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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

continued

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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles—*Continued from front cover*

Vol. 48, No. 3—June 2020

THE ANNALS OF STATISTICS Vol. 48, No. 3, pp. 1231–1874 June 2020

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 48, Number 3, June 2020. Published bimonthly by the Institute of Mathematical Statistics, 3163 Somerset Drive, Cleveland, Ohio 44122, USA. Periodicals postage paid at Cleveland, Ohio, and at additional mailing offices.

POSTMASTER: Send address changes to *The Annals of Statistics*, Institute of Mathematical Statistics, Dues and Subscriptions Office, 9650 Rockville Pike, Suite L 2310, Bethesda, Maryland 20814-3998, USA.

The Annals of Statistics
2020, Vol. 48, No. 3, 1231–1254
https://doi.org/10.1214/18-AOS1697

TESTING IN HIGH-DIMENSIONAL SPIKED MODELS

BY IAIN M. JOHNSTONE[1] AND ALEXEI ONATSKI[2]

[1]*Department of Statistics, Stanford University, imj@stanford.edu*
[2]*Faculty of Economics, University of Cambridge, ao319@cam.ac.uk*

We consider the five classes of multivariate statistical problems identified by James (*Ann. Math. Stat.* **35** (1964) 475–501), which together cover much of classical multivariate analysis, plus a simpler limiting case, symmetric matrix denoising. Each of James' problems involves the eigenvalues of $E^{-1}H$ where H and E are proportional to high-dimensional Wishart matrices. Under the null hypothesis, both Wisharts are central with identity covariance. Under the alternative, the noncentrality or the covariance parameter of H has a single eigenvalue, a spike, that stands alone. When the spike is smaller than a case-specific phase transition threshold, none of the sample eigenvalues separate from the bulk, making the testing problem challenging. Using a unified strategy for the six cases, we show that the log likelihood ratio processes parameterized by the value of the subcritical spike converge to Gaussian processes with logarithmic correlation. We then derive asymptotic power envelopes for tests for the presence of a spike.

REFERENCES

[1] ABRAMOWITZ, M. and STEGUN, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. *National Bureau of Standards Applied Mathematics Series* **55**. U.S. Government Printing Office, Washington, DC. MR0167642

[2] AZAÏS, J.-M. and WSCHEBOR, M. (2002). The distribution of the maximum of a Gaussian process: Rice method revisited. In *In and Out of Equilibrium* (*Mambucaba*, 2000). *Progress in Probability* **51** 321–348. Birkhäuser, Boston, MA. MR1901961

[3] BAI, Z. D. (1999). Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica* **9** 611–677. MR1711663

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

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

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

[7] BAI, Z. D. and YIN, Y. Q. (1988). Convergence to the semicircle law. *Ann. Probab.* **16** 863–875. MR0929083

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

[9] BAIK, J. and WANG, D. (2013). On the largest eigenvalue of a Hermitian random matrix model with spiked external source II: Higher rank cases. *Int. Math. Res. Not. IMRN* **2013** 3304–3370. MR3085761 https://doi.org/10.1093/imrn/rns136

[10] BAO, Z., HU, J., PAN, G. and ZHOU, W. (2014). Canonical correlation coefficients of high-dimensional normal vectors: Finite rank case. Available at arXiv:1407.7194.

[11] CAI, J.-F., CANDÈS, E. J. and SHEN, Z. (2010). A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* **20** 1956–1982. MR2600248 https://doi.org/10.1137/080738970

[12] DHARMAWANSA, P. and JOHNSTONE, I. M. (2014). Joint density of eigenvalues in spiked multivariate models. *Stat* **3** 240–249. MR4027340 https://doi.org/10.1002/sta4.58

MSC2010 subject classifications. Primary 62E20; secondary 62H15.

Key words and phrases. Likelihood ratio test, hypergeometric function, principal components analysis, canonical correlations, matrix denoising, multiple response regression.

[13] DHARMAWANSA, P., JOHNSTONE, I. M. and ONATSKI, A. (2014). Local asymptotic normality of the spectrum of high-dimensional spiked F-ratios. Available at arXiv:1411.3875.

[14] DOBRIBAN, E. (2017). Sharp detection in PCA under correlations: All eigenvalues matter. *Ann. Statist.* **45** 1810–1833. MR3670197 https://doi.org/10.1214/16-AOS1514

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

[16] FUJIKOSHI, Y., ULYANOV, V. V. and SHIMIZU, R. (2010). *Multivariate Statistics: High-Dimensional and Large-Sample Approximations. Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2640807 https://doi.org/10.1002/9780470539873

[17] FYODOROV, Y. V., KHORUZHENKO, B. A. and SIMM, N. J. (2016). Fractional Brownian motion with Hurst index $H = 0$ and the Gaussian unitary ensemble. *Ann. Probab.* **44** 2980–3031. MR3531684 https://doi.org/10.1214/15-AOP1039

[18] JAMES, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Stat.* **35** 475–501. MR0181057 https://doi.org/10.1214/aoms/1177703550

[19] JOHNSTONE, I. M. and ONATSKI, A. (2020). Supplement to "Testing in high-dimensional spiked models." https://doi.org/10.1214/18-AOS1697SUPP

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

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

[22] MAÏDA, M. (2007). Large deviations for the largest eigenvalue of rank one deformations of Gaussian ensembles. *Electron. J. Probab.* **12** 1131–1150. MR2336602 https://doi.org/10.1214/EJP.v12-438

[23] MEHTA, M. L. (2004). *Random Matrices*, 3rd ed. *Pure and Applied Mathematics* (*Amsterdam*) **142**. Elsevier/Academic Press, Amsterdam. MR2129906

[24] MUIRHEAD, R. J. (1978). Latent roots and matrix variates: A review of some asymptotic results. *Ann. Statist.* **6** 5–33. MR0458719

[25] MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory. Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0652932

[26] 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

[27] NADAKUDITI, R. R. and SILVERSTEIN, J. W. (2010). Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples. *IEEE J. Sel. Top. Signal Process.* **4** 468–480.

[28] OLVER, F. W. J. (1997). *Asymptotics and Special Functions*. A K Peters, Natick, MA.

[29] OLVER, F. W. J., LOZIER, D. W., BOISVERT, R. F. and CLARK, C. W., eds. (2010). *NIST Handbook of Mathematical Functions*. Cambridge Univ. Press, Cambridge. MR2723248

[30] ONATSKI, A. (2007). Asymptotics of the principal components estimator of large factor models with weak factors and i.i.d. Gaussian noise. Manuscript, University of Cambridge.

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

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

[33] PARIS, R. B. (2013). Asymptotics of the Gauss hypergeometric function with large parameters, I. *J. Class. Anal.* **2** 183–203. MR3322252 https://doi.org/10.7153/jca-02-15

[34] PARIS, R. B. (2013). Asymptotics of the Gauss hypergeometric function with large parameters, II. *J. Class. Anal.* **3** 1–15. MR3322253 https://doi.org/10.7153/jca-03-01

[35] PASSEMIER, D., MCKAY, M. R. and CHEN, Y. (2015). Asymptotic linear spectral statistics for spiked Hermitian random matrices. *J. Stat. Phys.* **160** 120–150. MR3357970 https://doi.org/10.1007/s10955-015-1233-x

[36] SHABALIN, A. A. and NOBEL, A. B. (2013). Reconstruction of a low-rank matrix in the presence of Gaussian noise. *J. Multivariate Anal.* **118** 67–76. MR3054091 https://doi.org/10.1016/j.jmva.2013.03.005

[37] 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

[38] WATSON, G. N. (1944). *A Treatise on the Theory of Bessel Functions*. Cambridge Univ. Press, Cambridge; The Macmillan Company, New York. MR0010746

[39] ZHENG, S. (2012). Central limit theorems for linear spectral statistics of large dimensional F-matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 444–476. MR2954263 https://doi.org/10.1214/11-AIHP414

The Annals of Statistics
2020, Vol. 48, No. 3, 1255–1280
https://doi.org/10.1214/18-AOS1798

LIMITING LAWS FOR DIVERGENT SPIKED EIGENVALUES AND LARGEST NONSPIKED EIGENVALUE OF SAMPLE COVARIANCE MATRICES

BY T. TONY CAI[1], XIAO HAN[2] AND GUANGMING PAN[3]

[1]*Department of Statistics, The Wharton School, University of Pennsylvania, tcai@wharton.upenn.edu*
[2]*University of Science and Technology of China and Nanyang Technological University, xhan011@e.ntu.edu.sg*
[3]*School of Physical & Mathematical Sciences, Nanyang Technological University, gmpan@ntu.edu.sg*

We study the asymptotic distributions of the spiked eigenvalues and the largest nonspiked eigenvalue of the sample covariance matrix under a general covariance model with divergent spiked eigenvalues, while the other eigenvalues are bounded but otherwise arbitrary. The limiting normal distribution for the spiked sample eigenvalues is established. It has distinct features that the asymptotic mean relies on not only the population spikes but also the nonspikes and that the asymptotic variance in general depends on the population eigenvectors. In addition, the limiting Tracy–Widom law for the largest nonspiked sample eigenvalue is obtained.

Estimation of the number of spikes and the convergence of the leading eigenvectors are also considered. The results hold even when the number of the spikes diverges. As a key technical tool, we develop a central limit theorem for a type of random quadratic forms where the random vectors and random matrices involved are dependent. This result can be of independent interest.

REFERENCES

[1] ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York. MR0771294

[2] BAI, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* **71** 135–171. MR1956857 https://doi.org/10.1111/1468-0262.00392

[3] BAI, J. and NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70** 191–221. MR1926259 https://doi.org/10.1111/1468-0262.00273

[4] BAI, J. and NG, S. (2008). Large dimensional factor analysis. *Found. Trends Econ.* **3** 89–163.

[5] 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

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

[7] BAI, Z. D., YASUNORI, F. and KWOK, P. C. (2017). High-dimensional consistency of AIC and BIC for estimating the number of significant components in principal component analysis. *Ann. Statist.* To appear.

[8] 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

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

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

MSC2010 subject classifications. Primary 62H25, 60B20; secondary 60F05, 62H10.

Key words and phrases. Extreme eigenvalues, factor model, principal component analysis, sample covariance matrix, spiked covariance matrix model, Tracy–Widom distribution.

[11] BAO, Z. G., PAN, G. M. and ZHOU, W. (2014). Local density of the spectrum on the edge for sample covariance matrices with general population.

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

[13] 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

[14] 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

[15] CAI, T. T, HAN, X. and PAN, G. (2019). Supplement to "Limiting laws for divergent spiked eigenvalues and largest nonspiked eigenvalue of sample covariance matrices." https://doi.org/10.1214/18-AOS1798SUPP.

[16] 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

[17] CHAMBERLAIN, G. and ROTHSCHILD, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica* **51** 1281–1304. MR0736050 https://doi.org/10.2307/1912275

[18] CHEN, B. B. and PAN, G. M. (2012). Convergence of the largest eigenvalue of normalized sample covariance matrices when p and n both tend to infinity with their ratio converging to zero. *Bernoulli* **18** 1405–1420. MR2995802 https://doi.org/10.3150/11-BEJ381

[19] DING, X. (2015). Convergence of sample eigenvectors of spiked population model. *Comm. Statist. Theory Methods* **44** 3825–3840. MR3397166 https://doi.org/10.1080/03610926.2013.833240

[20] 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

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

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

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

[24] JIANG, T. (2005). Maxima of entries of Haar distributed matrices. *Probab. Theory Related Fields* **131** 121–144. MR2105046 https://doi.org/10.1007/s00440-004-0376-5

[25] 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

[26] 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

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

[28] 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

[29] 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

[30] ONATSKI, A. (2009). Testing hypotheses about the numbers of factors in large factor models. *Econometrica* **77** 1447–1479. MR2561070 https://doi.org/10.3982/ECTA6964

[31] ONATSKI, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Rev. Econ. Stat.* **92** 1004–1016.

[32] 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

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

[34] SHEN, D., SHEN, H., ZHU, H. and MARRON, J. S. (2013). Surprising asymptotic conical structure in critical sample eigen-directions. Available at http://arxiv.org/abs/1303.6171.

[35] STOCK, J. H. and WATSON, M. W. (2002). Forecasting using principal components from a large number of predictors. *J. Amer. Statist. Assoc.* **97** 1167–1179. MR1951271 https://doi.org/10.1198/016214502388618960

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

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

The Annals of Statistics
2020, Vol. 48, No. 3, 1281–1303
https://doi.org/10.1214/19-AOS1847

POST HOC CONFIDENCE BOUNDS ON FALSE POSITIVES USING REFERENCE FAMILIES

BY GILLES BLANCHARD[1], PIERRE NEUVIAL[2] AND ETIENNE ROQUAIN[3]

[1]*Institut für Mathematik, Universität Potsdam, gilles.blanchard@math.uni-potsdam.de*
[2]*Institut de Mathématiques de Toulouse, UMR 5219, Université de Toulouse, CNRS, pierre.neuvial@math.univ-toulouse.fr*
[3]*Sorbonne Université, Sorbonne Paris Cité, CNRS, Laboratoire de Probabilités Statistique et Modélisation, LPSM, etienne.roquain@upmc.fr*

We follow a post hoc, "user-agnostic" approach to false discovery control in a large-scale multiple testing framework, as introduced by Genovese and Wasserman [*J. Amer. Statist. Assoc.* **101** (2006) 1408–1417], Goeman and Solari [*Statist. Sci.* **26** (2011) 584–597]: the statistical guarantee on the number of correct rejections must hold for any set of candidate items, possibly selected by the user after having seen the data. To this end, we introduce a novel point of view based on a family of reference rejection sets and a suitable criterion, namely the joint familywise error rate over that family (JER for short). First, we establish how to derive post hoc bounds from a given JER control and analyze some general properties of this approach. We then develop procedures for controlling the JER in the case where reference regions are p-value level sets. These procedures adapt to dependencies and to the unknown quantity of signal (via a step-down principle). We also show interesting connections to confidence envelopes of Meinshausen [*Scand. J. Stat.* **33** (2006) 227–237]; Genovese and Wasserman [*J. Amer. Statist. Assoc.* **101** (2006) 1408–1417], the closed testing based approach of Goeman and Solari [*Statist. Sci.* **26** (2011) 584–597] and to the higher criticism of Donoho and Jin [*Ann. Statist.* **32** (2004) 962–994]. Our theoretical statements are supported by numerical experiments.

REFERENCES

ARLOT, S., BLANCHARD, G. and ROQUAIN, E. (2010). Some nonasymptotic results on resampling in high dimension. II. Multiple tests. *Ann. Statist.* **38** 83–99. MR2589317 https://doi.org/10.1214/08-AOS668

BACHOC, F., PREINERSTORFER, D. and STEINBERGER, L. (2019). Uniformly valid confidence intervals post-model-selection. *Ann. Statist.* To appear.

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

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

BENJAMINI, Y. and YEKUTIELI, D. (2005). False discovery rate-adjusted multiple confidence intervals for selected parameters. *J. Amer. Statist. Assoc.* **100** 71–93. With comments and a rejoinder by the authors. MR2156820 https://doi.org/10.1198/016214504000001907

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. (2019). R package sansSouci version 0.8.0. https://github.com/pneuvial/sanssouci.

BLANCHARD, G., NEUVIAL, P. and ROQUAIN, E. (2019). Supplement to "Post hoc confidence bounds on false positives using reference families." https://doi.org/10.1214/19-AOS1847SUPP.

MSC2010 subject classifications. Primary 62G10; secondary 62H15.

Key words and phrases. Post hoc inference, multiple testing, Simes inequality, family-wise error rate, step-down algorithm, dependence, higher criticism.

BÜHLMANN, P. and MANDOZZI, J. (2014). High-dimensional variable screening and bias in subsequent inference, with an empirical comparison. *Comput. Statist.* **29** 407–430. MR3261821 https://doi.org/10.1007/s00180-013-0436-3

DEZEURE, R., BÜHLMANN, P., MEIER, L. and MEINSHAUSEN, N. (2015). High-dimensional inference: Confidence intervals, p-values and R-software `hdi`. *Statist. Sci.* **30** 533–558. MR3432840 https://doi.org/10.1214/15-STS527

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. and VAN DER LAAN, M. J. (2008). *Multiple Testing Procedures with Applications to Genomics. Springer Series in Statistics*. Springer, New York. MR2373771 https://doi.org/10.1007/978-0-387-49317-6

DURAND, G., BLANCHARD, G., NEUVIAL, P. and ROQUAIN, E. (2018). Post hoc false positive control for spatially structured hypotheses. Preprint. Available at arXiv:1807.01470.

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

FITHIAN, W., SUN, D. and TAYLOR, J. (2014). Optimal Inference after model selection. Preprint. Available at arXiv:1410.2597.

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

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. (2019). Permutation-based simultaneous confidence bounds for the false discovery proportion. *Biometrika*. To appear.

KATSEVICH, E. and RAMDAS, A. (2018). Towards "simultaneous selective inference": Post-hoc bounds on the false discovery proportion. Preprint. Available at arXiv:1803.06790.

LEE, J. D., SUN, D. L., SUN, Y. and TAYLOR, J. E. (2016). Exact post-selection inference, with application to the lasso. *Ann. Statist.* **44** 907–927. MR3485948 https://doi.org/10.1214/15-AOS1371

LI, W. (2012). Volcano plots in analyzing differential expressions with mrna microarrays. *Journal of Bioinformatics and Computational Biology* **10** 1231003.

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. and BÜHLMANN, P. (2005). Lower bounds for the number of false null hypotheses for multiple testing of associations under general dependence structures. *Biometrika* **92** 893–907. MR2234193 https://doi.org/10.1093/biomet/92.4.893

R CORE TEAM (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

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

SIMES, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73** 751–754. MR0897872 https://doi.org/10.1093/biomet/73.3.751

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

VAN DER LAAN, M. J., DUDOIT, S. and POLLARD, K. S. (2004). Augmentation procedures for control of the generalized family-wise error rate and tail probabilities for the proportion of false positives. *Stat. Appl. Genet. Mol. Biol.* **3** Art. 15, 27. MR2101464 https://doi.org/10.2202/1544-6115.1042

WESTFALL, P. H. and YOUNG, S. S. (1993). *Resampling-Based Multiple Testing. Examples and Methods for P-Value Adjustment*. Wiley.

WOO, C.-W., KRISHNAN, A. and WAGER, T. D. (2014). Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations. *NeuroImage* **91** 412–419.

The Annals of Statistics
2020, Vol. 48, No. 3, 1304–1328
https://doi.org/10.1214/19-AOS1848

DISTRIBUTION AND CORRELATION-FREE TWO-SAMPLE TEST OF HIGH-DIMENSIONAL MEANS

BY KAIJIE XUE[1] AND FANG YAO[2]

[1]*School of Statistics and Data Science, Nankai University, kaijie@nankai.edu.cn*

[2]*Department of Probability and Statistics, School of Mathematical Sciences, Center for Statistical Science, Peking University, fyao@math.pku.edu.cn*

We propose a two-sample test for high-dimensional means that requires neither distributional nor correlational assumptions, besides some weak conditions on the moments and tail properties of the elements in the random vectors. This two-sample test based on a nontrivial extension of the one-sample central limit theorem (*Ann. Probab.* **45** (2017) 2309–2352) provides a practically useful procedure with rigorous theoretical guarantees on its size and power assessment. In particular, the proposed test is easy to compute and does not require the independently and identically distributed assumption, which is allowed to have different distributions and arbitrary correlation structures. Further desired features include weaker moments and tail conditions than existing methods, allowance for highly unequal sample sizes, consistent power behavior under fairly general alternative, data dimension allowed to be exponentially high under the umbrella of such general conditions. Simulated and real data examples have demonstrated favorable numerical performance over existing methods.

REFERENCES

[1] AYYALA, D. N., PARK, J. and ROY, A. (2017). Mean vector testing for high-dimensional dependent observations. *J. Multivariate Anal.* **153** 136–155. MR3578843 https://doi.org/10.1016/j.jmva.2016.09.012

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

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

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

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

[6] CHEN, X. (2018). Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications. *Ann. Statist.* **46** 642–678. MR3782380 https://doi.org/10.1214/17-AOS1563

[7] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. MR3161448 https://doi.org/10.1214/13-AOS1161

[8] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* **162** 47–70. MR3350040 https://doi.org/10.1007/s00440-014-0565-9

[9] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. MR3693963 https://doi.org/10.1214/16-AOP1113

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

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

MSC2010 subject classifications. 62H05, 62F05.

Key words and phrases. High-dimensional central limit theorem, Kolmogorov distance, multiplier bootstrap, power function.

[12] HU, J., BAI, Z., WANG, C. and WANG, W. (2017). On testing the equality of high dimensional mean vectors with unequal covariance matrices. *Ann. Inst. Statist. Math.* **69** 365–387. MR3611524 https://doi.org/10.1007/s10463-015-0543-8

[13] HUSSAIN, L., AZIZ, W., NADEEM, S. A., SHAH, S. A. and MAJID, A. (2015). Electroencephalography (EEG) analysis of alcoholic and control subjects using multiscale permutation entropy. *J. Multidiscip. Eng. Sci. Technol.* **1** 3159–0040.

[14] PARK, J. and AYYALA, D. N. (2013). A test for the mean vector in large dimension and small samples. *J. Statist. Plann. Inference* **143** 929–943. MR3011304 https://doi.org/10.1016/j.jspi.2012.11.001

[15] SHEN, Y. and LIN, Z. (2015). An adaptive test for the mean vector in large-p-small-n problems. *Comput. Statist. Data Anal.* **89** 25–38. MR3349665 https://doi.org/10.1016/j.csda.2015.03.004

[16] SRIVASTAVA, M. S. (2007). Multivariate theory for analyzing high dimensional data. *J. Japan Statist. Soc.* **37** 53–86. MR2392485 https://doi.org/10.14490/jjss.37.53

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

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

[19] SRIVASTAVA, M. S. and KUBOKAWA, T. (2013). Tests for multivariate analysis of variance in high dimension under non-normality. *J. Multivariate Anal.* **115** 204–216. MR3004555 https://doi.org/10.1016/j.jmva.2012.10.011

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

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

[22] XUE, K. and YAO, F. (2019). Supplement to "Distribution and correlation-free two-sample test of high-dimensional means." https://doi.org/10.1214/19-AOS1848SUPP.

[23] YAGI, A. and SEO, T. (2014). A test for mean vector and simultaneous confidence intervals with three-step monotone missing data. *Amer. J. Math. Management Sci.* **33** 161–175.

[24] YAMADA, T. and HIMENO, T. (2015). Testing homogeneity of mean vectors under heteroscedasticity in high-dimension. *J. Multivariate Anal.* **139** 7–27. MR3349477 https://doi.org/10.1016/j.jmva.2015.02.005

[25] ZHANG, J. and PAN, M. (2016). A high-dimension two-sample test for the mean using cluster subspaces. *Comput. Statist. Data Anal.* **97** 87–97. MR3447038 https://doi.org/10.1016/j.csda.2015.12.004

[26] ZHANG, X. (2015). Testing high dimensional mean under sparsity. Preprint. Available at arXiv:1509.08444v2.

[27] ZHAO, J. (2017). A new test for the mean vector in large dimension and small samples. *Comm. Statist. Simulation Comput.* **46** 6115–6128. MR3740770 https://doi.org/10.1080/03610918.2016.1197244

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

[29] ZHU, Y. and BRADIC, J. (2016). Two-sample testing in non-sparse high-dimensional linear models. Preprint. Available at arXiv:1610.04580v1.

The Annals of Statistics
2020, Vol. 48, No. 3, 1329–1347
https://doi.org/10.1214/19-AOS1849

JUST INTERPOLATE: KERNEL "RIDGELESS" REGRESSION CAN GENERALIZE

BY TENGYUAN LIANG[1] AND ALEXANDER RAKHLIN[2]

[1]*Econometrics and Statistics, Booth School of Business, University of Chicago, tengyuan.liang@chicagobooth.edu*
[2]*Center for Statistics & IDSS, Massachusetts Institute of Technology, rakhlin@mit.edu*

In the absence of explicit regularization, Kernel "Ridgeless" Regression with nonlinear kernels has the potential to fit the training data perfectly. It has been observed empirically, however, that such interpolated solutions can still generalize well on test data. We isolate a phenomenon of implicit regularization for minimum-norm interpolated solutions which is due to a combination of high dimensionality of the input data, curvature of the kernel function and favorable geometric properties of the data such as an eigenvalue decay of the empirical covariance and kernel matrices. In addition to deriving a data-dependent upper bound on the out-of-sample error, we present experimental evidence suggesting that the phenomenon occurs in the MNIST dataset.

REFERENCES

[1] ALVAREZ, M. A., ROSASCO, L. and LAWRENCE, N. D. (2012). Kernels for vector-valued functions: A review. *Found. Trends Mach. Learn.* **4** 195–266.

[2] BARTLETT, P. L. and MENDELSON, S. (2002). Rademacher and Gaussian complexities: Risk bounds and structural results. *J. Mach. Learn. Res.* **3** 463–482. MR1984026 https://doi.org/10.1162/153244303321897690

[3] BELKIN, M. (2018). Approximation beats concentration? an approximation view on inference with smooth radial kernels. ArXiv preprint. Available at arXiv:1801.03437.

[4] BELKIN, M., HSU, D. and MITRA, P. (2018). Overfitting or perfect fitting? Risk bounds for classification and regression rules that interpolate. ArXiv preprint. Available at arXiv:1806.05161.

[5] BELKIN, M., MA, S. and MANDAL, S. (2018). To understand deep learning we need to understand kernel learning. ArXiv preprint. Available at arXiv:1802.01396.

[6] BELKIN, M., RAKHLIN, A. and TSYBAKOV, A. B. (2018). Does data interpolation contradict statistical optimality? ArXiv preprint. Available at arXiv:1806.09471.

[7] BOSE, A., CHATTERJEE, S. and GANGOPADHYAY, S. (2003). Limiting spectral distributions of large dimensional random matrices. *J. Indian Statist. Assoc.* **41** 221–259. MR2101995

[8] 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

[9] CRESSIE, N. (1990). The origins of kriging. *Math. Geol.* **22** 239–252. MR1047810 https://doi.org/10.1007/BF00889887

[10] CUCKER, F. and SMALE, S. (2002). Best choices for regularization parameters in learning theory: On the bias-variance problem. *Found. Comput. Math.* **2** 413–428. MR1930945 https://doi.org/10.1007/s102080010030

[11] DE VITO, E., CAPONNETTO, A. and ROSASCO, L. (2005). Model selection for regularized least-squares algorithm in learning theory. *Found. Comput. Math.* **5** 59–85. MR2125691 https://doi.org/10.1007/s10208-004-0134-1

[12] DOU, X. and LIANG, T. (2019). Training neural networks as learning data-adaptive kernels: Provable representation and approximation benefits. ArXiv preprint. Available at arXiv:1901.07114.

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

[14] EVGENIOU, T., PONTIL, M. and POGGIO, T. (2000). Regularization networks and support vector machines. *Adv. Comput. Math.* **13** 1–50. MR1759187 https://doi.org/10.1023/A:1018946025316

MSC2010 subject classifications. 68Q32, 62G08.

Key words and phrases. Minimum-norm interpolation, reproducing kernel Hilbert spaces, implicit regularization, high dimensionality, data-dependent bounds, kernel methods, spectral decay.

[15] GOLUB, G. H., HEATH, M. and WAHBA, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* **21** 215–223. MR0533250 https://doi.org/10.2307/1268518

[16] GUNASEKAR, S., WOODWORTH, B. E., BHOJANAPALLI, S., NEYSHABUR, B. and SREBRO, N. (2017). Implicit regularization in matrix factorization. In *Advances in Neural Information Processing Systems* 6151–6159.

[17] 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

[18] KOLTCHINSKII, V. and BEZNOSOVA, O. (2005). Exponential convergence rates in classification. In *Learning Theory*. *Lecture Notes in Computer Science* **3559** 295–307. Springer, Berlin. MR2203269 https://doi.org/10.1007/11503415_20

[19] LECUN, Y., CORTES, C. and BURGES, C. J. (2010). Mnist handwritten digit database. AT&T Labs [Online]. Available at http://yann.Lecun.Com/exdb/mnist.

[20] LI, Y., MA, T. and ZHANG, H. (2017). Algorithmic regularization in over-parameterized matrix recovery. ArXiv preprint. Available at arXiv:1712.09203.

[21] LIANG, T. and RAKHLIN, A. (2019). Supplement to "Just interpolate: Kernel "Ridgeless" regression can generalize". https://doi.org/10.1214/19-AOS1849SUPP.

[22] NEYSHABUR, B., TOMIOKA, R. and SREBRO, N. (2014). In search of the real inductive bias: On the role of implicit regularization in deep learning. ArXiv preprint. Available at arXiv:1412.6614.

[23] PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A. et al. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12** 2825–2830. MR2854348

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

[25] SMOLA, A. J. and SCHÖLKOPF, B. (1998). *Learning with Kernels* **4**. Citeseer.

[26] VAPNIK, V. N. (1998). *Statistical Learning Theory*. *Adaptive and Learning Systems for Signal Processing, Communications, and Control*. Wiley, New York. MR1641250

[27] VOVK, V. (2013). Kernel ridge regression. In *Empirical Inference* 105–116. Springer, Heidelberg. MR3236860 https://doi.org/10.1007/978-3-642-41136-6_11

[28] WAHBA, G. (1990). *Spline Models for Observational Data*. *CBMS-NSF Regional Conference Series in Applied Mathematics* **59**. SIAM, Philadelphia, PA. MR1045442 https://doi.org/10.1137/1.9781611970128

[29] 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

[30] YIN, Y. Q., BAI, Z. D. and KRISHNAIAH, P. R. (1988). On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Related Fields* **78** 509–521. MR0950344 https://doi.org/10.1007/BF00353874

[31] ZHANG, C., BENGIO, S., HARDT, M., RECHT, B. and VINYALS, O. (2016). Understanding deep learning requires rethinking generalization. ArXiv preprint. Available at arXiv:1611.03530.

The Annals of Statistics
2020, Vol. 48, No. 3, 1348–1382
https://doi.org/10.1214/19-AOS1850

BRIDGING THE GAP BETWEEN CONSTANT STEP SIZE STOCHASTIC GRADIENT DESCENT AND MARKOV CHAINS

BY AYMERIC DIEULEVEUT[1], ALAIN DURMUS[2] AND FRANCIS BACH[3]

[1]*CMAP, École Polytechnique (X), aymeric.dieuleveut@polytechnique.edu*
[2]*CMLA - École normale supérieure Paris-Saclay, CNRS, Université Paris-Saclay, alain.durmus@cmla.ens-cachan.fr*
[3]*INRIA - Département d'informatique de l'ENS, École normale supérieure, CNRS, PSL Research University, francis.bach@inria.fr*

We consider the minimization of a strongly convex objective function given access to unbiased estimates of its gradient through stochastic gradient descent (SGD) with constant step size. While the detailed analysis was only performed for quadratic functions, we provide an explicit asymptotic expansion of the moments of the averaged SGD iterates that outlines the dependence on initial conditions, the effect of noise and the step size, as well as the lack of convergence in the general (nonquadratic) case. For this analysis we bring tools from Markov chain theory into the analysis of stochastic gradient. We then show that Richardson–Romberg extrapolation may be used to get closer to the global optimum, and we show empirical improvements of the new extrapolation scheme.

REFERENCES

[1] ABDULLE, A., VILMART, G. and ZYGALAKIS, K. C. (2014). High order numerical approximation of the invariant measure of ergodic SDEs. *SIAM J. Numer. Anal.* **52** 1600–1622. MR3229658 https://doi.org/10.1137/130935616

[2] AGUECH, R., MOULINES, E. and PRIOURET, P. (2000). On a perturbation approach for the analysis of stochastic tracking algorithms. *SIAM J. Control Optim.* **39** 872–899. MR1786334 https://doi.org/10.1137/S0363012998333852

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

[4] BACH, F. and MOULINES, E. (2011). Non-asymptotic analysis of stochastic approximation algorithms for machine learning. In *Advances on Neural Information Processing Systems* (*NIPS*) 451–459.

[5] BACH, F. and MOULINES, E. (2013). Non-strongly-convex smooth stochastic approximation with convergence rate $O(1/n)$. In *Advances in Neural Information Processing Systems* (*NIPS*).

[6] BENAIM, M. (1996). A dynamical system approach to stochastic approximations. *SIAM J. Control Optim.* **34** 437–472. MR1377706 https://doi.org/10.1137/S0363012993253534

[7] BENVENISTE, A., MÉTIVIER, M. and PRIOURET, P. (1990). *Adaptive Algorithms and Stochastic Approximations*. *Applications of Mathematics* (*New York*) **22**. Springer, Berlin. Translated from the French by Stephen S. Wilson. MR1082341 https://doi.org/10.1007/978-3-642-75894-2

[8] BERTSEKAS, D. P. (1999). *Nonlinear Programming*, 2nd ed. *Athena Scientific Optimization and Computation Series*. Athena Scientific, Belmont, MA. MR3444832

[9] BOTTOU, L. and BOUSQUET, O. (2008). The tradeoffs of large scale learning. In *Advances in Neural Information Processing Systems* (*NIPS*).

[10] BOUTON, C. and PAGÈS, G. (1997). About the multidimensional competitive learning vector quantization algorithm with constant gain. *Ann. Appl. Probab.* **7** 679–710. MR1459266 https://doi.org/10.1214/aoap/1034801249

[11] CHEE, J. and TOULIS, P. (2017). Convergence diagnostics for stochastic gradient descent with constant step size. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, (*AISTATS*).

[12] CHEN, C., DING, N. and CARIN, L. (2015). On the convergence of stochastic gradient MCMC algorithms with high-order integrators. In *Advances on Neural Information Processing Systems* (*NIPS*) 2269–2277.

MSC2010 subject classifications. Primary 62L20; secondary 90C15, 93E35.
Key words and phrases. Stochastic gradient descent, Markov chains.

[13] CHEN, X., LEE, J. D., TONG, X. T. and ZHANG, Y. (2016). Statistical inference for model parameters in stochastic gradient descent. ArXiv preprint. Available at arXiv:1610.08637.

[14] DALALYAN, A. S. (2017). Theoretical guarantees for approximate sampling from smooth and log-concave densities. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 651–676. MR3641401 https://doi.org/10.1111/rssb.12183

[15] DÉFOSSEZ, A. and BACH, F. (2015). Averaged least-mean-squares: Bias-variance trade-offs and optimal sampling distributions. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, (*AISTATS*).

[16] DIEULEVEUT, A. and BACH, F. (2016). Nonparametric stochastic approximation with large step-sizes. *Ann. Statist.* **44** 1363–1399. MR3519927 https://doi.org/10.1214/15-AOS1391

[17] DIEULEVEUT, A., DURMUS, A. and BACH, F. (2019). Supplement to "Bridging the gap between constant step size stochastic gradient descent and Markov Chains." https://doi.org/10.1214/19-AOS1850SUPP.

[18] DIEULEVEUT, A., FLAMMARION, N. and BACH, F. (2017). Harder, better, faster, stronger convergence rates for least-squares regression. *J. Mach. Learn. Res.* **18** 101. MR3725440

[19] DURMUS, A. and MOULINES, É. (2017). Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *Ann. Appl. Probab.* **27** 1551–1587. MR3678479 https://doi.org/10.1214/16-AAP1238

[20] DURMUS, A., ŞIMŞEKLI, U., MOULINES, E., BADEAU, R. and RICHARD, G. (2016). Stochastic gradient Richardson–Romberg Markov chain Monte Carlo. In *Advances in Neural Information Processing Systems* 2047–2055.

[21] FORT, J.-C. and PAGÈS, G. (1996). Convergence of stochastic algorithms: From the Kushner–Clark theorem to the Lyapounov functional method. *Adv. in Appl. Probab.* **28** 1072–1094. MR1418247 https://doi.org/10.2307/1428165

[22] FORT, J.-C. and PAGÈS, G. (1999). Asymptotic behavior of a Markovian stochastic algorithm with constant step. *SIAM J. Control Optim.* **37** 1456–1482. MR1710228 https://doi.org/10.1137/S0363012997328610

[23] FREIDLIN, M. I. and WENTZELL, A. D. (1998). *Random Perturbations of Dynamical Systems*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **260**. Springer, New York. MR1652127 https://doi.org/10.1007/978-1-4612-0611-8

[24] GLYNN, P. W. and MEYN, S. P. (1996). A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.* **24** 916–931. MR1404536 https://doi.org/10.1214/aop/1039639370

[25] HARTMAN, P. (2002). *Ordinary Differential Equations*: *Second Edition*. *Classics in Applied Mathematics* **38**. SIAM, Philadelphia, PA. Corrected reprint of the second (1982) edition [Birkhäuser, Boston, MA; MR0658490 (83e:34002)], With a foreword by Peter Bates. MR1929104 https://doi.org/10.1137/1.9780898719222

[26] HE, K., ZHANG, X., REN, S. and SUN, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 770–778.

[27] JAIN, P., KAKADE, S. M., KIDAMBI, R., NETRAPALLI, P. and SIDFORD, A. (2018). Accelerating stochastic gradient descent. In *Proceedings of the International Conference on Learning Theory* (*COLT*).

[28] JAIN, P., KAKADE, S. M., KIDAMBI, R., NETRAPALLI, P. and SIDFORD, A. (2017). Parallelizing stochastic gradient descent for least squares regression: Mini-batching, averaging, and model misspecification. *J. Mach. Learn. Res.* **18** 223. MR3827111

[29] KRIZHEVSKY, A., SUTSKEVER, I. and HINTON, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (*NIPS*) 1097–1105.

[30] KUSHNER, H. J. and CLARK, D. S. (1978). *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. *Applied Mathematical Sciences* **26**. Springer, New York-Berlin. MR0499560

[31] LAN, G. (2012). An optimal method for stochastic composite optimization. *Math. Program.* **133** 365–397. MR2921104 https://doi.org/10.1007/s10107-010-0434-y

[32] LEVY, E. (2006). Why do partitions occur in Faa di Bruno's chain rule for higher derivatives? Technical Report 0602183.

[33] LJUNG, L. (1977). Analysis of recursive stochastic algorithms. *IEEE Trans. Automat. Control* **AC-22** 551–575. MR0465458 https://doi.org/10.1109/tac.1977.1101561

[34] LJUNG, L., PFLUG, G. and WALK, H. (1992). *Stochastic Approximation and Optimization of Random Systems*. *DMV Seminar* **17**. Birkhäuser, Basel. MR1162311 https://doi.org/10.1007/978-3-0348-8609-3

[35] MANDT, S., HOFFMAN, M. and BLEI, D. M. (2016). A variational analysis of stochastic gradient algorithms. In *International Conference on Machine Learning* 354–363.

[36] MANDT, S., HOFFMAN, M. D. and BLEI, D. M. (2017). Stochastic gradient descent as approximate Bayesian inference. *J. Mach. Learn. Res.* **18** 134. MR3763768

[37] MATTINGLY, J. C., STUART, A. M. and HIGHAM, D. J. (2002). Ergodicity for SDEs and approximations: Locally Lipschitz vector fields and degenerate noise. *Stochastic Process. Appl.* **101** 185–232. MR1931266 https://doi.org/10.1016/S0304-4149(02)00150-3

[38] MÉTIVIER, M. and PRIOURET, P. (1984). Applications of a Kushner and Clark lemma to general classes of stochastic algorithms. *IEEE Trans. Inform. Theory* **30** 140–151. MR0807052 https://doi.org/10.1109/TIT.1984.1056894

[39] MÉTIVIER, M. and PRIOURET, P. (1987). Théorèmes de convergence presque sure pour une classe d'algorithmes stochastiques à pas décroissant. *Probab. Theory Related Fields* **74** 403–428. MR0873887 https://doi.org/10.1007/BF00699098

[40] MEYN, S. and TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2509253 https://doi.org/10.1017/CBO9780511626630

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

[42] NEDIĆ, A. and BERTSEKAS, D. (2001). Convergence rate of incremental subgradient algorithms. In *Stochastic Optimization*: *Algorithms and Applications* (*Gainesville*, *FL*, 2000). *Appl. Optim.* **54** 223–264. Kluwer Academic, Dordrecht. MR1835501 https://doi.org/10.1007/978-1-4757-6594-6_11

[43] NEEDELL, D., WARD, R. and SREBRO, N. (2014). Stochastic gradient descent, weighted sampling, and the randomized Kaczmarz algorithm. In *Advances in Neural Information Processing Systems* (*NIPS*), 1017–1025.

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

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

[46] NESTEROV, Y. and VIAL, J.-P. (2008). Confidence level solutions for stochastic programming. *Automatica J. IFAC* **44** 1559–1568. MR2531843 https://doi.org/10.1016/j.automatica.2008.01.017

[47] PFLUG, G. C. (1986). Stochastic minimization with constant step-size: Asymptotic laws. *SIAM J. Control Optim.* **24** 655–666. MR0846373 https://doi.org/10.1137/0324039

[48] 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

[49] PRIOURET, P. and VERETENNIKOV, A. Y. (1998). A remark on the stability of the LMS tracking algorithm. *Stoch. Anal. Appl.* **16** 119–129. MR1603888 https://doi.org/10.1080/07362999808809521

[50] RAKHLIN, A., SHAMIR, O. and SRIDHARAN, K. (2011). Making gradient descent optimal for strongly convex stochastic optimization. In *Proceedings of the* 29*th International Conference on Machine Learning*.

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

[52] RUPPERT, D. (1988). Efficient estimations from a slowly convergent Robbins–Monro process Technical report, Cornell University Operations Research and Industrial Engineering.

[53] SHALEV-SHWARTZ, S., SHAMIR, O., SREBRO, N. and SRIDHARAN, K. (2009). Stochastic convex optimization. In *Proceedings of the International Conference on Learning Theory* (*COLT*).

[54] SHALEV-SHWARTZ, S., SINGER, Y. and SREBRO, N. (2007). Pegasos: Primal estimated sub-GrAdient SOlver for SVM. In *Proceedings of the International Conference on Machine Learning*, *ICML* 807–814.

[55] SHAMIR, O. and ZHANG, T. (2013). Stochastic gradient descent for non-smooth optimization: Convergence results and optimal averaging schemes. In *Proceedings of the International Conference on Machine Learning*.

[56] STOER, J. and BULIRSCH, R. (2002). *Introduction to Numerical Analysis*, 3rd ed. *Texts in Applied Mathematics* **12**. Springer, New York. MR1923481 https://doi.org/10.1007/978-0-387-21738-3

[57] SU, W. J. and ZHU, Y. (2018). Uncertainty quantification for online learning and stochastic approximation via hierarchical incremental gradient descent. ArXiv preprint. Available at arXiv:1802.04876.

[58] TADIĆ, V. B. and DOUCET, A. (2017). Asymptotic bias of stochastic gradient search. *Ann. Appl. Probab.* **27** 3255–3304. MR3737925 https://doi.org/10.1214/16-AAP1272

[59] TALAY, D. and TUBARO, L. (1990). Expansion of the global error for numerical schemes solving stochastic differential equations. *Stoch. Anal. Appl.* **8** 483–509. MR1091544 https://doi.org/10.1080/07362999008809220

[60] 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

[61] WELLING, M. and TEH, Y. W. (2011). Bayesian learning via Stochastic Gradient Langevin Dynamics. In *Proceedings of the International Conference on Machine Learning* (*ICML*) 681–688.

[62] ZHU, D. L. and MARCOTTE, P. (1996). Co-coercivity and its role in the convergence of iterative schemes for solving variational inequalities. *SIAM J. Optim.* **6** 714–726. MR1402202 https://doi.org/10.1137/S1052623494250415

The Annals of Statistics
2020, Vol. 48, No. 3, 1383–1408
https://doi.org/10.1214/19-AOS1851

NONPARAMETRIC STATISTICAL INFERENCE FOR DRIFT VECTOR FIELDS OF MULTI-DIMENSIONAL DIFFUSIONS

BY RICHARD NICKL[1] AND KOLYAN RAY[2]

[1]*Statistical Laboratory, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, r.nickl@statslab.cam.ac.uk*

[2]*Department of Mathematics, King's College London, kolyan.ray@kcl.ac.uk*

The problem of determining a periodic Lipschitz vector field $b = (b_1, \dots, b_d)$ from an observed trajectory of the solution $(X_t : 0 \le t \le T)$ of the multi-dimensional stochastic differential equation

$$dX_t = b(X_t)\,dt + dW_t, \quad t \ge 0,$$

where W_t is a standard d-dimensional Brownian motion, is considered. Convergence rates of a penalised least squares estimator, which equals the maximum a posteriori (MAP) estimate corresponding to a high-dimensional Gaussian product prior, are derived. These results are deduced from corresponding contraction rates for the associated posterior distributions. The rates obtained are optimal up to log-factors in L^2-loss in any dimension, and also for supremum norm loss when $d \le 4$. Further, when $d \le 3$, nonparametric Bernstein–von Mises theorems are proved for the posterior distributions of b. From this, we deduce functional central limit theorems for the implied estimators of the invariant measure μ_b. The limiting Gaussian process distributions have a covariance structure that is asymptotically optimal from an information-theoretic point of view.

REFERENCES

[1] ABRAHAM, K. (2018). Nonparametric Bayesian posterior contraction rates for scalar diffusions with high-frequency data. *Bernoulli*. To appear. Available at arXiv:1802.05635.

[2] AECKERLE-WILLEMS, C. and STRAUCH, C. (2018). Concentration of scalar ergodic diffusions and some statistical implications. Preprint. Available at arXiv:1807.11331.

[3] AECKERLE-WILLEMS, C. and STRAUCH, C. (2018). Sup-norm adaptive simultaneous drift estimation for ergodic diffusions. Preprint. Available at arXiv:1808.10660.

[4] AGAPIOU, S., LARSSON, S. and STUART, A. M. (2013). Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems. *Stochastic Process. Appl.* **123** 3828–3860. MR3084161 https://doi.org/10.1016/j.spa.2013.05.001

[5] BASS, R. F. (2011). *Stochastic Processes. Cambridge Series in Statistical and Probabilistic Mathematics* **33**. Cambridge Univ. Press, Cambridge. MR2856623 https://doi.org/10.1017/CBO9780511997044

[6] BESKOS, A., PAPASPILIOPOULOS, O., ROBERTS, G. O. and FEARNHEAD, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 333–382. MR2278331 https://doi.org/10.1111/j.1467-9868.2006.00552.x

[7] BLADT, M., FINCH, S. and SØRENSEN, M. (2016). Simulation of multivariate diffusion bridges. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 343–369. MR3454200 https://doi.org/10.1111/rssb.12118

[8] CASTILLO, I. (2014). On Bayesian supremum norm contraction rates. *Ann. Statist.* **42** 2058–2091. MR3262477 https://doi.org/10.1214/14-AOS1253

[9] CASTILLO, I. (2017). Pólya tree posterior distributions on densities. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 2074–2102. MR3729648 https://doi.org/10.1214/16-AIHP784

[10] 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

MSC2010 subject classifications. Primary 62G20; secondary 62F15, 65N21.

Key words and phrases. Penalised least squares estimator, asymptotics of nonparametric Bayes procedures, Bernstein–von Mises theorem, uncertainty quantification.

[11] 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

[12] CASTILLO, I. and ROUSSEAU, J. (2015). A Bernstein–von Mises theorem for smooth functionals in semiparametric models. *Ann. Statist.* **43** 2353–2383. MR3405597 https://doi.org/10.1214/15-AOS1336

[13] DALALYAN, A. (2005). Sharp adaptive estimation of the drift function for ergodic diffusions. *Ann. Statist.* **33** 2507–2528. MR2253093 https://doi.org/10.1214/009053605000000615

[14] DALALYAN, A. and REISS, M. (2006). Asymptotic statistical equivalence for scalar ergodic diffusions. *Probab. Theory Related Fields* **134** 248–282. MR2222384 https://doi.org/10.1007/s00440-004-0416-1

[15] 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

[16] DASHTI, M., LAW, K. J. H., STUART, A. M. and VOSS, J. (2013). MAP estimators and their consistency in Bayesian nonparametric inverse problems. *Inverse Probl.* **29** 095017, 27. MR3104933 https://doi.org/10.1088/0266-5611/29/9/095017

[17] 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

[18] DUDLEY, R. M. (2014). *Uniform Central Limit Theorems*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **142**. Cambridge Univ. Press, New York. MR3445285

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

[20] 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

[21] GINÉ, E. and NICKL, R. (2008). Uniform central limit theorems for kernel density estimators. *Probab. Theory Related Fields* **141** 333–387. MR2391158 https://doi.org/10.1007/s00440-007-0087-9

[22] GINÉ, E. and NICKL, R. (2009). An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation. *Probab. Theory Related Fields* **143** 569–596. MR2475673 https://doi.org/10.1007/s00440-008-0137-y

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

[24] GIORDANO, M. and KEKKONEN, H. (2018). Bernstein–von Mises theorems and uncertainty quantification for linear inverse problems. Preprint. Available at arXiv:1811.04058.

[25] GUGUSHVILI, S. and SPREIJ, P. (2014). Nonparametric Bayesian drift estimation for multidimensional stochastic differential equations. *Lith. Math. J.* **54** 127–141. MR3212631 https://doi.org/10.1007/s10986-014-9232-1

[26] HORN, R. A. and JOHNSON, C. R. (2013). *Matrix Analysis*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2978290

[27] KNAPIK, B. and SALOMOND, J.-B. (2018). A general approach to posterior contraction in nonparametric inverse problems. *Bernoulli* **24** 2091–2121. MR3757524 https://doi.org/10.3150/16-BEJ921

[28] 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

[29] KUTOYANTS, Y. A. (2004). *Statistical Inference for Ergodic Diffusion Processes*. *Springer Series in Statistics*. Springer, London. MR2144185 https://doi.org/10.1007/978-1-4471-3866-2

[30] LÖCHERBACH, E., LOUKIANOVA, D. and LOUKIANOV, O. (2011). Penalized nonparametric drift estimation for a continuously observed one-dimensional diffusion process. *ESAIM Probab. Stat.* **15** 197–216. MR2870512 https://doi.org/10.1051/ps/2009016

[31] MONARD, F., NICKL, R. and PATERNAIN, G. P. (2019). Efficient nonparametric Bayesian inference for X-ray transforms. *Ann. Statist.* **47** 1113–1147. MR3909962 https://doi.org/10.1214/18-AOS1708

[32] NICKL, R. and RAY, K. (2019). Supplement to "Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions." https://doi.org/10.1214/19-AOS1851SUPP.

[33] NICKL, R. and SÖHL, J. (2017). Bernstein–von Mises theorems for statistical inverse problems II: Compound Poisson processes. Preprint. Available at arXiv:1709.07752.

[34] NICKL, R. and SÖHL, J. (2017). Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions. *Ann. Statist.* **45** 1664–1693. MR3670192 https://doi.org/10.1214/16-AOS1504

[35] NICKL, R. and SÖHL, J. (2019). Bernstein–von Mises theorems for statistical inverse problems II: Compound Poisson processes. *Electron. J. Stat.* **13** 3513–3571. MR4013745 https://doi.org/10.1214/19-ejs1609

[36] NICKL, R., VAN DE GEER, S. and WANG, S. (2018). Convergence rates for penalised least squares estimators in PDE-constrained regression problems. Preprint. Available at arXiv:1809.08818.

[37] PAPASPILIOPOULOS, O., POKERN, Y., ROBERTS, G. O. and STUART, A. M. (2012). Nonparametric estimation of diffusions: A differential equations approach. *Biometrika* **99** 511–531. MR2966767 https://doi.org/10.1093/biomet/ass034

[38] PARDOUX, E. and VERETENNIKOV, A. Y. (2001). On the Poisson equation and diffusion approximation. I. *Ann. Probab.* **29** 1061–1085. MR1872736 https://doi.org/10.1214/aop/1015345596

[39] POKERN, Y., STUART, A. M. and VAN ZANTEN, J. H. (2013). Posterior consistency via precision operators for Bayesian nonparametric drift estimation in SDEs. *Stochastic Process. Appl.* **123** 603–628. MR3003365 https://doi.org/10.1016/j.spa.2012.08.010

[40] RAY, K. (2013). Bayesian inverse problems with non-conjugate priors. *Electron. J. Stat.* **7** 2516–2549. MR3117105 https://doi.org/10.1214/13-EJS851

[41] RAY, K. (2014). Asymptotic theory for Bayesian nonparametric procedures in inverse problems. Ph.D. thesis, University of Cambridge.

[42] 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

[43] REVUZ, D. and YOR, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **293**. Springer, Berlin. MR1725357 https://doi.org/10.1007/978-3-662-06400-9

[44] RIVOIRARD, V. and ROUSSEAU, J. (2012). Bernstein–von Mises theorem for linear functionals of the density. *Ann. Statist.* **40** 1489–1523. MR3015033 https://doi.org/10.1214/12-AOS1004

[45] RUNST, T. and SICKEL, W. (1996). *Sobolev Spaces of Fractional Order, Nemytskij Operators, and Nonlinear Partial Differential Equations. De Gruyter Series in Nonlinear Analysis and Applications* **3**. de Gruyter, Berlin. MR1419319 https://doi.org/10.1515/9783110812411

[46] SCHAUER, M., VAN DER MEULEN, F. and VAN ZANTEN, H. (2017). Guided proposals for simulating multi-dimensional diffusion bridges. *Bernoulli* **23** 2917–2950. MR3648050 https://doi.org/10.3150/16-BEJ833

[47] SCHMEISSER, H.-J. and TRIEBEL, H. (1987). *Topics in Fourier Analysis and Function Spaces. A Wiley-Interscience Publication*. Wiley, Chichester. MR0891189

[48] SCHMISSER, E. (2013). Penalized nonparametric drift estimation for a multidimensional diffusion process. *Statistics* **47** 61–84. MR3023015 https://doi.org/10.1080/02331888.2011.591931

[49] STRAUCH, C. (2015). Sharp adaptive drift estimation for ergodic diffusions: The multivariate case. *Stochastic Process. Appl.* **125** 2562–2602. MR3332848 https://doi.org/10.1016/j.spa.2015.02.003

[50] STRAUCH, C. (2016). Exact adaptive pointwise drift estimation for multidimensional ergodic diffusions. *Probab. Theory Related Fields* **164** 361–400. MR3449393 https://doi.org/10.1007/s00440-014-0614-4

[51] 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

[52] VAN WAAIJ, J. and VAN ZANTEN, H. (2016). Gaussian process methods for one-dimensional diffusions: Optimal rates and adaptation. *Electron. J. Stat.* **10** 628–645. MR3471991 https://doi.org/10.1214/16-EJS1117

[53] VAN DER MEULEN, F. and SCHAUER, M. (2017). Bayesian estimation of discretely observed multi-dimensional diffusion processes using guided proposals. *Electron. J. Stat.* **11** 2358–2396. MR3656495 https://doi.org/10.1214/17-EJS1290

[54] VAN DER MEULEN, F. and VAN ZANTEN, H. (2013). Consistent nonparametric Bayesian inference for discretely observed scalar diffusions. *Bernoulli* **19** 44–63. MR3019485 https://doi.org/10.3150/11-BEJ385

[55] VAN DER MEULEN, F. H., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2006). Convergence rates of posterior distributions for Brownian semimartingale models. *Bernoulli* **12** 863–888. MR2265666 https://doi.org/10.3150/bj/1161614950

[56] 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

[57] VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2008). Rates of contraction of posterior distributions based on Gaussian process priors. *Ann. Statist.* **36** 1435–1463. MR2418663 https://doi.org/10.1214/009053607000000613

[58] VERSHYNIN, R. (2010). Introduction to the non-asymptotic analysis of random matrices. Preprint. Available at arXiv:1011.3027.

The Annals of Statistics
2020, Vol. 48, No. 3, 1409–1431
https://doi.org/10.1214/19-AOS1852

ROBUST INFERENCE WITH KNOCKOFFS

BY RINA FOYGEL BARBER[1], EMMANUEL J. CANDÈS[2] AND RICHARD J. SAMWORTH[3]

[1]*Department of Statistics, The University of Chicago, rina@uchicago.edu*
[2]*Departments of Statistics and Mathematics, Stanford University, candes@stanford.edu*
[3]*Statistical Laboratory, University of Cambridge, r.samworth@statslab.cam.ac.uk*

We consider the variable selection problem, which seeks to identify important variables influencing a response Y out of many candidate features $X_1, \dots, X_p$. We wish to do so while offering finite-sample guarantees about the fraction of false positives—selected variables X_j that in fact have no effect on Y after the other features are known. When the number of features p is large (perhaps even larger than the sample size n), and we have no prior knowledge regarding the type of dependence between Y and X, the model-X knockoffs framework nonetheless allows us to select a model with a guaranteed bound on the false discovery rate, as long as the distribution of the feature vector $X = (X_1, \dots, X_p)$ is exactly known. This model selection procedure operates by constructing "knockoff copies" of each of the p features, which are then used as a control group to ensure that the model selection algorithm is not choosing too many irrelevant features. In this work, we study the practical setting where the distribution of X can only be estimated, rather than known exactly, and the knockoff copies of the X_j's are therefore constructed somewhat incorrectly. Our results, which are free of any modeling assumption whatsoever, show that the resulting model selection procedure incurs an inflation of the false discovery rate that is proportional to our errors in estimating the distribution of each feature X_j conditional on the remaining features $\{X_k : k \neq j\}$. The model-X knockoffs framework is therefore robust to errors in the underlying assumptions on the distribution of X, making it an effective method for many practical applications, such as genome-wide association studies, where the underlying distribution on the features $X_1, \dots, X_p$ is estimated accurately but not known exactly.

REFERENCES

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

BARBER, R. F, CANDÈS, E. J and SAMWORTH, R. J (2019). Supplement to "Robust inference with knockoffs." https://doi.org/10.1214/19-AOS1852SUPP.

BARBER, R. F. and CANDÈS, E. J. (2019). A knockoff filter for high-dimensional selective inference. *Ann. Statist.* (to appear).

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

CANDÈS, E., FAN, Y., JANSON, L. and LV, J. (2018). Panning for gold: 'model-X' knockoffs for high dimensional controlled variable selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 551–577. MR3798878 https://doi.org/10.1111/rssb.12265

FAN, Y., DEMIRKAYA, E., LI, G. and LV, J. (2019). RANK: Large-scale inference with graphical nonlinear knockoffs. *J. Amer. Statist. Assoc.* (to appear). https://doi.org/10.1080/01621459.2018.1546589.

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

MSC2010 subject classifications. 62F03, 62F35, 62G10, 62G35.

Key words and phrases. Knockoffs, variable selection, false discovery rate (FDR), high-dimensional regression, robustness.

FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441.

HOWIE, B., FUCHSBERGER, C., STEPHENS, M., MARCHINI, J. and ABECASIS, G. R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44** 955–959.

JANSON, L. (2017). A model-free approach to high-dimensional inference Ph.D. thesis, Stanford University.

LI, N. and STEPHENS, M. (2003). Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* **165** 2213–2233.

MARCHINI, J. and HOWIE, B. (2010). Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11** 499–511.

QIN, Z. S., NIU, T. and LIU, J. S. (2002). Partition-ligation-expectation-maximization algorithm for haplotype inference with single-nucleotide polymorphisms. *Am. J. Hum. Genet.* **71** 1242–1247.

RAVIKUMAR, P., WAINWRIGHT, M. J., RASKUTTI, G. and YU, B. (2011). High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electron. J. Stat.* **5** 935–980. MR2836766 https://doi.org/10.1214/11-EJS631

ROMANO, Y., SESIA, M. and CANDÈS, E. J. (2018). Deep knockoffs. ArXiv preprint. Available at arXiv:1811.06687.

SESIA, M., SABATTI, C. and CANDÈS, E. J. (2019). Gene hunting with hidden Markov model knockoffs. *Biometrika* **106** 1–18. MR3912377 https://doi.org/10.1093/biomet/asy033

STEPHENS, M., SMITH, N. J. and DONNELLY, P. (2001). A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68** 978–989.

STOREY, J. D., TAYLOR, J. E. and SIEGMUND, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: A unified approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 187–205. MR2035766 https://doi.org/10.1111/j.1467-9868.2004.00439.x

YUAN, M. and LIN, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94** 19–35. MR2367824 https://doi.org/10.1093/biomet/asm018

ZHANG, K., DENG, M., CHEN, T., WATERMAN, M. S. and SUN, F. (2002). A dynamic programming algorithm for haplotype block partitioning. *Proc. Natl. Acad. Sci. USA* **99** 7335–7339.

The Annals of Statistics
2020, Vol. 48, No. 3, 1432–1451
https://doi.org/10.1214/19-AOS1853

NONPARAMETRIC BAYESIAN ANALYSIS OF THE COMPOUND POISSON PRIOR FOR SUPPORT BOUNDARY RECOVERY

BY MARKUS REISS[1] AND JOHANNES SCHMIDT-HIEBER[2]

[1]*Institut für Mathematik, Humboldt-Universität zu Berlin, mreiss@math.hu-berlin.de*
[2]*Department of Applied Mathematics, University of Twente, a.j.schmidt-hieber@utwente.nl*

Given data from a Poisson point process with intensity $(x, y) \mapsto n\mathbf{1}(f(x) \leq y)$, frequentist properties for the Bayesian reconstruction of the support boundary function f are derived. We mainly study compound Poisson process priors with fixed intensity proving that the posterior contracts with nearly optimal rate for monotone support boundaries and adapts to Hölder smooth boundaries. We then derive a limiting shape result for a compound Poisson process prior and a function space with increasing parameter dimension. It is shown that the marginal posterior of the mean functional performs an automatic bias correction and contracts with a faster rate than the MLE. In this case, $(1-\alpha)$-credible sets are also asymptotic $(1-\alpha)$-confidence intervals. As a negative result, it is shown that the frequentist coverage of credible sets is lost for linear functions f outside the function class.

REFERENCES

[1] BONTEMPS, D. (2011). Bernstein–von Mises theorems for Gaussian regression with increasing number of regressors. *Ann. Statist.* **39** 2557–2584. MR2906878 https://doi.org/10.1214/11-AOS912

[2] CASTILLO, I. and ROUSSEAU, J. (2015). A Bernstein–von Mises theorem for smooth functionals in semiparametric models. *Ann. Statist.* **43** 2353–2383. MR3405597 https://doi.org/10.1214/15-AOS1336

[3] CASTILLO, I., SCHMIDT-HIEBER, J. and VAN DER VAART, A. (2015). Bayesian linear regression with sparse priors. *Ann. Statist.* **43** 1986–2018. MR3375874 https://doi.org/10.1214/15-AOS1334

[4] CASTILLO, I. and VAN DER VAART, A. (2012). Needles and straw in a haystack: Posterior concentration for possibly sparse sequences. *Ann. Statist.* **40** 2069–2101. MR3059077 https://doi.org/10.1214/12-AOS1029

[5] 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

[6] CHERNOZHUKOV, V. and HONG, H. (2004). Likelihood estimation and inference in a class of nonregular econometric models. *Econometrica* **72** 1445–1480. MR2077489 https://doi.org/10.1111/j.1468-0262.2004.00540.x

[7] CHIPMAN, H. A., GEORGE, E. I. and MCCULLOCH, R. E. (2010). BART: Bayesian additive regression trees. *Ann. Appl. Stat.* **4** 266–298. MR2758172 https://doi.org/10.1214/09-AOAS285

[8] CORAM, M. and LALLEY, S. P. (2006). Consistency of Bayes estimators of a binary regression function. *Ann. Statist.* **34** 1233–1269. MR2278357 https://doi.org/10.1214/009053606000000236

[9] DENISON, D. G. T., MALLICK, B. K. and SMITH, A. F. M. (1998). A Bayesian CART algorithm. *Biometrika* **85** 363–377. MR1649118 https://doi.org/10.1093/biomet/85.2.363

[10] EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (2003). *Modelling Extremal Events*: *For Insurance and Finance*. *Applications of Mathematics* (*New York*) **33**. Springer, New York. MR1458613 https://doi.org/10.1007/978-3-642-33483-2

[11] 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

[12] GAO, C., HAN, F. and ZHANG, C.-H. (2017). On estimation of isotonic piecewise constant signals. Preprint. Available at arXiv:1705.06386. *Ann. Statist.* (to appear).

MSC2010 subject classifications. Primary 62C10, 62G05; secondary 60G55.

Key words and phrases. Frequentist Bayes analysis, posterior contraction, Bernstein–von Mises theorem, Poisson point process, boundary detection, compound Poisson process, subordinator prior.

[13] GHOSAL, S. (1999). Asymptotic normality of posterior distributions in high-dimensional linear models. *Bernoulli* **5** 315–331. MR1681701 https://doi.org/10.2307/3318438

[14] GHOSAL, S. (2000). Asymptotic normality of posterior distributions for exponential families when the number of parameters tends to infinity. *J. Multivariate Anal.* **74** 49–68. MR1790613 https://doi.org/10.1006/jmva.1999.1874

[15] GHOSAL, S., GHOSH, J. K. and SAMANTA, T. (1995). On convergence of posterior distributions. *Ann. Statist.* **23** 2145–2152. MR1389869 https://doi.org/10.1214/aos/1034713651

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

[17] GIJBELS, I., MAMMEN, E., PARK, B. U. and SIMAR, L. (1999). On estimation of monotone and concave frontier functions. *J. Amer. Statist. Assoc.* **94** 220–228. MR1689226 https://doi.org/10.2307/2669696

[18] HOLMES, C. C. and HEARD, N. A. (2003). Generalized monotonic regression using random change points. *Stat. Med.* **22** 623–638.

[19] JIRAK, M., MEISTER, A. and REISS, M. (2014). Adaptive function estimation in nonparametric regression with one-sided errors. *Ann. Statist.* **42** 1970–2002. MR3262474 https://doi.org/10.1214/14-AOS1248

[20] KIM, Y. and LEE, J. (2004). A Bernstein–von Mises theorem in the nonparametric right-censoring model. *Ann. Statist.* **32** 1492–1512. MR2089131 https://doi.org/10.1214/009053604000000526

[21] KLEIJN, B. and KNAPIK, B. (2012). Semiparametric posterior limits under local asymptotic exponentiality. Preprint. Available at arXiv:1210.6204.

[22] KOROSTELËV, A. P. and TSYBAKOV, A. B. (1993). *Minimax Theory of Image Reconstruction*. *Lecture Notes in Statistics* **82**. Springer, New York. MR1226450 https://doi.org/10.1007/978-1-4612-2712-0

[23] LI, M. and GHOSAL, S. (2017). Bayesian detection of image boundaries. *Ann. Statist.* **45** 2190–2217. MR3718166 https://doi.org/10.1214/16-AOS1523

[24] MARIUCCI, E., RAY, K. and SZABO, B. (2017). A Bayesian nonparametric approach to log-concave density estimation. Preprint. Available at arXiv:1703.09531.

[25] MEISTER, A. and REISS, M. (2013). Asymptotic equivalence for nonparametric regression with non-regular errors. *Probab. Theory Related Fields* **155** 201–229. MR3010397 https://doi.org/10.1007/s00440-011-0396-x

[26] PANOV, M. and SPOKOINY, V. (2015). Finite sample Bernstein–von Mises theorem for semiparametric problems. *Bayesian Anal.* **10** 665–710. MR3420819 https://doi.org/10.1214/14-BA926

[27] REISS, M. and SCHMIDT-HIEBER, J. (2017). Posterior contraction rates for support boundary recovery. Preprint. Available at arXiv:1703.08358.

[28] REISS, M. and SCHMIDT-HIEBER, J. (2019). Supplement to "Nonparametric Bayesian analysis of the compound Poisson prior for support boundary recovery." https://doi.org/10.1214/19-AOS1853SUPP.

[29] REISS, M. and SELK, L. (2017). Efficient estimation of functionals in nonparametric boundary models. *Bernoulli* **23** 1022–1055. MR3606758 https://doi.org/10.3150/15-BEJ768

[30] RIVOIRARD, V. and ROUSSEAU, J. (2012). Bernstein–von Mises theorem for linear functionals of the density. *Ann. Statist.* **40** 1489–1523. MR3015033 https://doi.org/10.1214/12-AOS1004

[31] ROCKOVA, V. and VAN DER PAS, S. (2017). Posterior concentration for Bayesian regression trees and their ensembles. Preprint. Available at arXiv:1708.08734. *Ann. Statist.* (to appear).

[32] SALOMOND, J.-B. (2014). Concentration rate and consistency of the posterior distribution for selected priors under monotonicity constraints. *Electron. J. Stat.* **8** 1380–1404. MR3263126 https://doi.org/10.1214/14-EJS929

[33] SATO, K. (2013). *Lévy Processes and Infinitely Divisible Distributions*. *Cambridge Studies in Advanced Mathematics* **68**. Cambridge Univ. Press, Cambridge. MR3185174

[34] SCRICCIOLO, C. (2007). On rates of convergence for Bayesian density estimation. *Scand. J. Stat.* **34** 626–642. MR2368802 https://doi.org/10.1111/j.1467-9469.2006.00540.x

[35] SIMON, T. (2004). Small ball estimates in p-variation for stable processes. *J. Theoret. Probab.* **17** 979–1002. MR2105744 https://doi.org/10.1007/s10959-004-0586-x

[36] 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

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

The Annals of Statistics
2020, Vol. 48, No. 3, 1452–1474
https://doi.org/10.1214/19-AOS1854

ENTRYWISE EIGENVECTOR ANALYSIS OF RANDOM MATRICES WITH LOW EXPECTED RANK

BY EMMANUEL ABBE[1], JIANQING FAN[2,*], KAIZHENG WANG[2,**] AND YIQIAO ZHONG[2,†]

[1]*PACM and Department of EE, Princeton University, eabbe@princeton.edu*
[2]*Department of ORFE, Princeton University, *jqfan@princeton.edu; **kaizheng@princeton.edu; †yiqiaoz@princeton.edu*

Recovering low-rank structures via eigenvector perturbation analysis is a common problem in statistical machine learning, such as in factor analysis, community detection, ranking, matrix completion, among others. While a large variety of bounds are available for average errors between empirical and population statistics of eigenvectors, few results are tight for entrywise analyses, which are critical for a number of problems such as community detection.

This paper investigates entrywise behaviors of eigenvectors for a large class of random matrices whose expectations are low rank, which helps settle the conjecture in Abbe, Bandeira and Hall (2014) that the spectral algorithm achieves exact recovery in the stochastic block model without any trimming or cleaning steps. The key is a first-order approximation of eigenvectors under the ℓ_∞ norm:

$$u_k \approx \frac{Au_k^*}{\lambda_k^*},$$

where $\{u_k\}$ and $\{u_k^*\}$ are eigenvectors of a random matrix A and its expectation $\mathbb{E}A$, respectively. The fact that the approximation is both tight and linear in A facilitates sharp comparisons between u_k and u_k^*. In particular, it allows for comparing the signs of u_k and u_k^* even if $\|u_k - u_k^*\|_\infty$ is large. The results are further extended to perturbations of eigenspaces, yielding new ℓ_∞-type bounds for synchronization ($\mathbb{Z}_2$-spiked Wigner model) and noisy matrix completion.

REFERENCES

ABBE, E. (2017). Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** 177. MR3827065

ABBE, E., BANDEIRA, A. S. and HALL, G. (2014). Exact recovery in the stochastic block model. arXiv preprint arXiv:1405.3267.

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

ABBE, E. and SANDON, C. (2015). Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In 2015 *IEEE 56th Annual Symposium on Foundations of Computer Science—FOCS* 2015 670–688. IEEE Computer Soc., Los Alamitos, CA. MR3473334

ABBE, E. and SANDON, C. (2018). Proof of the achievability conjectures for the general stochastic block model. *Comm. Pure Appl. Math.* **71** 1334–1406. MR3812075 https://doi.org/10.1002/cpa.21719

ABBE, E., BANDEIRA, A. S., BRACHER, A. and SINGER, A. (2014). Decoding binary node labels from censored edge measurements: Phase transition and efficient recovery. *IEEE Trans. Netw. Sci. Eng.* **1** 10–22. MR3349181 https://doi.org/10.1109/TNSE.2014.2368716

ABBE, E., FAN, J., WANG, K. and ZHONG, Y. (2019). Supplement to "Entrywise eigenvector analysis of random matrices with low expected rank." https://doi.org/10.1214/19-AOS1854SUPP.

MSC2010 subject classifications. Primary 62H25; secondary 60B20, 62H12.

Key words and phrases. Eigenvector perturbation, spectral analysis, synchronization, community detection, matrix completion, low-rank structures, random matrices.

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

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

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

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

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

BANKS, J., MOORE, C. NEEMAN, J., and NETRAPALLI, P. (2016). Information-theoretic thresholds for community detection in sparse networks. In *Conference on Learning Theory*.

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

BICKEL, P. J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70** 428–434. MR0386168

BORDENAVE, C., LELARGE, M. and MASSOULIÉ, L. (2015). Non-backtracking spectrum of random graphs: Community detection and non-regular Ramanujan graphs. In 2015 *IEEE 56th Annual Symposium on Foundations of Computer Science—FOCS* 2015 1347–1357. IEEE Computer Soc., Los Alamitos, CA. MR3473374

CANDÈS, E. J., LI, X. and SOLTANOLKOTABI, M. (2015). Phase retrieval via Wirtinger flow: Theory and algorithms. *IEEE Trans. Inform. Theory* **61** 1985–2007. MR3332993 https://doi.org/10.1109/TIT.2015.2399924

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

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

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

CANDÈS, E. J., LI, X., MA, Y. and WRIGHT, J. (2011). Robust principal component analysis? *J. ACM* **58** 11. MR2811000 https://doi.org/10.1145/1970392.1970395

CAPE, J., TANG, M. and PRIEBE, C. E. (2017). The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. ArXiv preprint. Available at arXiv:1705.10735.

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

CHEN, Y., 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

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*.

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

CUCURINGU, M., LIPMAN, Y. and SINGER, A. (2012). Sensor network localization by eigenvector synchronization over the Euclidean group. *ACM Trans. Sens. Netw.* **8**. 19.

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

DECELLE, A., KRZAKALA, F., MOORE, C. and ZDEBOROVÁ, L. (2011). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* **84** 066106.

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

ELDRIDGE, J., BELKIN, M. and WANG, Y. (2018). Unperturbed: Spectral analysis beyond Davis–Kahan. In *Algorithmic Learning Theory* 2018. *Proc. Mach. Learn. Res.* (*PMLR*) **83** 38. MR3857310

FAN, J., WANG, W. and ZHONG, Y. (2017). An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. *J. Mach. Learn. Res.* **18** 207. MR3827095

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

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** 60. MR3687603

GE, R., LEE, J. D. and MA, T. (2016). Matrix completion has no spurious local minimum. In *Advances in Neural Information Processing Systems*.

GIRIDHAR, A. and KUMAR, P. R. (2006). Distributed clock synchronization over wireless networks: Algorithms and analysis. In *Decision and Control*, 2006 45*th IEEE Conference on* IEEE.

GROSS, D. (2011). Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inform. Theory* **57** 1548–1566. MR2815834 https://doi.org/10.1109/TIT.2011.2104999

GUÉDON, O. and VERSHYNIN, R. (2016). Community detection in sparse networks via Grothendieck's inequality. *Probab. Theory Related Fields* **165** 1025–1049. MR3520025 https://doi.org/10.1007/s00440-015-0659-z

HAJEK, B., WU, Y. and XU, J. (2016). Achieving exact cluster recovery threshold via semidefinite programming: Extensions. *IEEE Trans. Inform. Theory* **62** 5918–5937. MR3552431 https://doi.org/10.1109/TIT.2016.2594812

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

JAIN, P. and NETRAPALLI, P. (2015). Fast exact matrix completion with finite samples. In *Conference on Learning Theory*.

JAIN, P., NETRAPALLI, P. and SANGHAVI, S. (2013). Low-rank matrix completion using alternating minimization (extended abstract). In *STOC'*13*—Proceedings of the* 2013 *ACM Symposium on Theory of Computing* 665–674. ACM, New York. MR3210828 https://doi.org/10.1145/2488608.2488693

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

KESHAVAN, R. H., MONTANARI, A. and OH, S. (2010a). Matrix completion from a few entries. *IEEE Trans. Inform. Theory* **56** 2980–2998. MR2683452 https://doi.org/10.1109/TIT.2010.2046205

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

KESHAVAN, R. H. and OH, S. (2009). A gradient descent algorithm on the Grassman manifold for matrix completion. ArXiv preprint. Available at arXiv:0910.5260.

KOLTCHINSKII, V. and LOUNICI, K. (2016). Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 1976–2013. MR3573302 https://doi.org/10.1214/15-AIHP705

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

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

KRZAKALA, F., MOORE, C., MOSSEL, E., NEEMAN, J., SLY, A., ZDEBOROVÁ, L. and ZHANG, P. (2013). Spectral redemption in clustering sparse networks. *Proc. Natl. Acad. Sci. USA* **110** 20935–20940. MR3174850 https://doi.org/10.1073/pnas.1312486110

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

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

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

MA, C., WANG, K., CHI, Y. and CHEN, Y. (2017). Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion and blind deconvolution. ArXiv preprint. Available at arXiv:1711.10467.

MASSOULIÉ, L. (2014). Community detection thresholds and the weak Ramanujan property. In *STOC'*14*—Proceedings of the* 2014 *ACM Symposium on Theory of Computing* 694–703. ACM, New York. MR3238997

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

MONTANARI, A. and SEN, S. (2016). Semidefinite programs on sparse random graphs and their application to community detection. In *STOC'*16*—Proceedings of the* 48*th Annual ACM SIGACT Symposium on Theory of Computing* 814–827. ACM, New York. MR3536616

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

MOSSEL, E., NEEMAN, J. and SLY, A. (2018). A proof of the block model threshold conjecture. *Combinatorica* **38** 665–708. MR3876880 https://doi.org/10.1007/s00493-016-3238-8

NG, A. Y., JORDAN, M. I. and WEISS, Y. (2002). On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*.

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

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

RAYLEIGH, J. W. S. (1896). *The Theory of Sound*, **2**. Macmillan.

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

ROSEN, D. M., CARLONE, L., BANDEIRA, A. S. and LEONARD, J. J. (2016). A certifiably correct algorithm for synchronization over the special Euclidean group. ArXiv preprint. Available at arXiv:1611.00128.

SCHRÖDINGER, E. (1926). Quantisierung als eigenwertproblem. *Ann. Phys.* **385** 437–490.

SHI, J. and MALIK, J. (2000). Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **22** 888–905.

SHKOLNISKY, Y. and SINGER, A. (2012). Viewing direction estimation in cryo-EM using synchronization. *SIAM J. Imaging Sci.* **5** 1088–1110. MR3022188 https://doi.org/10.1137/120863642

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

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

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

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

TRON, R. and VIDAL, R. (2009). Distributed image-based 3-D localization of camera sensor networks. In *Decision and Control*, 2009 *Held Jointly with the* 2009 28*th Chinese Control Conference*. *CDC/CCC* 2009. *Proceedings of the* 48*th IEEE Conference on* IEEE.

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

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

XIA, D. and ZHOU, F. (2017). The ℓ_∞ perturbation of HOSVD and low rank tensor denoising. ArXiv preprint. Available at arXiv:1707.01207.

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

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

ZHANG, A. Y. and ZHOU, H. H. (2016). Minimax rates of community detection in stochastic block models. *Ann. Statist.* **44** 2252–2280. MR3546450 https://doi.org/10.1214/15-AOS1428

ZHONG, Y. (2017). Eigenvector under random perturbation: A nonasymptotic Rayleigh–Schrödinger theory. ArXiv preprint. Available at arXiv:1702.00139.

ZHONG, Y. and BOUMAL, N. (2018). Near-optimal bounds for phase synchronization. *SIAM J. Optim.* **28** 989–1016. MR3782406 https://doi.org/10.1137/17M1122025

The Annals of Statistics
2020, Vol. 48, No. 3, 1475–1497
https://doi.org/10.1214/19-AOS1855

CONCENTRATION OF TEMPERED POSTERIORS AND OF THEIR VARIATIONAL APPROXIMATIONS

BY PIERRE ALQUIER[1] AND JAMES RIDGWAY[2]

[1]*RIKEN Center for Advanced Intelligence Project, pierre.alquier.stat@gmail.com*
[2]*Capital Fund Management, james.lp.ridgway@gmail.com*

While Bayesian methods are extremely popular in statistics and machine learning, their application to massive data sets is often challenging, when possible at all. The classical MCMC algorithms are prohibitively slow when both the model dimension and the sample size are large. Variational Bayesian methods aim at approximating the posterior by a distribution in a tractable family $\mathcal{F}$. Thus, MCMC are replaced by an optimization algorithm which is orders of magnitude faster. VB methods have been applied in such computationally demanding applications as collaborative filtering, image and video processing or NLP to name a few. However, despite nice results in practice, the theoretical properties of these approximations are not known. We propose a general oracle inequality that relates the quality of the VB approximation to the prior π and to the structure of $\mathcal{F}$. We provide a simple condition that allows to derive rates of convergence from this oracle inequality. We apply our theory to various examples. First, we show that for parametric models with log-Lipschitz likelihood, Gaussian VB leads to efficient algorithms and consistent estimators. We then study a high-dimensional example: matrix completion, and a nonparametric example: density estimation.

REFERENCES

[1] ALQUIER, P., COTTET, V., CHOPIN, N. and ROUSSEAU, J. (2014). Bayesian matrix completion: Prior specification. Preprint. Available at arXiv:1406.1440.

[2] ALQUIER, P., COTTET, V. and LECUÉ, G. (2019). Estimation bounds and sharp oracle inequalities of regularized procedures with Lipschitz loss functions. *Ann. Statist.* **47** 2117–2144. MR3953446 https://doi.org/10.1214/18-AOS1742

[3] ALQUIER, P. and RIDGWAY, J. (2020). Supplement to "Concentration of tempered posteriors and of their variational approximations." https://doi.org/10.1214/19-AOS1855SUPP.

[4] ALQUIER, P., RIDGWAY, J. and CHOPIN, N. (2016). On the properties of variational approximations of Gibbs posteriors. *J. Mach. Learn. Res.* **17** Paper No. 239, 41. MR3595173

[5] BABACAN, S. D., LUESSI, M. and MOLINA, R. and KATSAGGELOS, A. K. (2011). Low-rank matrix completion by variational sparse Bayesian learning. In *IEEE International Conference on Audio, Speech and Signal Processing* 2188–2191. Prague.

[6] BHATTACHARYA, A., PATI, D. and YANG, Y. (2019). Bayesian fractional posteriors. *Ann. Statist.* **47** 39–66. MR3909926 https://doi.org/10.1214/18-AOS1712

[7] BISHOP, C. M. (2006). *Pattern Recognition and Machine Learning. Information Science and Statistics*. Springer, New York. MR2247587 https://doi.org/10.1007/978-0-387-45528-0

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

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

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

[11] CATONI, O. (2004). *Statistical Learning Theory and Stochastic Optimization. Lecture Notes in Math.* **1851**. Springer, Berlin. MR2163920 https://doi.org/10.1007/b99352

MSC2010 subject classifications. Primary 62G15; secondary 62C10, 62C20, 62F25, 62G25.

Key words and phrases. Concentration of the posterior, rate of convergence, variational approximation, PAC-Bayesian bounds.

[12] CATONI, O. (2007). *Pac-Bayesian Supervised Classification*: *The Thermodynamics of Statistical Learning*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **56**. IMS, Beachwood, OH. MR2483528

[13] CHOPIN, N. and RIDGWAY, J. (2017). Leave Pima Indians alone: Binary regression as a benchmark for Bayesian computation. *Statist. Sci.* **32** 64–87. MR3634307 https://doi.org/10.1214/16-STS581

[14] COTTET, V. and ALQUIER, P. (2018). 1-bit matrix completion: PAC—Bayesian analysis of a variational approximation. *Mach. Learn.* **107** 579–603. MR3761297 https://doi.org/10.1007/s10994-017-5667-z

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

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

[17] GIBBS, M. N. and MACKAY, D. J. C. (2000). Variational Gaussian process classifiers. *IEEE Trans. Neural Netw. Learn. Syst.* **11** 1458–1464.

[18] GRÜNWALD, P. (2012). The safe Bayesian: Learning the learning rate via the mixability gap. In *Algorithmic Learning Theory*. *Lecture Notes in Computer Science* **7568** 169–183. Springer, Heidelberg. MR3042889 https://doi.org/10.1007/978-3-642-34106-9_16

[19] GRÜNWALD, P. and VAN OMMEN, T. (2017). Inconsistency of Bayesian inference for misspecified linear models, and a proposal for repairing it. *Bayesian Anal.* **12** 1069–1103. MR3724979 https://doi.org/10.1214/17-BA1085

[20] GRÜNWALD, P. D. and MEHTA, N. A. (2016). Fast rates with unbounded losses. Preprint. Available at arXiv:1605.00252.

[21] HOFFMAN, M. D., BLEI, D. M., WANG, C. and PAISLEY, J. (2013). Stochastic variational inference. *J. Mach. Learn. Res.* **14** 1303–1347. MR3081926

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

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

[24] LAWRENCE, N. D. and URTASUN, R. (2009). Non-linear matrix factorization with Gaussian processes. In *Proceedings of the* 26*th Annual International Conference on Machine Learning* 601–608. ACM, New York.

[25] LI, X. and ZHENG, Y. (2009). Patch-based video processing: A variational Bayesian approach. *IEEE Transactions on Circuits and Systems for Video Technology* **19** 27–40.

[26] LIM, Y. J. and TEH, Y. W. (2007). Variational Bayesian approach to movie rating prediction. *In Proceedings of KDD Cup and Workshop* **7** 15–21.

[27] MAI, T. T. and ALQUIER, P. (2015). A Bayesian approach for noisy matrix completion: Optimal rate under general sampling distribution. *Electron. J. Stat.* **9** 823–841. MR3331862 https://doi.org/10.1214/15-EJS1020

[28] MARSDEN, A. and BACALLADO, S. (2017). Sequential matrix completion. Preprint. Available at arXiv:1710.08045.

[29] OPPER, M. and ARCHAMBEAU, C. (2009). The variational Gaussian approximation revisited. *Neural Comput.* **21** 786–792. MR2478318 https://doi.org/10.1162/neco.2008.08-07-592

[30] PAISLEY, J. and CARIN, L. (2010). A nonparametric Bayesian model for kernel matrix completion. In *Proceedings of ICASSP* 2010, Dallas, USA.

[31] ROUSSEAU, J. (2016). On the frequentist properties of Bayesian nonparametric methods. *Annual Review of Statistics and Its Application* **3** 211–231.

[32] SALAKHUTDINOV, R. and MNIH, A. (2008). Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. In *Proceedings of the* 25*th International Conference on Machine Learning* 880–887. ACM, New York.

[33] SHALEV-SHWARTZ, S. (2012). Online learning and online convex optimization. *Found. Trends Mach. Learn.* **4** 107–194.

[34] SUZUKI, T. (2015). Convergence rate of Bayesian tensor estimator and its minimax optimality. In *ICML* 1273–1282.

[35] TITSIAS, M. and LÁZARO-GREDILLA, M. (2014). Doubly stochastic variational Bayes for non-conjugate inference. In *Proceedings of the* 31*st International Conference on Machine Learning* (*ICML*-14) 1971–1979.

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

[37] VAN ERVEN, T. and HARREMOËS, P. (2014). Rényi divergence and Kullback–Leibler divergence. *IEEE Trans. Inform. Theory* **60** 3797–3820. MR3225930 https://doi.org/10.1109/TIT.2014.2320500

[38] WANG, B. and TITTERINGTON, D. M. (2004). Convergence and asymptotic normality of variational Bayesian approximations for exponential family models with missing values. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence* 577–584. AUAI Press.

[39] WANG, Y. and BLEI, D. M. Frequentist consistency of variational Bayes. *J. Amer. Statist. Assoc*. To appear. https://doi.org/10.1080/01621459.2018.1473776.

[40] YANG, Y., PATI, D. and BHATTACHARYA, A. In α-variational inference with statistical guarantees. *Ann. Statist*. To appear.

[41] ZHANG, F. and GAO, C. (2018). Convergence rates of variational posterior distributions. Preprint. Available at arXiv:1712.02519.

[42] ZHOU, M., WANG, C., CHEN, M., PAISLEY, J., DUNSON, D. and CARIN, L. (2010). Nonparametric Bayesian matrix completion. In *Proc. IEEE SAM*.

The Annals of Statistics
2020, Vol. 48, No. 3, 1498–1513
https://doi.org/10.1214/19-AOS1856

ROBUST AND RATE-OPTIMAL GIBBS POSTERIOR INFERENCE ON THE BOUNDARY OF A NOISY IMAGE

BY NICHOLAS SYRING[1] AND RYAN MARTIN[2]

[1]*Department of Mathematics and Statistics, Washington University in St. Louis, nasyring@wustl.edu*
[2]*Department of Statistics, North Carolina State University, rgmarti3@ncsu.edu*

Detection of an image boundary when the pixel intensities are measured with noise is an important problem in image segmentation. From a statistical point of view, a challenge is that likelihood-based methods require modeling the pixel intensities inside and outside the image boundary, even though these distributions are typically not of interest. Since misspecification of the pixel intensity distributions can negatively affect inference on the image boundary, it would be desirable to avoid this modeling step altogether. Toward this, we develop a robust Gibbsian approach that constructs a posterior distribution for the image boundary directly, without modeling the pixel intensities. We prove that the Gibbs posterior concentrates asymptotically at the minimax optimal rate, adaptive to the boundary smoothness. Monte Carlo computation of the Gibbs posterior is straightforward, and simulation results show that the corresponding inference is more accurate than that based on existing Bayesian methodology.

REFERENCES

ANAM, S., UCHINO, E. and SUETAKE, N. (2013). Image boundary detection using the modified level set method and a diffusion filter. *Proc. Comput. Sci.* **22** 192–200.

BISSIRI, P. G., HOLMES, C. C. and WALKER, S. G. (2016). A general framework for updating belief distributions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 1103–1130. MR3557191 https://doi.org/10.1111/rssb.12158

CATONI, O. (2007). *PAC—Bayesian Supervised Classification: The Thermodynamics of Statistical Learning. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **56**. IMS, Beachwood, OH. MR2483528

DENISON, D. G. T., MALLICK, B. K. and SMITH, A. F. M. (1998). Automatic Bayesian curve fitting. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 333–350. MR1616029 https://doi.org/10.1111/1467-9868.00128

DIMATTEO, I., GENOVESE, C. R. and KASS, R. E. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika* **88** 1055–1071. MR1872219 https://doi.org/10.1093/biomet/88.4.1055

FITZPATRICK, M. C., PREISSER, E. L., PORTER, A., ELKINTON, J., WALLER, L. A., CARLIN, B. P. and ELLISON, A. M. (2010). Ecological boundary detection using Bayesian areal wombling. *Ecology* **91** 3448–3455.

GREEN, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82** 711–732. MR1380810 https://doi.org/10.1093/biomet/82.4.711

GU, K., PATI, D. and DUNSON, D. B. (2014). Bayesian multiscale modeling of closed curves in point clouds. *J. Amer. Statist. Assoc.* **109** 1481–1494. MR3293605 https://doi.org/10.1080/01621459.2014.934825

JIANG, W. and TANNER, M. A. (2008). Gibbs posterior for variable selection in high-dimensional classification and data mining. *Ann. Statist.* **36** 2207–2231. MR2458185 https://doi.org/10.1214/07-AOS547

LI, M. and GHOSAL, S. (2017). Bayesian detection of image boundaries. *Ann. Statist.* **45** 2190–2217. MR3718166 https://doi.org/10.1214/16-AOS1523

LI, C., XU, C., GUI, C. and FOX, M. D. (2010). Distance regularized level set evolution and its application to image segmentation. *IEEE Trans. Image Process.* **19** 3243–3254. MR2789710 https://doi.org/10.1109/TIP.2010.2069690

LIANG, S., BANERJEE, S. and CARLIN, B. P. (2009). Bayesian wombling for spatial point processes. *Biometrics* **65** 1243–1253. MR2756512 https://doi.org/10.1111/j.1541-0420.2009.01203.x

LU, H. and CARLIN, B. P. (2004). Bayesian areal wombling for geographical boundary analysis. *Geogr. Anal.* **37** 265–285.

MSC2010 subject classifications. Primary 62G20; secondary 62G05.

Key words and phrases. Adaptation, boundary detection, likelihood-free inference, model misspecification, posterior concentration rate.

MA, H. and CARLIN, B. P. (2007). Bayesian multivariate areal wombling for multiple disease boundary analysis. *Bayesian Anal*. **2** 281–302. MR2312282 https://doi.org/10.1214/07-BA211

MAINI, R. and AGGARWAL, H. (2009). Study and comparison of various image edge detection techniques. *IJIP* **3** 1–11.

MAMMEN, E. and TSYBAKOV, A. B. (1995). Asymptotical minimax recovery of sets with smooth boundaries. *Ann*. *Statist*. **23** 502–524. MR1332579 https://doi.org/10.1214/aos/1176324533

MARTIN, D. R., FOWLKES, C. C. and MALIK, J. (2004). Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Trans*. *Pattern Anal*. *Mach*. *Intell*. **26** 530–549.

SCHUMAKER, L. L. (1981). *Spline Functions*: *Basic Theory*. Wiley, New York. MR0606200

SHEN, W. and GHOSAL, S. (2015). Adaptive Bayesian procedures using random series priors. *Scand*. *J*. *Stat*. **42** 1194–1213. MR3426318 https://doi.org/10.1111/sjos.12159

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

SYRING, N. and LI, M. (2017a). BayesBD: Bayesian inference for image boundaries. R package version 1.2.

SYRING, N. and LI, M. (2017b). BayesBD: An R package for Bayesian inference on image boundaries. *R J*. **9** 149–162.

SYRING, N. and MARTIN, R. (2017). Gibbs posterior inference on the minimum clinically important difference. *J*. *Statist*. *Plann*. *Inference* **187** 67–77. MR3638043 https://doi.org/10.1016/j.jspi.2017.03.001

SYRING, N. and MARTIN, R. (2019). Calibrating general posterior credible regions. *Biometrika* **106** 479–486.

YUAN, W., CHIN, K. S., HUA, M., DONG, G. and WANG, C. (2016). Shape classification of wear particles by image boundary analysis using machine learning algorithms. *Mech*. *Syst*. *Signal Pr*. **72–73** 346–358.

ZHANG, T. (2006). Information-theoretic upper and lower bounds for statistical estimation. *IEEE Trans*. *Inform*. *Theory* **52** 1307–1321. MR2241190 https://doi.org/10.1109/TIT.2005.864439

ZIOU, D. and TABBONE, S. (1998). Edge detection techniques—an overview. *Pattern Recognition and Image Analysis C/C of Raspoznavaniye Obrazov I Analiz Izobrazhenii* **8** 537–559.

The Annals of Statistics
2020, Vol. 48, No. 3, 1514–1538
https://doi.org/10.1214/19-AOS1857

THE HARDNESS OF CONDITIONAL INDEPENDENCE TESTING AND THE GENERALISED COVARIANCE MEASURE

BY RAJEN D. SHAH[1] AND JONAS PETERS[2]

[1]*Statistical Laboratory, University of Cambridge, r.shah@statslab.cam.ac.uk*
[2]*Department of Mathematical Sciences, University of Copenhagen, jonas.peters@math.ku.dk*

It is a common saying that testing for conditional independence, that is, testing whether whether two random vectors X and Y are independent, given Z, is a hard statistical problem if Z is a continuous random variable (or vector). In this paper, we prove that conditional independence is indeed a particularly difficult hypothesis to test for. Valid statistical tests are required to have a size that is smaller than a pre-defined significance level, and different tests usually have power against a different class of alternatives. We prove that a valid test for conditional independence does not have power against *any* alternative.

Given the nonexistence of a uniformly valid conditional independence test, we argue that tests must be designed so their suitability for a particular problem may be judged easily. To address this need, we propose in the case where X and Y are univariate to nonlinearly regress X on Z, and Y on Z and then compute a test statistic based on the sample covariance between the residuals, which we call the generalised covariance measure (GCM). We prove that validity of this form of test relies almost entirely on the weak requirement that the regression procedures are able to estimate the conditional means X given Z, and Y given Z, at a slow rate. We extend the methodology to handle settings where X and Y may be multivariate or even high dimensional. While our general procedure can be tailored to the setting at hand by combining it with any regression technique, we develop the theoretical guarantees for kernel ridge regression. A simulation study shows that the test based on GCM is competitive with state of the art conditional independence tests. Code is available as the R package `GeneralisedCovarianceMeasure` on CRAN.

REFERENCES

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

[2] BAHADUR, R. R. and SAVAGE, L. J. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *Ann. Math. Stat.* **27** 1115–1122. MR0084241 https://doi.org/10.1214/aoms/1177728077

[3] BALAKRISHNAN, S. and WASSERMAN, L. (2019). Hypothesis testing for densities and high-dimensional multinomials: Sharp local minimax rates. *Ann. Statist.* **47** 1893–1927. MR3953439 https://doi.org/10.1214/18-AOS1729

[4] BARRON, A. R. (1989). Uniformly powerful goodness of fit tests. *Ann. Statist.* **17** 107–124. MR0981439 https://doi.org/10.1214/aos/1176347005

[5] BERGSMA, W. P. (2004). Testing conditional independence for continuous random variables. EURANDOM-report 2004-049.

[6] BERRETT, T. B. and SAMWORTH, R. J. (2019). Nonparametric independence testing via mutual information. *Biometrika* **106** 547–566. MR3992389 https://doi.org/10.1093/biomet/asz024

[7] BERRETT, T. B., WANG, Y., BARBER, R. F. and SAMWORTH, R. J. (2018). The conditional permutation test. ArXiv, arxiv:1807.05405.

MSC2010 subject classifications. Primary 62G10; secondary 62G08.

Key words and phrases. Conditional independence, hypothesis testing, testability, wild bootstrap, kernel ridge regression.

[8] BERTANHA, M. and MOREIRA, M. J. (2018). Impossible inference in econometrics: Theory and applications. ArXiv , arxiv:1612.02024v3.

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

[10] 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

[11] CANAY, I. A., SANTOS, A. and SHAIKH, A. M. (2013). On the testability of identification in some nonparametric models with endogeneity. *Econometrica* **81** 2535–2559. MR3138554 https://doi.org/10.3982/ECTA10851

[12] CHEN, T. and GUESTRIN, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the* 22*nd ACM SigKDD International Conference on Knowledge Discovery and Data Mining* 785–794. ACM.

[13] CHEN, T., HE, T., BENESTY, M., KHOTILOVICH, V. and TANG, Y. (2018). xgboost: Extreme gradient boosting. R package version 0.6.4.1.

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

[15] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. MR3161448 https://doi.org/10.1214/13-AOS1161

[16] DAUDIN, J.-J. (1980). Partial association measures and an application to qualitative regression. *Biometrika* **67** 581–590. MR0601095 https://doi.org/10.1093/biomet/67.3.581

[17] DAWID, A. P. (1979). Conditional independence in statistical theory. *J. Roy. Statist. Soc. Ser. B* **41** 1–31. MR0535541

[18] DOBRA, A., HANS, C., JONES, B., NEVINS, J. R., YAO, G. and WEST, M. (2004). Sparse graphical models for exploring gene expression data. *J. Multivariate Anal.* **90** 196–212. MR2064941 https://doi.org/10.1016/j.jmva.2004.02.009

[19] DUFOUR, J.-M. (2003). Identification, weak instruments, and statistical inference in econometrics. *Can. J. Econ.* **36** 767–808.

[20] FAN, J., FENG, Y. and XIA, L. (2015). A projection based conditional dependence measure with applications to high-dimensional undirected graphical models. ArXiv, arxiv:1501.01617.

[21] FISHER, R. A. (1920). A mathematical examination of the methods of determining the accuracy of an observation by the mean error and by the mean square error. *Mon. Not. R. Astron. Soc.* **80** 758–770.

[22] FISHER, R. A. (1934). Two new properties of mathematical likelihood. *Proc. R. Soc. A* **114** 285–307.

[23] FISHER, R. A. (1935). *The Design of Experiments*. Oliver and Boyd, Edinburgh, UK.

[24] FUKUMIZU, K., GRETTON, A., SUN, X. and SCHÖLKOPF, B. (2008). Kernel measures of conditional dependence. In *Advances in Neural Information Processing Systems* 20 (*NIPS*) 489–496. MIT Press.

[25] GRETTON, A., FUKUMIZU, K., TEO, C. H., SONG, L., SCHÖLKOPF, B. and SMOLA, A. (2008). A kernel statistical test of independence. In *Advances in Neural Information Processing Systems* 20 (*NIPS*) 585–592. MIT Press.

[26] HEINZE-DEML, C., PETERS, J. and MEINSHAUSEN, N. (2017). Invariant causal prediction for nonlinear models. *J. Causal Inference*. To appear. ArXiv, arxiv:1706.08576.

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

[28] JENSEN, J. L. and SØRENSEN, M. (2017). Statistical principles: A first course (lecture notes), Aarhus and Copenhagen, Denmark, available upon request.

[29] KIM, S. (2015). ppcor: Partial and semi-partial (part) correlation. R package version 1.1.

[30] KOLLER, D. and FRIEDMAN, N. (2009). *Probabilistic Graphical Models*: *Principles and Techniques*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR2778120

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

[32] LAURITZEN, S. L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. Oxford University Press, New York. MR1419991

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

[34] LI, L., TCHETGEN TCHETGEN, E., VAN DER VAART, A. and ROBINS, J. M. (2011). Higher order inference on a treatment effect under low regularity conditions. *Statist. Probab. Lett.* **81** 821–828. MR2793749 https://doi.org/10.1016/j.spl.2011.02.030

[35] MARKOWETZ, F. and SPANG, R. (2007). Inferring cellular networks—a review. *BMC Bioinform*. **8 Suppl 6** S5. https://doi.org/10.1186/1471-2105-8-S6-S5

[36] NEWEY, W. K. and ROBINS, J. (2018). Cross-fitting and fast remainder rates for semiparametric estimation. ArXiv, arxiv:1801.09138.

[37] PEARL, J. (2009). *Causality*: *Models*, *Reasoning*, *and Inference*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2548166 https://doi.org/10.1017/CBO9780511803161

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

[39] PETERS, J., JANZING, D. and SCHÖLKOPF, B. (2017). *Elements of Causal Inference*: *Foundations and Learning Algorithms*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3822088

[40] RAMSEY, J. D. (2014). A scalable conditional Independence test for nonlinear, non-Gaussian data. ArXiv, arxiv:1401.5031.

[41] REN, Z., SUN, T., ZHANG, C.-H. and ZHOU, H. H. (2015). Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *Ann. Statist*. **43** 991–1026. MR3346695 https://doi.org/10.1214/14-AOS1286

[42] ROBINS, J., LI, L., TCHETGEN TCHETGEN, E. and VAN DER VAART, A. (2008). Higher order influence functions and minimax estimation of nonlinear functionals. In *Probability and Statistics*: *Essays in Honor of David A. Freedman* 335–421. IMS.

[43] ROBINS, J., TCHETGEN TCHETGEN, E., LI, L. and VAN DER VAART, A. (2009). Semiparametric minimax rates. *Electron. J. Stat*. **3** 1305–1321. MR2566189 https://doi.org/10.1214/09-EJS479

[44] ROBINS, J. M., LI, L., MUKHERJEE, R., TCHETGEN, E. T. and VAN DER VAART, A. (2017). Minimax estimation of a functional on a structured high-dimensional model. *Ann. Statist*. **45** 1951–1987. MR3718158 https://doi.org/10.1214/16-AOS1515

[45] ROMANO, J. P. (2004). On non-parametric testing, the uniform behaviour of the t-test, and related problems. *Scand. J. Stat*. **31** 567–584. MR2101540 https://doi.org/10.1111/j.1467-9469.2004.00407.x

[46] SHAH, R. D. and BÜHLMANN, P. (2018). Goodness-of-fit tests for high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **80** 113–135. MR3744714 https://doi.org/10.1111/rssb.12234

[47] SHAH, R. D. and PETERS, J. (2020). Supplement to "The hardness of conditional independence testing and the generalised covariance measure." https://doi.org/10.1214/19-AOS1857SUPP.

[48] SPIRTES, P., GLYMOUR, C. and SCHEINES, R. (2000). *Causation*, *Prediction*, *and Search*, 2nd ed. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. With additional material by David Heckerman, Christopher Meek, Gregory F. Cooper and Thomas Richardson, A Bradford Book. MR1815675

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

[50] WEAVER, N. (2013). *Measure Theory and Functional Analysis*. World Scientific Co. Pte. Ltd., Hackensack, NJ. MR3185963 https://doi.org/10.1142/8813

[51] ZHANG, K., PETERS, J., JANZING, D. and SCHÖLKOPF, B. (2011). Kernel-based conditional independence test and application in causal discovery. In *Proceedings of the* 27*th Annual Conference on Uncertainty in Artificial Intelligence* (*UAI*) 804–813. AUAI Press.

[52] ZHENG, W. and VAN DER LAAN, M. J. (2011). Cross-validated targeted minimum-loss-based estimation. In *Targeted Learning*. *Springer Ser. Statist*. 459–474. Springer, New York. MR2867139 https://doi.org/10.1007/978-1-4419-9782-1_27

The Annals of Statistics
2020, Vol. 48, No. 3, 1539–1566
https://doi.org/10.1214/19-AOS1858

SOME THEORETICAL PROPERTIES OF GANS

BY GÉRARD BIAU[1,*], BENOÎT CADRE[2], MAXIME SANGNIER[1,**] AND UGO TANIELIAN[1,†]

[1]*Laboratoire de Probabilités, Statistique et Modélisation, Sorbonne Université,* [*]*gerard.biau@sorbonne-universite.fr;* [**]*maxime.sangnier@sorbonne-universite.fr;* [†]*u.tanielian@criteo.com*

[2]*ENS Rennes, Univ Rennes, IRMAR, benoit.cadre@ens-rennes.fr*

Generative Adversarial Networks (GANs) are a class of generative algorithms that have been shown to produce state-of-the-art samples, especially in the domain of image creation. The fundamental principle of GANs is to approximate the unknown distribution of a given data set by optimizing an objective function through an adversarial game between a family of generators and a family of discriminators. In this paper, we offer a better theoretical understanding of GANs by analyzing some of their mathematical and statistical properties. We study the deep connection between the adversarial principle underlying GANs and the Jensen–Shannon divergence, together with some optimality characteristics of the problem. An analysis of the role of the discriminator family via approximation arguments is also provided. In addition, taking a statistical point of view, we study the large sample properties of the estimated distribution and prove in particular a central limit theorem. Some of our results are illustrated with simulated examples.

REFERENCES

ABADI, M., AGARWAL, A., BARHAM, P., BREVDO, E., CHEN, Z., CITRO, C., CORRADO, G. S., DAVIS, A., DEAN, J. et al. (2015). TensorFlow: A system for large-scale machine learning. Available at https://www.tensorflow.org/. Software available from tensorflow.org.

ANGLES, T. and MALLAT, S. (2018). Generative networks as inverse problems with scattering transforms. In *International Conference on Learning Representations*.

ARJOVSKY, M. and BOTTOU, L. (2017). Towards principled methods for training generative adversarial networks. In *International Conference on Learning Representations*.

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

BOJANOWSKI, P., JOULIN, A., LOPEZ-PAZ, D. and SZLAM, A. (2018). Optimizing the latent space of generative networks. In *Proceedings of the 35th International Conference on Machine Learning* (J. Dy and A. Krause, eds.) **80** 600–609. Proceedings of Machine Learning Research.

DEVROYE, L. (1997). Universal smoothing factor selection in density estimation: Theory and practice. *TEST* **6** 223–320. MR1616896 https://doi.org/10.1007/BF02564701

DONAHUE, C., LIPTON, Z. C., BALSUBRAMANI, A. and MCAULEY, J. (2018). Semantically decomposing the latent spaces of generative adversarial networks. In *International Conference on Learning Representations*.

DZIUGAITE, G. K., ROY, D. M. and GHAHRAMANI, Z. (2015). Training generative neural networks via maximum mean discrepancy optimization. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence* (M. Meila and T. Heskes, eds.) 258–267. AUAI Press, Arlington.

ENDRES, D. M. and SCHINDELIN, J. E. (2003). A new metric for probability distributions. *IEEE Trans. Inform. Theory* **49** 1858–1860. MR1985590 https://doi.org/10.1109/TIT.2003.813506

GOODFELLOW, I. (2016). NIPS 2016 tutorial: Generative adversarial networks. Available at arXiv:1701.00160.

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

MSC2010 subject classifications. Primary 62F12; secondary 68T01.

Key words and phrases. Generative models, adversarial principle, Jensen–Shannon divergence, neural networks, central limit theorem.

LIANG, T. (2018). On how well generative adversarial networks learn densities: Nonparametric and parametric results. Available at arXiv:1811.03179.

LIPTON, Z. C. and TRIPATHI, S. (2017). Precise recovery of latent vectors from generative adversarial networks. Available at arXiv:1702.04782.

LIU, S., BOUSQUET, O. and CHAUDHURI, K. (2017). Approximation and convergence properties of generative adversarial learning. In *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, eds.) **30** 5551–5559. Curran Associates, Red Hook.

MCDIARMID, C. (1989). On the method of bounded differences. In *Surveys in Combinatorics*, 1989 (*Norwich*, 1989). *London Mathematical Society Lecture Note Series* **141** 148–188. Cambridge Univ. Press, Cambridge. MR1036755

MIRZA, M. and OSINDERO, S. (2014). Conditional generative adversarial nets. Available at arXiv:1411.1784.

NOWOZIN, S., CSEKE, B. and TOMIOKA, R. (2016). f-GAN: Training generative neural samplers using variational divergence minimization. In *Advances in Neural Information Processing Systems* (D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon and R. Garnett, eds.) **29** 271–279. Curran Associates, Red Hook.

RADFORD, A., METZ, L. and CHINTALA, S. (2016). Unsupervised representation learning with deep convolutional generative adversarial networks. In *International Conference on Learning Representations*.

SALIMANS, T., GOODFELLOW, I., ZAREMBA, W., CHEUNG, V., RADFORD, A. and CHEN, X. (2016). Improved techniques for training GANs. In *Advances in Neural Information Processing Systems* (D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon and R. Garnett, eds.) **29** 2234–2242. Curran Associates, Red Hook.

SRIVASTAVA, A., VALKOZ, L., RUSSELL, C., GUTMANN, M. U. and VEEGAN, C. S. (2017). Reducing mode collapse in GANs using implicit variational learning. In *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, eds.) **30** 3308–3318. Curran Associates, Red Hook.

VAN HANDEL, R. (2016). *Probability in High Dimension*. APC 550 Lecture Notes, Princeton University.

ZHANG, P., LIU, Q., ZHOU, D., XU, T. and HE, X. (2018). On the discriminative-generalization tradeoff in GANs. In *International Conference on Learning Representations*.

The Annals of Statistics
2020, Vol. 48, No. 3, 1567–1592
https://doi.org/10.1214/19-AOS1859

ON POST DIMENSION REDUCTION STATISTICAL INFERENCE

BY KYONGWON KIM[1,*], BING LI[1,**], ZHOU YU[2] AND LEXIN LI[3]

[1]*Department of Statistics, Pennsylvania State University,* [*]*kbk5206@psu.edu;* [**]*bxl9@psu.edu*
[2]*KLATASDS-MOE, School of Statistics, East China Normal University, zyu@stat.ecnu.edu.cn*
[3]*Department of Biostatistics and Epidemiology, University of California, Berkeley, lexinli@berkeley.edu*

The methodologies of sufficient dimension reduction have undergone extensive developments in the past three decades. However, there has been a lack of systematic and rigorous development of post dimension reduction inference, which has seriously hindered its applications. The current common practice is to treat the estimated sufficient predictors as the true predictors and use them as the starting point of the downstream statistical inference. However, this naive inference approach would grossly overestimate the confidence level of an interval, or the power of a test, leading to the distorted results. In this paper, we develop a general and comprehensive framework of post dimension reduction inference, which can accommodate any dimension reduction method and model building method, as long as their corresponding influence functions are available. Within this general framework, we derive the influence functions and present the explicit post reduction formulas for the combinations of numerous dimension reduction and model building methods. We then develop post reduction inference methods for both confidence interval and hypothesis testing. We investigate the finite-sample performance of our procedures by simulations and a real data analysis.

REFERENCES

[1] BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models* **2**. Springer, New York. MR1623559

[2] BONDELL, H. D. and LI, L. (2009). Shrinkage inverse regression estimation for model-free variable selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 287–299. MR2655534 https://doi.org/10.1111/j.1467-9868.2008.00686.x

[3] CHEN, X., ZOU, C. and COOK, R. D. (2010). Coordinate-independent sparse sufficient dimension reduction and variable selection. *Ann. Statist.* **38** 3696–3723. MR2766865 https://doi.org/10.1214/10-AOS826

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

[5] COOK, R. D. (1998). Principal Hessian directions revisited. *J. Amer. Statist. Assoc.* **93** 84–100. MR1614584 https://doi.org/10.2307/2669605

[6] COOK, R. D. and NI, L. (2005). Sufficient dimension reduction via inverse regression: A minimum discrepancy approach. *J. Amer. Statist. Assoc.* **100** 410–428. MR2160547 https://doi.org/10.1198/016214504000001501

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

[8] FERNHOLZ, L. T. (2012). *Von Mises Calculus for Statistical Functionals. Lecture Notes in Statistics* **19**. Springer, New York. MR0713611 https://doi.org/10.1007/978-1-4612-5604-5

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

[10] HANSEN, L. P., HEATON, J. and YARON, A. (1996). Finite-sample properties of some alternative GMM estimators. *J. Bus. Econom. Statist.* **14** 262–280.

[11] HE, X., FU, B. and FUNG, W. (2003). Median regression for longitudinal data. *Stat. Med.* **22** 3655–3669.

MSC2010 subject classifications. Primary 62G08; secondary 62H99.

Key words and phrases. Central subspace, directional regression, estimating equations, generalized method of moment, influence function, sliced inverse regression, Von Mises expansion.

[12] HYVÄRINEN, A., KARHUNEN, J. and OJA, E. (2004). *Independent Component Analysis* **46**. Wiley, New York.

[13] KIM, K., LI, B., YU, Z. and LI, L. (2020). Supplement to "On post dimension reduction statistical inference." https://doi.org/10.1214/19-AOS1859SUPP.

[14] LI, B. (1993). A deviance function for the quasi-likelihood method. *Biometrika* **80** 741–753. MR1282783 https://doi.org/10.1093/biomet/80.4.741

[15] LI, B. (2018). *Sufficient Dimension Reduction*: *Methods and Applications with R*. *Monographs on Statistics and Applied Probability* **161**. CRC Press, Boca Raton, FL. MR3838449 https://doi.org/10.1201/9781315119427

[16] LI, B., ARTEMIOU, A. and LI, L. (2011). Principal support vector machines for linear and nonlinear sufficient dimension reduction. *Ann. Statist.* **39** 3182–3210. MR3012405 https://doi.org/10.1214/11-AOS932

[17] LI, B. and SONG, J. (2017). Nonlinear sufficient dimension reduction for functional data. *Ann. Statist.* **45** 1059–1095. MR3662448 https://doi.org/10.1214/16-AOS1475

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

[19] LI, B., ZHA, H. and CHIAROMONTE, F. (2005). Contour regression: A general approach to dimension reduction. *Ann. Statist.* **33** 1580–1616. MR2166556 https://doi.org/10.1214/009053605000000192

[20] LI, G.-R., ZHU, L.-P. and ZHU, L.-X. (2010). Adaptive confidence region for the direction in semiparametric regressions. *J. Multivariate Anal.* **101** 1364–1377. MR2609498 https://doi.org/10.1016/j.jmva.2010.02.002

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

[22] LI, K.-C. (1992). On principal Hessian directions for data visualization and dimension reduction: Another application of Stein's lemma. *J. Amer. Statist. Assoc.* **87** 1025–1039. MR1209564

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

[24] LI, L. (2018). Sufficient dimension reduction. In *Wiley StatsRef*: *Statistics Reference Online* 1–8. American Cancer Society, New York. https://doi.org/10.1002/9781118445112.stat08042.

[25] LUO, W., LI, B. and YIN, X. (2014). On efficient dimension reduction with respect to a statistical functional of interest. *Ann. Statist.* **42** 382–412. MR3189490 https://doi.org/10.1214/13-AOS1195

[26] MA, Y. and ZHU, L. (2012). A semiparametric approach to dimension reduction. *J. Amer. Statist. Assoc.* **107** 168–179. MR2949349 https://doi.org/10.1080/01621459.2011.646925

[27] MA, Y. and ZHU, L. (2013). Efficient estimation in sufficient dimension reduction. *Ann. Statist.* **41** 250–268. MR3059417 https://doi.org/10.1214/12-AOS1072

[28] MA, Y. and ZHU, L. (2014). On estimation efficiency of the central mean subspace. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 885–901. MR3271171 https://doi.org/10.1111/rssb.12044

[29] MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models*. *Monographs on Statistics and Applied Probability*. CRC Press, London. MR3223057 https://doi.org/10.1007/978-1-4899-3242-6

[30] OWEN, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18** 90–120. MR1041387 https://doi.org/10.1214/aos/1176347494

[31] OWEN, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75** 237–249. MR0946049 https://doi.org/10.1093/biomet/75.2.237

[32] PARK, C. and LINDSAY, B. G. (1999). Robust estimation and tests based on quadratic inference function. Technical report.

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

[34] SHAO, Y., COOK, R. D. and WEISBERG, S. (2007). Marginal tests with sliced average variance estimation. *Biometrika* **94** 285–296. MR2331487 https://doi.org/10.1093/biomet/asm021

[35] SMOLA, A. J. and SCHÖLKOPF, B. (2004). A tutorial on support vector regression. *Stat. Comput.* **14** 199–222. MR2086398 https://doi.org/10.1023/B:STCO.0000035301.49549.88

[36] 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

[37] WANG, H. J. and WANG, L. (2009). Locally weighted censored quantile regression. *J. Amer. Statist. Assoc.* **104** 1117–1128. MR2562007 https://doi.org/10.1198/jasa.2009.tm08230

[38] WANG, Q. and YIN, X. (2008). A nonlinear multi-dimensional variable selection method for high dimensional data: Sparse MAVE. *Comput. Statist. Data Anal.* **52** 4512–4520. MR2432477 https://doi.org/10.1016/j.csda.2008.03.003

[39] XIA, Y. (2007). A constructive approach to the estimation of dimension reduction directions. *Ann. Statist.* **35** 2654–2690. MR2382662 https://doi.org/10.1214/009053607000000352

[40] XIA, Y., TONG, H., LI, W. K. and ZHU, L.-X. (2002). An adaptive estimation of dimension reduction space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 363–410. MR1924297 https://doi.org/10.1111/1467-9868.03411

[41] YIN, X., LI, B. and COOK, R. D. (2008). Successive direction extraction for estimating the central subspace in a multiple-index regression. *J. Multivariate Anal.* **99** 1733–1757. MR2444817 https://doi.org/10.1016/j.jmva.2008.01.006

[42] ZHU, L.-P., LI, L., LI, R. and ZHU, L.-X. (2011). Model-free feature screening for ultrahigh-dimensional data. *J. Amer. Statist. Assoc.* **106** 1464–1475. MR2896849 https://doi.org/10.1198/jasa.2011.tm10563

[43] ZHU, L.-X. and FANG, K.-T. (1996). Asymptotics for kernel estimate of sliced inverse regression. *Ann. Statist.* **24** 1053–1068. MR1401836 https://doi.org/10.1214/aos/1032526955

The Annals of Statistics
2020, Vol. 48, No. 3, 1593–1614
https://doi.org/10.1214/19-AOS1860

STATISTICAL AND COMPUTATIONAL LIMITS FOR SPARSE MATRIX DETECTION

BY T. TONY CAI[1] AND YIHONG WU[2]

[1]*Department of Statistics, The Wharton School, University of Pennsylvania, tcai@wharton.upenn.edu*
[2]*Department of Statistics and Data Science, Yale University, yihong.wu@yale.edu*

This paper investigates the fundamental limits for detecting a high-dimensional sparse matrix contaminated by white Gaussian noise from both the statistical and computational perspectives. We consider $p \times p$ matrices whose rows and columns are individually k-sparse. We provide a tight characterization of the statistical and computational limits for sparse matrix detection, which precisely describe when achieving optimal detection is easy, hard or impossible, respectively. Although the sparse matrices considered in this paper have no apparent submatrix structure and the corresponding estimation problem has no computational issue at all, the detection problem has a surprising computational barrier when the sparsity level k exceeds the cubic root of the matrix size p: attaining the optimal detection boundary is computationally at least as hard as solving the planted clique problem.

The same statistical and computational limits also hold in the sparse covariance matrix model, where each variable is correlated with at most k others. A key step in the construction of the statistically optimal test is a structural property of sparse matrices, which can be of independent interest.

REFERENCES

[1] ALON, N., ANDONI, A., KAUFMAN, T., MATULEF, K., RUBINFELD, R. and XIE, N. (2007). Testing k-wise and almost k-wise independence. In *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing* 496–505. ACM, New York. MR2402475 https://doi.org/10.1145/1250790.1250863

[2] ALON, N., KRIVELEVICH, M. and SUDAKOV, B. (1998). Finding a large hidden clique in a random graph. In *Proceedings of the Eighth International Conference "Random Structures and Algorithms" (Poznan*, 1997) **13** 457–466. MR1662795 https://doi.org/10.1002/(SICI)1098-2418(199810/12)13:3/4<457::AID-RSA14>3.3.CO;2-K

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

[4] ARIAS-CASTRO, E., BUBECK, S. and LUGOSI, G. (2015). Detecting positive correlations in a multivariate sample. *Bernoulli* **21** 209–241. MR3322317 https://doi.org/10.3150/13-BEJ565

[5] ARIAS-CASTRO, E., CANDÈS, E. J., HELGASON, H. and ZEITOUNI, O. (2008). Searching for a trail of evidence in a maze. *Ann. Statist.* **36** 1726–1757. MR2435454 https://doi.org/10.1214/07-AOS526

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

[7] ARIAS-CASTRO, E., DONOHO, D. L. and HUO, X. (2005). Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Inform. Theory* **51** 2402–2425. MR2246369 https://doi.org/10.1109/TIT.2005.850056

[8] ARIAS-CASTRO, E. and VERZELEN, N. (2014). Community detection in dense random networks. *Ann. Statist.* **42** 940–969. MR3210992 https://doi.org/10.1214/14-AOS1208

[9] BALAKRISHNAN, S., KOLAR, M., RINALDO, A., SINGH, A. and WASSERMAN, L. (2011). Statistical and computational tradeoffs in biclustering. In *NIPS* 2011 *Workshop on Computational Trade-Offs in Statistical Learning*.

[10] BERTHET, Q. and RIGOLLET, P. (2013). Complexity theoretic lower bounds for sparse principal component detection. *JMLR*: *Workshop Conf. Proc.* **30** 1046–1066.

MSC2010 subject classifications. Primary 62H15; secondary 62C20.
Key words and phrases. Minimax rates, computational limits, sparse covariance matrix, sparse detection.

[11] 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

[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] BUTUCEA, C. and INGSTER, Y. I. (2013). Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli* **19** 2652–2688. MR3160567 https://doi.org/10.3150/12-BEJ470

[14] 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

[15] CAI, T. T., JENG, X. J. and JIN, J. (2011). Optimal detection of heterogeneous and heteroscedastic mixtures. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 629–662. MR2867452 https://doi.org/10.1111/j.1467-9868.2011.00778.x

[16] CAI, T. T., LIANG, T. and RAKHLIN, A. (2017). Computational and statistical boundaries for submatrix localization in a large noisy matrix. *Ann. Statist.* **45** 1403–1430. MR3670183 https://doi.org/10.1214/16-AOS1488

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

[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 WU, Y. (2020). Supplement to "Statistical and computational limits for sparse matrix detection." https://doi.org/10.1214/19-AOS1860SUPP.

[20] CAI, T. T. and YUAN, M. (2014). Minimax rate optimal detection of very short signal segments. Preprint. Available at arXiv:1407.2812.

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

[22] CANDÈS, E. J., LI, X., MA, Y. and WRIGHT, J. (2011). Robust principal component analysis? *J. ACM* **58** Art. 11, 37. MR2811000 https://doi.org/10.1145/1970392.1970395

[23] CHEN, S. X., ZHANG, L.-X. and ZHONG, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** 810–819. MR2724863 https://doi.org/10.1198/jasa.2010.tm09560

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

[25] 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

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

[27] FAN, J., RIGOLLET, P. and WANG, W. (2015). Estimation of functionals of sparse covariance matrices. *Ann. Statist.* **43** 2706–2737. MR3405609 https://doi.org/10.1214/15-AOS1357

[28] FEDOTOV, A. A., HARREMOËS, P. and TOPSØE, F. (2003). Refinements of Pinsker's inequality. *IEEE Trans. Inform. Theory* **49** 1491–1498. MR1984937 https://doi.org/10.1109/TIT.2003.811927

[29] FELDMAN, V., GRIGORESCU, E., REYZIN, L., VEMPALA, S. S. and XIAO, Y. (2013). Statistical algorithms and a lower bound for detecting planted cliques. In *STOC'13—Proceedings of the* 2013 *ACM Symposium on Theory of Computing* 655–664. ACM, New York. MR3210827 https://doi.org/10.1145/2488608.2488692

[30] GAO, C., MA, Z. and ZHOU, H. H. (2017). Sparse CCA: Adaptive estimation and computational barriers. *Ann. Statist.* **45** 2074–2101. MR3718162 https://doi.org/10.1214/16-AOS1519

[31] GOLUB, G. H. and VAN LOAN, C. F. (1996). *Matrix Computations*, 3rd ed. *Johns Hopkins Studies in the Mathematical Sciences*. Johns Hopkins Univ. Press, Baltimore, MD. MR1417720

[32] HAJEK, B., WU, Y. and XU, J. (2015). Computational lower bounds for community detection on random graphs. In *Proceedings of COLT* 2015 899–928. Available at arXiv:1406.6625.

[33] HAZAN, E. and KRAUTHGAMER, R. (2011). How hard is it to approximate the best Nash equilibrium? *SIAM J. Comput.* **40** 79–91. MR2765712 https://doi.org/10.1137/090766991

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

[35] INGSTER, Y. I. and SUSLINA, I. A. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*. *Lecture Notes in Statistics* **169**. Springer, New York. MR1991446 https://doi.org/10.1007/978-0-387-21580-8

[36] JERRUM, M. (1992). Large cliques elude the Metropolis process. *Random Structures Algorithms* **3** 347–359. MR1179827 https://doi.org/10.1002/rsa.3240030402

[37] KLOPP, O. and TSYBAKOV, A. B. (2015). Estimation of matrices with row sparsity. *Probl. Inf. Transm.* **51** 335–348. Translation of *Problemy Peredachi Informatsii* **51** (2015), no. 4, 32–46. MR3449576 https://doi.org/10.1134/s0032946015040031

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

[39] MA, Z. and WU, Y. (2015). Computational barriers in minimax submatrix detection. *Ann. Statist.* **43** 1089–1116. MR3346698 https://doi.org/10.1214/14-AOS1300

[40] MA, Z. and WU, Y. (2015). Volume ratio, sparsity, and minimaxity under unitarily invariant norms. *IEEE Trans. Inform. Theory* **61** 6939–6956. MR3430731 https://doi.org/10.1109/TIT.2015.2487541

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

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

[43] RUDELSON, M. and VERSHYNIN, R. (2007). Sampling from large matrices: An approach through geometric functional analysis. *J. ACM* **54** Art. 21, 19. MR2351844 https://doi.org/10.1145/1255443.1255449

[44] 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

[45] WANG, T., BERTHET, Q. and SAMWORTH, R. J. (2016). Statistical and computational trade-offs in estimation of sparse principal components. *Ann. Statist.* **44** 1896–1930. MR3546438 https://doi.org/10.1214/15-AOS1369

[46] YANG, D., MA, Z. and BUJA, A. (2016). Rate optimal denoising of simultaneously sparse and low rank matrices. *J. Mach. Learn. Res.* **17** Paper No. 92, 27. MR3543498

The Annals of Statistics
2020, Vol. 48, No. 3, 1615–1647
https://doi.org/10.1214/19-AOS1861

SEGMENTATION AND ESTIMATION OF CHANGE-POINT MODELS: FALSE POSITIVE CONTROL AND CONFIDENCE REGIONS

BY XIAO FANG[1], JIAN LI[2] AND DAVID SIEGMUND[3]

[1]*Department of Statistics, The Chinese University of Hong Kong, xfang@sta.cuhk.edu.hk*
[2]*Adobe Systems, jianli0124@gmail.com*
[3]*Department of Statistics, Stanford University, siegmund@stanford.edu*

To segment a sequence of independent random variables at an unknown number of change-points, we introduce new procedures that are based on thresholding the likelihood ratio statistic, and give approximations for the probability of a false positive error when there are no change-points. We also study confidence regions based on the likelihood ratio statistic for the change-points and joint confidence regions for the change-points and the parameter values. Applications to segment array CGH data are discussed.

REFERENCES

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

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

CHAN, H. P. and CHEN, H. (2017). Multi-sequence segmentation via score and higher-criticism tests. Available at arXiv:1706.07586v1.

CHURCHILL, G. A. (1989). Stochastic models for heterogeneous DNA sequences. *Bull. Math. Biol.* **51** 79–94. MR0978904 https://doi.org/10.1016/S0092-8240(89)80049-7

DU, C., KAO, C.-L. M. and KOU, S. C. (2016). Stepwise signal extraction via marginal likelihood. *J. Amer. Statist. Assoc.* **111** 314–330. MR3494662 https://doi.org/10.1080/01621459.2015.1006365

DÜMBGEN, L. and SPOKOINY, V. G. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29** 124–152. MR1833961 https://doi.org/10.1214/aos/996986504

ELHAIK, E., GRAUR, D. and JOSIĆ, K. (2010). Comparative testing of DNA segmentation algorithms using benchmark simulations. *Mol. Biol. Evol.* **27** 1015–1024.

FANG, X., LI, J. and SIEGMUND, D. (2020). Supplement to "Segmentation and estimation of change-point models: False positive control and confidence regions." https://doi.org/10.1214/19-AOS1861SUPP.

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

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

HAO, N., NIU, Y. S. and ZHANG, H. (2013). Multiple change-point detection via a screening and ranking algorithm. *Statist. Sinica* **23** 1553–1572. MR3222810

LAI, W. R., JOHNSON, M. D., KUCHERLAPATI, R. and PARK, P. J. (2005). Comparative analysis of algorithms for identifying amplifications and deletions in array CGH data. *Bioinformatics* **21** 3763–3770.

NIU, Y. S. and ZHANG, H. (2012). The screening and ranking algorithm to detect DNA copy number variations. *Ann. Appl. Stat.* **6** 1306–1326. MR3012531 https://doi.org/10.1214/12-AOAS539

OLSHEN, A. B., VENKATRAMAN, E. S., LUCITO, R. and WIGLER, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5** 557–572.

PICARD, F., ROBIN, S., LAVIELLE, M., VAISSE, C. and DAUDIN, J. J. (2005). A statistical approach for array CGH data analysis. *BMC Bioinform.* **6** 27.

POLLACK, J. R., PEROU, C. M., ALIZADEH, A. A., EISEN, M. B., PERGAMENSCHIKOV, A., WILLIAMS, C. F., JEFFREY, S. S., BOTSTEIN, D. and BROWN, P. O. (1999). Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* **23** 41–46.

MSC2010 subject classifications. 62G05, 62G15.

Key words and phrases. Array CGH analysis, change-points, confidence regions, exponential families, likelihood ratio statistics.

POLLACK, J. R., SØRLIE, T., PEROU, C. M., REES, C. A., JEFFREY, S. S., LONNING, P. E., TIBSHIRANI, R., BOTSTEIN, D., BØRRESEN-DALE, A. L. et al. (2002). Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc. Natl. Acad. Sci. USA* **99** 12963–12968.

ROBBINS, M. W., GALLAGHER, C. M. and LUND, R. B. (2016). A general regression changepoint test for time series data. *J. Amer. Statist. Assoc.* **111** 670–683. MR3538696 https://doi.org/10.1080/01621459.2015.1029130

SCHWARTZMAN, A., GAVRILOV, Y. and ADLER, R. J. (2011). Multiple testing of local maxima for detection of peaks in 1D. *Ann. Statist.* **39** 3290–3319. MR3012409 https://doi.org/10.1214/11-AOS943

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

SIEGMUND, D. (1988a). Approximate tail probabilities for the maxima of some random fields. *Ann. Probab.* **16** 487–501. MR0929059

SIEGMUND, D. (1988b). Confidence sets in change-point problems. *Int. Stat. Rev.* **56** 31–48. MR0963139 https://doi.org/10.2307/1403360

SIEGMUND, D. and YAKIR, B. (2000). Tail probabilities for the null distribution of scanning statistics. *Bernoulli* **6** 191–213. MR1748719 https://doi.org/10.2307/3318574

SIEGMUND, D. and YAKIR, B. (2007). *The Statistics of Gene Mapping. Statistics for Biology and Health*. Springer, New York. MR2301277

SIEGMUND, D. O., ZHANG, N. R. and YAKIR, B. (2011). False discovery rate for scanning statistics. *Biometrika* **98** 979–985. MR2860337 https://doi.org/10.1093/biomet/asr057

SNIJDERS, A. M., FRIDLYAND, J., MANS, D. A., SEGRAVES, R., JAIN, A. N., PINKEL, D. and ALBERTSONN, D. G. (2003). Shaping of tumor and drug-resistant genomes by instability and selection. *Oncogene* **22** 4370–4379.

TU, I. and SIEGMUND, D. (1999). The maximum of a function of a Markov chain and application to linkage analysis. *Adv. in Appl. Probab.* **31** 510–531. MR1724565 https://doi.org/10.1239/aap/1029955145

VOSTRIKOVA, L. (1981). Detecting 'disorder' in multidimensional random processes. *Sov. Math., Dokl.* **24** 55–59.

WORSLEY, K. J. (1986). Confidence regions and test for a change-point in a sequence of exponential family random variables. *Biometrika* **73** 91–104. MR0836437 https://doi.org/10.1093/biomet/73.1.91

YAKIR, B. (2013). *Extremes in Random Fields: A Theory and Its Applications. Wiley Series in Probability and Statistics*. Wiley, Chichester. MR3241226 https://doi.org/10.1002/9781118720608

ZHANG, Y. and LIU, J. S. (2011). Fast and accurate approximation to significance tests in genome-wide association studies. *J. Amer. Statist. Assoc.* **106** 846–857. MR2894742 https://doi.org/10.1198/jasa.2011.ap10657

ZHANG, N. R. and SIEGMUND, D. O. (2007). A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics* **63** 22–32. MR2345571 https://doi.org/10.1111/j.1541-0420.2006.00662.x

ZHANG, N. R., SIEGMUND, D. O., JI, H. and LI, J. Z. (2010). Detecting simultaneous changepoints in multiple sequences. With supplementary data available online. *Biometrika* **97** 631–645. MR2672488 https://doi.org/10.1093/biomet/asq025

ZHANG, N. R., YAKIR, B., XIA, L. C. and SIEGMUND, D. (2016). Scan statistics on Poisson random fields with applications in genomics. *Ann. Appl. Stat.* **10** 726–755. MR3528358 https://doi.org/10.1214/15-AOAS892

ZHAO, X., LI, C., PAEZ, J. G., CHIN, K., JÄNNE, P. A., CHEN, T.-H., GIRARD, L., MINNA, J., CHRISTIANI, D. et al. (2004). An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Res.* **64** 3060–3071.

The Annals of Statistics
2020, Vol. 48, No. 3, 1648–1664
https://doi.org/10.1214/19-AOS1862

ROBUST COVARIANCE ESTIMATION UNDER $L_4 - L_2$ NORM EQUIVALENCE

BY SHAHAR MENDELSON[1] AND NIKITA ZHIVOTOVSKIY[2]

[1]*Mathematical Sciences Institute, Australian National University, shahar.mendelson@upmc.fr*
[2]*Higher School of Economics, nikita.zhivotovskiy@phystech.edu*

Let X be a centered random vector taking values in $\mathbb{R}^d$ and let $\Sigma = \mathbb{E}(X \otimes X)$ be its covariance matrix. We show that if X satisfies an $L_4 - L_2$ norm equivalence (sometimes referred to as the bounded kurtosis assumption), there is a covariance estimator $\hat{\Sigma}$ that exhibits almost the same performance one would expect had X been a Gaussian vector. The procedure also improves the current state-of-the-art regarding high probability bounds in the sub-Gaussian case (sharp results were only known in expectation or with constant probability).

In both scenarios the new bounds do not depend explicitly on the dimension d, but rather on the effective rank of the covariance matrix Σ.

REFERENCES

[1] CATONI, O. (2012). Challenging the empirical mean and empirical variance: A deviation study. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 1148–1185. MR3052407 https://doi.org/10.1214/11-AIHP454

[2] 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

[3] KOTZ, S. and NADARAJAH, S. (2004). *Multivariate t Distributions and Their Applications*. Cambridge Univ. Press, Cambridge. MR2038227 https://doi.org/10.1017/CBO9780511550683

[4] LATAŁA, R. and WOJTASZCZYK, J. O. (2008). On the infimum convolution inequality. *Studia Math.* **189** 147–187. MR2449135 https://doi.org/10.4064/sm189-2-5

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

[6] LUGOSI, G. and MENDELSON, S. (2019). Near-optimal mean estimators with respect to general norms. *Probab. Theory Related Fields*. To appear.

[7] MENDELSON, S. (2010). Empirical processes with a bounded ψ_1 diameter. *Geom. Funct. Anal.* **20** 988–1027. MR2729283 https://doi.org/10.1007/s00039-010-0084-5

[8] MENDELSON, S. (2016). Upper bounds on product and multiplier empirical processes. *Stochastic Process. Appl.* **126** 3652–3680. MR3565471 https://doi.org/10.1016/j.spa.2016.04.019

[9] MENDELSON, S., PAJOR, A. and TOMCZAK-JAEGERMANN, N. (2007). Reconstruction and subgaussian operators in asymptotic geometric analysis. *Geom. Funct. Anal.* **17** 1248–1282. MR2373017 https://doi.org/10.1007/s00039-007-0618-7

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

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

[12] MINSKER, S. and WEI, X. (2017). Estimation of the covariance structure of heavy-tailed distributions. In *NIPS*.

[13] MINSKER, S. and WEI, X. (2018). Robust modifications of U-statistics and applications to covariance estimation problems. Preprint. Available at arXiv:1801.05565.

[14] NEMIROVSKY, A. S. and YUDIN, D. B. (1983). *Problem Complexity and Method Efficiency in Optimization. A Wiley-Interscience Publication*. Wiley, New York. MR0702836

[15] SO, W. (1990). Facial structures of Schatten p-norms. *Linear Multilinear Algebra* **27** 207–212. MR1064897 https://doi.org/10.1080/03081089008818012

MSC2010 subject classifications. Primary 62G35; secondary 62G15.
Key words and phrases. Covariance estimation, robust estimation, median of means.

[16] TALAGRAND, M. (2014). *Upper and Lower Bounds for Stochastic Processes*: *Modern Methods and Classical Problems*. **60**. Springer, Heidelberg. MR3184689 https://doi.org/10.1007/978-3-642-54075-2

[17] TROPP, J. (2015). An introduction to matrix concentration inequalities. *Found. Trends Mach. Learn.* **8** 1–230.

The Annals of Statistics
2020, Vol. 48, No. 3, 1665–1691
https://doi.org/10.1214/19-AOS1863

ROBUST INFERENCE VIA MULTIPLIER BOOTSTRAP

BY XI CHEN[1] AND WEN-XIN ZHOU[2]

[1]*Stern School of Business, New York University, xchen3@stern.nyu.edu*
[2]*Department of Mathematics, University of California, San Diego, wez243@ucsd.edu*

This paper investigates the theoretical underpinnings of two fundamental statistical inference problems, the construction of confidence sets and large-scale simultaneous hypothesis testing, in the presence of heavy-tailed data. With heavy-tailed observation noise, finite sample properties of the least squares-based methods, typified by the sample mean, are suboptimal both theoretically and empirically. In this paper, we demonstrate that the adaptive Huber regression, integrated with the multiplier bootstrap procedure, provides a useful robust alternative to the method of least squares. Our theoretical and empirical results reveal the effectiveness of the proposed method, and highlight the importance of having inference methods that are robust to heavy tailedness.

REFERENCES

ARLOT, S., BLANCHARD, G. and ROQUAIN, E. (2010). Some nonasymptotic results on resampling in high dimension. I. Confidence regions. *Ann. Statist.* **38** 51–82. MR2589316 https://doi.org/10.1214/08-AOS667

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

BARRAS, L., SCAILLET, O. and WERMERS, R. (2010). False discoveries in mutual fund performance: Measuring luck in estimated alphas. *J. Finance* **65** 179–216.

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

BERK, J. B. and GREEN, R. C. (2004). Mutual fund flows and performance in rational markets. *J. Polit. Econ.* **112** 1269–1295.

BROWNLEES, C., JOLY, E. and LUGOSI, G. (2015). Empirical risk minimization for heavy-tailed losses. *Ann. Statist.* **43** 2507–2536. MR3405602 https://doi.org/10.1214/15-AOS1350

CATONI, O. (2012). Challenging the empirical mean and empirical variance: A deviation study. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 1148–1185. MR3052407 https://doi.org/10.1214/11-AIHP454

CATONI, O. and GIULINI, L. (2017). Dimension free PAC-Bayesian bounds for matrices, vectors, and linear least squares regression. Technical Report.

CHATTERJEE, S. and BOSE, A. (2005). Generalized bootstrap for estimating equations. *Ann. Statist.* **33** 414–436. MR2157808 https://doi.org/10.1214/009053604000000904

CHEN, X. and ZHOU, W.-X (2020). Supplement to "Robust inference via multiplier bootstrap." https://doi.org/10.1214/19-AOS1863SUPP.

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. MR3161448 https://doi.org/10.1214/13-AOS1161

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Anti-concentration and honest, adaptive confidence bands. *Ann. Statist.* **42** 1787–1818. MR3262468 https://doi.org/10.1214/14-AOS1235

DELAIGLE, A., HALL, P. and JIN, J. (2011). Robustness and accuracy of methods for high dimensional data analysis based on Student's t-statistic. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 283–301. MR2815777 https://doi.org/10.1111/j.1467-9868.2010.00761.x

DESAI, K. H. and STOREY, J. D. (2012). Cross-dimensional inference of dependent high-dimensional data. *J. Amer. Statist. Assoc.* **107** 135–151. MR2949347 https://doi.org/10.1080/01621459.2011.645777

DEVROYE, L., LERASLE, M., LUGOSI, G. and OLIVEIRA, R. I. (2016). Sub-Gaussian mean estimators. *Ann. Statist.* **44** 2695–2725. MR3576558 https://doi.org/10.1214/16-AOS1440

MSC2010 subject classifications. Primary 62F35, 62F40; secondary 62J05, 62J15.

Key words and phrases. Confidence set, heavy-tailed data, multiple testing, multiplier bootstrap, robust regression, Wilks' theorem.

DUDOIT, S. and VAN DER LAAN, M. J. (2008). *Multiple Testing Procedures with Applications to Genomics*. *Springer Series in Statistics*. Springer, New York. MR2373771 https://doi.org/10.1007/978-0-387-49317-6

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

FAMA, E. F. and FRENCH, K. R. (1993). Common risk factors in the returns on stocks and bonds. *J. Financ. Econ.* **33** 3–56.

FAN, J., HALL, P. and YAO, Q. (2007). To how many simultaneous hypothesis tests can normal, Student's t or bootstrap calibration be applied? *J. Amer. Statist. Assoc.* **102** 1282–1288. MR2372536 https://doi.org/10.1198/016214507000000969

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

FAN, J., LI, Q. and WANG, Y. (2017). Estimation of high dimensional mean regression in the absence of symmetry and light tail assumptions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 247–265. MR3597972 https://doi.org/10.1111/rssb.12166

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

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

GIULINI, I. (2017). Robust PCA and pairs of projections in a Hilbert space. *Electron. J. Stat.* **11** 3903–3926. MR3714302 https://doi.org/10.1214/17-EJS1343

HAHN, M. G., KUELBS, J. and WEINER, D. C. (1990). The asymptotic joint distribution of self-normalized censored sums and sums of squares. *Ann. Probab.* **18** 1284–1341. MR1062070

HSU, D. and SABATO, S. (2016). Loss minimization and parameter estimation with heavy tails. *J. Mach. Learn. Res.* **17** 18. MR3491112

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

HUBER, P. J. and RONCHETTI, E. M. (2009). *Robust Statistics*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2488795 https://doi.org/10.1002/9780470434697

LAN, W. and DU, L. (2019). A factor-adjusted multiple testing procedure with application to mutual fund selection. *J. Bus. Econom. Statist.* **37** 147–157. MR3910232 https://doi.org/10.1080/07350015.2017.1294078

LEPSKIĬ, O. V. (1991). Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Teor. Veroyatn. Primen.* **36** 645–659. MR1147167 https://doi.org/10.1137/1136085

LINTNER, J. (1965). The valuation of risk assets and the selection of risky investment in stock portfolios and capital budgets. *Rev. Econ. Stat.* **47** 13–37.

LIU, W. and SHAO, Q.-M. (2014). Phase transition and regularized bootstrap in large-scale t-tests with false discovery rate control. *Ann. Statist.* **42** 2003–2025. MR3262475 https://doi.org/10.1214/14-AOS1249

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

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

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

QI, L. and SUN, D. (1999). A survey of some nonsmooth equations and smoothing Newton methods. In *Progress in Optimization*. *Appl. Optim.* **30** 121–146. Kluwer Academic, Dordrecht. MR1719516 https://doi.org/10.1007/978-1-4613-3285-5_7

SHARPE, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *J. Finance* **19** 425–442.

SPOKOINY, V. and ZHILOVA, M. (2015). Bootstrap confidence sets under model misspecification. *Ann. Statist.* **43** 2653–2675. MR3405607 https://doi.org/10.1214/15-AOS1355

SUN, Q., ZHOU, W.-X. and FAN, J. (2019). Adaptive Huber regression. Technical Report.

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

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

WANG, J., ZHAO, Q., HASTIE, T. and OWEN, A. B. (2017). Confounder adjustment in multiple hypothesis testing. *Ann. Statist.* **45** 1863–1894. MR3718155 https://doi.org/10.1214/16-AOS1511

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

ZHILOVA, M. (2016). Non-classical Berry–Esseen inequality and accuracy of the weighted bootstrap. Technical Report.

ZHOU, W.-X., BOSE, K., FAN, J. and LIU, H. (2018). A new perspective on robust M-estimation: Finite sample theory and applications to dependence-adjusted multiple testing. *Ann. Statist.* **46** 1904–1931. MR3845005 https://doi.org/10.1214/17-AOS1606

The Annals of Statistics
2020, Vol. 48, No. 3, 1692–1717
https://doi.org/10.1214/19-AOS1864

ON THE OPTIMAL RECONSTRUCTION OF PARTIALLY OBSERVED FUNCTIONAL DATA

BY ALOIS KNEIP* AND DOMINIK LIEBL**

*Statistische Abteilung, University of Bonn, *akneip@uni-bonn.de; **dliebl@uni-bonn.de*

We propose a new reconstruction operator that aims to recover the missing parts of a function given the observed parts. This new operator belongs to a new, very large class of functional operators which includes the classical regression operators as a special case. We show the optimality of our reconstruction operator and demonstrate that the usually considered regression operators generally cannot be optimal reconstruction operators. Our estimation theory allows for autocorrelated functional data and considers the practically relevant situation in which each of the n functions is observed at m_i, $i = 1, \dots, n$, discretization points. We derive rates of consistency for our nonparametric estimation procedures using a double asymptotic. For data situations, as in our real data application where m_i is considerably smaller than n, we show that our functional principal components based estimator can provide better rates of convergence than conventional nonparametric smoothing methods.

REFERENCES

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

CAI, T. T. and HALL, P. (2006). Prediction in functional linear regression. *Ann. Statist.* **34** 2159–2179. MR2291496 https://doi.org/10.1214/009053606000000830

CARDOT, H., MAS, A. and SARDA, P. (2007). CLT in functional linear regression models. *Probab. Theory Related Fields* **138** 325–361. MR2299711 https://doi.org/10.1007/s00440-006-0025-2

CLUDIUS, J., HERMANN, H., MATTHES, F. C. and GRAICHEN, V. (2014). The merit order effect of wind and photovoltaic electricity generation in Germany 2008–2016: Estimation and distributional implications. *Energy Econ.* **44** 302–313.

DELAIGLE, A. and HALL, P. (2013). Classification using censored functional data. *J. Amer. Statist. Assoc.* **108** 1269–1283. MR3174707 https://doi.org/10.1080/01621459.2013.824893

DELAIGLE, A. and HALL, P. (2016). Approximating fragmented functional data by segments of Markov chains. *Biometrika* **103** 779–799. MR3620439 https://doi.org/10.1093/biomet/asw040

DELAIGLE, A., HALL, P., HUANG, W. and KNEIP, A. (2018). Estimating the covariance of fragmented and other incompletely observed functional data. Working Paper.

DESCARY, M.-H. and PANARETOS, V. M. (2019). Recovering covariance from functional fragments. *Biometrika* **106** 145–160. MR3912388 https://doi.org/10.1093/biomet/asy055

FANONE, E., GAMBA, A. and PROKOPCZUK, M. (2013). The case of negative day-ahead electricity prices. *Energy Econ.* **35** 22–34.

GOLDBERG, Y., RITOV, Y. and MANDELBAUM, A. (2014). Predicting the continuation of a function with applications to call center data. *J. Statist. Plann. Inference* **147** 53–65. MR3151845 https://doi.org/10.1016/j.jspi.2013.11.006

GROMENKO, O., KOKOSZKA, P. and SOJKA, J. (2017). Evaluation of the cooling trend in the ionosphere using functional regression with incomplete curves. *Ann. Appl. Stat.* **11** 898–918. MR3693551 https://doi.org/10.1214/17-AOAS1022

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

HALL, P., MÜLLER, H.-G. and WANG, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34** 1493–1517. MR2278365 https://doi.org/10.1214/009053606000000272

MSC2010 subject classifications. 62M20, 62H25, 62G05, 62G08.
Key words and phrases. Functional data analysis, functional principal components, incomplete functions.

HIRTH, L. (2013). The market value of variable renewables: The effect of solar wind power variability on their relative price. *Energy Econ*. **38** 218–236.

HORVÁTH, L. and KOKOSZKA, P. (2012). *Inference for Functional Data with Applications*. *Springer Series in Statistics*. Springer, New York. MR2920735 https://doi.org/10.1007/978-1-4614-3655-3

KNEIP, A. and LIEBL, D. (2020). Supplement to "On the optimal reconstruction of partially observed functional data." https://doi.org/10.1214/19-AOS1864SUPP.

KRAUS, D. (2015). Components and completion of partially observed functional data. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **77** 777–801. MR3382597 https://doi.org/10.1111/rssb.12087

LI, Y. and HSING, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist*. **38** 3321–3351. MR2766854 https://doi.org/10.1214/10-AOS813

LIEBL, D. (2013). Modeling and forecasting electricity spot prices: A functional data perspective. *Ann. Appl. Stat*. **7** 1562–1592. MR3127959 https://doi.org/10.1214/13-AOAS652

LIEBL, D. (2019). Nonparametric testing for differences in electricity prices: The case of the Fukushima nuclear accident. *Ann. Appl. Stat*. **13** 1128–1146. MR3963565 https://doi.org/10.1214/18-AOAS1230

NICOLOSI, M. (2010). Wind power integration and power system flexibility—an empirical analysis of extreme events in Germany under the new negative price regime. *Energy Policy* **38** 7257–7268.

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

RICE, J. A. and SILVERMAN, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B* **53** 233–243. MR1094283

WEIGT, H. (2009). Germany's wind energy: The potential for fossil capacity replacement and cost saving. *Appl. Energy* **86** 1857–1863.

WERON, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *Int. J. Forecast*. **30** 1030–1081.

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005a). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc*. **100** 577–590. MR2160561 https://doi.org/10.1198/016214504000001745

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005b). Functional linear regression analysis for longitudinal data. *Ann. Statist*. **33** 2873–2903. MR2253106 https://doi.org/10.1214/009053605000000660

ZHANG, J.-T. and CHEN, J. (2007). Statistical inferences for functional data. *Ann. Statist*. **35** 1052–1079. MR2341698 https://doi.org/10.1214/009053606000001505

ZHANG, X. and WANG, J.-L. (2016). From sparse to dense functional data and beyond. *Ann. Statist*. **44** 2281–2321. MR3546451 https://doi.org/10.1214/16-AOS1446

The Annals of Statistics
2020, Vol. 48, No. 3, 1718–1741
https://doi.org/10.1214/19-AOS1865

LARGE SAMPLE PROPERTIES OF PARTITIONING-BASED SERIES ESTIMATORS

BY MATIAS D. CATTANEO[1], MAX H. FARRELL[2] AND YINGJIE FENG[3]

[1]*Department of Operations Research and Financial Engineering, Princeton University, cattaneo@princeton.edu*
[2]*Booth School of Business, University of Chicago, max.farrell@chicagobooth.edu*
[3]*Department of Politics, Princeton University, yingjief@princeton.edu*

We present large sample results for partitioning-based least squares nonparametric regression, a popular method for approximating conditional expectation functions in statistics, econometrics and machine learning. First, we obtain a general characterization of their leading asymptotic bias. Second, we establish integrated mean squared error approximations for the point estimator and propose feasible tuning parameter selection. Third, we develop pointwise inference methods based on undersmoothing and robust bias correction. Fourth, employing different coupling approaches, we develop uniform distributional approximations for the undersmoothed and robust bias-corrected t-statistic processes and construct valid confidence bands. In the univariate case, our uniform distributional approximations require seemingly minimal rate restrictions and improve on approximation rates known in the literature. Finally, we apply our general results to three partitioning-based estimators: splines, wavelets and piecewise polynomials. The Supplemental Appendix includes several other general and example-specific technical and methodological results. A companion R package is provided.

REFERENCES

[1] AGARWAL, G. G. and STUDDEN, W. J. (1980). Asymptotic integrated mean square error using least squares and bias minimizing splines. *Ann. Statist.* **8** 1307–1325. MR0594647

[2] BELLONI, A., CHERNOZHUKOV, V., CHETVERIKOV, D. and FERNANDEZ-VAL, I. (2019). Conditional quantile processes based on series or many regressors. *J. Econometrics*. To appear.

[3] BELLONI, A., CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Some new asymptotic theory for least squares series: Pointwise and uniform results. *J. Econometrics* **186** 345–366. MR3343791 https://doi.org/10.1016/j.jeconom.2015.02.014

[4] BHATIA, R. (2013). *Matrix Analysis*. Springer, New York. MR1477662 https://doi.org/10.1007/978-1-4612-0653-8

[5] BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. and STONE, C. J. (1984). *Classification and Regression Trees*. *Wadsworth Statistics/Probability Series*. Wadsworth Advanced Books and Software, Belmont, CA. MR0726392

[6] CALONICO, S., CATTANEO, M. D. and FARRELL, M. H. (2018). On the effect of bias estimation on coverage accuracy in nonparametric inference. *J. Amer. Statist. Assoc.* **113** 767–779. MR3832225 https://doi.org/10.1080/01621459.2017.1285776

[7] CALONICO, S., CATTANEO, M. D. and FARRELL, M. H. (2019). Coverage error optimal confidence intervals for local polynomial regression. arXiv:1808.01398.

[8] CALONICO, S., CATTANEO, M. D. and TITIUNIK, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* **82** 2295–2326. MR3301169 https://doi.org/10.3982/ECTA11757

[9] CALONICO, S., CATTANEO, M. D. and TITIUNIK, R. (2015). Optimal data-driven regression discontinuity plots. *J. Amer. Statist. Assoc.* **110** 1753–1769. MR3449071 https://doi.org/10.1080/01621459.2015.1017578

MSC2010 subject classifications. Primary 62E20, 62G08, 62G20; secondary 62M99.

Key words and phrases. Nonparametric regression, series methods, sieve methods, robust bias correction, uniform inference, strong approximation, tuning parameter selection.

[10] CATTANEO, M. D., CRUMP, R. K., FARRELL, M. H. and FENG, Y. (2019). On binscatter. arXiv:1902.09608.

[11] CATTANEO, M. D., CRUMP, R. K., FARRELL, M. H. and SCHAUMBURG, E. (2019). Characteristic-sorted portfolios: Estimation and inference. *Rev. Econ. Stat*. To appear.

[12] CATTANEO, M. D. and FARRELL, M. H. (2011). Efficient estimation of the dose-response function under ignorability using subclassification on the covariates. In *Missing Data Methods*: *Cross-Sectional Methods and Applications*. *Adv. Econom*. **27** 93–127. Emerald Group Publ, Ltd., Bingley. MR3495703

[13] CATTANEO, M. D. and FARRELL, M. H. (2013). Optimal convergence rates, Bahadur representation, and asymptotic normality of partitioning estimators. *J. Econometrics* **174** 127–143. MR3045024 https://doi.org/10.1016/j.jeconom.2013.02.002

[14] CATTANEO, M. D., FARRELL, M. H. and FENG, Y. (2019). `lspartition`: Partitioning-based least squares regression. *R J*. To appear.

[15] CATTANEO, M. D., FARRELL, M. H. and FENG, Y. (2020). Supplement to "Large sample properties of partitioning-based series estimators." https://doi.org/10.1214/19-AOS1865SUPP.

[16] CHEN, X. and CHRISTENSEN, T. M. (2015). Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. *J. Econometrics* **188** 447–465. MR3383220 https://doi.org/10.1016/j.jeconom.2015.03.010

[17] CHEN, X. and CHRISTENSEN, T. M. (2018). Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression. *Quant. Econ*. **9** 39–84. MR3789729 https://doi.org/10.3982/QE722

[18] 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

[19] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Anti-concentration and honest, adaptive confidence bands. *Ann. Statist*. **42** 1787–1818. MR3262468 https://doi.org/10.1214/14-AOS1235

[20] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* **162** 47–70. MR3350040 https://doi.org/10.1007/s00440-014-0565-9

[21] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2016). Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings. *Stochastic Process. Appl*. **126** 3632–3651. MR3565470 https://doi.org/10.1016/j.spa.2016.04.009

[22] CHERNOZHUKOV, V., LEE, S. and ROSEN, A. M. (2013). Intersection bounds: Estimation and inference. *Econometrica* **81** 667–737. MR3043345 https://doi.org/10.3982/ECTA8718

[23] DAVYDOV, O. (2001). Stable local bases for multivariate spline spaces. *J. Approx. Theory* **111** 267–297. MR1849550 https://doi.org/10.1006/jath.2001.3577

[24] EGGERMONT, P. P. B. and LARICCIA, V. N. (2009). *Maximum Penalized Likelihood Estimation*. *Volume II*: *Regression*. *Springer Series in Statistics*. Springer, Dordrecht. MR2817245 https://doi.org/10.1007/b12285

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

[26] 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

[27] HUANG, J. Z. (1998). Projection estimation in multiple regression with application to functional ANOVA models. *Ann. Statist*. **26** 242–272. MR1611780 https://doi.org/10.1214/aos/1030563984

[28] HUANG, J. Z. (2003). Local asymptotics for polynomial spline regression. *Ann. Statist*. **31** 1600–1635. MR2012827 https://doi.org/10.1214/aos/1065705120

[29] JURINSKIĬ, V. V. (1978). On the error of the Gaussian approximation for convolutions. *Theory Probab. Appl*. **22** 236–247. MR0517490

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

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

[32] NEWEY, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *J. Econometrics* **79** 147–168. MR1457700 https://doi.org/10.1016/S0304-4076(97)00011-0

[33] NOBEL, A. (1996). Histogram regression estimation using data-dependent partitions. *Ann. Statist*. **24** 1084–1105. MR1401839 https://doi.org/10.1214/aos/1032526958

[34] RUPPERT, D., WAND, M. P. and CARROLL, R. J. (2009). *Semiparametric Regression*. Cambridge Univ. Press, New York. MR1998720 https://doi.org/10.1017/CBO9780511755453

[35] SAKHANENKO, A. I. (1985). Convergence rate in the invariance principle for non-identically distributed variables with exponential moments. In *Advances in Probability Theory*: *Limit Theorems for Sums of Random Variables* 2–73.

[36] SAKHANENKO, A. I. (1991). On the accuracy of normal approximation in the invariance principle [translation of Trudy Inst. Mat. (Novosibirsk) **13** (1989), Asimptot. Analiz Raspred. Sluch. Protsess., 40–66; MR1037248 (91d:60082)] **1** 58–91. Siberian Adv. Math., 4, Siberian Advances in Mathematics. MR1138005

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

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

[39] TUKEY, J. W. (1961). Curves as parameters, and touch estimation. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I* 681–694. Univ. California Press, Berkeley, CA. MR0132677

[40] ZAĬTSEV, A. YU. (2013). The accuracy of strong Gaussian approximation for sums of independent random vectors. *Russian Math. Surveys* **68** 721–761. MR3154816 https://doi.org/10.1070/rm2013v068n04abeh004851

[41] ZHAI, A. (2018). A high-dimensional CLT in W2 distance with near optimal convergence rate. *Probab. Theory Relat. Fields* **3-4** 821–845.

[42] ZHANG, H. and SINGER, B. H. (2010). *Recursive Partitioning and Applications*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2674991 https://doi.org/10.1007/978-1-4419-6824-1

[43] ZHOU, S., SHEN, X. and WOLFE, D. A. (1998). Local asymptotics for regression splines and confidence regions. *Ann. Statist.* **26** 1760–1782. MR1673277 https://doi.org/10.1214/aos/1024691356

[44] ZHOU, S. and WOLFE, D. A. (2000). On derivative estimation in spline regression. *Statist. Sinica* **10** 93–108. MR1742102

The Annals of Statistics
2020, Vol. 48, No. 3, 1742–1769
https://doi.org/10.1214/19-AOS1866

STATISTICAL INFERENCE IN TWO-SAMPLE SUMMARY-DATA MENDELIAN RANDOMIZATION USING ROBUST ADJUSTED PROFILE SCORE

BY QINGYUAN ZHAO[1], JINGSHU WANG[2], GIBRAN HEMANI[3], JACK BOWDEN[4] AND DYLAN S. SMALL[5]

[1]*Statistical Laboratory, University of Cambridge, qyzhao@statslab.cam.ac.uk*
[2]*Department of Statistics, University of Chicago, jingshuw@galton.uchicago.edu*
[3]*MRC Integrative Epidemiology Unit, University of Bristol, g.hemani@bristol.ac.uk*
[4]*University of Exeter Medical School, j.bowden2@exeter.ac.uk*
[5]*Department of Statistics, The Wharton School, University of Pennsylvania, dsmall@wharton.upenn.edu*

Mendelian randomization (MR) is a method of exploiting genetic variation to unbiasedly estimate a causal effect in presence of unmeasured confounding. MR is being widely used in epidemiology and other related areas of population science. In this paper, we study statistical inference in the increasingly popular two-sample summary-data MR design. We show a linear model for the observed associations approximately holds in a wide variety of settings when all the genetic variants satisfy the exclusion restriction assumption, or in genetic terms, when there is no pleiotropy. In this scenario, we derive a maximum profile likelihood estimator with provable consistency and asymptotic normality. However, through analyzing real datasets, we find strong evidence of both systematic and idiosyncratic pleiotropy in MR, echoing the omnigenic model of complex traits that is recently proposed in genetics. We model the systematic pleiotropy by a random effects model, where no genetic variant satisfies the exclusion restriction condition exactly. In this case, we propose a consistent and asymptotically normal estimator by adjusting the profile score. We then tackle the idiosyncratic pleiotropy by robustifying the adjusted profile score. We demonstrate the robustness and efficiency of the proposed methods using several simulated and real datasets.

REFERENCES

[1] 1000 GENOMES PROJECT CONSORTIUM, AUTON, A., BROOKS, L. D., DURBIN, R. M., GARRISON, E. P., KANG, H. M., KORBEL, J. O., MARCHINI, J. L., MCCARTHY, S. et al. (2015). A global reference for human genetic variation. *Nature* **526** 68–74. https://doi.org/10.1038/nature15393

[2] ANDERSON, T. W. and RUBIN, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Ann. Math. Stat.* **20** 46–63. MR0028546 https://doi.org/10.1214/aoms/1177730090

[3] BAIOCCHI, M., CHENG, J. and SMALL, D. S. (2014). Instrumental variable methods for causal inference. *Stat. Med.* **33** 2297–2340. MR3257582 https://doi.org/10.1002/sim.6128

[4] BARNDORFF-NIELSEN, O. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70** 343–365. MR0712023 https://doi.org/10.1093/biomet/70.2.343

[5] BEKKER, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica* **62** 657–681. MR1281697 https://doi.org/10.2307/2951662

[6] BOUND, J., JAEGER, D. A. and BAKER, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *J. Amer. Statist. Assoc.* **90** 443–450.

[7] BOWDEN, J., DAVEY SMITH, G. and BURGESS, S. (2015). Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44** 512–525.

MSC2010 subject classifications. Primary 65J05; secondary 46N60, 62F35.

Key words and phrases. Causal inference, limited information maximum likelihood, weak instruments, errors in variables, path analysis, pleiotropy effects.

[8] BOWDEN, J., DAVEY SMITH, G., HAYCOCK, P. C. and BURGESS, S. (2016). Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40** 304–314.

[9] BOWDEN, J., DEL GRECO M., F., MINELLI, C., DAVEY SMITH, G., SHEEHAN, N. and THOMPSON, J. (2017). A framework for the investigation of pleiotropy in two-sample summary data mendelian randomization. *Stat. Med.* **36** 1783–1802. MR3648622 https://doi.org/10.1002/sim.7221

[10] BOWDEN, J., FABIOLA DEL GRECO, M., MINELLI, C., LAWLOR, D., SHEEHAN, N., THOMPSON, J. and SMITH, G. D. (2019). Improving the accuracy of two-sample summary-data Mendelian randomization: Moving beyond the NOME assumption. *Int. J. Epidemiol.* **48** 728–742. https://doi.org/10.1093/ije/dyy258

[11] BOWDEN, J., SPILLER, W., DEL-GRECO, F., SHEEHAN, N., THOMPSON, J., MINELLI, C. and SMITH, G. D. (2018). Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression. *Int. J. Epidemiol.* **47** 1264–1278. https://doi.org/10.1093/ije/dyy101

[12] BOYLE, E. A., LI, Y. I. and PRITCHARD, J. K. (2017). An expanded view of complex traits: From polygenic to omnigenic. *Cell* **169** 1177–1186.

[13] BURGESS, S., BUTTERWORTH, A. and THOMPSON, S. G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37** 658–665.

[14] BURGESS, S., SCOTT, R. A., TIMPSON, N. J., SMITH, G. D., THOMPSON, S. G. and CONSORTIUM, E.-I. (2015). Using published data in Mendelian randomization: A blueprint for efficient identification of causal risk factors. *Eur. J. Epidemiol.* **30** 543–552.

[15] CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A. and CRAINICEANU, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd ed. *Monographs on Statistics and Applied Probability* **105**. CRC Press/CRC, Boca Raton, FL. MR2243417 https://doi.org/10.1201/9781420010138

[16] CLARIVATE ANALYTICS (2017). Web of Science Topic: Mendelian Randomization. Available at http://www.webofknowledge.com.

[17] COX, D. R. and REID, N. (1987). Parameter orthogonality and approximate conditional inference. *J. Roy. Statist. Soc. Ser. B* **49** 1–39. MR0893334

[18] DAVEY SMITH, G. and EBRAHIM, S. (2003). "Mendelian randomization": Can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* **32** 1–22.

[19] DAVEY SMITH, G. and HEMANI, G. (2014). Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **23** R89–R98.

[20] DAVEY SMITH, G., LAWLOR, D. A., HARBORD, R., TIMPSON, N., DAY, I. and EBRAHIM, S. (2007). Clustered environments and randomized genes: A fundamental distinction between conventional and genetic epidemiology. *PLoS Med.* **4** e352.

[21] DIDELEZ, V. and SHEEHAN, N. (2007). Mendelian randomization as an instrumental variable approach to causal inference. *Stat. Methods Med. Res.* **16** 309–330. MR2395652 https://doi.org/10.1177/0962280206077743

[22] FISHER, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford Univ. Press, Oxford.

[23] GUO, Z., KANG, H., CAI, T. T. and SMALL, D. S. (2018). Confidence intervals for causal effects with invalid instruments by using two-stage hard thresholding with voting. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 793–815. MR3849344 https://doi.org/10.1111/rssb.12275

[24] HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393. MR0362657

[25] HANSEN, C., HAUSMAN, J. and NEWEY, W. (2008). Estimation with many instrumental variables. *J. Bus. Econom. Statist.* **26** 398–422. MR2459342 https://doi.org/10.1198/073500108000000024

[26] HARTWIG, F. P., DAVEY SMITH, G. and BOWDEN, J. (2017). Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* **46** 1985–1998. https://doi.org/10.1093/ije/dyx102

[27] HAYCOCK, P. C., BURGESS, S., WADE, K. H., BOWDEN, J., RELTON, C. and SMITH, G. D. (2016). Best (but oft-forgotten) practices: The design, analysis, and interpretation of Mendelian randomization studies. *Am. J. Clin. Nutr.* **103** 965–978.

[28] HEMANI, G., ZHENG, J., ELSWORTH, B., WADE, K. H., HABERLAND, V., BAIRD, D., LAURIN, C., BURGESS, S., BOWDEN, J. et al. (2018). The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7** e34408. https://doi.org/10.7554/eLife.34408

[29] HERNÁN, M. A. and ROBINS, J. M. (2006). Instruments for causal inference: An epidemiologist's dream? *Epidemiology* **17** 360–372.

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

[31] IOANNIDIS, J. P., TRIKALINOS, T. A. and KHOURY, M. J. (2006). Implications of small effect sizes of individual genetic variants on the design and interpretation of genetic association studies of complex diseases. *Am. J. Epidemiol.* **164** 609–614.

[32] KANG, H., ZHANG, A., CAI, T. T. and SMALL, D. S. (2016). Instrumental variables estimation with some invalid instruments and its application to mendelian randomization. *J. Amer. Statist. Assoc.* **111** 132–144. MR3494648 https://doi.org/10.1080/01621459.2014.994705

[33] KATAN, M. (1986). Apoupoprotein E isoforms, serum cholesterol, and cancer. *Lancet* **327** 507–508.

[34] LI, S. (2017). Mendelian randomization when many instruments are invalid: Hierarchical empirical Bayes estimation. Available at arXiv:1706.01389.

[35] LOCKE, A. E., KAHALI, B., BERNDT, S. I., JUSTICE, A. E., PERS, T. H., DAY, F. R., POWELL, C., VEDANTAM, S., BUCHKOVICH, M. L. et al. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518** 197–206.

[36] MARONNA, R. A., MARTIN, R. D. and YOHAI, V. J. (2006). *Robust Statistics*: *Theory and Methods*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR2238141 https://doi.org/10.1002/0470010940

[37] MCCULLAGH, P. and TIBSHIRANI, R. (1990). A simple method for the adjustment of profile likelihoods. *J. Roy. Statist. Soc. Ser. B* **52** 325–344. MR1064420

[38] MURPHY, S. A. and VAN DER VAART, A. W. (1996). Likelihood inference in the errors-in-variables model. *J. Multivariate Anal.* **59** 81–108. MR1424904 https://doi.org/10.1006/jmva.1996.0055

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

[40] PACINI, D. and WINDMEIJER, F. (2016). Robust inference for the two-sample 2SLS estimator. *Econom. Lett.* **146** 50–54. MR3542584 https://doi.org/10.1016/j.econlet.2016.06.033

[41] PARK, J.-H., WACHOLDER, S., GAIL, M. H., PETERS, U., JACOBS, K. B., CHANOCK, S. J. and CHATTERJEE, N. (2010). Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat. Genet.* **42** 570–575.

[42] PEARL, J. (2009). *Causality*: *Models*, *Reasoning*, *and Inference*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2548166 https://doi.org/10.1017/CBO9780511803161

[43] PURCELL, S. PLINK (software V1.07). http://pngu.mgh.harvard.edu/purcell/plink/.

[44] PURCELL, S., NEALE, B., TODD-BROWN, K., THOMAS, L., FERREIRA, M. A., BENDER, D., MALLER, J., SKLAR, P., DE BAKKER, P. I. et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81** 559–575.

[45] ROTHENBERG, T. J. (1971). Identification in parametric models. *Econometrica* **39** 577–591. MR0436944 https://doi.org/10.2307/1913267

[46] SHI, H., KICHAEV, G. and PASANIUC, B. (2016). Contrasting the genetic architecture of 30 complex traits from summary association data. *Am. J. Hum. Genet.* **99** 139–153.

[47] SOLOVIEFF, N., COTSAPAS, C., LEE, P. H., PURCELL, S. M. and SMOLLER, J. W. (2013). Pleiotropy in complex traits: Challenges and strategies. *Nat. Rev. Genet.* **14** 483–495.

[48] STEARNS, F. W. (2010). One hundred years of pleiotropy: A retrospective. *Genetics* **186** 767–773.

[49] STOCK, J. H. and YOGO, M. (2005). Asymptotic distributions of instrumental variables statistics with many instruments. In *Identification and Inference for Econometric Models* 109–120. Cambridge Univ. Press, Cambridge. MR2232141 https://doi.org/10.1017/CBO9780511614491.007

[50] TCHETGEN TCHETGEN, E. J., SUN, B. and WALTER, S. (2017). The GENIUS approach to robust Mendelian randomization inference. Available at arXiv:1709.07779.

[51] VAN KIPPERSLUIS, H. and RIETVELD, C. A. (2017). Pleiotropy-robust Mendelian randomization. *Int. J. Epidemiol.* In press.

[52] VERBANCK, M., CHEN, C.-Y., NEALE, B. and DO, R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50** 693.

[53] WALD, A. (1940). The fitting of straight lines if both variables are subject to error. *Ann. Math. Stat.* **11** 285–300. MR0002739 https://doi.org/10.1214/aoms/1177731868

[54] WRIGHT, P. G. (1928). *Tariff on Animal and Vegetable Oils*. MacMillan, New York.

[55] WRIGHT, S. (1968). *Evolution and the Genetics of Populations*, *Volume* 1: *Genetic and Biometric Foundations* **1**. Univ. Chicago Press, Chicago.

[56] YOHAI, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *Ann. Statist.* **15** 642–656. MR0888431 https://doi.org/10.1214/aos/1176350366

[57] ZHAO, Q., WANG, J., HEMANI, G., BOWDEN, J. and SMALL, D. S. (2020). Supplement to "Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score." https://doi.org/10.1214/19-AOS1866SUPP.

The Annals of Statistics
2020, Vol. 48, No. 3, 1770–1788
https://doi.org/10.1214/19-AOS1867

LOCAL UNCERTAINTY SAMPLING FOR LARGE-SCALE MULTICLASS LOGISTIC REGRESSION

BY LEI HAN[1], KEAN MING TAN[2], TING YANG[3] AND TONG ZHANG[4]

[1]*Tencent AI Lab, Tencent Technology (Shenzhen) Company Limited, leihan.cs@gmail.com*
[2]*Department of Statistics, University of Michigan, keanming@umich.edu*
[3]*Yelp Inc., ting.yang.pku@gmail.com*
[4]*Department of Mathematics and Department of Computer Science, Hong Kong University of Science and Technology, tongzhang@tongzhang-ml.org*

A major challenge for building statistical models in the big data era is that the available data volume far exceeds the computational capability. A common approach for solving this problem is to employ a subsampled dataset that can be handled by available computational resources. We propose a general subsampling scheme for large-scale multiclass logistic regression and examine the variance of the resulting estimator. We show that asymptotically, the proposed method always achieves a smaller variance than that of the uniform random sampling. Moreover, when the classes are conditionally imbalanced, significant improvement over uniform sampling can be achieved. Empirical performance of the proposed method is evaluated and compared to other methods via both simulated and real-world datasets, and these results match and confirm our theoretical analysis.

REFERENCES

[1] ABE, N., ZADROZNY, B. and LANGFORD, J. (2004). An iterative method for multi-class cost-sensitive learning. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 3–11.

[2] ANDERSON, J. A. (1972). Separate sample logistic discrimination. *Biometrika* **59** 19–35. MR0345332 https://doi.org/10.1093/biomet/59.1.19

[3] ATKESON, C. G., MOORE, A. W. and SCHAAL, S. (1997). Locally weighted learning for control. In *Lazy Learning* 75–113. Springer, Berlin.

[4] BRESLOW, N. (1982). Design and analysis of case-control studies. *Annu. Rev. Public Health* **3** 29–54. https://doi.org/10.1146/annurev.pu.03.050182.000333

[5] CHAWLA, N. V., JAPKOWICZ, N. and KOTCZ, A. (2004). Editorial: Special issue on learning from imbalanced data sets. *ACM SIGKDD Explor. Newsl.* **6** 1–6.

[6] CORTES, C., MANSOUR, Y. and MOHRI, M. (2010). Learning bounds for importance weighting. In *Advances in Neural Information Processing Systems* 442–450.

[7] CORTES, C., MOHRI, M., RILEY, M. and ROSTAMIZADEH, A. (2008). Sample selection bias correction theory. In *Algorithmic Learning Theory. Lecture Notes in Computer Science* **5254** 38–53. Springer, Berlin. MR2540648 https://doi.org/10.1007/978-3-540-87987-9_8

[8] DHILLON, P., LU, Y., FOSTER, D. P. and UNGAR, L. (2013). New subsampling algorithms for fast least squares regression. In *Advances in Neural Information Processing Systems* 360–368.

[9] FITHIAN, W. and HASTIE, T. (2014). Local case-control sampling: Efficient subsampling in imbalanced data sets. *Ann. Statist.* **42** 1693–1724. MR3257627 https://doi.org/10.1214/14-AOS1220

[10] FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2000). Additive logistic regression: A statistical view of boosting. *Ann. Statist.* **28** 337–407. MR1790002 https://doi.org/10.1214/aos/1016218223

[11] HAN, L., TAN, K. M., YANG, T. and ZHANG, T. (2020). Supplement to "Local uncertainty sampling for large-scale multiclass logistic regression." https://doi.org/10.1214/19-AOS1867SUPP.

[12] HE, H. and GARCIA, E. A. (2009). Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.* **21** 1263–1284.

MSC2010 subject classifications. Primary 62D05; secondary 62F10, 62F12.
Key words and phrases. Sampling, large-scale, multiclass logistic regression.

[13] 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

[14] KIM, H.-C., PANG, S., JE, H.-M., KIM, D. and BANG, S. Y. (2002). Pattern classification using support vector machine ensemble. In *Proceedings of the International Conference on Pattern Recognition* **2** 160–163.

[15] KING, G. and ZENG, L. (2001). Logistic regression in rare events data. *Polit. Anal.* **9** 137–163.

[16] LECUN, Y., BOTTOU, L., BENGIO, Y. and HAFFNER, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* **86** 2278–2324.

[17] MANTEL, N. and HAENSZEL, W. (1959). Statistical aspects of the analysis of data from retrospective studies. *J. Natl. Cancer Inst.* **22** 719–748.

[18] MINEIRO, P. and KARAMPATZIAKIS, N. (2013). Loss-proportional subsampling for subsequent ERM. Preprint. Available at arXiv:1306.1840.

[19] SCOTT, A. and WILD, C. (2002). On the robustness of weighted methods for fitting models to case-control data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 207–219. MR1904701 https://doi.org/10.1111/1467-9868.00333

[20] SCOTT, A. J. and WILD, C. J. (1986). Fitting logistic models under case-control or choice based sampling. *J. Roy. Statist. Soc. Ser. B* **48** 170–182. MR0867995

[21] SCOTT, A. J. and WILD, C. J. (1991). Fitting logistic regression models in stratified case-control studies. *Biometrics* **47** 497–510. MR1132540 https://doi.org/10.2307/2532141

[22] TAN, A. C., GILBERT, D. and DEVILLE, Y. (2003). Multi-class protein fold classification using a new ensemble machine learning approach. *Genome Inform.* **14** 206–217.

[23] WEBB, S., CAVERLEE, J. and PU, C. (2006). Introducing the Webb Spam Corpus: Using email spam to identify Web spam automatically. In *Proceedings of the Third Conference on Email and Anti-Spam*.

[24] WIDODO, A. and YANG, B.-S. (2007). Support vector machine in machine condition monitoring and fault diagnosis. *Mech. Syst. Signal Process.* **21** 2560–2574.

[25] XIE, Y. and MANSKI, C. F. (1989). The logit model and response-based samples. *Sociol. Methods Res.* **17** 283–302.

[26] ZADROZNY, B. (2004). Learning and evaluating classifiers under sample selection bias. In *Proceedings of the International Conference on Machine Learning* 114.

[27] ZHANG, T. and OLES, F. (2000). The value of unlabeled data for classification problems. In *Proceedings of the International Conference on Machine Learning* 1191–1198.

The Annals of Statistics
2020, Vol. 48, No. 3, 1789–1814
https://doi.org/10.1214/19-AOS1868

LOCAL NEAREST NEIGHBOUR CLASSIFICATION WITH APPLICATIONS TO SEMI-SUPERVISED LEARNING

BY TIMOTHY I. CANNINGS[1], THOMAS B. BERRETT[2,*] AND RICHARD J. SAMWORTH[2,**]

[1]*School of Mathematics, University of Edinburgh, timothy.cannings@ed.ac.uk*
[2]*Statistical Laboratory, Centre for Mathematical Sciences, University of Cambridge, [*]t.berrett@statslab.cam.ac.uk; [**]r.samworth@statslab.cam.ac.uk*

We derive a new asymptotic expansion for the global excess risk of a local-k-nearest neighbour classifier, where the choice of k may depend upon the test point. This expansion elucidates conditions under which the dominant contribution to the excess risk comes from the decision boundary of the optimal Bayes classifier, but we also show that if these conditions are not satisfied, then the dominant contribution may arise from the tails of the marginal distribution of the features. Moreover, we prove that, provided the d-dimensional marginal distribution of the features has a finite ρth moment for some $\rho > 4$ (as well as other regularity conditions), a local choice of k can yield a rate of convergence of the excess risk of $O(n^{-4/(d+4)})$, where n is the sample size, whereas for the standard k-nearest neighbour classifier, our theory would require $d \geq 5$ and $\rho > 4d/(d-4)$ finite moments to achieve this rate. These results motivate a new k-nearest neighbour classifier for semi-supervised learning problems, where the unlabelled data are used to obtain an estimate of the marginal feature density, and fewer neighbours are used for classification when this density estimate is small. Our worst-case rates are complemented by a minimax lower bound, which reveals that the local, semi-supervised k-nearest neighbour classifier attains the minimax optimal rate over our classes for the excess risk, up to a subpolynomial factor in n. These theoretical improvements over the standard k-nearest neighbour classifier are also illustrated through a simulation study.

REFERENCES

ABRAMSON, I. S. (1982). On bandwidth variation in kernel estimates—a square root law. *Ann. Statist.* **10** 1217–1223. MR0673656

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

BERRETT, T. B. and SAMWORTH, R. J. (2019a). Efficient two-sample functional estimation and the super-oracle phenomenon. https://arxiv.org/abs/1904.09347.

BERRETT, T. B. and SAMWORTH, R. J. (2019b). Nonparametric independence testing via mutual information. *Biometrika* **106** 547–566. MR3992389 https://doi.org/10.1093/biomet/asz024

BERRETT, T. B., SAMWORTH, R. J. and YUAN, M. (2019). Efficient multivariate entropy estimation via k-nearest neighbour distances. *Ann. Statist.* **47** 288–318. MR3909934 https://doi.org/10.1214/18-AOS1688

BIAU, G., CÉROU, F. and GUYADER, A. (2010). On the rate of convergence of the bagged nearest neighbor estimate. *J. Mach. Learn. Res.* **11** 687–712. MR2600626

BIAU, G. and DEVROYE, L. (2015). *Lectures on the Nearest Neighbor Method. Springer Series in the Data Sciences*. Springer, Cham. MR3445317 https://doi.org/10.1007/978-3-319-25388-6

BOUCHERON, S., BOUSQUET, O. and LUGOSI, G. (2005). Theory of classification: A survey of some recent advances. *ESAIM Probab. Stat.* **9** 323–375. MR2182250 https://doi.org/10.1051/ps:2005018

BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities*. Oxford Univ. Press, Oxford. MR3185193 https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

MSC2010 subject classifications. 62G20.

Key words and phrases. Classification problems, nearest neighbours, nonparametric classification, semi-supervised learning.

BREIMAN, L., MEISEL, W. and PURCELL, E. (1977). Variable kernel estimates of multivariate densities. *Technometrics* **19** 135–144.

CANNINGS, T. I, BERRETT, T. B and SAMWORTH, R. J (2020). Supplement to "Local nearest neighbour classification with applications to semi-supervised learning." https://doi.org/10.1214/19-AOS1868SUPP.

CELISSE, A. and MARY-HUARD, T. (2018). Theoretical analysis of cross-validation for estimating the risk of the k-nearest neighbor classifier. *J. Mach. Learn. Res.* **19** 58. MR3899760

CHAPELLE, O., ZIEN, A. and SCHÖLKOPF, B., eds. (2006). *Semi-Supervised Learning*. MIT Press, Cambridge MA.

CHAUDHURI, K. and DASGUPTA, S. (2014). Rates of convergence for nearest neighbor classification. *Adv. Neural Inf. Process. Syst.* **27** 3437–3445.

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

DEVROYE, L., GYÖRFI, L. and LUGOSI, G. (1996). *A Probabilistic Theory of Pattern Recognition. Applications of Mathematics* (*New York*) **31**. Springer, New York. MR1383093 https://doi.org/10.1007/978-1-4612-0711-5

FIX, E. and HODGES, J. L. (1951). Discriminatory analysis—nonparametric discrimination: Consistency properties Technical Report number 4, USAF School of Aviation Medicine, Randolph Field, TX.

FIX, E. and HODGES, J. L. (1989). Discriminatory analysis—nonparametric discrimination: Consistency properties. *Int. Stat. Rev.* **57** 238–247.

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

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

GINÉ, E. and SANG, H. (2010). Uniform asymptotics for kernel density estimators with variable bandwidths. *J. Nonparametr. Stat.* **22** 773–795. MR2682221 https://doi.org/10.1080/10485250903483331

HALL, P. and KANG, K.-H. (2005). Bandwidth choice for nonparametric classification. *Ann. Statist.* **33** 284–306. MR2157804 https://doi.org/10.1214/009053604000000959

HALL, P., PARK, B. U. and SAMWORTH, R. J. (2008). Choice of neighbor order in nearest-neighbor classification. *Ann. Statist.* **36** 2135–2152. MR2458182 https://doi.org/10.1214/07-AOS537

HECKEL, R. and BÖLCSKEI, H. (2015). Robust subspace clustering via thresholding. *IEEE Trans. Inform. Theory* **61** 6320–6342. MR3418967 https://doi.org/10.1109/TIT.2015.2472520

KOZACHENKO, L. F. and LEONENKO, N. N. (1987). A statistical estimate for the entropy of a random vector. *Problemy Peredachi Informatsii* **23** 9–16. MR0908626

KULKARNI, S. R. and POSNER, S. E. (1995). Rates of convergence of nearest neighbor estimation under arbitrary sampling. *IEEE Trans. Inform. Theory* **41** 1028–1039. MR1366756 https://doi.org/10.1109/18.391248

LOFTSGAARDEN, D. O. and QUESENBERRY, C. P. (1965). A nonparametric estimate of a multivariate density function. *Ann. Math. Stat.* **36** 1049–1051. MR0176567 https://doi.org/10.1214/aoms/1177700079

MACK, Y. P. (1983). Rate of strong uniform convergence of k-NN density estimates. *J. Statist. Plann. Inference* **8** 185–192. MR0720150 https://doi.org/10.1016/0378-3758(83)90037-X

MACK, Y. P. and ROSENBLATT, M. (1979). Multivariate k-nearest neighbor density estimates. *J. Multivariate Anal.* **9** 1–15. MR0530638 https://doi.org/10.1016/0047-259X(79)90065-4

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

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

SCHILLING, M. F. (1986). Multivariate two-sample tests based on nearest neighbors. *J. Amer. Statist. Assoc.* **81** 799–806. MR0860514

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics. Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York. MR0838963

STONE, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–645. MR0443204

The Annals of Statistics
2020, Vol. 48, No. 3, 1815–1847
https://doi.org/10.1214/19-AOS1869

AN ADAPTABLE GENERALIZATION OF HOTELLING'S T^2 TEST IN HIGH DIMENSION

BY HAORAN LI[1,*], ALEXANDER AUE[1,**], DEBASHIS PAUL[1,†], JIE PENG[1,‡] AND PEI WANG[2]

[1]*Department of Statistics, University of California, Davis,* [*]*hrli@ucdavis.edu;* [**]*aaue@ucdavis.edu;* [†]*debpaul@ucdavis.edu;* [‡]*jiepeng@ucdavis.edu*

[2]*Icahn Institute, Department of Genetics and Genomic Sciences, School of Medicine at Mount Sinai, pei.wang@mssm.edu*

We propose a two-sample test for detecting the difference between mean vectors in a high-dimensional regime based on a ridge-regularized Hotelling's T^2. To choose the regularization parameter, a method is derived that aims at maximizing power within a class of local alternatives. We also propose a composite test that combines the optimal tests corresponding to a specific collection of local alternatives. Weak convergence of the stochastic process corresponding to the ridge-regularized Hotelling's T^2 is established and used to derive the cut-off values of the proposed test. Large sample properties are verified for a class of sub-Gaussian distributions. Through an extensive simulation study, the composite test is shown to compare favorably against a host of existing two-sample test procedures in a wide range of settings. The performance of the proposed test procedures is illustrated through an application to a breast cancer data set where the goal is to detect the pathways with different DNA copy number alterations across breast cancer subtypes.

REFERENCES

ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. *Wiley Publications in Statistics*. Wiley, New York. MR0091588

BAI, Z., CHEN, J. and YAO, J. (2010). On estimation of the population spectral distribution from a high-dimensional sample covariance matrix. *Aust. N. Z. J. Stat.* **52** 423–437. MR2791528 https://doi.org/10.1111/j.1467-842X.2010.00590.x

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

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

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

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

BERGAMASCHI, A., KIM, Y. H., WANG, P., SØRLIE, T., HERNANDEZ-BOUSSARD, T., LONNING, P. E., TIBSHIRANI, R., BØRRESEN-DALE, A. L. and POLLACK, J. R. (2006). Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. *Genes Chromosomes Cancer* **45** 1033–40.

BISWAS, M. and GHOSH, A. K. (2014). A nonparametric two-sample test applicable to high dimensional data. *J. Multivariate Anal.* **123** 160–171. MR3130427 https://doi.org/10.1016/j.jmva.2013.09.004

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

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

MSC2010 subject classifications. Primary 62J99; secondary 60B20.

Key words and phrases. Asymptotic property, covariance matrix, Hotelling's T^2 statistic, hypothesis testing, locally most powerful tests, random matrix theory.

CHANG, J., ZHOU, W. and ZHOU, W. X. (2014). Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity—An alternative road to high dimensional tests. Preprint. Available at arXiv:1406.1939.

CHATTERJEE, S. (2009). Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields* **143** 1–40. MR2449121 https://doi.org/10.1007/s00440-007-0118-6

CHEN, S. X., LI, J. and ZHONG, P. (2014). Two-sample tests for high dimensional means with thresholding and data transformation. Preprint. Available at arXiv:1410.2848.

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

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

CREIGHTON, C. J. (2012). The molecular profile of luminal B breast cancer. *Biologics* **6** 289–297. https://doi.org/10.2147/BTT.S29923

DONG, K., PANG, H., TONG, T. and GENTON, M. G. (2016). Shrinkage-based diagonal Hotelling's tests for high-dimensional small sample size data. *J. Multivariate Anal.* **143** 127–142. MR3431423 https://doi.org/10.1016/j.jmva.2015.08.022

EL KAROUI, N. and KÖSTERS (2011). Geometric sensitivity of random matrix results: Consequences for shrinkage estimators of covariance and related statistical methods. Preprint. Available at arXiv:1105.1404.

ELLIS, M. J., GILLETTE, M., CARR, S. A., PAULOVICH, A. G., SMITH, R. D., RODLAND, K. K., TOWNSEND, R. R., KINSINGER, C., MESRI, M. et al. (2013). Connecting genomic alterations to cancer biology with proteomics: The NCI Clinical Proteomic Tumor Analysis Consortium. *Cancer Discov.* **3** 1108–1112.

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

GRETTON, A., SRIPERUMBUDUR, B., SEJDINOVIC, D., STRATHMANN, H., BALAKRISHNAN, S., PONTIL, M. and FUKUMIZU, K. (2012). Optimal kernel choice for large-scale two-sample tests. *Adv. Neural Inf. Process. Syst.* **25** 1205–1213.

GUO, B. and CHEN, S. X. (2016). Tests for high dimensional generalized linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 1079–1102. MR3557190 https://doi.org/10.1111/rssb.12152

LAMY, P.-J., FINA, F., BASCOUL-MOLLEVI, C., LABERENNE, A.-C., MARTIN, P.-M., OUAFIK, L. and JACOT, W. (2011). Quantification and clinical relevance of gene amplification at chromosome 17q12-q21 in human epidermal growth factor receptor 2-amplified breast cancers. *Breast Cancer Res.* **13** R15. https://doi.org/10.1186/bcr2824

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

LI, H., AUE, A., PAUL, D., PENG, J. and WANG, P. (2020). Supplement to "An adaptable generalization of Hotelling's T^2 test in high dimension." https://doi.org/10.1214/19-AOS1869SUPP.

LIU, H., AUE, A. and PAUL, D. (2015). On the Marčenko–Pastur law for linear time series. *Ann. Statist.* **43** 675–712. MR3319140 https://doi.org/10.1214/14-AOS1294

LOPES, M. E., JACOB, L. and WAINWRIGHT, M. J. (2011). A more powerful two-sample test in high dimensions using random projection. *Adv. Neural Inf. Process. Syst.* 1206–1214.

MERTINS, P., MANI, D. R., RUGGLES, K. V., GILLETTE, M. A., CLAUSER, K. R., WANG, P., WANG, X., QIAO, J. W., CAO, S. et al. (2016). Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534** 55–62. https://doi.org/10.1038/nature18003

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory. Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0652932

CANCER GENOME ATLAS NETWORK (2012). Comprehensive molecular portraits of human breast tumours. *Nature* **490** 61–70. https://doi.org/10.1038/nature11412

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

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

PAULOVICH, A. G., BILLHEIMER, D., HAM, A. J., VEGA-MONTOTO, L., RUDNICK, P. A., TABB, D. L., WANG, P. et al. (2010). Interlaboratory study characterizing a yeast performance standard for benchmarking LC-MS platform performance. *Mol. Cell. Proteomics* **9** 242–254.

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

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

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

TRAN, B. and BEDARD, P. L. (2011). Luminal-B breast cancer and novel therapeutic targets. *Breast Cancer Res.* **13** 221. https://doi.org/10.1186/bcr2904

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

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

The Annals of Statistics
2020, Vol. 48, No. 3, 1848–1874
https://doi.org/10.1214/19-AOS1846

GRID: A VARIABLE SELECTION AND STRUCTURE DISCOVERY METHOD FOR HIGH DIMENSIONAL NONPARAMETRIC REGRESSION

BY FRANCESCO GIORDANO[1,*], SOUMENDRA NATH LAHIRI[2] AND MARIA LUCIA PARRELLA[1,†]

[1]*Department of Economics and Statistics, University of Salerno,* [*]*giordano@unisa.it;* [†]*mparrella@unisa.it*
[2]*Department of Statistics, North Carolina State University, snlahiri@ncsu.edu*

We consider nonparametric regression in high dimensions where only a relatively small subset of a large number of variables are relevant and may have nonlinear effects on the response. We develop methods for variable selection, structure discovery and estimation of the true low-dimensional regression function, allowing any degree of interactions among the relevant variables that need not be specified *a-priori*. The proposed method, called the GRID, combines empirical likelihood based marginal testing with the local linear estimation machinery in a novel way to select the relevant variables. Further, it provides a simple graphical tool for identifying the low dimensional nonlinear structure of the regression function. Theoretical results establish consistency of variable selection and structure discovery, and also Oracle risk property of the GRID estimator of the regression function, allowing the dimension d of the covariates to grow with the sample size n at the rate $d = O(n^a)$ for any $a \in (0, \infty)$ and the number of relevant covariates r to grow at a rate $r = O(n^\gamma)$ for some $\gamma \in (0, 1)$ under some regularity conditions that, in particular, require finiteness of certain absolute moments of the error variables depending on a. Finite sample properties of the GRID are investigated in a moderately large simulation study.

REFERENCES

BERTIN, K. and LECUÉ, G. (2008). Selection of variables and dimension reduction in high-dimensional nonparametric regression. *Electron. J. Stat.* **2** 1224–1241. MR2461900 https://doi.org/10.1214/08-EJS327

CHANG, J., TANG, C. Y. and WU, Y. (2013). Marginal empirical likelihood and sure independence feature screening. *Ann. Statist.* **41** 2123–2148. MR3127860 https://doi.org/10.1214/13-AOS1139

CHEN, S. X. and QIN, Y. S. (2000). Empirical likelihood confidence intervals for local linear smoothers. *Biometrika* **87** 946–953. MR1813987 https://doi.org/10.1093/biomet/87.4.946

CHEN, S. X. and VAN KEILEGOM, I. (2009). A review on empirical likelihood methods for regression. *TEST* **18** 415–447. MR2566404 https://doi.org/10.1007/s11749-009-0159-5

CHOI, N. H., LI, W. and ZHU, J. (2010). Variable selection with the strong heredity constraint and its oracle property. *J. Amer. Statist. Assoc.* **105** 354–364. With supplementary material available online. MR2656056 https://doi.org/10.1198/jasa.2010.tm08281

COMMINGES, L. and DALALYAN, A. S. (2012). Tight conditions for consistency of variable selection in the context of high dimensionality. *Ann. Statist.* **40** 2667–2696. MR3097616 https://doi.org/10.1214/12-AOS1046

FAN, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.* **87** 998–1004. MR1209561

FAN, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.* **21** 196–216. MR1212173 https://doi.org/10.1214/aos/1176349022

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

FAN, J. and JIANG, J. (2005). Nonparametric inferences for additive models. *J. Amer. Statist. Assoc.* **100** 890–907. MR2201017 https://doi.org/10.1198/016214504000001439

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

MSC2010 subject classifications. Primary 62G08, 62G20; secondary 62G10, 62H15.
Key words and phrases. Empirical likelihood, marginal testing, variable selection consistency.

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

FAN, J., GASSER, T., GIJBELS, I., BROCKMANN, M. and ENGEL, J. (1997). Local polynomial regression: Optimal kernels and asymptotic minimax efficiency. *Ann. Inst. Statist. Math.* **49** 79–99. MR1450693 https://doi.org/10.1023/A:1003162622169

FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2007). Sparse inverse covarinace estimation with the graphical lasso. *Biostatistics* **9** 432–441.

GIORDANO, F., LAHIRI, S. N. and PARRELLA, M. L. (2020). Supplement to "GRID: A variable selection and structure discovery method for high dimensional nonparametric regression." https://doi.org/10.1214/19-AOS1846SUPP

GIORDANO, F. and PARRELLA, M. L. (2016). Bias-corrected inference for multivariate nonparametric regression: Model selection and oracle property. *J. Multivariate Anal.* **143** 71–93. MR3431420 https://doi.org/10.1016/j.jmva.2015.08.016

HALL, P. and MILLER, H. (2009). Using generalized correlation to effect variable selection in very high dimensional problems. *J. Comput. Graph. Statist.* **18** 533–550. MR2751640 https://doi.org/10.1198/jcgs.2009.08041

JAMES, G. M., RADCHENKO, P. and LV, J. (2009). DASSO: Connections between the Dantzig selector and lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 127–142. MR2655526 https://doi.org/10.1111/j.1467-9868.2008.00668.x

LAFFERTY, J. and WASSERMAN, L. (2008). Rodeo: Sparse, greedy nonparametric regression. *Ann. Statist.* **36** 28–63. MR2387963 https://doi.org/10.1214/009053607000000811

LIN, Y. and ZHANG, H. H. (2006). Component selection and smoothing in multivariate nonparametric regression. *Ann. Statist.* **34** 2272–2297. MR2291500 https://doi.org/10.1214/009053606000000722

MEIER, L., VAN DE GEER, S. and BÜHLMANN, P. (2009). High-dimensional additive modeling. *Ann. Statist.* **37** 3779–3821. MR2572443 https://doi.org/10.1214/09-AOS692

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

OPSOMER, J. D. (2000). Asymptotic properties of backfitting estimators. *J. Multivariate Anal.* **73** 166–179. MR1763322 https://doi.org/10.1006/jmva.1999.1868

OWEN, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75** 237–249. MR0946049 https://doi.org/10.1093/biomet/75.2.237

OWEN, A. B. (2001). *Empirical Likelihood*. CRC Press/CRC, Boca Raton, FL.

PRAKASA RAO, B. L. S. (1983). *Nonparametric Functional Estimation. Probability and Mathematical Statistics*. Academic Press, New York. MR0740865

QIN, J. and LAWLESS, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22** 300–325. MR1272085 https://doi.org/10.1214/aos/1176325370

RADCHENKO, P. and JAMES, G. M. (2010). Variable selection using adaptive nonlinear interaction structures in high dimensions. *J. Amer. Statist. Assoc.* **105** 1541–1553. MR2796570 https://doi.org/10.1198/jasa.2010.tm10130

RAVIKUMAR, P., LAFFERTY, J., LIU, H. and WASSERMAN, L. (2009). Sparse additive models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 1009–1030. MR2750255 https://doi.org/10.1111/j.1467-9868.2009.00718.x

RUPPERT, D. and WAND, M. P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.* **22** 1346–1370. MR1311979 https://doi.org/10.1214/aos/1176325632

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

STONE, C. J., HANSEN, M. H., KOOPERBERG, C. and TRUONG, Y. K. (1997). Polynomial splines and their tensor products in extended linear modeling. *Ann. Statist.* **25** 1371–1470. With discussion and a rejoinder by the authors and Jianhua Z. Huang. MR1463561 https://doi.org/10.1214/aos/1031594728

STORLIE, C. B., BONDELL, H. D., REICH, B. J. and ZHANG, H. H. (2011). Surface estimation, variable selection, and the nonparametric oracle property. *Statist. Sinica* **21** 679–705. MR2829851 https://doi.org/10.5705/ss.2011.030a

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

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

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

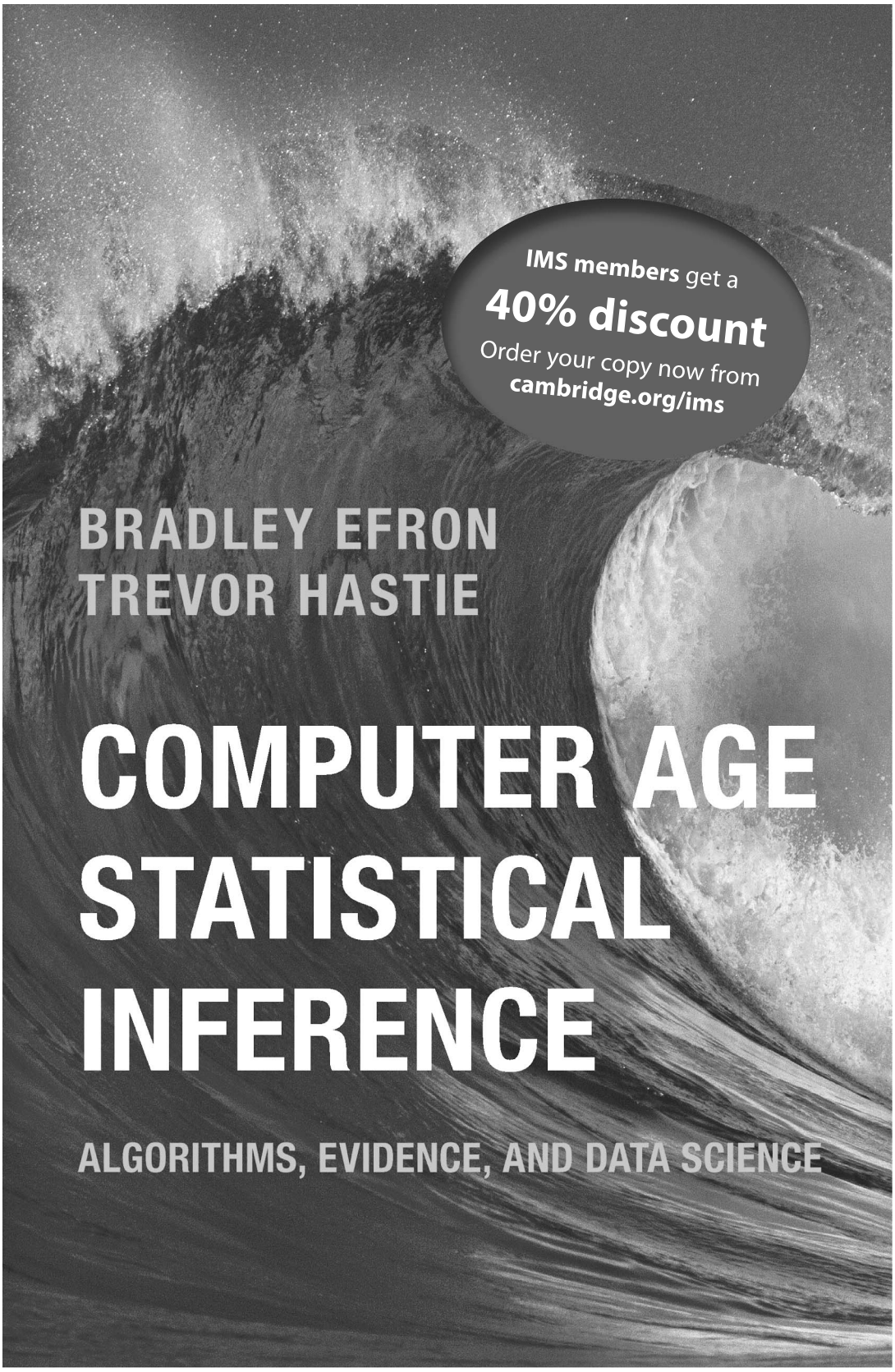
YANG, Y. and TOKDAR, S. T. (2015). Minimax-optimal nonparametric regression in high dimensions. *Ann. Statist.* **43** 652–674. MR3319139 https://doi.org/10.1214/14-AOS1289

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

ZHANG, H. H., CHENG, G. and LIU, Y. (2011). Linear or nonlinear? Automatic structure discovery for partially linear models. *J. Amer. Statist. Assoc.* **106** 1099–1112. MR2894767 https://doi.org/10.1198/jasa.2011.tm10281

ZHANG, J. and LIU, A. (2003). Local polynomial fitting based on empirical likelihood. *Bernoulli* **9** 579–605. MR1996271 https://doi.org/10.3150/bj/1066223270

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


IMS members get a
40% discount
Order your copy now from
cambridge.org/ims
BRADLEY EFRON
TREVOR HASTIE
COMPUTER AGE STATISTICAL INFERENCE
ALGORITHMS, EVIDENCE, AND DATA SCIENCE