER, W., BRANNATH, W. and POSCH, M. (2009). A graphical approach to sequentially rejective multiple test procedures. *Stat. Med.* **28** 586–604. MR2655732 https://doi.org/10.1002/sim.3495

CAI, G. and SARKAR, S. K. (2008). Modified Simes' critical values under independence. *Statist. Probab. Lett.* **78** 1362–1368. MR2528334 https://doi.org/10.1016/j.spl.2007.12.018

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

MSC2020 subject classifications. 62F03.

Key words and phrases. Familywise error rate, selective inference, simultaneous inference, multiple testing, higher criticism.

DOBRIBAN, E. (2020). Fast closed testing for exchangeable local tests. *Biometrika* **107** 761–768. MR4138990 https://doi.org/10.1093/biomet/asz082

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

DUDOIT, S., VAN DER LAAN, M. J. and POLLARD, K. S. (2004). Multiple testing. I. Single-step procedures for control of general type I error rates. *Stat. Appl. Genet. Mol. Biol.* **3** Art. 13, 71. MR2101462 https://doi.org/10.2202/1544-6115.1040

DURAND, G., BLANCHARD, G., NEUVIAL, P. and ROQUAIN, E. (2020). Post hoc false positive control for structured hypotheses. *Scand. J. Stat.* **47** 1114–1148. https://doi.org/10.1111/sjos.12453

EBRAHIMPOOR, M., SPITALI, P., HETTNE, K., TSONAKA, R. and GOEMAN, J. (2020). Simultaneous enrichment analysis of all possible gene-sets: Unifying self-contained and competitive methods. *Brief. Bioinform.* **21** 1302–1312. https://doi.org/10.1093/bib/bbz074

FARCOMENI, A. (2009). Generalized augmentation to control the false discovery exceedance in multiple testing. *Scand. J. Stat.* **36** 501–517. MR2549707 https://doi.org/10.1111/j.1467-9469.2008.00633.x

FINNER, H. and STRASSBURGER, K. (2002). The partitioning principle: A powerful tool in multiple decision theory. *Ann. Statist.* **30** 1194–1213. MR1926174 https://doi.org/10.1214/aos/1031689023

FINOS, L. and FARCOMENI, A. (2011). k-FWER control without p-value adjustment, with application to detection of genetic determinants of multiple sclerosis in Italian twins. *Biometrics* **67** 174–181. MR2898829 https://doi.org/10.1111/j.1541-0420.2010.01443.x

GE, Y. and LI, X. (2012). Control of the false discovery proportion for independently tested null hypotheses. *J. Probab. Stat.* Art. ID 320425, 19. MR2934965 https://doi.org/10.1155/2012/320425

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

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

GOEMAN, J. J. and FINOS, L. (2012). The inheritance procedure: Multiple testing of tree-structured hypotheses. *Stat. Appl. Genet. Mol. Biol.* **11** Art. 11, 20. MR2924200 https://doi.org/10.1515/1544-6115.1554

GOEMAN, J. J, HEMERIK, J. and SOLARI, A. (2021). Supplement to "Only closed testing procedures are admissible for controlling false discovery proportions." https://doi.org/10.1214/20-AOS1999SUPP

GOEMAN, J. J. and MANSMANN, U. (2008). Multiple testing on the directed acyclic graph of gene ontology. *Bioinformatics* **24** 537–544.

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

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

GOEMAN, J. J. and SOLARI, A. (2014). Multiple hypothesis testing in genomics. *Stat. Med.* **33** 1946–1978. MR3257576 https://doi.org/10.1002/sim.6082

GOEMAN, J. J., MEIJER, R. J., KREBS, T. J. P. and SOLARI, A. (2019). Simultaneous control of all false discovery proportions in large-scale multiple hypothesis testing. *Biometrika* **106** 841–856. MR4046036 https://doi.org/10.1093/biomet/asz041

GONTSCHARUK, V., LANDWEHR, S. and FINNER, H. (2016). Goodness of fit tests in terms of local levels with special emphasis on higher criticism tests. *Bernoulli* **22** 1331–1363. MR3474818 https://doi.org/10.3150/14-BEJ694

GOU, J. and TAMHANE, A. C. (2014). On generalized Simes critical constants. *Biom. J.* **56** 1035–1054. MR3270109 https://doi.org/10.1002/bimj.201300258

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

GUO, W. and RAO, M. B. (2010). On stepwise control of the generalized familywise error rate. *Electron. J. Stat.* **4** 472–485. MR2657378 https://doi.org/10.1214/08-EJS320

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

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

HEMERIK, J., SOLARI, A. and GOEMAN, J. J. (2019). Permutation-based simultaneous confidence bounds for the false discovery proportion. *Biometrika* **106** 635–649. MR3992394 https://doi.org/10.1093/biomet/asz021

HENNING, K. S. S. and WESTFALL, P. H. Closed testing in pharmaceutical research: Historical and recent developments. *Stat. Biopharm. Res.* **7** 126–147. https://doi.org/10.1080/19466315.2015.1004270

HOMMEL, G. and HOFFMANN, T. (1988). Controlled uncertainty. In *Multiple Hypothesenprüfung/Multiple Hypotheses Testing* 154–161. Springer, Berlin.

JANSON, L. and SU, W. (2016). Familywise error rate control via knockoffs. *Electron. J. Stat.* **10** 960–975. MR3486422 https://doi.org/10.1214/16-EJS1129

JAVANMARD, A. and MONTANARI, A. (2018). Online rules for control of false discovery rate and false discovery exceedance. *Ann. Statist.* **46** 526–554. MR3782376 https://doi.org/10.1214/17-AOS1559

KATSEVICH, E. and RAMDAS, A. (2020). Simultaneous high-probability bounds on the false discovery proportion in structured, regression, and online settings. *Ann. Statist.* **48** 3465–3487. https://doi.org/doi:10.1214/19-AOS1938

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

LEHMANN, E. L. and ROMANO, J. P. (2005a). Generalizations of the familywise error rate. *Ann. Statist.* **33** 1138–1154. MR2195631 https://doi.org/10.1214/009053605000000084

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

MARCUS, R., PERITZ, E. and GABRIEL, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* **63** 655–660. MR0468056 https://doi.org/10.1093/biomet/63.3.655

MEIJER, R. J. and GOEMAN, J. J. (2015a). Multiple testing of gene sets from gene ontology: Possibilities and pitfalls. *Brief. Bioinform.* **17** 808–818.

MEIJER, R. J. and GOEMAN, J. J. (2015b). A multiple testing method for hypotheses structured in a directed acyclic graph. *Biom. J.* **57** 123–143. MR3298222 https://doi.org/10.1002/bimj.201300253

MEIJER, R. J., KREBS, T. J. P. and GOEMAN, J. J. (2015). A region-based multiple testing method for hypotheses ordered in space or time. *Stat. Appl. Genet. Mol. Biol.* **14** 1–19. MR3305943 https://doi.org/10.1515/sagmb-2013-0075

MEINSHAUSEN, N. (2006). False discovery control for multiple tests of association under general dependence. *Scand. J. Stat.* **33** 227–237. MR2279639 https://doi.org/10.1111/j.1467-9469.2005.00488.x

MEINSHAUSEN, N. (2008). Hierarchical testing of variable importance. *Biometrika* **95** 265–278. MR2521583 https://doi.org/10.1093/biomet/asn007

MEINSHAUSEN, N. and RICE, J. (2006). Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *Ann. Statist.* **34** 373–393. MR2275246 https://doi.org/10.1214/009053605000000741

POLINE, J.-B. and MAZOYER, B. M. (1993). Analysis of individual positron emission tomography activation maps by detection of high signal-to-noise-ratio pixel clusters. *J. Cereb. Blood Flow Metab.* **13** 425–437.

ROMANO, J. P. and SHAIKH, A. M. (2006). Stepup procedures for control of generalizations of the familywise error rate. *Ann. Statist.* **34** 1850–1873. MR2283720 https://doi.org/10.1214/009053606000000461

ROMANO, J. P., SHAIKH, A. and WOLF, M. (2011). Consonance and the closure method in multiple testing. *Int. J. Biostat.* **7** Art. 12, 27. MR2775079 https://doi.org/10.2202/1557-4679.1300

ROSENBLATT, J. D., FINOS, L., WEEDA, W. D., SOLARI, A. and GOEMAN, J. J. (2018). All-resolutions inference for brain imaging. *NeuroImage* **181** 786–796. https://doi.org/10.1016/j.neuroimage.2018.07.060

ROSENBLUM, M., LIU, H. and YEN, E.-H. (2014). Optimal tests of treatment effects for the overall population and two subpopulations in randomized trials, using sparse linear programming. *J. Amer. Statist. Assoc.* **109** 1216–1228. MR3265692 https://doi.org/10.1080/01621459.2013.879063

ROSSET, S., HELLER, R., PAINSKY, A. and AHARONI, E. (2018). Optimal Procedures for Multiple Testing Problems. Preprint. Available at arXiv:1804.10256.

SARKAR, S. K. (2007). Stepup procedures controlling generalized FWER and generalized FDR. *Ann. Statist.* **35** 2405–2420. MR2382652 https://doi.org/10.1214/009053607000000398

SHAFFER, J. P. (1986). Modified sequentially rejective multiple test procedures. *J. Amer. Statist. Assoc.* **81** 826–831.

SONNEMANN, E. (1982). Allgemeine Lösungen Multipler Testprobleme. Universität Bern. Institut für Mathematische Statistik und Versicherungslehre.

SONNEMANN, E. (2008). General solutions to multiple testing problems. *Biom. J.* **50** 641–656. MR2542333 https://doi.org/10.1002/bimj.200810462

SONNEMANN, E. and FINNER, H. (1988). Vollständigkeitssätze für multiple Testprobleme. In *Multiple Hypothesenprüfung/Multiple Hypotheses Testing* 121–135. Springer, Berlin.

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

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

WESTFALL, P. and YOUNG, S. (1993). *Resampling-Based Multiple Testing*. Wiley, New York.