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

THE ANNALS *of* STATISTICS

An Official Journal of the Institute of Mathematical Statistics

Articles

THE ANNALS OF STATISTICS Vol. 51, No. 2, pp. 391–877 April 2023

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

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

IMS OFFICERS

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

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

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

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

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

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

IMS EDITORS

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

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

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

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

Statistical Science. *Editor*: Moulinath Banerjee, Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA

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

The Annals of Statistics **[ISSN 0090-5364 (print); ISSN 2168-8966 (online)],** Volume 51, Number 2, April 2023. Published bimonthly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, OH 44094, USA. Periodicals postage paid at Cleveland, Ohio, and at additional mailing offices.

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

Copyright © 2023 by the Institute of Mathematical Statistics
Printed in the United States of America

The Annals of Statistics
2023, Vol. 51, No. 2, 391–436
https://doi.org/10.1214/22-AOS2243
© Institute of Mathematical Statistics, 2023

DEBIASING CONVEX REGULARIZED ESTIMATORS AND INTERVAL ESTIMATION IN LINEAR MODELS

BY PIERRE C. BELLEC[a] AND CUN-HUI ZHANG[b]

Department of Statistics, Rutgers University, [a]*pierre.bellec@rutgers.edu,* [b]*czhang@stat.rutgers.edu*

New upper bounds are developed for the L_2 distance between $\xi/\mathrm{Var}[\xi]^{1/2}$ and linear and quadratic functions of $z \sim N(\mathbf{0}, \boldsymbol{I}_n)$ for random variables of the form $\xi = z^\top f(z) - \mathrm{div}\, f(z)$. The linear approximation yields a central limit theorem when the squared norm of $f(z)$ dominates the squared Frobenius norm of $\nabla f(z)$ in expectation.

Applications of this normal approximation are given for the asymptotic normality of debiased estimators in linear regression with correlated design and convex penalty in the regime $p/n \le \gamma$ for constant $\gamma \in (0, \infty)$. For the estimation of linear functions $\langle \boldsymbol{a}_0, \boldsymbol{\beta} \rangle$ of the unknown coefficient vector $\boldsymbol{\beta}$, this analysis leads to asymptotic normality of the debiased estimate for most normalized directions $\boldsymbol{a}_0$, where "most" is quantified in a precise sense. This asymptotic normality holds for any convex penalty if $\gamma < 1$ and for any strongly convex penalty if $\gamma \ge 1$. In particular, the penalty needs not be separable or permutation invariant. By allowing arbitrary regularizers, the results vastly broaden the scope of applicability of debiasing methodologies to obtain confidence intervals in high dimensions. In the absence of strong convexity for $p > n$, asymptotic normality of the debiased estimate is obtained for the Lasso and the group Lasso under additional conditions. For general convex penalties, our analysis also provides prediction and estimation error bounds of independent interest.

REFERENCES

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

[2] BELLEC, P. and TSYBAKOV, A. (2017). Bounds on the prediction error of penalized least squares estimators with convex penalty. In *Modern Problems of Stochastic Analysis and Statistics*. *Springer Proc. Math. Stat.* **208** 315–333. Springer, Cham. MR3747672 https://doi.org/10.1007/978-3-319-65313-6_13

[3] BELLEC, P. C. (2018). The noise barrier and the large signal bias of the lasso and other convex estimators. Available at: arXiv:1804.01230, https://arxiv.org/pdf/1804.01230.pdf.

[4] BELLEC, P. C. (2020). Out-of-sample error estimate for robust m-estimators with convex penalty. arXiv preprint. Available at arXiv:2008.11840.

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

[6] BELLEC, P. C. and SHEN, Y. (2022). Derivatives and residual distribution of regularized m-estimators with application to adaptive tuning. In *Conference on Learning Theory* 1912–1947. PMLR.

[7] BELLEC, P. C. and ZHANG, C.-H. (2021). Second-order Stein: SURE for SURE and other applications in high-dimensional inference. *Ann. Statist.* **49** 1864–1903. MR4319234 https://doi.org/10.1214/20-aos2005

[8] BELLEC, P. C. and ZHANG, C.-H. (2022). De-biasing the lasso with degrees-of-freedom adjustment. *Bernoulli* **28** 713–743. MR4389062 https://doi.org/10.3150/21-BEJ1348

[9] BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2014). High-dimensional methods and inference on structural and treatment effects. *J. Econ. Perspect.* **28** 29–50.

[10] 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

MSC2020 subject classifications. Primary 62H12, 62G15; secondary 62F35, 62J07.

Key words and phrases. Bias correction, central limit theorem, confidence intervals, convex regularization, Gaussian Poincaré inequality, high-dimensional linear models, Lasso, Stein's formula, variance estimation.

[11] BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. MR2533469 https://doi.org/10.1214/08-AOS620
[12] BLANCHARD, J. D., CARTIS, C. and TANNER, J. (2011). Compressed sensing: How sharp is the restricted isometry property? *SIAM Rev.* **53** 105–125. MR2785881 https://doi.org/10.1137/090748160
[13] BLUMENSATH, T. and DAVIES, M. E. (2009). Iterative hard thresholding for compressed sensing. *Appl. Comput. Harmon. Anal.* **27** 265–274. MR2559726 https://doi.org/10.1016/j.acha.2009.04.002
[14] BOGACHEV, V. I. (1998). *Gaussian Measures. Mathematical Surveys and Monographs* **62**. Amer. Math. Soc., Providence, RI. MR1642391 https://doi.org/10.1090/surv/062
[15] BU, Z., KLUSOWSKI, J., RUSH, C. and SU, W. (2019). Algorithmic analysis and statistical estimation of slope via approximate message passing. In *Advances in Neural Information Processing Systems* 9361–9371.
[16] CAI, T. T. and GUO, Z. (2017). Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *Ann. Statist.* **45** 615–646. MR3650395 https://doi.org/10.1214/16-AOS1461
[17] CELENTANO, M. and MONTANARI, A. (2022). Fundamental barriers to high-dimensional regression with convex penalties. *Ann. Statist.* **50** 170–196. MR4382013 https://doi.org/10.1214/21-aos2100
[18] CELENTANO, M., MONTANARI, A. and WEI, Y. (2020). The lasso with general gaussian designs with applications to hypothesis testing. arXiv preprint. Available at arXiv:2007.13716.
[19] 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
[20] 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
[21] DONOHO, D. and MONTANARI, A. (2016). High dimensional robust M-estimation: Asymptotic variance via approximate message passing. *Probab. Theory Related Fields* **166** 935–969. MR3568043 https://doi.org/10.1007/s00440-015-0675-z
[22] EDELMAN, A. (1988). Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.* **9** 543–560. MR0964668 https://doi.org/10.1137/0609045
[23] EL KAROUI, N., BEAN, D., BICKEL, P. J., LIM, C. and YU, B. (2013). On robust regression with high-dimensional predictors. *Proc. Natl. Acad. Sci. USA* **110** 14557–14562.
[24] 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
[25] FENG, L. and ZHANG, C.-H. (2019). Sorted concave penalized regression. *Ann. Statist.* **47** 3069–3098. MR4025735 https://doi.org/10.1214/18-AOS1759
[26] JAVANMARD, A. and MONTANARI, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909. MR3277152
[27] JAVANMARD, A. and MONTANARI, A. (2014). Hypothesis testing in high-dimensional regression under the Gaussian random design model: Asymptotic theory. *IEEE Trans. Inf. Theory* **60** 6522–6554. MR3265038 https://doi.org/10.1109/TIT.2014.2343629
[28] JAVANMARD, A. and MONTANARI, A. (2018). Debiasing the Lasso: Optimal sample size for Gaussian designs. *Ann. Statist.* **46** 2593–2622. MR3851749 https://doi.org/10.1214/17-AOS1630
[29] LAURENT, B. and MASSART, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* **28** 1302–1338. MR1805785 https://doi.org/10.1214/aos/1015957395
[30] LECUÉ, G. and MENDELSON, S. (2018). Regularization and the small-ball method I: Sparse recovery. *Ann. Statist.* **46** 611–641. MR3782379 https://doi.org/10.1214/17-AOS1562
[31] LEI, L., BICKEL, P. J. and EL KAROUI, N. (2018). Asymptotics for high dimensional regression M-estimates: Fixed design results. *Probab. Theory Related Fields* **172** 983–1079. MR3877551 https://doi.org/10.1007/s00440-017-0824-7
[32] LOUREIRO, B., GERBELOT, C., CUI, H., GOLDT, S., KRZAKALA, F., MEZARD, M. and ZDEBOROVÁ, L. (2021). Learning curves of generic features maps for realistic datasets with a teacher-student model. *Adv. Neural Inf. Process. Syst.* **34** 18137–18151.
[33] MIOLANE, L. and MONTANARI, A. (2021). The distribution of the Lasso: Uniform control over sparse balls and adaptive parameter tuning. *Ann. Statist.* **49** 2313–2335. MR4319252 https://doi.org/10.1214/20-aos2038
[34] MOHAMED, N. (2020). Scaled minimax optimality in high-dimensional linear regression: A non-convex algorithmic regularization approach. arXiv preprint. Available at arXiv:2008.12236.
[35] NICULESCU, C. P. and PERSSON, L.-E. (2006). *Convex Functions and Their Applications. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC* **23**. Springer, New York. MR2178902 https://doi.org/10.1007/0-387-31077-0
[36] SILVERSTEIN, J. W. (1985). The smallest eigenvalue of a large-dimensional Wishart matrix. *Ann. Probab.* **13** 1364–1368. MR0806232

[37] STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. MR0630098

[38] SUN, T. and ZHANG, C.-H. (2012). Scaled sparse linear regression. *Biometrika* **99** 879–898. MR2999166 https://doi.org/10.1093/biomet/ass043

[39] SUN, T. and ZHANG, C.-H. (2013). Sparse matrix inversion with scaled lasso. *J. Mach. Learn. Res.* **14** 3385–3418. MR3144466

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

[41] TALAGRAND, M. (2011). *Mean Field Models for Spin Glasses. Volume I: Basic Examples. Ergebnisse der Mathematik und Ihrer Grenzgebiete*. 3. *Folge. A Series of Modern Surveys in Mathematics* [*Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics*] **54**. Springer, Berlin. MR2731561 https://doi.org/10.1007/978-3-642-15202-3

[42] THRAMPOULIDIS, C., ABBASI, E. and HASSIBI, B. (2015). Lasso with non-linear measurements is equivalent to one with linear measurements. In *Advances in Neural Information Processing Systems* 3420–3428.

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

[44] TIBSHIRANI, R. J. (2013). The lasso problem and uniqueness. *Electron. J. Stat.* **7** 1456–1490. MR3066375 https://doi.org/10.1214/13-EJS815

[45] VAITER, S., DELEDALLE, C., PEYRÉ, G., FADILI, J. and DOSSAL, C. (2012). The degrees of freedom of the group lasso. arXiv preprint. Available at arXiv:1205.1481.

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

[47] 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

[48] 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

[49] ZHANG, C.-H. (2011). Statistical inference for high-dimensional data. *Math. Forsch. Oberwolfach* **48** 28–31.

[50] ZHANG, C.-H. and HUANG, J. (2008). The sparsity and bias of the LASSO selection in high-dimensional linear regression. *Ann. Statist.* **36** 1567–1594. MR2435448 https://doi.org/10.1214/07-AOS520

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

The Annals of Statistics
2023, Vol. 51, No. 2, 437–463
https://doi.org/10.1214/22-AOS2252
© Institute of Mathematical Statistics, 2023

CONDITIONAL SEQUENTIAL MONTE CARLO IN HIGH DIMENSIONS

BY AXEL FINKE[1,a] AND ALEXANDRE H. THIERY[2,b]

[1]*Department of Mathematical Sciences, Loughborough University,* [a]*a.finke@lboro.ac.uk*
[2]*Department of Statistics and Applied Probability, National University of Singapore,* [b]*a.h.thiery@nus.edu.sg*

The *iterated conditional sequential Monte Carlo (i-CSMC)* algorithm from Andrieu, Doucet and Holenstein (*J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** (2010) 269–342) is an MCMC approach for efficiently sampling from the joint posterior distribution of the T latent states in challenging time-series models, for example, in nonlinear or non-Gaussian state-space models. It is also the main ingredient in *particle Gibbs* samplers which infer unknown model parameters alongside the latent states. In this work, we first prove that the i-CSMC algorithm suffers from a curse of dimension in the dimension of the states, D: it breaks down unless the number of samples ('particles'), N, proposed by the algorithm grows exponentially with D. Then we present a novel 'local' version of the algorithm which proposes particles using Gaussian random-walk moves that are suitably scaled with D. We prove that this *iterated random-walk conditional sequential Monte Carlo (i-RW-CSMC)* algorithm avoids the curse of dimension: for arbitrary N, its acceptance rates and expected squared jumping distance converge to nontrivial limits as $D \to \infty$. If $T = N = 1$, our proposed algorithm reduces to a Metropolis–Hastings or Barker's algorithm with Gaussian random-walk moves and we recover the well-known scaling limits for such algorithms.

REFERENCES

AGRAWAL, S., VATS, D., ŁATUSZYŃSKI, K. and ROBERTS, G. O. (2021). Optimal scaling of MCMC beyond Metropolis. Available at arXiv:2104.02020. https://doi.org/10.48550/arXiv.2104.02020

ANDRIEU, C., DOUCET, A. and HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 269–342. MR2758115 https://doi.org/10.1111/j.1467-9868.2009.00736.x

WHITELEY, N. (2010). Contribution to the discussion on 'Particle Markov chain Monte Carlo methods' by Andrieu, C., Doucet, A., and Holenstein, R. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72** 306–307.

ANDRIEU, C., LEE, A. and VIHOLA, M. (2018). Uniform ergodicity of the iterated conditional SMC and geometric ergodicity of particle Gibbs samplers. *Bernoulli* **24** 842–872. MR3706778 https://doi.org/10.3150/15-BEJ785

BARKER, A. A. (1965). Monte Carlo calculations of the radial distribution functions for a proton–electron plasma. *Australian Journal of Physics* **18** 119–134. https://doi.org/10.1071/PH650119

BÉDARD, M., DOUC, R. and MOULINES, E. (2012). Scaling analysis of multiple-try MCMC methods. *Stochastic Process. Appl.* **122** 758–786. MR2891436 https://doi.org/10.1016/j.spa.2011.11.004

BÉDARD, M. and MIREUTA, M. (2013). On the empirical efficiency of local MCMC algorithms with pools of proposals. *Canad. J. Statist.* **41** 657–678. MR3146003 https://doi.org/10.1002/cjs.11196

BROWN, S., JENKINS, P. A., JOHANSEN, A. M. and KOSKELA, J. (2021). Simple conditions for convergence of sequential Monte Carlo genealogies with applications. *Electron. J. Probab.* **26** Paper No. 1, 22. MR4216514 https://doi.org/10.1214/20-ejp561

CAPPÉ, O., MOULINES, E. and RYDÉN, T. (2005). *Inference in Hidden Markov Models. Springer Series in Statistics*. Springer, New York. MR2159833

CHOPIN, N. and SINGH, S. S. (2015). On particle Gibbs sampling. *Bernoulli* **21** 1855–1883. MR3352064 https://doi.org/10.3150/14-BEJ629

CHRISTEN, J. A. and FOX, C. (2005). Markov chain Monte Carlo using an approximation. *J. Comput. Graph. Statist.* **14** 795–810. MR2211367 https://doi.org/10.1198/106186005X76983

MSC2020 subject classifications. Primary 65C05; secondary 60J05, 65C35, 65C40.

Key words and phrases. High dimensions, curse of dimension, Markov chain Monte Carlo, particle filter, state-space model.

CRESSIE, N. and WIKLE, C. K. (2015). *Statistics for Spatio-temporal Data*. Wiley, Hoboken, NJ.
DEL MORAL, P., KOHN, R. and PATRAS, F. (2016). On particle Gibbs samplers. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 1687–1733. MR3573292 https://doi.org/10.1214/15-AIHP695
DELMAS, J.-F. and JOURDAIN, B. (2009). Does waste recycling really improve the multi-proposal Metropolis-Hastings algorithm? An analysis based on control variates. *J. Appl. Probab.* **46** 938–959. MR2582699 https://doi.org/10.1239/jap/1261670681
DHAENE, J., WANG, S., YOUNG, V. and GOOVAERTS, M. J. (2000). Comonotonicity and maximal stop-loss premiums. *Schweiz. Aktuarver. Mitt.* **2** 99–113. MR1821441
DOUC, R., CAPPÉ, O. and MOULINES, E. (2005). Comparison of resampling schemes for particle filtering. In *Proceedings of the* 4*th International Symposium on Image and Signal Processing and Analysis* 64–69. IEEE. https://doi.org/10.1109/ISPA.2005.195385
DOUCET, A., GODSILL, S. J. and ANDRIEU, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statist. Comput.* **10** 197–208. https://doi.org/10.1023/A:1008935410038
DVORETZKY, A. (1972). Asymptotic normality for sums of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (*Univ. California*, *Berkeley*, *Calif.*, 1970/1971), *Vol. II*: *Probability Theory* 513–535. Univ. California Press, Berkeley, CA. MR0415728
FEARNHEAD, P. (1998). Sequential Monte Carlo methods in filter theory Ph.D. thesis Department of Statistics, Univ. Oxford UK.
FEARNHEAD, P. and CLIFFORD, P. (2003). On-line inference for hidden Markov models via particle filters. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **65** 887–899. MR2017876 https://doi.org/10.1111/1467-9868.00421
FINKE, A., DOUCET, A. and JOHANSEN, A. M. (2016). On embedded hidden Markov models and particle Markov chain Monte Carlo methods. Available at arXiv:1610.08962. https://doi.org/10.48550/arXiv.1610.08962
FINKE, A. and SINGH, S. S. (2017). Approximate smoothing and parameter estimation in high-dimensional state-space models. *IEEE Trans. Signal Process.* **65** 5982–5994. MR3722975 https://doi.org/10.1109/TSP.2017.2733504
FINKE, A. and THIERY, A. H. (2023). Supplement to "Conditional sequential Monte Carlo in high dimensions." https://doi.org/10.1214/22-AOS2252SUPP
FRENKEL, D. (2004). Speed-up of Monte Carlo simulations by sampling of rejected states. *Proc. Natl. Acad. Sci. USA* **101** 17571–17575. https://doi.org/10.1073/pnas.0407950101
GUARNIERO, P., JOHANSEN, A. M. and LEE, A. (2017). The iterated auxiliary particle filter. *J. Amer. Statist. Assoc.* **112** 1636–1647. MR3750887 https://doi.org/10.1080/01621459.2016.1222291
HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109. MR3363437 https://doi.org/10.1093/biomet/57.1.97
KONG, A., LIU, J. S. and WONG, W. H. (1994). Sequential imputations and Bayesian missing data problems. *J. Amer. Statist. Assoc.* **89** 278–288. https://doi.org/10.1080/01621459.1994.10476469
LEE, A., SINGH, S. S. and VIHOLA, M. (2020). Coupled conditional backward sampling particle filter. *Ann. Statist.* **48** 3066–3089. MR4152635 https://doi.org/10.1214/19-AOS1922
LINDSTEN, F., DOUC, R. and MOULINES, E. (2015). Uniform ergodicity of the particle Gibbs sampler. *Scand. J. Stat.* **42** 775–797. MR3391692 https://doi.org/10.1111/sjos.12136
LINDSTEN, F., JORDAN, M. I. and SCHÖN, T. B. (2012). Ancestor sampling for particle Gibbs. In *Proceedings of the* 2012 *Conference on Neural Information Processing Systems*.
LIU, J. S. (1996). Peskun's theorem and a modified discrete-state Gibbs sampler. *Biometrika* **83** 681–682. MR1423883 https://doi.org/10.1093/biomet/83.3.681
MALORY, S. J. (2021). *Bayesian Inference for Stochastic Processes*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—Lancaster University (United Kingdom). MR4336613
METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics* **21** 1087–1092. https://doi.org/10.1063/1.1699114
MURPHY, J. and GODSILL, S. J. (2015). Blocked particle Gibbs schemes for high dimensional interacting systems. *IEEE Journal of Selected Topics in Signal Processing* **10** 328–342. https://doi.org/10.1109/JSTSP.2015.2509940
NEAL, P. and ROBERTS, G. (2011). Optimal scaling of random walk Metropolis algorithms with non-Gaussian proposals. *Methodol. Comput. Appl. Probab.* **13** 583–601. MR2822397 https://doi.org/10.1007/s11009-010-9176-9
NEAL, R. M. (2003). Markov Chain sampling for non-linear state space models using embedded hidden Markov models. *ArXiv Mathematics E-prints*. https://doi.org/10.48550/arXiv.math/0305039
NEAL, R. M., BEAL, M. J. and ROWEIS, S. T. (2004). Inferring state sequences for non-linear systems with embedded hidden Markov models. *Adv. Neural Inf. Process. Syst.* **16** 401–408.

PESKUN, P. H. (1973). Optimum Monte-Carlo sampling using Markov chains. *Biometrika* **60** 607–612. MR0362823 https://doi.org/10.1093/biomet/60.3.607

PITT, M. K. and SHEPHARD, N. (1999). Filtering via simulation: Auxiliary particle filters. *J. Amer. Statist. Assoc.* **94** 590–599. MR1702328 https://doi.org/10.2307/2670179

REBESCHINI, P. and VAN HANDEL, R. (2015). Can local particle filters beat the curse of dimensionality? *Ann. Appl. Probab.* **25** 2809–2866. MR3375889 https://doi.org/10.1214/14-AAP1061

ROBERTS, G. O., GELMAN, A. and GILKS, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* **7** 110–120. MR1428751 https://doi.org/10.1214/aoap/1034625254

SCHWEDES, T. and CALDERHEAD, B. (2018). Quasi Markov chain Monte Carlo methods. Available at arXiv:1807.00070. https://doi.org/10.48550/arXiv.1807.00070

SHERLOCK, C. and ROBERTS, G. (2009). Optimal scaling of the random walk Metropolis on elliptically symmetric unimodal targets. *Bernoulli* **15** 774–798. MR2555199 https://doi.org/10.3150/08-BEJ176

SHESTOPALOFF, A. and DOUCET, A. (2019). Replica Conditional Sequential Monte Carlo. In *International Conference on Machine Learning* 5749–5757.

SHESTOPALOFF, A. Y. and NEAL, R. M. (2013). MCMC for non-linear state space models using ensembles of latent sequences. *ArXiv E-prints*. https://doi.org/10.48550/arXiv.1305.0320

SHESTOPALOFF, A. Y. and NEAL, R. M. (2018). Sampling latent states for high-dimensional non-linear state space models with the embedded HMM method. *Bayesian Anal.* **13** 797–822. MR3807867 https://doi.org/10.1214/17-BA1077

SINGH, S. S., LINDSTEN, F. and MOULINES, E. (2017). Blocking strategies and stability of particle Gibbs samplers. *Biometrika* **104** 953–969. MR3737314 https://doi.org/10.1093/biomet/asx051

TJELMELAND, H. (2004). Using all Metropolis–Hastings proposals to estimate mean values. Preprint No. 4/2004. Norwegian Univ. Science and Technology, Trondheim, Norway.

VAN LEEUWEN, P. J. (2009). Particle filtering in geophysical systems. *Monthly Weather Review* **137** 4089–4114. https://doi.org/10.1175/2009MWR2835.1

YANG, J., ROBERTS, G. O. and ROSENTHAL, J. S. (2020). Optimal scaling of random-walk metropolis algorithms on general target distributions. *Stochastic Process. Appl.* **130** 6094–6132. MR4140028 https://doi.org/10.1016/j.spa.2020.05.004

YANG, S., CHEN, Y., BERNTON, E. and LIU, J. S. (2018). On parallelizable Markov chain Monte Carlo algorithms with waste-recycling. *Stat. Comput.* **28** 1073–1081. MR3835634 https://doi.org/10.1007/s11222-017-9780-4

The Annals of Statistics
2023, Vol. 51, No. 2, 464–486
https://doi.org/10.1214/22-AOS2254
© Institute of Mathematical Statistics, 2023

INTERACTIVE VERSUS NONINTERACTIVE LOCALLY DIFFERENTIALLY PRIVATE ESTIMATION: TWO ELBOWS FOR THE QUADRATIC FUNCTIONAL

BY CRISTINA BUTUCEA[1,a], ANGELIKA ROHDE[2,b] AND LUKAS STEINBERGER[3,c]

[1]*CREST, ENSAE, IP Paris,* [a]*cristina.butucea@ensae.fr*
[2]*Institute of Mathematics, Albert-Ludwigs-Universität Freiburg,* [b]*angelika.rohde@stochastik.uni-freiburg.de*
[3]*Department of Statistics and OR, University of Vienna,* [c]*lukas.steinberger@univie.ac.at*

Local differential privacy has recently received increasing attention from the statistics community as a valuable tool to protect the privacy of individual data owners without the need of a trusted third party. Similar to the classical notion of randomized response, the idea is that data owners randomize their true information locally and only release the perturbed data. Many different protocols for such local perturbation procedures can be designed. In most estimation problems studied in the literature so far, however, no significant difference in terms of minimax risk between purely noninteractive protocols and protocols that allow for some amount of interaction between individual data providers could be observed. In this paper, we show that for estimating the integrated square of a density, sequentially interactive procedures improve substantially over the best possible noninteractive procedure in terms of minimax rate of estimation.

In particular, in the noninteractive scenario we identify an elbow in the minimax rate at $s = \frac{3}{4}$, whereas in the sequentially interactive scenario the elbow is at $s = \frac{1}{2}$. This is markedly different from both, the case of direct observations, where the elbow is well known to be at $s = \frac{1}{4}$, as well as from the case where Laplace noise is added to the original data, where an elbow at $s = \frac{9}{4}$ is obtained.

We also provide adaptive estimators that achieve the optimal rate up to log-factors, we draw connections to nonparametric goodness-of-fit testing and estimation of more general integral functionals and conduct a series of numerical experiments. The fact that a particular locally differentially private, but interactive, mechanism improves over the simple noninteractive one is also of great importance for practical implementations of local differential privacy.

REFERENCES

ACHARYA, J., CANONNE, C. L., SUN, Z. and TYAGI, H. (2022). The role of interactivity in structured estimation. Preprint. Available at arXiv:2203.06870.

BERRETT, T. and BUTUCEA, C. (2020). Locally private non-asymptotic testing of discrete distributions is faster using interactive mechanisms. *NeurIPS* **34**.

BICKEL, P. J. and RITOV, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhyā Ser. A* **50** 381–393. MR1065550

BIRGÉ, L. and MASSART, P. (1995). Estimation of integral functionals of a density. *Ann. Statist.* **23** 11–29. MR1331653 https://doi.org/10.1214/aos/1176324452

BUTUCEA, C. (2007). Goodness-of-fit testing and quadratic functional estimation from indirect observations. *Ann. Statist.* **35** 1907–1930. MR2363957 https://doi.org/10.1214/009053607000000118

BUTUCEA, C. and ISSARTEL, Y. (2021). Locally differentially private estimation of nonlinear functionals of discrete distributions. *NeurIPS* **34**.

MSC2020 subject classifications. Primary 62G05; secondary 62G10, 62C20.

Key words and phrases. Local differential privacy, quadratic functional, minimax estimation, rate of convergence, nonparametric estimation.

BUTUCEA, C., ROHDE, A. and STEINBERGER, L. (2023). Supplement to "Interactive versus noninteractive locally differentially private estimation: Two elbows for the quadratic functional." https://doi.org/10.1214/22-AOS2254SUPP

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

CAI, T. T. and LOW, M. G. (2005). Nonquadratic estimators of a quadratic functional. *Ann. Statist.* **33** 2930–2956. MR2253108 https://doi.org/10.1214/009053605000000147

CAI, T. T. and LOW, M. G. (2006). Optimal adaptive estimation of a quadratic functional. *Ann. Statist.* **34** 2298–2325. MR2291501 https://doi.org/10.1214/009053606000000849

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

COLLIER, O., COMMINGES, L. and TSYBAKOV, A. B. (2017). Minimax estimation of linear and quadratic functionals on sparsity classes. *Ann. Statist.* **45** 923–958. MR3662444 https://doi.org/10.1214/15-AOS1432

DEVORE, R. A., JAWERTH, B. and POPOV, V. (1992). Compression of wavelet decompositions. *Amer. J. Math.* **114** 737–785. MR1175690 https://doi.org/10.2307/2374796

DE LA PEÑA, V. H. and MONTGOMERY-SMITH, S. J. (1995). Decoupling inequalities for the tail probabilities of multivariate U-statistics. *Ann. Probab.* **23** 806–816. MR1334173

DONOHO, D. L. and NUSSBAUM, M. (1990). Minimax quadratic estimation of a quadratic functional. *J. Complexity* **6** 290–323. MR1081043 https://doi.org/10.1016/0885-064X(90)90025-9

DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2013a). Local privacy and statistical minimax rates. In 2013 *IEEE* 54*th Annual Symposium on Foundations of Computer Science—FOCS* 2013 429–438. IEEE Computer Soc., Los Alamitos, CA. MR3246246 https://doi.org/10.1109/FOCS.2013.53

DUCHI, J., JORDAN, M. I. and WAINWRIGHT, M. J. (2013b). Local privacy and minimax bounds: Sharp rates for probability estimation. In *Adv. Neural Inf. Process. Syst.* 1529–1537.

DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2014). Local privacy, data processing inequalities, and statistical minimax rates. Preprint. Available at arXiv:1302.3203.

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

DUCHI, J. C. and RUAN, F. (2018). The right complexity measure in locally private estimation: It is not the Fisher information. Preprint. Available at arXiv:1806.05756v1.

DUCHI, J. C. and RUAN, F. (2020). The right complexity measure in locally private estimation: It is not the Fisher information. Available at https://arxiv.org/abs/1806.05756v3.

DWORK, C., MCSHERRY, F., NISSIM, K. and SMITH, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography* (S. Halevi and T. Rabin, eds.). *Lecture Notes in Computer Science* **3876** 265–284. Springer, Berlin. MR2241676 https://doi.org/10.1007/11681878_14

EVFIMIEVSKI, A., GEHRKE, J. and SRIKANT, R. (2003). Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* 211–222. ACM, New York.

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

HÄRDLE, W., KERKYACHARIAN, G., PICARD, D. and TSYBAKOV, A. (1998). *Wavelets, Approximation, and Statistical Applications*. *Lecture Notes in Statistics* **129**. Springer, New York. MR1618204 https://doi.org/10.1007/978-1-4612-2222-4

HOUDRÉ, C. and REYNAUD-BOURET, P. (2003). Exponential inequalities, with constants, for U-statistics of order two. In *Stochastic Inequalities and Applications*. *Progress in Probability* **56** 55–69. Birkhäuser, Basel. MR2073426

JOSEPH, M., MAO, J. and ROTH, A. (2020). Exponential separations in local differential privacy. In *Proceedings of the* 2020 *ACM-SIAM Symposium on Discrete Algorithms* 515–527. SIAM, Philadelphia, PA. MR4141213

JOSEPH, M., MAO, J., NEEL, S. and ROTH, A. (2019). The role of interactivity in local differential privacy. In 2019 *IEEE* 60*th Annual Symposium on Foundations of Computer Science* 94–105. IEEE Comput. Soc. Press, Los Alamitos, CA. MR4228162 https://doi.org/10.1109/FOCS.2019.00015

KASIVISWANATHAN, S. P., LEE, H. K., NISSIM, K., RASKHODNIKOVA, S. and SMITH, A. (2011). What can we learn privately? *SIAM J. Comput.* **40** 793–826. MR2823508 https://doi.org/10.1137/090756090

KLEMELÄ, J. (2006). Sharp adaptive estimation of quadratic functionals. *Probab. Theory Related Fields* **134** 539–564. MR2214904 https://doi.org/10.1007/s00440-005-0447-2

LAM-WEIL, J., LAURENT, B. and LOUBES, J.-M. (2022). Minimax optimal goodness-of-fit testing for densities and multinomials under a local differential privacy constraint. *Bernoulli* **28** 579–600. MR4337717 https://doi.org/10.3150/21-bej1358

LAURENT, B. (2005). Adaptive estimation of a quadratic functional of a density by model selection. *ESAIM Probab. Stat*. **9** 1–18. MR2148958 https://doi.org/10.1051/ps:2005001

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

RITOV, Y. and BICKEL, P. J. (1990). Achieving information bounds in non and semiparametric models. *Ann. Statist*. **18** 925–938. MR1056344 https://doi.org/10.1214/aos/1176347633

ROHDE, A. and STEINBERGER, L. (2020). Geometrizing rates of convergence under local differential privacy constraints. *Ann. Statist*. **48** 2646–2670. MR4152116 https://doi.org/10.1214/19-AOS1901

SMITH, A. (2008). Efficient, differentially private point estimators. Preprint. Available at arXiv:0809.4794.

SMITH, A. (2011). Privacy-preserving statistical estimation with optimal convergence rates. In *STOC'*11—*Proceedings of the* 43*rd ACM Symposium on Theory of Computing* 813–821. ACM, New York. MR2932032 https://doi.org/10.1145/1993636.1993743

STEINBERGER, L. (2023). Efficiency in local differential privacy. Preprint. Available at arXiv:2301.10600.

WARNER, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc*. **60** 63–69. https://doi.org/10.1080/01621459.1965.10480775

WASSERMAN, L. and ZHOU, S. (2010). A statistical framework for differential privacy. *J. Amer. Statist. Assoc*. **105** 375–389. MR2656057 https://doi.org/10.1198/jasa.2009.tm08651

YE, M. and BARG, A. (2017). Asymptotically optimal private estimation under mean square loss. Preprint. Available at arXiv:1708.00059.

The Annals of Statistics
2023, Vol. 51, No. 2, 487–518
https://doi.org/10.1214/23-AOS2256
© Institute of Mathematical Statistics, 2023

NONLINEAR INDEPENDENT COMPONENT ANALYSIS FOR DISCRETE-TIME AND CONTINUOUS-TIME SIGNALS

BY ALEXANDER SCHELL[a] AND HARALD OBERHAUSER[b]

Mathematical Institute, University of Oxford, [a]*schell@maths.ox.ac.uk,* [b]*oberhauser@maths.ox.ac.uk*

We study the classical problem of recovering a multidimensional source signal from observations of nonlinear mixtures of this signal. We show that this recovery is possible (up to a permutation and monotone scaling of the source's original component signals) if the mixture is due to a sufficiently differentiable and invertible but otherwise arbitrarily nonlinear function and the component signals of the source are statistically independent with 'non-degenerate' second-order statistics. The latter assumption requires the source signal to meet one of three regularity conditions which essentially ensure that the source is sufficiently far away from the nonrecoverable extremes of being deterministic or constant in time. These assumptions, which cover many popular time series models and stochastic processes, allow us to reformulate the initial problem of nonlinear blind source separation as a simple-to-state problem of optimisation-based function approximation. We propose to solve this approximation problem by minimizing a novel type of objective function that efficiently quantifies the mutual statistical dependence between multiple stochastic processes via cumulant-like statistics. This yields a scalable and direct new method for nonlinear Independent Component Analysis with widely applicable theoretical guarantees and for which our experiments indicate good performance.

REFERENCES

[1] ALMEIDA, L. B. (2003). MISEP—linear and nonlinear ICA based on mutual information. *J. Mach. Learn. Res.* **4** 1297–1318.

[2] ARDIZZONE, L., KRUSE, J., WIRKERT, S., RAHNER, D., PELLEGRINI, E. W., KLESSEN, R. S., MAIER-HEIN, L., ROTHER, C. and KÖTHE, U. (2018). Analyzing Inverse Problems With Invertible Neural Networks. Published at ICLR 2019. Preprint. Available at arXiv:1808.04730.

[3] BACH, F. R. and JORDAN, M. I. (2003). Kernel independent component analysis. *J. Mach. Learn. Res.* **3** 1–48. MR1966051 https://doi.org/10.1162/153244303768966085

[4] BELL, A. J. and SEJNOWSKI, T. (1995). An information maximisation approach to blind separation and blind deconvolution. *Neural Comput.* **7** 1129–1159.

[5] BELOUCHRANI, A., MERAIM, K. A., CARDOSO, J. F. and MOULINES, E. (1997). A blind source separation technique using second-order statistics. *IEEE Trans. Signal Process.* **45** 434–444.

[6] BONNIER, P. and OBERHAUSER, H. (2020). Signature cumulants, ordered partitions, and independence of stochastic processes. *Bernoulli* **26** 2727–2757. MR4140527 https://doi.org/10.3150/20-BEJ1205

[7] BRAKEL, P. and BENGIO, Y. (2017). Learning independent features with adversarial nets for non-linear ICA. Preprint. Available at arXiv:1710.05050 [stat.ML].

[8] CARDOSO, J. F. (1999). High-order contrasts for independent component analysis. *Neural Comput.* **11** 157–192. https://doi.org/10.1162/089976699300016863

[9] CARDOSO, J. F. and SOULOUMIAC, A. (1993). Blind beamforming for non Gaussian signals. *IEE Proc. F* **140** 362–370.

[10] CHEN, X. and FAN, Y. (2006). Estimation of copula-based semiparametric time series models. *J. Econometrics* **130** 307–335. MR2211797 https://doi.org/10.1016/j.jeconom.2005.03.004

[11] CHEVYREV, I. and LYONS, T. (2016). Characteristic functions of measures on geometric rough paths. *Ann. Probab.* **44** 4049–4082. MR3572331 https://doi.org/10.1214/15-AOP1068

MSC2020 subject classifications. Primary 62H25, 62M99; secondary 62H05, 60L10, 62M45, 62R10.

Key words and phrases. Blind source separation, independent component analysis, inverse problem, latent variable model, statistical independence, unsupervised learning, statistical inference for stochastic processes, functional data analysis, nonlinear BSS, nonlinear ICA.

[12] CHEVYREV, I. and OBERHAUSER, H. (2018). Signature Moments to Characterize Laws of Stochastic Processes. Preprint. Available at arXiv:1810.1097.
[13] COLMENAREJO, L. and PREISS, R. (2020). Signatures of paths transformed by polynomial maps. *Beitr. Algebra Geom.* **61** 695–717. MR4160818 https://doi.org/10.1007/s13366-020-00493-9
[14] COMON, P. (1994). Independent component analysis, a new concept? *Signal Process.* **36** 287–314.
[15] COMON, P. and JUTTEN, C., eds. (2010). *Handbook of Blind Source Separation*: *Independent Component Analysis and Applications*. Academic Press, San Diego.
[16] CRANMER, K., BREHMER, J. and LOUPPE, G. (2020). The frontier of simulation-based inference. *Proc. Natl. Acad. Sci. USA* **117** 30055–30062. MR4263287 https://doi.org/10.1073/pnas.1912789117
[17] DARMOIS, G. (1953). Analyse générale des liaisons stochastiques. Etude particulière de l'analyse factorielle linéaire. *Rev. Inst. Int. Stat.* **21** 2–8. MR0061322
[18] DARSOW, W. F., NGUYEN, B. and OLSEN, E. T. (1992). Copulas and Markov processes. *Illinois J. Math.* **36** 600–642. MR1215798
[19] DING, H., WANG, Y., YANG, Z. and PFEIFFER, O. (2019). Nonlinear blind source separation and fault feature extraction method for mining machine diagnosis. *Appl. Sci.* **9** 1852.
[20] EMBRECHTS, P., MCNEIL, A. J. and STRAUMANN, D. (2002). Correlation and dependence in risk management: Properties and pitfalls. In *Risk Management*: *Value at Risk and Beyond* (*Cambridge*, 1998) 176–223. Cambridge Univ. Press, Cambridge. MR1892190 https://doi.org/10.1017/CBO9780511615337.008
[21] EMURA, T., LONG, T.-H. and SUN, L.-H. (2017). R routines for performing estimation and statistical process control under copula-based time series models. *Comm. Statist. Simulation Comput.* **46** 3067–3087. MR3640123 https://doi.org/10.1080/03610918.2015.1073303
[22] ERIKSSON, J. and KOIVUNEN, V. (2004). Identifiability, separability and uniqueness of linear ICA models. *IEEE Signal Process. Lett.* **11** 601–604.
[23] FRIZ, P. K. and VICTOIR, N. B. (2010). *Multidimensional Stochastic Processes as Rough Paths*: *Theory and Applications*. *Cambridge Studies in Advanced Mathematics* **120**. Cambridge Univ. Press, Cambridge. MR2604669 https://doi.org/10.1017/CBO9780511845079
[24] GRETTON, A., HERBRICH, R., SMOLA, A., BOUSQUET, O. and SCHÖLKOPF, B. (2005). Kernel methods for measuring independence. *J. Mach. Learn. Res.* **6** 2075–2129. MR2249882
[25] HARMELING, S., ZIEHE, A., KAWANABE, M. and MÜLLER, K. R. (2003). Kernel-based nonlinear blind source separation. *Neural Comput.* **15** 1089–1124.
[26] HASTIE, T. and TIBSHIRANI, R. (2003). Independent component analysis through product density estimation. *Adv. Neural Inf. Process. Syst.* **15** 649–656.
[27] HE, Q. P. and WANG, J. (2018). Statistical process monitoring as a big data analytics tool for smart manufacturing. *J. Process. Control* **67** 35–43.
[28] HJELM, R. D. et al. (2018). Learning deep representations by mutual information estimation and maximization. Preprint. Available at arXiv:1808.06670v5 [stat.ML].
[29] HYVÄRINEN, A. (1997). Independent Component Analysis by Minimization of Mutual Information. Technical Report (Report A46), Helsiniki Univ. Technology.
[30] HYVÄRINEN, A. (1998). New approximations of differential entropy for independent component analysis and projection pursuit. *Adv. Neural Inf. Process. Syst.* 273–279.
[31] HYVÄRINEN, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.* **10** 626–634.
[32] HYVÄRINEN, A. (2013). Independent component analysis: Recent advances. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **371** 20110534. MR3005670 https://doi.org/10.1098/rsta.2011.0534
[33] HYVÄRINEN, A., KARHUNEN, J. and OJA, E. (2001). *Independent Component Analysis*. Wiley, New York.
[34] HYVÄRINEN, A. and MORIOKA, H. (2016). Unsupervised feature extraction by time-contrastive learning and nonlinear ICA. *NeurIPS*2016 3765–3773.
[35] HYVÄRINEN, A. and MORIOKA, H. (2017). Nonlinear ICA of temporally dependent stationary sources. *PMLR* **54** 460–469. Supplementary Material at http://proceedings.mlr.press/v54/hyvarinen17a/hyvarinen17a-supp.pdf.
[36] HYVÄRINEN, A. and PAJUNEN, P. (1999). Nonlinear independent component analysis: Existence and uniqueness results. *Neural Netw.* **12** 429–439.
[37] HYVÄRINEN, A., SASAKI, H. and TURNER, R. (2019). Nonlinear ICA using auxiliary variables and generalized contrastive learning. *AISTATS*.
[38] İCAN, Ö. and ÇELIK, T. B. (2017). Stock market prediction performance of neural networks: A literature review. *Int. J. Econ. Finance* **9** 100–108.
[39] KHEMAKHEM, I., KINGMA, D. P., MONTI, R. P. and HYVÄRINEN, A. (2020). Variational autoencoders and nonlinear ICA: A unifying framework. *AISTATS*2020.

[40] KHOSHNEVIS, S. A. and SANKAR, R. (2019). Applications of higher order statistics in electroencephalography signal processing: A comprehensive survey. *IEEE Rev. Biomed. Eng.* **13** 169–183.
[41] KIDGER, P. and LYONS, T. Signatory: Differentiable computations of the signature and logsignature transforms, on both CPU and GPU. *ICLR* 2021.
[42] LYONS, T. J., CARUANA, M. and LÉVY, T. (2007). *Differential Equations Driven by Rough Paths. Lecture Notes in Math.* **1908**. Springer, Berlin. MR2314753
[43] MIETTINEN, J., NORDHAUSEN, K. and TASKINEN, S. (2017). Blind source separation based on joint diagonalization in R: The packages JADE and BSSasymp. *J. Stat. Softw.* **76**.
[44] MOULINES, E., CARDOSO, J. F. and GASSIAT, E. (1997). Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models. *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing* (*ICASSP'*97) 3617–3620.
[45] NELSEN, R. B. (2006). *An Introduction to Copulas*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2197664 https://doi.org/10.1007/s11229-005-3715-x
[46] NOÉ, F., TKATCHENKO, A., MÜLLER, K.-R. and CLEMENTI, C. (2020). Machine learning for molecular simulation. *Annu. Rev. Phys. Chem.* **71** 361–390. https://doi.org/10.1146/annurev-physchem-042018-052331
[47] PHAM, D. T. and GARRAT, P. (1997). Blind separation of mixture of independent sources through a quasi-maximum likelihood approach. *IEEE Trans. Signal Process.* **45** 1712–1725.
[48] RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning. Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR2514435
[49] REIERSØL, O. (1950). Identifiability of a linear relation between variables which are subject to error. *Econometrica* **18** 375–389. MR0038054 https://doi.org/10.2307/1907835
[50] ROGERS, L. C. G. and WILLIAMS, D. (2000). *Diffusions, Markov Processes, and Martingales. Vol.* 2. *Cambridge Mathematical Library*. Cambridge Univ. Press, Cambridge. MR1780932 https://doi.org/10.1017/CBO9781107590120
[51] SAMWORTH, R. J. and YUAN, M. (2012). Independent component analysis via nonparametric maximum likelihood estimation. *Ann. Statist.* **40** 2973–3002. MR3097966 https://doi.org/10.1214/12-AOS1060
[52] SCHELL, A. (2021). SigNICA. GitHub repository, https://github.com/alexander-schell/SigNICA.git Code for Section 8 (Jupyter Notebooks and Python files).
[53] SCHELL, A. and OBERHAUSER, H. (2023). Supplement to "Nonlinear Independent Component Analysis For Discrete-Time and Continuous-Time Signals." https://doi.org/10.1214/23-AOS2256SUPP
[54] SKITOVIČ, V. P. (1953). On a property of the normal distribution. *Dokl. Akad. Nauk SSSR* **89** 217–219. MR0055597
[55] TAN, Y., WANG, J. and ZURADA, J. M. (2001). Nonlinear blind source separation using a radial basis function network. *IEEE Trans. Neural Netw.* **12** 124–134.

The Annals of Statistics
2023, Vol. 51, No. 2, 519–546
https://doi.org/10.1214/23-AOS2257
© Institute of Mathematical Statistics, 2023

LOCAL CONVEXITY OF THE TAP FREE ENERGY AND AMP CONVERGENCE FOR $\mathbb{Z}_2$-SYNCHRONIZATION

BY MICHAEL CELENTANO[1,a], ZHOU FAN[2,c] AND SONG MEI[1,b]

[1]*Department of Statistics, University of California, Berkeley,* [a]*mcelentano@berkeley.edu,* [b]*songmei@berkeley.edu*
[2]*Department of Statistics and Data Science, Yale University,* [c]*zhou.fan@yale.edu*

We study mean-field variational Bayesian inference using the TAP approach, for $\mathbb{Z}_2$-synchronization as a prototypical example of a high-dimensional Bayesian model. We show that for any signal strength $\lambda > 1$ (the weak-recovery threshold), there exists a unique local minimizer of the TAP free energy functional near the mean of the Bayes posterior law. Furthermore, the TAP free energy in a local neighborhood of this minimizer is strongly convex. Consequently, a natural-gradient/mirror-descent algorithm achieves linear convergence to this minimizer from a local initialization, which may be obtained by a constant number of iterations of Approximate Message Passing (AMP). This provides a rigorous foundation for variational inference in high dimensions via minimization of the TAP free energy.

We also analyze the finite-sample convergence of AMP, showing that AMP is asymptotically stable at the TAP minimizer for any $\lambda > 1$, and is linearly convergent to this minimizer from a spectral initialization for sufficiently large λ. Such a guarantee is stronger than results obtainable by state evolution analyses, which only describe a fixed number of AMP iterations in the infinite-sample limit.

Our proofs combine the Kac–Rice formula and Sudakov–Fernique Gaussian comparison inequality to analyze the complexity of critical points that satisfy strong convexity and stability conditions within their local neighborhoods.

REFERENCES

[1] ALQUIER, P. and RIDGWAY, J. (2020). Concentration of tempered posteriors and of their variational approximations. *Ann. Statist.* **48** 1475–1497. MR4124331 https://doi.org/10.1214/19-AOS1855

[2] AMARI, S.-I. (1998). Natural gradient works efficiently in learning. *Neural Comput.* **10** 251–276.

[3] AROUS, G. B., BOURGADE, P. and MCKENNA, B. (2021). Landscape complexity beyond invariance and the elastic manifold. Preprint. Available at arXiv:2105.05051.

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

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

[6] AUFFINGER, A. and JAGANNATH, A. (2019). Thouless–Anderson–Palmer equations for generic p-spin glasses. *Ann. Probab.* **47** 2230–2256. MR3980920 https://doi.org/10.1214/18-AOP1307

[7] AUGERI, F. (2020). Nonlinear large deviation bounds with applications to Wigner matrices and sparse Erdős–Rényi graphs. *Ann. Probab.* **48** 2404–2448. MR4152647 https://doi.org/10.1214/20-AOP1427

[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] BANDEIRA, A. S., CHEN, Y., LEDERMAN, R. R. and SINGER, A. (2020). Non-unique games over compact groups and orientation estimation in cryo-EM. *Inverse Probl.* **36** 064002, 39 pp. MR4105332 https://doi.org/10.1088/1361-6420/ab7d2c

MSC2020 subject classifications. 62C10.

Key words and phrases. Variational inference, $\mathbb{Z}_2$ synchronization, TAP free energy, approximate message passing, natural gradient descent, nonconvex optimization, landscape analysis.

[10] BARBIER, J., DIA, M., MACRIS, N., KRZAKALA, F., LESIEUR, T. and ZDEBOROVÁ, L. (2016). Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. In *Neural Information Processing Systems*.
[11] BASAK, A. and MUKHERJEE, S. (2017). Universality of the mean-field for the Potts model. *Probab. Theory Related Fields* **168** 557–600. MR3663625 https://doi.org/10.1007/s00440-016-0718-0
[12] BASKERVILLE, N. P., KEATING, J. P., MEZZADRI, F. and NAJNUDEL, J. (2021). The loss surfaces of neural networks with general activation functions. *J. Stat. Mech. Theory Exp.* **2021** Paper No. 064001, 71 pp. MR4309626 https://doi.org/10.1088/1742-5468/abfa1e
[13] BASKERVILLE, N. P., KEATING, J. P., MEZZADRI, F. and NAJNUDEL, J. (2022). A spin glass model for the loss surfaces of generative adversarial networks. *J. Stat. Phys.* **186** Paper No. 29, 45 pp. MR4368914 https://doi.org/10.1007/s10955-022-02875-w
[14] BAUSCHKE, H. H., BOLTE, J. and TEBOULLE, M. (2017). A descent lemma beyond Lipschitz gradient continuity: First-order methods revisited and applications. *Math. Oper. Res.* **42** 330–348. MR3651994 https://doi.org/10.1287/moor.2016.0817
[15] BAYATI, M., LELARGE, M. and MONTANARI, A. (2015). Universality in polytope phase transitions and message passing algorithms. *Ann. Appl. Probab.* **25** 753–822. MR3313755 https://doi.org/10.1214/14-AAP1010
[16] BAYATI, M. and MONTANARI, A. (2011). The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inf. Theory* **57** 764–785. MR2810285 https://doi.org/10.1109/TIT.2010.2094817
[17] BECK, A. and TEBOULLE, M. (2003). Mirror descent and nonlinear projected subgradient methods for convex optimization. *Oper. Res. Lett.* **31** 167–175. MR1967286 https://doi.org/10.1016/S0167-6377(02)00231-6
[18] BELIUS, D. and KISTLER, N. (2019). The TAP-Plefka variational principle for the spherical SK model. *Comm. Math. Phys.* **367** 991–1017. MR3943486 https://doi.org/10.1007/s00220-019-03304-y
[19] BEN AROUS, G., MEI, S., MONTANARI, A. and NICA, M. (2019). The landscape of the spiked tensor model. *Comm. Pure Appl. Math.* **72** 2282–2330. MR4011861 https://doi.org/10.1002/cpa.21861
[20] BERTHIER, R., MONTANARI, A. and NGUYEN, P.-M. (2020). State evolution for approximate message passing with non-separable functions. *Inf. Inference* **9** 33–79. MR4079177 https://doi.org/10.1093/imaiai/iay021
[21] BICKEL, P., CHOI, D., CHANG, X. and ZHANG, H. (2013). Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *Ann. Statist.* **41** 1922–1943. MR3127853 https://doi.org/10.1214/13-AOS1124
[22] BINGHAM, E., CHEN, J. P., JANKOWIAK, M., OBERMEYER, F., PRADHAN, N., KARALETSOS, T., SINGH, R., SZERLIP, P., HORSFALL, P. et al. (2019). Pyro: Deep universal probabilistic programming. *J. Mach. Learn. Res.* **20** 973–978.
[23] BLEI, D. M. (2012). Probabilistic topic models. *Commun. ACM* **55** 77–84.
[24] 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
[25] BLEI, D. M., NG, A. Y. and JORDAN, M. I. (2003). Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3** 993–1022.
[26] BOLTHAUSEN, E. (2014). An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model. *Comm. Math. Phys.* **325** 333–366. MR3147441 https://doi.org/10.1007/s00220-013-1862-3
[27] BOLTHAUSEN, E. (2019). A Morita type proof of the replica-symmetric formula for SK. In *Statistical Mechanics of Classical and Disordered Systems. Springer Proc. Math. Stat.* **293** 63–93. Springer, Cham. MR4015008 https://doi.org/10.1007/978-3-030-29077-1_4
[28] BRAY, A. and MOORE, M. A. (1980). Metastable states in spin glasses. *J. Phys. C, Solid State Phys.* **13** L469.
[29] BRAY, A., MOORE, M. A. and YOUNG, A. P. (1984). Weighted averages of TAP solutions and Parisi's $q(x)$. *J. Phys. C, Solid State Phys.* **17** L155.
[30] CARBONETTO, P. and STEPHENS, M. (2012). Scalable variational inference for Bayesian variable selection in regression, and its accuracy in genetic association studies. *Bayesian Anal.* **7** 73–107. MR2896713 https://doi.org/10.1214/12-BA703
[31] CAVAGNA, A., GIARDINA, I., PARISI, G. and MÉZARD, M. (2003). On the formal equivalence of the TAP and thermodynamic methods in the SK model. *J. Phys. A* **36** 1175–1194. MR1960081 https://doi.org/10.1088/0305-4470/36/5/301
[32] CELENTANO, M. (2022). Sudakov–Fernique post-AMP, and a new proof of the local convexity of the TAP free energy. Preprint. Available at arXiv:2208.09550.

[33] CELENTANO, M., FAN, Z. and MEI, S. (2023). Supplement to "Local convexity of the TAP free energy and AMP convergence for $\mathbb{Z}_2$-synchronization." https://doi.org/10.1214/23-AOS2257SUPP
[34] CELENTANO, M., MONTANARI, A. and WEI, Y. (2020). The Lasso with general Gaussian designs with applications to hypothesis testing. Preprint. Available at arXiv:2007.13716.
[35] CHATTERJEE, S. (2010). Spin glasses and Stein's method. *Probab. Theory Related Fields* **148** 567–600. MR2678899 https://doi.org/10.1007/s00440-009-0240-8
[36] CHATTERJEE, S. and DEMBO, A. (2016). Nonlinear large deviations. *Adv. Math.* **299** 396–450. MR3519474 https://doi.org/10.1016/j.aim.2016.05.017
[37] CHEN, W.-K. and LAM, W.-K. (2021). Universality of approximate message passing algorithms. *Electron. J. Probab.* **26** Paper No. 36, 44 pp. MR4235487 https://doi.org/10.1214/21-EJP604
[38] CHEN, W.-K. and PANCHENKO, D. (2018). On the TAP free energy in the mixed p-spin models. *Comm. Math. Phys.* **362** 219–252. MR3833609 https://doi.org/10.1007/s00220-018-3143-7
[39] CHEN, W.-K., PANCHENKO, D. and SUBAG, E. (2018). Generalized TAP free energy. *Comm. Pure Appl. Math.*
[40] CHÉRIEF-ABDELLATIF, B.-E. (2019). Consistency of ELBO maximization for model selection. In *Symposium on Advances in Approximate Bayesian Inference*. *Proc. Mach. Learn. Res.* (*PMLR*) **96** 11–31. PMLR. MR3980812
[41] CRISANTI, A., LEUZZI, L., PARISI, G. and RIZZO, T. (2003). Complexity in the Sherrington–Kirkpatrick model in the annealed approximation. *Phys. Rev. B* **68** 174401.
[42] CRISANTI, A., LEUZZI, L. and RIZZO, T. (2005). Complexity in mean-field spin-glass models: Ising p-spin. *Phys. Rev. B* **71** 094202.
[43] DE DOMINICIS, C. and YOUNG, A. P. (1983). Weighted averages and order parameters for the infinite range Ising spin glass. *J. Phys. A* **16** 2063–2075. MR0712998
[44] 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
[45] DESHPANDE, Y. and MONTANARI, A. (2014). Information-theoretically optimal sparse PCA. In 2014 *IEEE International Symposium on Information Theory* 2197–2201. IEEE, New York.
[46] DING, J. and SUN, N. (2019). Capacity lower bound for the Ising perceptron. In *STOC'*19*—Proceedings of the* 51*st Annual ACM SIGACT Symposium on Theory of Computing* 816–827. ACM, New York. MR4003386 https://doi.org/10.1145/3313276.3316383
[47] DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2009). Message-passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci.* **106** 18914–18919.
[48] DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2010). Message passing algorithms for compressed sensing: I. motivation and construction. In 2010 *IEEE Information Theory Workshop on Information Theory* (*ITW* 2010, *Cairo*) 1–5. IEEE, New York.
[49] DRAGOMIR, R. A., EVEN, M. and HENDRIKX, H. (2021). Fast stochastic Bregman gradient methods: Sharp analysis and variance reduction. In *International Conference on Machine Learning* 2815–2825. PMLR.
[50] DUDEJA, R., SEN, S. and LU, Y. M. (2022). Spectral universality of regularized linear regression with nearly deterministic sensing matrices. Preprint. Available at arXiv:2208.02753.
[51] EL ALAOUI, A., MONTANARI, A. and SELLKE, M. (2022). Sampling from the Sherrington–Kirkpatrick Gibbs measure via algorithmic stochastic localization. In 2022 *IEEE* 63*rd Annual Symposium on Foundations of Computer Science—FOCS* 2022 323–334. IEEE Computer Soc., Los Alamitos, CA. MR4537214
[52] ELDAN, R. (2018). Gaussian-width gradient complexity, reverse log-Sobolev inequalities and nonlinear large deviations. *Geom. Funct. Anal.* **28** 1548–1596. MR3881829 https://doi.org/10.1007/s00039-018-0461-z
[53] FAN, Z., MEI, S. and MONTANARI, A. (2021). TAP free energy, spin glasses and variational inference. *Ann. Probab.* **49** 1–45. MR4203332 https://doi.org/10.1214/20-AOP1443
[54] FAN, Z. and WU, Y. (2021). The replica-symmetric free energy for Ising spin glasses with orthogonally invariant couplings. Preprint. Available at arXiv:2105.02797.
[55] FENG, O. Y., VENKATARAMANAN, R., RUSH, C., SAMWORTH, R. J. et al. (2022). A unifying tutorial on approximate message passing. *Found. Trends Mach. Learn.* **15** 335–536.
[56] FERNIQUE, X. (1975). Regularité des trajectoires des fonctions aléatoires gaussiennes. In *École D'Été de Probabilités de Saint-Flour, IV*-1974. *Lecture Notes in Math.* **480** 1–96. Springer, Berlin. MR0413238
[57] FYODOROV, Y. V. (2004). Complexity of random energy landscapes, glass transition, and absolute value of the spectral determinant of random matrices. *Phys. Rev. Lett.* **92** 240601, 4 pp. MR2115095 https://doi.org/10.1103/PhysRevLett.92.240601

[58] GAUCHER, S. and KLOPP, O. (2021). Optimality of variational inference for stochasticblock model with missing links. *Adv. Neural Inf. Process. Syst.* **34** 19947–19959.
[59] GHORBANI, B., JAVADI, H. and MONTANARI, A. (2019). An instability in variational inference for topic models. In *International Conference on Machine Learning* 2221–2231. PMLR.
[60] GORDON, Y. (1985). Some inequalities for Gaussian processes and applications. *Israel J. Math.* **50** 265–289. MR0800188 https://doi.org/10.1007/BF02759761
[61] GUTMAN, D. H. and PEÑA, J. F. (2023). Perturbed Fenchel duality and first-order methods. *Math. Program.* **198** 443–469. MR4550955 https://doi.org/10.1007/s10107-022-01779-7
[62] HALL, P., ORMEROD, J. T. and WAND, M. P. (2011). Theory of Gaussian variational approximation for a Poisson mixed model. *Statist. Sinica* **21** 369–389. MR2796867
[63] HALL, P., PHAM, T., WAND, M. P. and WANG, S. S. J. (2011). Asymptotic normality and valid inference for Gaussian variational approximation. *Ann. Statist.* **39** 2502–2532. MR2906876 https://doi.org/10.1214/11-AOS908
[64] HAN, Q. and SHEN, Y. (2022). Universality of regularized regression estimators in high dimensions. Preprint. Available at arXiv:2206.07936.
[65] HANZELY, F., RICHTÁRIK, P. and XIAO, L. (2021). Accelerated Bregman proximal gradient methods for relatively smooth convex optimization. *Comput. Optim. Appl.* **79** 405–440. MR4254648 https://doi.org/10.1007/s10589-021-00273-8
[66] HU, H. and LU, Y. M. (2020). Universality laws for high-dimensional learning with random features. Preprint. Available at arXiv:2009.07669.
[67] JAIN, V., KOEHLER, F. and MOSSEL, E. (2018). The mean-field approximation: Information inequalities, algorithms, and complexity. In *Conference on Learning Theory* 1326–1347. PMLR.
[68] JAVANMARD, A. and MONTANARI, A. (2013). State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Inf. Inference* **2** 115–144. MR3311445 https://doi.org/10.1093/imaiai/iat004
[69] 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
[70] 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
[71] JORDAN, M. I., GHAHRAMANI, Z., JAAKKOLA, T. S. and SAUL, L. K. (1999). An introduction to variational methods for graphical models. *Mach. Learn.* **37** 183–233.
[72] KABASHIMA, Y. (2003). A CDMA multiuser detection algorithm on the basis of belief propagation. *J. Phys. A* **36** 11111–11121. MR2025247 https://doi.org/10.1088/0305-4470/36/43/030
[73] KAHANE, J.-P. (1986). Une inégalité du type de Slepian et Gordon sur les processus gaussiens. *Israel J. Math.* **55** 109–110. MR0858463 https://doi.org/10.1007/BF02772698
[74] KRZAKALA, F., MANOEL, A., TRAMEL, E. W. and ZDEBOROVÁ, L. (2014). Variational free energies for compressed sensing. In 2014 *IEEE International Symposium on Information Theory* 1499–1503. IEEE, New York.
[75] KRZAKALA, F., XU, J. and ZDEBOROVÁ, L. (2016). Mutual information in rank-one matrix estimation. In 2016 *IEEE Information Theory Workshop* (*ITW*) 71–75. IEEE, New York.
[76] 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
[77] LESIEUR, T., KRZAKALA, F. and ZDEBOROVÁ, L. (2015). Phase transitions in sparse PCA. In 2015 *IEEE International Symposium on Information Theory* (*ISIT*) 1635–1639. IEEE, New York.
[78] LI, G. and WEI, Y. (2022). A non-asymptotic framework for approximate message passing in spiked models. Preprint. Available at arXiv:2208.03313.
[79] LIANG, P., PETROV, S., JORDAN, M. I. and KLEIN, D. (2007). The infinite PCFG using hierarchical Dirichlet processes. In *Proceedings of the* 2007 *Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (*EMNLP-CoNLL*) 688–697.
[80] LU, H., FREUND, R. M. and NESTEROV, Y. (2018). Relatively smooth convex optimization by first-order methods, and applications. *SIAM J. Optim.* **28** 333–354. MR3759881 https://doi.org/10.1137/16M1099546
[81] MAILLARD, A., AROUS, G. B. and BIROLI, G. (2020). Landscape complexity for the empirical risk of generalized linear models. In *Mathematical and Scientific Machine Learning* 287–327. PMLR.
[82] MEI, S., BAI, Y. and MONTANARI, A. (2018). The landscape of empirical risk for nonconvex losses. *Ann. Statist.* **46** 2747–2774. MR3851754 https://doi.org/10.1214/17-AOS1637
[83] MINKA, T., WINN, J., GUIVER, J., WEBSTER, S., ZAYKOV, Y., YANGEL, B., SPENGLER, A. and BRONSKILL, J. (2014). Infer NET 2.6. Microsoft Research Cambridge. Available at http://research.microsoft.com/infernet.

[84] MINKA, T. P. (2001). A family of algorithms for approximate Bayesian inference, PhD thesis, Massachusetts Institute of Technology. MR2717007
[85] MIOLANE, L. and MONTANARI, A. (2021). The distribution of the Lasso: Uniform control over sparse balls and adaptive parameter tuning. *Ann. Statist.* **49** 2313–2335. MR4319252 https://doi.org/10.1214/20-aos2038
[86] MONTANARI, A. (2012). Graphical models concepts in compressed sensing. In *Compressed Sensing* 394–438. Cambridge Univ. Press, Cambridge. MR2963574 https://doi.org/10.1017/CBO9780511794308.010
[87] MONTANARI, A. and NGUYEN, P.-M. (2017). Universality of the elastic net error. In 2017 *IEEE International Symposium on Information Theory* (*ISIT*) 2338–2342. IEEE, New York.
[88] MONTANARI, A. and RICHARD, E. (2016). Non-negative principal component analysis: Message passing algorithms and sharp asymptotics. *IEEE Trans. Inf. Theory* **62** 1458–1484. MR3472260 https://doi.org/10.1109/TIT.2015.2457942
[89] MONTANARI, A. and SAEED, B. (2022). Universality of empirical risk minimization. Preprint. Available at arXiv:2202.08832.
[90] 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 https://doi.org/10.1145/2897518.2897548
[91] MONTANARI, A. and VENKATARAMANAN, R. (2021). Estimation of low-rank matrices via approximate message passing. *Ann. Statist.* **49** 321–345. MR4206680 https://doi.org/10.1214/20-AOS1958
[92] MUKHERJEE, S. S., SARKAR, P., WANG, Y. and YAN, B. (2018). Mean field for the stochastic blockmodel: Optimization landscape and convergence issues. *Adv. Neural Inf. Process. Syst.* **31**.
[93] NEMIROVSKY, A. S. and YUDIN, D. B. (1983). *Problem Complexity and Method Efficiency in Optimization*. *Wiley-Interscience Series in Discrete Mathematics*. Wiley, New York. MR0702836
[94] OYMAK, S., THRAMPOULIDIS, C. and HASSIBI, B. (2013). The squared-error of generalized lasso: A precise analysis. In 2013 51*st Annual Allerton Conference on Communication, Control, and Computing* (*Allerton*) 1002–1009. IEEE, New York.
[95] PEARL, J. (1982). Reverend Bayes on inference engines: A distributed hierarchical approach. In *Proceedings of the Second AAAI Conference on Artificial Intelligence* 133–136.
[96] PÉCHÉ, S. (2006). The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probab. Theory Related Fields* **134** 127–173. MR2221787 https://doi.org/10.1007/s00440-005-0466-z
[97] PLEFKA, T. (1982). Convergence condition of the TAP equation for the infinite-ranged Ising spin glass model. *J. Phys. A* **15** 1971–1978. MR0663708
[98] PLUMMER, S., PATI, D. and BHATTACHARYA, A. (2020). Dynamics of coordinate ascent variational inference: A case study in 2D Ising models. *Entropy* **22** Paper No. 1263, 33 pp. MR4222066 https://doi.org/10.3390/e22111263
[99] QIU, J. and SEN, S. (2022). The TAP free energy for high-dimensional linear regression. Preprint. Available at arXiv:2203.07539.
[100] RAJ, A., STEPHENS, M. and PRITCHARD, J. K. (2014). fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics* **197** 573–589. https://doi.org/10.1534/genetics.114.164350
[101] RANGAN, S. (2011). Generalized approximate message passing for estimation with random linear mixing. In 2011 *IEEE International Symposium on Information Theory Proceedings* 2168–2172. IEEE, New York.
[102] RANGAN, S. and FLETCHER, A. K. (2012). Iterative estimation of constrained rank-one matrices in noise. In 2012 *IEEE International Symposium on Information Theory Proceedings* 1246–1250. IEEE, New York.
[103] RANGAN, S., FLETCHER, A. K., SCHNITER, P. and KAMILOV, U. S. (2017). Inference for generalized linear models via alternating directions and Bethe free energy minimization. *IEEE Trans. Inf. Theory* **63** 676–697. MR3599966 https://doi.org/10.1109/TIT.2016.2619373
[104] RANGAN, S., SCHNITER, P. and FLETCHER, A. K. (2019). Vector approximate message passing. *IEEE Trans. Inf. Theory* **65** 6664–6684. MR4009222 https://doi.org/10.1109/TIT.2019.2916359
[105] RAY, K. and SZABÓ, B. (2022). Variational Bayes for high-dimensional linear regression with sparse priors. *J. Amer. Statist. Assoc.* **117** 1270–1281. MR4480711 https://doi.org/10.1080/01621459.2020.1847121
[106] RUSH, C. and VENKATARAMANAN, R. (2016). Finite-sample analysis of approximate message passing. In 2016 *IEEE International Symposium on Information Theory* (*ISIT*) 755–759. https://doi.org/10.1109/ISIT.2016.7541400

[107] RUSH, C. and VENKATARAMANAN, R. (2018). Finite sample analysis of approximate message passing algorithms. *IEEE Trans. Inf. Theory* **64** 7264–7286. MR3876443 https://doi.org/10.1109/TIT.2018.2816681

[108] SAADE, A., KRZAKALA, F. and ZDEBOROVÁ, L. (2014). Spectral clustering of graphs with the Bethe Hessian. In *Neural Information Processing Systems*.

[109] 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

[110] SLEPIAN, D. (1962). The one-sided barrier problem for Gaussian noise. *Bell Syst. Tech. J.* **41** 463–501. MR0133183 https://doi.org/10.1002/j.1538-7305.1962.tb02419.x

[111] STOJNIC, M. (2013). A framework to characterize performance of lasso algorithms. Preprint. Available at arXiv:1303.7291.

[112] SUBAG, E. (2017). The complexity of spherical p-spin models—A second moment approach. *Ann. Probab.* **45** 3385–3450. MR3706746 https://doi.org/10.1214/16-AOP1139

[113] SUBAG, E. (2021). The free energy of spherical pure p-spin models—Computation from the TAP approach. Preprint. Available at arXiv:2101.04352.

[114] SUDAKOV, V. N. (1971). Gaussian random processes, and measures of solid angles in Hilbert space. *Dokl. Akad. Nauk SSSR* **197** 43–45. MR0288832

[115] SUDAKOV, V. N. (1979). Geometric problems in the theory of infinite-dimensional probability distributions. *Proc. Steklov Inst. Math.* **2** i–v, 1–178. Cover to cover translation of Trudy Mat. Inst. Steklov **141** (1976). MR0530375

[116] SUN, J., QU, Q. and WRIGHT, J. (2018). A geometric analysis of phase retrieval. *Found. Comput. Math.* **18** 1131–1198. MR3857907 https://doi.org/10.1007/s10208-017-9365-9

[117] TALAGRAND, M. (2011). *Mean Field Models for Spin Glasses. Volume I: Basic Examples. Ergebnisse der Mathematik und Ihrer Grenzgebiete*. 3. *Folge. A Series of Modern Surveys in Mathematics* [*Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics*] **54**. Springer, Berlin. MR2731561 https://doi.org/10.1007/978-3-642-15202-3

[118] THOULESS, D. J., ANDERSON, P. W. and PALMER, R. G. (1977). Solution of 'solvable model of a spin glass'. *Philos. Mag.* **35** 593–601.

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

[120] TRAN, D., KUCUKELBIR, A., DIENG, A. B., RUDOLPH, M., LIANG, D. and BLEI, D. M. (2016). Edward: A library for probabilistic modeling, inference, and criticism. Preprint. Available at arXiv:1610.09787.

[121] WAINWRIGHT, M. J. and JORDAN, M. I. (2008). *Graphical Models, Exponential Families, and Variational Inference*. Now Publishers, Hanover.

[122] WANG, T., ZHONG, X. and FAN, Z. (2022). Universality of approximate message passing algorithms and tensor networks. Preprint. Available at arXiv:2206.13037.

[123] WANG, Y. and BLEI, D. M. (2019). Frequentist consistency of variational Bayes. *J. Amer. Statist. Assoc.* **114** 1147–1161. MR4011769 https://doi.org/10.1080/01621459.2018.1473776

[124] YAN, J. (2020). Nonlinear large deviations: Beyond the hypercube. *Ann. Appl. Probab.* **30** 812–846. MR4108123 https://doi.org/10.1214/19-AAP1516

[125] YANG, Y., PATI, D. and BHATTACHARYA, A. (2020). α-variational inference with statistical guarantees. *Ann. Statist.* **48** 886–905. MR4102680 https://doi.org/10.1214/19-AOS1827

[126] YEDIDIA, J. S., FREEMAN, W. T. and WEISS, Y. (2003). Understanding belief propagation and its generalizations. In *Exploring Artificial Intelligence in the New Millennium* **8** 236–239.

[127] ZHANG, A. Y. and ZHOU, H. H. (2020). Theoretical and computational guarantees of mean field variational inference for community detection. *Ann. Statist.* **48** 2575–2598. MR4152113 https://doi.org/10.1214/19-AOS1898

[128] ZHANG, F. and GAO, C. (2020). Convergence rates of variational posterior distributions. *Ann. Statist.* **48** 2180–2207. MR4134791 https://doi.org/10.1214/19-AOS1883

The Annals of Statistics
2023, Vol. 51, No. 2, 547–572
https://doi.org/10.1214/23-AOS2259
© Institute of Mathematical Statistics, 2023

CANONICAL NOISE DISTRIBUTIONS AND PRIVATE HYPOTHESIS TESTS

BY JORDAN AWAN[1,a] AND SALIL VADHAN[2,b]

[1]*Department of Statistics, Purdue University,* [a]*jawan@purdue.edu*
[2]*Harvard John A. Paulson School of Engineering and Applied Sciences,* [b]*salil_vadhan@harvard.edu*

f-DP has recently been proposed as a generalization of differential privacy allowing a lossless analysis of composition, post-processing, and privacy amplification via subsampling. In the setting of f-DP, we propose the concept of a *canonical noise distribution* (CND), the first mechanism designed for an arbitrary f-DP guarantee. The notion of CND captures whether an additive privacy mechanism perfectly matches the privacy guarantee of a given f. We prove that a CND always exists, and give a construction that produces a CND for any f. We show that private hypothesis tests are intimately related to CNDs, allowing for the release of private p-values at no additional privacy cost, as well as the construction of uniformly most powerful (UMP) tests for binary data, within the general f-DP framework.

We apply our techniques to the problem of difference-of-proportions testing, and construct a UMP unbiased (UMPU) "semiprivate" test which upper bounds the performance of any f-DP test. Using this as a benchmark, we propose a private test based on the inversion of characteristic functions, which allows for optimal inference on the two population parameters and is nearly as powerful as the semiprivate UMPU. When specialized to the case of $(\epsilon, 0)$-DP, we show empirically that our proposed test is more powerful than any $(\epsilon/\sqrt{2})$-DP test and has more accurate type I errors than the classic normal approximation test.

REFERENCES

ABADI, M., CHU, A., GOODFELLOW, I., MCMAHAN, H. B., MIRONOV, I., TALWAR, K. and ZHANG, L. (2016). Deep learning with differential privacy. In *Proceedings of the* 2016 *ACM SIGSAC Conference on Computer and Communications Security* 308–318.

ACHARYA, J., SUN, Z. and ZHANG, H. (2018). Differentially private testing of identity and closeness of discrete distributions. *Adv. Neural Inf. Process. Syst.* **31**.

AKRITAS, M. (2015). *Probability and Statistics with R*. Pearson, New York.

ALIAKBARPOUR, M., DIAKONIKOLAS, I. and RUBINFELD, R. (2018). Differentially private identity and equivalence testing of discrete distributions. In *International Conference on Machine Learning* 169–178. PMLR, Stockholm, Sweden.

AWAN, J. and CAI, Z. (2020). One step to efficient synthetic data. arXiv preprint. Available at arXiv:2006.02397.

AWAN, J. and DONG, J. (2022). Log-concave and multivariate canonical noise distributions for differential privacy. *Adv. Neural Inf. Process. Syst.* **35**.

AWAN, J. and SLAVKOVIĆ, A. (2018). Differentially private uniformly most powerful tests for binomial data. *Adv. Neural Inf. Process. Syst.* **31** 4208–4218.

AWAN, J. and SLAVKOVIĆ, A. (2020). Differentially private inference for binomial data. *J. Priv. Confid.* **10**.

AWAN, J. and SLAVKOVIĆ, A. (2021). Structure and sensitivity in differential privacy: Comparing K-norm mechanisms. *J. Amer. Statist. Assoc.* **116** 935–954. MR4270035 https://doi.org/10.1080/01621459.2020.1773831

AWAN, J. and VADHAN, S. (2023). Supplement to "Canonical noise distributions and private hypothesis tests." https://doi.org/10.1214/23-AOS2259SUPPA, https://doi.org/10.1214/23-AOS2259SUPPB

BARRIENTOS, A. F., REITER, J. P., MACHANAVAJJHALA, A. and CHEN, Y. (2019). Differentially private significance tests for regression coefficients. *J. Comput. Graph. Statist.* **28** 440–453. MR3974892 https://doi.org/10.1080/10618600.2018.1538881

BLACKWELL, D. (1950). Comparison of experiments. Technical report, Howard Univ., Washington, United States.

MSC2020 subject classifications. Primary 68P27; secondary 62F03.
Key words and phrases. Differential privacy, uniformly most powerful test, frequentist inference.

BUN, M. and STEINKE, T. (2016). Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography. Part I. Lecture Notes in Computer Science* **9985** 635–658. Springer, Berlin. MR3591832 https://doi.org/10.1007/978-3-662-53641-4_24

BUN, M., DWORK, C., ROTHBLUM, G. N. and STEINKE, T. (2018). Composable and versatile privacy via truncated CDP. In *STOC'*18*—Proceedings of the* 50*th Annual ACM SIGACT Symposium on Theory of Computing* 74–86. ACM, New York. MR3826235

BUN, M., KAMATH, G., STEINKE, T. and WU, S. Z. (2019). Private hypothesis selection. *Adv. Neural Inf. Process. Syst.* **32**.

CAI, B., DASKALAKIS, C. and KAMATH, G. (2017). Priv'it: Private and sample efficient identity testing. In *International Conference on Machine Learning* 635–644. PMLR, Sydney, Australia.

CANONNE, C. L., KAMATH, G., MCMILLAN, A., SMITH, A. and ULLMAN, J. (2019). The structure of optimal private tests for simple hypotheses. In *STOC'*19*—Proceedings of the* 51*st Annual ACM SIGACT Symposium on Theory of Computing* 310–321. ACM, New York. MR4003341

CASELLA, G. and BERGER, R. L. (2002). *Statistical Inference*. Duxbury, N. Scituate.

CHAUDHURI, K., MONTELEONI, C. and SARWATE, A. D. (2011). Differentially private empirical risk minimization. *J. Mach. Learn. Res.* **12** 1069–1109. MR2786918

COLQUHOUN, D. (2017). The reproducibility of research and the misinterpretation of p-values. *R. Soc. Open Sci.* **4** 171085. MR3744762 https://doi.org/10.1098/rsos.171085

DANG, X., KEETON, S. L. and PENG, H. (2009). A unified approach for analyzing exchangeable binary data with applications to developmental toxicity studies. *Stat. Med.* **28** 2580–2604. MR2750310 https://doi.org/10.1002/sim.3638

DONG, J., ROTH, A. and SU, W. J. (2022). Gaussian differential privacy. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 3–37. MR4400389

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

DWORK, C. and LEI, J. (2009). Differential privacy and robust statistics. In *STOC'*09*—Proceedings of the* 2009 *ACM International Symposium on Theory of Computing* 371–380. ACM, New York. MR2780083

DWORK, C. and ROTH, A. (2014). The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.* **9** 211–407. MR3254020 https://doi.org/10.1561/0400000042

DWORK, C., MCSHERRY, F., NISSIM, K. and SMITH, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography. Lecture Notes in Computer Science* **3876** 265–284. Springer, Berlin. MR2241676 https://doi.org/10.1007/11681878_14

DWORK, C., NAOR, M., REINGOLD, O., ROTHBLUM, G. N. and VADHAN, S. (2009). On the complexity of differentially private data release. In *STOC'*09*—Proceedings of the* 2009 *ACM International Symposium on Theory of Computing* 381–390. ACM, New York. MR2780084

FOG, A. (2008). Sampling methods for Wallenius' and Fisher's noncentral hypergeometric distributions. *Comm. Statist. Simulation Comput.* **37** 241–257. MR2422884 https://doi.org/10.1080/03610910701790236

GABOARDI, M., LIM, H., ROGERS, R. and VADHAN, S. (2016). Differentially private chi-squared hypothesis testing: Goodness of fit and independence testing. In *Proceedings of the* 33*rd International Conference on Machine Learning* (M. F. Balcan and K. Q. Weinberger, eds.). *Proceedings of Machine Learning Research* **48** 2111–2120. PMLR, New York.

GABOARDI, M. and ROGERS, R. (2018). Local private hypothesis testing: Chi-square tests. In *Proceedings of the* 35*th International Conference on Machine Learning* (J. Dy and A. Krause, eds.). *Proceedings of Machine Learning Research* **80** 1626–1635. PMLR, Stockholm, Sweden.

GAO, J., GONG, R. and YU, F.-Y. (2022). Subspace differential privacy. In *Proceedings of the AAAI Conference on Artificial Intelligence* **36** 3986–3995.

GENG, Q. and VISWANATH, P. (2015). The optimal noise-adding mechanism in differential privacy. *IEEE Trans. Inf. Theory* **62** 925–951. MR3455907 https://doi.org/10.1109/TIT.2015.2504967

GHOSH, A., ROUGHGARDEN, T. and SUNDARARAJAN, M. (2012). Universally utility-maximizing privacy mechanisms. *SIAM J. Comput.* **41** 1673–1693. MR3029267 https://doi.org/10.1137/09076828X

HALL, R., RINALDO, A. and WASSERMAN, L. (2013). Differential privacy for functions and functional data. *J. Mach. Learn. Res.* **14** 703–727. MR3033345

HARDT, M. and TALWAR, K. (2010). On the geometry of differential privacy. In *STOC'*10*—Proceedings of the* 2010 *ACM International Symposium on Theory of Computing* 705–714. ACM, New York. MR2743320

HARKNESS, W. L. (1965). Properties of the extended hypergeometric distribution. *Ann. Math. Stat.* **36** 938–945. MR0182073 https://doi.org/10.1214/aoms/1177700066

KAIROUZ, P., OH, S. and VISWANATH, P. (2017). The composition theorem for differential privacy. *IEEE Trans. Inf. Theory* **63** 4037–4049. MR3677761 https://doi.org/10.1109/TIT.2017.2685505

KAKIZAKI, K., FUKUCHI, K. and SAKUMA, J. (2017). Differentially private chi-squared test by unit circle mechanism. In *International Conference on Machine Learning* 1761–1770. PMLR, Sydney, Australia.

KARWA, V. and VADHAN, S. P. (2017). Finite sample differentially private confidence intervals. Available at arXiv:1711.03908.

KARWA, V. and VADHAN, S. (2018). Private correspondence.

KIFER, D. and ROGERS, R. (2016). A new class of private chi-square tests. In *Proceedings of the* 20*th International Conference on Artificial Intelligence and Statistics*. *AISTATS* **17** 991–1000.

MIRONOV, I. (2017). Rényi differential privacy. In 2017 *IEEE* 30*th Computer Security Foundations Symposium* (*CSF*) 263–275. IEEE Press, New York.

MIRSHANI, A., REIMHERR, M. and SLAVKOVIĆ, A. (2019). Formal privacy for functional data with Gaussian perturbations. In *International Conference on Machine Learning* 4595–4604. PMLR, Long Beach, CA.

REIMHERR, M. and AWAN, J. (2019). Elliptical perturbations for differential privacy. *Adv. Neural Inf. Process. Syst*. **32**.

ROBINS, J. M., VAN DER VAART, A. and VENTURA, V. (2000). Asymptotic distribution of p values in composite null models. *J. Amer. Statist. Assoc*. **95** 1143–1156. MR1804240 https://doi.org/10.2307/2669750

SCHERVISH, M. J. (2012). *Theory of Statistics*. Springer, Berlin.

SHEFFET, O. (2017). Differentially private ordinary least squares. In *Proceedings of the* 34*th International Conference on Machine Learning* (D. Precup and Y. W. Teh, eds.). *Proceedings of Machine Learning Research* **70** 3105–3114. PMLR, Sydney, Australia.

SOLEA, E. (2014). Differentially private hypothesis testing for normal random variables. Master's thesis, The Pennsylvania State Univ.

SURESH, A. T. (2021). Robust hypothesis testing and distribution estimation in Hellinger distance. In *International Conference on Artificial Intelligence and Statistics* 2962–2970. PMLR, San Diego, CA.

UHLER, C., SLAVKOVIĆ, A. and FIENBERG, S. (2013). Privacy-preserving data sharing for genome-wide association studies. *J. Priv. Confid*. **5**.

VU, D. and SLAVKOVIĆ, A. (2009). Differential privacy for clinical trial data: Preliminary evaluations. In 2009 *IEEE International Conference on Data Mining Workshops* 138–143. IEEE Press, New York.

WANG, Y., KIFER, D., LEE, J. and KARWA, V. (2018). Statistical approximating distributions under differential privacy. *J. Priv. Confid*. **8**.

WANG, Y., LEE, J. and KIFER, D. (2015). Revisiting differentially private hypothesis tests for categorical data. Available at arXiv:1511.03376.

WASSERMAN, L. and ZHOU, S. (2010). A statistical framework for differential privacy. *J. Amer. Statist. Assoc*. **105** 375–389. MR2656057 https://doi.org/10.1198/jasa.2009.tm08651

WASSERSTEIN, R. L. and LAZAR, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose [Editorial]. *Amer. Statist*. **70** 129–133. MR3511040 https://doi.org/10.1080/00031305.2016.1154108

ZHANG, J., ZHANG, Z., XIAO, X., YANG, Y. and WINSLETT, M. (2012). Functional mechanism: Regression analysis under differential privacy. *Proc. VLDB Endow.* **5**.

The Annals of Statistics
2023, Vol. 51, No. 2, 573–598
https://doi.org/10.1214/23-AOS2260
© Institute of Mathematical Statistics, 2023

ON THE DISJOINT AND SLIDING BLOCK MAXIMA METHOD FOR PIECEWISE STATIONARY TIME SERIES

BY AXEL BÜCHER[a] AND LEANDRA ZANGER[b]

Mathematical Institute, Heinrich-Heine-University Düsseldorf, [a]*axel.buecher@hhu.de,* [b]*leandra.zanger@hhu.de*

Modeling univariate block maxima by the generalized extreme value distribution constitutes one of the most widely applied approaches in extreme value statistics. It has recently been found that, for an underlying stationary time series, respective estimators may be improved by calculating block maxima in an overlapping way. A proof of concept is provided that the latter finding also holds in situations that involve certain piecewise stationarities. A weak convergence result for an empirical process of central interest is provided, and further details are examplarily worked out for the probability weighted moment estimator. Irrespective of the serial dependence, the asymptotic estimation variance is shown to be smaller for the new estimator. In extensive simulation experiments, the finite-sample variance was typically found to be smaller as well, while the bias stays approximately the same. The results are illustrated by Monte Carlo simulation experiments and are applied to a common situation involving temperature extremes in a changing climate.

REFERENCES

[1] BEIRLANT, J., GOEGEBEUR, Y., TEUGELS, J. and SEGERS, J. (2004). *Statistics of Extremes*: *Theory and Applications*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. With contributions from Daniel De Waal and Chris Ferro. MR2108013 https://doi.org/10.1002/0470012382

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

[3] BRADLEY, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** 107–144. Update of, and a supplement to, the 1986 original. MR2178042 https://doi.org/10.1214/154957805100000104

[4] BROCKWELL, P. J. and DAVIS, R. A. (1987). *Time Series*: *Theory and Methods*. *Springer Series in Statistics*. Springer, New York. MR0868859 https://doi.org/10.1007/978-1-4899-0004-3

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

[6] BÜCHER, A. and SEGERS, J. (2017). On the maximum likelihood estimator for the generalized extreme-value distribution. *Extremes* **20** 839–872. MR3737387 https://doi.org/10.1007/s10687-017-0292-6

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

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

[9] BÜCHER, A. and ZANGER, L. (2023). Supplement to "On the disjoint and sliding block maxima method for piecewise stationary time series." https://doi.org/10.1214/23-AOS2260SUPP

[10] CISSOKHO, Y. and KULIK, R. (2021). Estimation of cluster functionals for regularly varying time series: Sliding blocks estimators. *Electron. J. Stat.* **15** 2777–2831. MR4280158 https://doi.org/10.1214/21-ejs1843

[11] COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. *Springer Series in Statistics*. Springer London, Ltd., London. MR1932132 https://doi.org/10.1007/978-1-4471-3675-0

MSC2020 subject classifications. Primary 62G32, 62F12; secondary 62P12, 60G70.

Key words and phrases. Asymptotic normality, extreme value statistics, Marshall–Olkin distribution, return level, temperature extremes.

[12] COOLEY, D. (2013). Return periods and return levels under climate change. In *Extremes in a Changing Climate*. *Water Sci. Technol. Libr.* **65** 97–114. Springer, Dordrecht. MR3051154 https://doi.org/10.1007/978-94-007-4479-0_4

[13] DAVISON, A. C. and HINKLEY, D. V. (1997). *Bootstrap Methods and Their Application*. *Cambridge Series in Statistical and Probabilistic Mathematics* **1**. Cambridge Univ. Press, Cambridge. With 1 IBM-PC floppy disk (3.5 inch; HD). MR1478673 https://doi.org/10.1017/CBO9780511802843

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

[15] DEHLING, H. and PHILIPP, W. (2002). Empirical process techniques for dependent data. In *Empirical Process Techniques for Dependent Data* 3–113. Birkhäuser, Boston, MA. MR1958777 https://doi.org/10.1007/978-1-4612-0099-4_1

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

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

[18] DREES, H. and NEBLUNG, S. (2021). Asymptotics for sliding blocks estimators of rare events. *Bernoulli* **27** 1239–1269. MR4255233 https://doi.org/10.3150/20-bej1272

[19] EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (1997). *Modelling Extremal Events*: *For Insurance and Finance*. *Applications of Mathematics* (*New York*) **33**. Springer, Berlin. MR1458613 https://doi.org/10.1007/978-3-642-33483-2

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

[21] FISHER, R. A. and TIPPETT, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Math. Proc. Cambridge Philos. Soc.* **24** 180–190. https://doi.org/10.1017/S0305004100015681

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

[23] HOSKING, J. R. M. and WALLIS, J. R. (1997). *Regional Frequency Analysis*: *An Approach Based on L-Moments*. Cambridge Univ. Press, Cambridge.

[24] HOSKING, J. R. M., WALLIS, J. R. and WOOD, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* **27** 251–261. MR0797563 https://doi.org/10.2307/1269706

[25] KATZ, R. W., PARLANGE, M. B. and NAVEAU, P. (2002). Statistics of extremes in hydrology. *Adv. Water Resour.* **25** 1287–1304.

[26] KOJADINOVIC, I. and NAVEAU, P. (2017). Detecting distributional changes in samples of independent block maxima using probability weighted moments. *Extremes* **20** 417–450. MR3638372 https://doi.org/10.1007/s10687-016-0273-1

[27] KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. *Springer Series in Statistics*. Springer, New York. MR2724368 https://doi.org/10.1007/978-0-387-74978-5

[28] LANDWEHR, J. M., MATALAS, N. C. and WALLIS, J. R. (1979). Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resour. Res.* **15** 1055–1064. https://doi.org/10.1029/WR015i005p01055

[29] LEADBETTER, M. R. (1983). Extremes and local dependence in stationary sequences. *Z. Wahrsch. Verw. Gebiete* **65** 291–306. MR0722133 https://doi.org/10.1007/BF00532484

[30] OORSCHOT, J. and ZHOU, C. (2020). All block maxima method for estimating the extreme value index. arXiv preprints arXiv:2010.15950.

[31] PHILIP, S., KEW, S., VAN OLDENBORGH, G. J., OTTO, F., VAUTARD, R., VAN DER WIEL, K., KING, A., LOTT, F., ARRIGHI, J. et al. (2020). A protocol for probabilistic extreme event attribution analyses. *Adv. Stat. Climatol. Meteorol. Oceanogr.* **6** 177–203.

[32] PRESCOTT, P. and WALDEN, A. T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika* **67** 723–724. MR0601119 https://doi.org/10.1093/biomet/67.3.723

[33] SEN, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *J. Amer. Statist. Assoc.* **63** 1379–1389. MR0258201

[34] STEIN, M. L. (2017). Should annual maximum temperatures follow a generalized extreme value distribution? *Biometrika* **104** 1–16. MR3626483 https://doi.org/10.1093/biomet/asw070

[35] STOTT, P. A., CHRISTIDIS, N., OTTO, F. E. L., SUN, Y., VANDERLINDEN, J.-P., VAN OLDENBORGH, G. J., VAUTARD, R., VON STORCH, H., WALTON, P. et al. (2016). Attribution of extreme weather and climate-related events. *WIREs Climate Change* **7** 23–41.

[36] ZOU, N., VOLGUSHEV, S. and BÜCHER, A. (2021). Multiple block sizes and overlapping blocks for multivariate time series extremes. *Ann. Statist.* **49** 295–320. MR4206679 https://doi.org/10.1214/20-AOS1957

The Annals of Statistics
2023, Vol. 51, No. 2, 599–623
https://doi.org/10.1214/23-AOS2262
© Institute of Mathematical Statistics, 2023

LEARNING SPARSE GRAPHONS AND THE GENERALIZED KESTEN–STIGUM THRESHOLD

BY EMMANUEL ABBE[1,a], SHUANGPING LI[2,b] AND ALLAN SLY[3,c]

[1]*Institute of Mathematics, EPFL,* [a]*emmanuel.abbe@epfl.ch*
[2]*PACM, Princeton University,* [b]*sl31@princeton.edu*
[3]*Department of Mathematics, Princeton University,* [c]*allansly@princeton.edu*

The problem of learning graphons has attracted considerable attention across several scientific communities, with significant progress over the recent years in sparser regimes. Yet, the current techniques still require diverging degrees in order to succeed with efficient algorithms in the challenging cases where the local structure of the graph is homogeneous. This paper provides an efficient algorithm to learn graphons in the constant expected degree regime. The algorithm is shown to succeed in estimating the rank-k projection of a graphon in the L_2 metric if the top k eigenvalues of the graphon satisfy a generalized Kesten–Stigum condition.

REFERENCES

[1] ABBE, E. (2017). Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** Paper No. 177, 86. MR3827065

[2] ABBE, E., BOIX-ADSERÀ, E., RALLI, P. and SANDON, C. (2020). Graph powering and spectral robustness. *SIAM J. Math. Data Sci.* **2** 132–157. MR4065193 https://doi.org/10.1137/19M1257135

[3] ABBE, E., LI, S. and SLY, A. (2023). Supplement to "Learning sparse graphons and the generalized Kesten–Stigum threshold." https://doi.org/10.1214/23-AOS2262SUPP

[4] ABBE, E. and SANDON, C. (2015). Recovering communities in the general stochastic block model without knowing the parameters. In *Advances in Neural Information Processing Systems* (C. Cortes, N. Lawrence, D. Lee, M. Sugiyama and R. Garnett, eds.) **28**. Curran Associates, Red Hook, NY.

[5] ABBE, E. and SANDON, C. (2015). Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In 2015 *IEEE* 56*th Annual Symposium on Foundations of Computer Science—FOCS* 2015 670–688. IEEE Computer Soc., Los Alamitos, CA. MR3473334 https://doi.org/10.1109/FOCS.2015.47

[6] 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

[7] AIROLDI, E. M., COSTA, T. B. and CHAN, S. H. (2013). Stochastic blockmodel approximation of a graphon: Theory and consistent estimation. In *Advances in Neural Information Processing Systems* (C. J. Burges, L. Bottou, M. Welling, Z. Ghahramani and K. Q. Weinberger, eds.) **26**. Curran Associates, Red Hook, NY.

[8] AMINI, A. A., CHEN, A., BICKEL, P. J. and LEVINA, E. (2013). Pseudo-likelihood methods for community detection in large sparse networks. *Ann. Statist.* **41** 2097–2122. MR3127859 https://doi.org/10.1214/13-AOS1138

[9] ANANDKUMAR, A., GE, R., HSU, D. and KAKADE, S. M. (2014). A tensor approach to learning mixed membership community models. *J. Mach. Learn. Res.* **15** 2239–2312. MR3231594

[10] BICKEL, P. J. and CHEN, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA* **106** 21068–21073. https://doi.org/10.1073/pnas.0907096106

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

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

MSC2020 subject classifications. 62H22.
Key words and phrases. Inference on networks, graphon, spectral algorithm.

[13] BOPPANA, R. B. (1987). Eigenvalues and graph bisection: An average-case analysis. In 28*th Annual Symposium on Foundations of Computer Science* (*sfcs* 1987) 280–285. https://doi.org/10.1109/SFCS.1987.22

[14] BORDENAVE, C., LELARGE, M. and MASSOULIÉ, L. (2018). Nonbacktracking spectrum of random graphs: Community detection and nonregular Ramanujan graphs. *Ann. Probab.* **46** 1–71. MR3758726 https://doi.org/10.1214/16-AOP1142

[15] BORGS, C., CHAYES, J. T., COHN, H. and GANGULY, S. (2021). Consistent nonparametric estimation for heavy-tailed sparse graphs. *Ann. Statist.* **49** 1904–1930. MR4319235 https://doi.org/10.1214/20-aos1985

[16] BORGS, C., CHAYES, J. T., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2008). Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.* **219** 1801–1851. MR2455626 https://doi.org/10.1016/j.aim.2008.07.008

[17] BORGS, C., CHAYES, J. T., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2012). Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Ann. of Math.* (2) **176** 151–219. MR2925382 https://doi.org/10.4007/annals.2012.176.1.2

[18] BORGS, C., CHAYES, J. T. and SMITH, A. (2015). Private graphon estimation for sparse graphs. In *Proceedings of the* 28*th International Conference on Neural Information Processing Systems*, *Vol.* 1. *NIPS'*15 1369–1377. MIT Press, Cambridge, MA.

[19] BUI, T. N., CHAUDHURI, S., LEIGHTON, F. T. and SIPSER, M. (1987). Graph bisection algorithms with good average case behavior. *Combinatorica* **7** 171–191. MR0905164 https://doi.org/10.1007/BF02579448

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

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

[22] 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. https://doi.org/10.1103/PhysRevE.84.066106

[23] DYER, M. E. and FRIEZE, A. M. (1989). The solution of some random NP-hard problems in polynomial expected time. *J. Algorithms* **10** 451–489. MR1022107 https://doi.org/10.1016/0196-6774(89)90001-1

[24] GAO, C., LU, Y. and ZHOU, H. H. (2015). Rate-optimal graphon estimation. *Ann. Statist.* **43** 2624–2652. MR3405606 https://doi.org/10.1214/15-AOS1354

[25] GULIKERS, L., LELARGE, M. and MASSOULIÉ, L. (2017). Non-backtracking spectrum of degree-corrected stochastic block models. In 8*th Innovations in Theoretical Computer Science Conference*. *LIPIcs. Leibniz Int. Proc. Inform.* **67** Art. No. 44, 27. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR3754968

[26] GULIKERS, L., LELARGE, M. and MASSOULIÉ, L. (2018). An impossibility result for reconstruction in the degree-corrected stochastic block model. *Ann. Appl. Probab.* **28** 3002–3027. MR3847979 https://doi.org/10.1214/18-AAP1381

[27] HASHIMOTO, K. (1989). Zeta functions of finite graphs and representations of p-adic groups. In *Automorphic Forms and Geometry of Arithmetic Varieties*. *Adv. Stud. Pure Math.* **15** 211–280. Academic Press, Boston, MA. MR1040609 https://doi.org/10.2969/aspm/01510211

[28] 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

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

[30] KESTEN, H. and STIGUM, B. P. (1966). A limit theorem for multidimensional Galton–Watson processes. *Ann. Math. Stat.* **37** 1211–1223. MR0198552 https://doi.org/10.1214/aoms/1177699266

[31] KLOPP, O., TSYBAKOV, A. B. and VERZELEN, N. (2017). Oracle inequalities for network models and sparse graphon estimation. *Ann. Statist.* **45** 316–354. MR3611494 https://doi.org/10.1214/16-AOS1454

[32] KRIVELEVICH, M. and SUDAKOV, B. (2003). The largest eigenvalue of sparse random graphs. *Combin. Probab. Comput.* **12** 61–72. MR1967486 https://doi.org/10.1017/S0963548302005424

[33] 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

[34] 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

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

[36] MOSSEL, E., NEEMAN, J. and SLY, A. (2016). Belief propagation, robust reconstruction and optimal recovery of block models. *Ann. Appl. Probab.* **26** 2211–2256. MR3543895 https://doi.org/10.1214/15-AAP1145

[37] 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

[38] MOSSEL, E. and PERES, Y. (2003). Information flow on trees. *Ann. Appl. Probab.* **13** 817–844. MR1994038 https://doi.org/10.1214/aoap/1060202828

[39] 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

[40] STEPHAN, L. and MASSOULIÉ, L. (2022). Non-backtracking spectra of weighted inhomogeneous random graphs. *Math. Stat. Learn.* **5** 201–271. MR4526300 https://doi.org/10.4171/msl/34

[41] TALENTI, G. (1987). Recovering a function from a finite number of moments. *Inverse Probl.* **3** 501–517. MR0909407

[42] WOLFE, P. J. and OLHEDE, S. C. Nonparametric graphon estimation. Available at arXiv:1309.5936.

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

[44] XU, J., MASSOULIÉ, L. and LELARGE, M. (2014). Edge label inference in generalized stochastic block models: From spectral theory to impossibility results. In *Proceedings of the* 27*th Conference on Learning Theory* (M. F. Balcan, V. Feldman and C. Szepesvári, eds.). *Proceedings of Machine Learning Research* **35** 903–920. PMLR, Barcelona, Spain.

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

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

The Annals of Statistics
2023, Vol. 51, No. 2, 624–645
https://doi.org/10.1214/23-AOS2263
© Institute of Mathematical Statistics, 2023

ON SINGULAR VALUES OF DATA MATRICES WITH GENERAL INDEPENDENT COLUMNS

BY TIANXING MEI[1,a], CHEN WANG[1,b] AND JIANFENG YAO[2,c]

[1]*Department of Statistics and Actuarial Science*
The University of Hong Kong, [a]*meitx@hku.hk,* [b]*stacw@hku.hk*
[2]*School of Data Science*
The Chinese University of Hong Kong (Shenzhen), [c]*jeffyao@cuhk.edu.cn*

We analyze the singular values of a large $p \times n$ data matrix $\mathbf{X}_n = (\mathbf{x}_{n1}, \ldots, \mathbf{x}_{nn})$, where the columns $\{\mathbf{x}_{nj}\}$ are independent p-dimensional vectors, possibly with different distributions. Assuming that the covariance matrices $\mathbf{\Sigma}_{nj} = \mathrm{Cov}(\mathbf{x}_{nj})$ of the column vectors can be asymptotically simultaneously diagonalized, with appropriately converging spectra, we establish a limiting spectral distribution (LSD) for the singular values of $\mathbf{X}_n$ when both dimensions p and n grow to infinity in comparable magnitudes. Our matrix model goes beyond and includes many different types of sample covariance matrices in existing work, such as weighted sample covariance matrices, Gram matrices, and sample covariance matrices of a linear time series model. Furthermore, three applications of our general approach are developed. First, we obtain the existence and uniqueness of the LSD for realized covariance matrices of a multi-dimensional diffusion process with anisotropic time-varying co-volatility. Second, we derive the LSD for singular values of data matrices from a recent matrix-valued auto-regressive model. Finally, we also obtain the LSD for singular values of data matrices from a generalized finite mixture model.

REFERENCES

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

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

[3] BAI, Z. and ZHOU, W. (2008). Large sample covariance matrices without independence structures in columns. *Statist. Sinica* **18** 425–442. MR2411613

[4] CHEN, E. Y., TSAY, R. S. and CHEN, R. (2020). Constrained factor models for high-dimensional matrix-variate time series. *J. Amer. Statist. Assoc.* **115** 775–793. MR4107679 https://doi.org/10.1080/01621459.2019.1584899

[5] CHEN, R., XIAO, H. and YANG, D. (2021). Autoregressive models for matrix-valued time series. *J. Econometrics* **222** 539–560. MR4234832 https://doi.org/10.1016/j.jeconom.2020.07.015

[6] FRÜHWIRTH-SCHNATTER, S. (2006). *Finite Mixture and Markov Switching Models*. *Springer Series in Statistics*. Springer, New York. MR2265601

[7] GRAY, R. M. (2006). Toeplitz and circulant matrices: A review. *Found. Trends Commun. Inf. Theory* **2** 155–239.

[8] HACHEM, W., LOUBATON, P. and NAJIM, J. (2006). The empirical distribution of the eigenvalues of a Gram matrix with a given variance profile. *Ann. Inst. Henri Poincaré Probab. Stat.* **42** 649–670. MR2269232 https://doi.org/10.1016/j.anihpb.2005.10.001

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

MSC2020 subject classifications. Primary 62H10; secondary 60B20.

Key words and phrases. Large data matrix, large sample covariance matrices, singular value distribution, eigenvalue distribution, separable covariance matrix, realized covariance matrix, matrix-valued autoregressive model.

[10] JIN, B., WANG, C., BAI, Z. D., NAIR, K. K. and HARDING, M. (2014). Limiting spectral distribution of a symmetrized auto-cross covariance matrix. *Ann. Appl. Probab.* **24** 1199–1225. MR3199984 https://doi.org/10.1214/13-AAP945

[11] JIN, B., WANG, C., MIAO, B. and LO HUANG, M.-N. (2009). Limiting spectral distribution of large-dimensional sample covariance matrices generated by VARMA. *J. Multivariate Anal.* **100** 2112–2125. MR2543090 https://doi.org/10.1016/j.jmva.2009.06.011

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

[13] 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

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

[15] MCLACHLAN, G. and PEEL, D. (2000). *Finite Mixture Models*. *Wiley Series in Probability and Statistics*: *Applied Probability and Statistics*. Wiley Interscience, New York. MR1789474 https://doi.org/10.1002/0471721182

[16] MEI, T., WANG, C. and YAO, J. (2023). Supplement to "On singular values of data matrices with general independent columns." https://doi.org/10.1214/23-AOS2263SUPP

[17] PAJOR, A. and PASTUR, L. (2009). On the limiting empirical measure of eigenvalues of the sum of rank one matrices with log-concave distribution. *Studia Math.* **195** 11–29. MR2539559 https://doi.org/10.4064/sm195-1-2

[18] 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

[19] SILVERSTEIN, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.* **55** 331–339. MR1370408 https://doi.org/10.1006/jmva.1995.1083

[20] SILVERSTEIN, J. W. and BAI, Z. D. (1995). On the empirical distribution of eigenvalues of a class of large-dimensional random matrices. *J. Multivariate Anal.* **54** 175–192. MR1345534 https://doi.org/10.1006/jmva.1995.1051

[21] WANG, D., LIU, X. and CHEN, R. (2019). Factor models for matrix-valued high-dimensional time series. *J. Econometrics* **208** 231–248. MR3906969 https://doi.org/10.1016/j.jeconom.2018.09.013

[22] YAO, J. (2012). A note on a Marčenko–Pastur type theorem for time series. *Statist. Probab. Lett.* **82** 22–28. MR2863018 https://doi.org/10.1016/j.spl.2011.08.011

[23] YAO, J., ZHENG, S. and BAI, Z. (2015). *Large Sample Covariance Matrices and High-dimensional Data Analysis*. *Cambridge Series in Statistical and Probabilistic Mathematics* **39**. Cambridge Univ. Press, New York. MR3468554 https://doi.org/10.1017/CBO9781107588080

[24] ZHANG, L. (2006). Spectral analysis of large dimensional random matrices. Ph.D. thesis National Univ. Singapore.

[25] ZHENG, X. and LI, Y. (2011). On the estimation of integrated covariance matrices of high dimensional diffusion processes. *Ann. Statist.* **39** 3121–3151. MR3012403 https://doi.org/10.1214/11-AOS939

The Annals of Statistics
2023, Vol. 51, No. 2, 646–667
https://doi.org/10.1214/23-AOS2264
© Institute of Mathematical Statistics, 2023

OPTIMAL ESTIMATION AND COMPUTATIONAL LIMIT OF LOW-RANK GAUSSIAN MIXTURES

BY ZHONGYUAN LYU[a] AND DONG XIA[b]

Department of Mathematics, Hong Kong University of Science and Technology, [a]*zlyuab@connect.ust.hk,* [b]*madxia@ust.hk*

Structural matrix-variate observations routinely arise in diverse fields such as multilayer network analysis and brain image clustering. While data of this type have been extensively investigated with fruitful outcomes being delivered, the fundamental questions like its statistical optimality and computational limit are largely under-explored. In this paper, we propose a low-rank Gaussian mixture model (LrMM) assuming each matrix-valued observation has a planted low-rank structure. Minimax lower bounds for estimating the underlying low-rank matrix are established allowing a whole range of sample sizes and signal strength. Under a minimal condition on signal strength, referred to as the *information-theoretical limit* or *statistical limit*, we prove the minimax optimality of a maximum likelihood estimator which, in general, is computationally infeasible. If the signal is stronger than a certain threshold, called the *computational limit*, we design a computationally fast estimator based on spectral aggregation and demonstrate its minimax optimality. Moreover, when the signal strength is smaller than the computational limit, we provide evidences based on the low-degree likelihood ratio framework to claim that no polynomial-time algorithm can consistently recover the underlying low-rank matrix. Our results reveal multiple phase transitions in the minimax error rates and the statistical-to-computational gap. Numerical experiments confirm our theoretical findings. We further showcase the merit of our spectral aggregation method on the worldwide food trading dataset.

REFERENCES

[1] ADAMCZAK, R. (2008). A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.* **13** 1000–1034. MR2424985 https://doi.org/10.1214/EJP.v13-521

[2] BAKSHI, A., DIAKONIKOLAS, I., JIA, H., KANE, D. M., KOTHARI, P. K. and VEMPALA, S. S. (2022). Robustly learning mixtures of k arbitrary Gaussians. In *STOC '22—Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing* 1234–1247. ACM, New York. MR4490075

[3] BALAKRISHNAN, S., WAINWRIGHT, M. J. and YU, B. (2017). Statistical guarantees for the EM algorithm: From population to sample-based analysis. *Ann. Statist.* **45** 77–120. MR3611487 https://doi.org/10.1214/16-AOS1435

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

[5] BELKIN, M. and SINHA, K. (2010). Toward learning Gaussian mixtures with arbitrary separation. In *COLT* 407–419.

[6] CAI, T. T., MA, J. and ZHANG, L. (2019). CHIME: Clustering of high-dimensional Gaussian mixtures with EM algorithm and its optimality. *Ann. Statist.* **47** 1234–1267. MR3911111 https://doi.org/10.1214/18-AOS1711

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

[8] CATTELL, R. B. (1966). The scree test for the number of factors. *Multivar. Behav. Res.* **1** 245–276. https://doi.org/10.1207/s15327906mbr0102_10

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

MSC2020 subject classifications. Primary 62C20; secondary 62F30.
Key words and phrases. Gaussian mixtures, low-rank matrix, minimax optimality.

[10] CHEN, Y., MA, C., POOR, H. V. and CHEN, Y. (2021). Learning mixtures of low-rank models. *IEEE Trans. Inf. Theory* **67** 4613–4636. MR4306287 https://doi.org/10.1109/TIT.2021.3065700

[11] CHENG, C., WEI, Y. and CHEN, Y. (2021). Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise. *IEEE Trans. Inf. Theory* **67** 7380–7419. MR4345128 https://doi.org/10.1109/TIT.2021.3111828

[12] DAVIES, S., MAZUMDAR, A., PAL, S. and RASHTCHIAN, C. (2022). Lower bounds on the total variation distance between mixtures of two Gaussians. In *Algorithmic Learning Theory* 2022. *Proc. Mach. Learn. Res.* (*PMLR*) **167** 319–341. Proceedings of Machine Learning Research PMLR. MR4461160

[13] DAVIS, D., DIAZ, M. and WANG, K. (2021). Clustering a mixture of Gaussians with unknown covariance. arXiv preprint arXiv:2110.01602.

[14] DE DOMENICO, M., NICOSIA, V., ARENAS, A. and LATORA, V. (2015). Structural reducibility of multilayer networks. *Nat. Commun.* **6** 1–9.

[15] DOSS, N., WU, Y., YANG, P. and ZHOU, H. H. (2020). Optimal estimation of high-dimensional location Gaussian mixtures. arXiv preprint arXiv:2002.05818.

[16] GAO, X., SHEN, W., ZHANG, L., HU, J., FORTIN, N. J., FROSTIG, R. D. and OMBAO, H. (2021). Regularized matrix data clustering and its application to image analysis. *Biometrics* **77** 890–902. MR4320665 https://doi.org/10.1111/biom.13354

[17] GAVISH, M. and DONOHO, D. L. (2014). The optimal hard threshold for singular values is $4/\sqrt{3}$. *IEEE Trans. Inf. Theory* **60** 5040–5053. MR3245370 https://doi.org/10.1109/TIT.2014.2323359

[18] GE, R., HUANG, Q. and KAKADE, S. M. (2015). Learning mixtures of Gaussians in high dimensions [extended abstract]. In *STOC*'15—*Proceedings of the* 2015 *ACM Symposium on Theory of Computing* 761–770. ACM, New York. MR3388256 https://doi.org/10.1145/2746539.2746616

[19] GENOVESE, C. R. and WASSERMAN, L. (2000). Rates of convergence for the Gaussian mixture sieve. *Ann. Statist.* **28** 1105–1127. MR1810921 https://doi.org/10.1214/aos/1015956709

[20] GHOSAL, S. and VAN DER VAART, A. W. (2001). Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.* **29** 1233–1263. MR1873329 https://doi.org/10.1214/aos/1013203453

[21] GUPTA, A. K. and NAGAR, D. K. (2000). *Matrix Variate Distributions*. *Chapman & Hall/CRC Monographs and Surveys in Pure and Applied Mathematics* **104**. CRC Press/CRC, Boca Raton, FL. MR1738933

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

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

[24] HOPKINS, S. (2018). *Statistical Inference and the Sum of Squares Method*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.), Cornell Univ. MR3864930

[25] HU, W., SHEN, W., ZHOU, H. and KONG, D. (2020). Matrix linear discriminant analysis. *Technometrics* **62** 196–205. MR4095745 https://doi.org/10.1080/00401706.2019.1610069

[26] JING, B.-Y., LI, T., LYU, Z. and XIA, D. (2021). Community detection on mixture multilayer networks via regularized tensor decomposition. *Ann. Statist.* **49** 3181–3205. MR4352527 https://doi.org/10.1214/21-aos2079

[27] KOLTCHINSKII, V. and XIA, D. (2015). Optimal estimation of low rank density matrices. *J. Mach. Learn. Res.* **16** 1757–1792. MR3417798

[28] 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

[29] KUNISKY, D., WEIN, A. S. and BANDEIRA, A. S. (2022). Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *Mathematical Analysis, Its Applications and Computation. Springer Proc. Math. Stat.* **385** 1–50. Springer, Cham. MR4461037 https://doi.org/10.1007/978-3-030-97127-4_1

[30] LE, C. M., LEVIN, K. and LEVINA, E. (2018). Estimating a network from multiple noisy realizations. *Electron. J. Stat.* **12** 4697–4740. MR3894068 https://doi.org/10.1214/18-ejs1521

[31] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces*: *Isoperimetry and Processes*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete* (3) [*Results in Mathematics and Related Areas* (3)] **23**. Springer, Berlin. MR1102015 https://doi.org/10.1007/978-3-642-20212-4

[32] LEROUX, B. G. (1992). Consistent estimation of a mixing distribution. *Ann. Statist.* **20** 1350–1360. MR1186253 https://doi.org/10.1214/aos/1176348772

[33] LÖFFLER, M., WEIN, A. S. and BANDEIRA, A. S. (2022). Computationally efficient sparse clustering. *Inf. Inference* **11** 1255–1286. MR4526323 https://doi.org/10.1093/imaiai/iaac019

[34] LÖFFLER, M., ZHANG, A. Y. and ZHOU, H. H. (2021). Optimality of spectral clustering in the Gaussian mixture model. *Ann. Statist.* **49** 2506–2530. MR4338373 https://doi.org/10.1214/20-aos2044

[35] LYU, Z. and XIA, D. (2023). Supplement to "Optimal estimation and computational limit of low-rank Gaussian mixtures." https://doi.org/10.1214/23-AOS2264SUPP
[36] LYU, Z., XIA, D. and ZHANG, Y. (2021). Latent space model for higher-order networks and generalized tensor decomposition. arXiv preprint arXiv:2106.16042.
[37] MA, Z. and WU, Y. (2015). Volume ratio, sparsity, and minimaxity under unitarily invariant norms. *IEEE Trans. Inf. Theory* **61** 6939–6956. MR3430731 https://doi.org/10.1109/TIT.2015.2487541
[38] MAI, Q., ZHANG, X., PAN, Y. and DENG, K. (2022). A doubly enhanced EM algorithm for model-based tensor clustering. *J. Amer. Statist. Assoc.* **117** 2120–2134. MR4528493 https://doi.org/10.1080/01621459.2021.1904959
[39] MAUGIS, C. and MICHEL, B. (2011). A non asymptotic penalized criterion for Gaussian mixture model selection. *ESAIM Probab. Stat.* **15** 41–68. MR2870505 https://doi.org/10.1051/ps/2009004
[40] 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
[41] MOITRA, A. and VALIANT, G. (2010). Settling the polynomial learnability of mixtures of Gaussians. In 2010 *IEEE 51st Annual Symposium on Foundations of Computer Science—FOCS* 2010 93–102. IEEE Computer Soc., Los Alamitos, CA. MR3024779
[42] PAUL, S. and CHEN, Y. (2020). Spectral and matrix factorization methods for consistent community detection in multi-layer networks. *Ann. Statist.* **48** 230–250. MR4065160 https://doi.org/10.1214/18-AOS1800
[43] PEARSON, K. (1894). Contributions to the mathematical theory of evolution. *Philos. Trans. R. Soc. Lond. A* **185** 71–110.
[44] SANJEEV, A. and KANNAN, R. (2001). Learning mixtures of arbitrary Gaussians. In *Proceedings of the Thirty-Third Annual ACM Symposium on Theory of Computing* 247–257. ACM, New York. MR2120323 https://doi.org/10.1145/380752.380808
[45] SUN, W. W. and LI, L. (2019). Dynamic tensor clustering. *J. Amer. Statist. Assoc.* **114** 1894–1907. MR4047308 https://doi.org/10.1080/01621459.2018.1527701
[46] TOSH, C. and DASGUPTA, S. (2017). Maximum likelihood estimation for mixtures of spherical Gaussians is NP-hard. *J. Mach. Learn. Res.* **18** Paper No. 175, 11. MR3827063
[47] VAN DE GEER, S. (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.* **21** 14–44. MR1212164 https://doi.org/10.1214/aos/1176349013
[48] VAN DE GEER, S. (2000). *Empirical Processes in M-Estimation* **6**. Cambridge University Press.
[49] 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
[50] WANG, L., ZHANG, Z. and DUNSON, D. (2019). Common and individual structure of brain networks. *Ann. Appl. Stat.* **13** 85–112. MR3937422 https://doi.org/10.1214/18-AOAS1193
[51] WANG, X. and ZHU, H. (2017). Generalized scalar-on-image regression models via total variation. *J. Amer. Statist. Assoc.* **112** 1156–1168. MR3735367 https://doi.org/10.1080/01621459.2016.1194846
[52] WU, Y. and YANG, P. (2020). Optimal estimation of Gaussian mixtures via denoised method of moments. *Ann. Statist.* **48** 1981–2007. MR4134783 https://doi.org/10.1214/19-AOS1873
[53] WU, Y. and ZHOU, H. H. (2021). Randomly initialized EM algorithm for two-component Gaussian mixture achieves near optimality in $O(\sqrt{n})$ iterations. *Math. Stat. Learn.* **4** 143–220. MR4383733 https://doi.org/10.4171/msl/29
[54] XIA, D. (2021). Normal approximation and confidence region of singular subspaces. *Electron. J. Stat.* **15** 3798–3851. MR4298986 https://doi.org/10.1214/21-ejs1876
[55] XIA, D., YUAN, M. and ZHANG, C.-H. (2021). Statistically optimal and computationally efficient low rank tensor completion from noisy entries. *Ann. Statist.* **49** 76–99. MR4206670 https://doi.org/10.1214/20-AOS1942
[56] XIA, D. and ZHOU, F. (2019). The sup-norm perturbation of HOSVD and low rank tensor denoising. *J. Mach. Learn. Res.* **20** Paper No. 61, 42. MR3960915
[57] XU, J., HSU, D. and MALEKI, A. (2016). Global analysis of expectation maximization for mixtures of two Gaussians. arXiv preprint arXiv:1608.07630.
[58] ZADIK, I., SONG, M. J., WEIN, A. S. and BRUNA, J. (2021). Lattice-based methods surpass sum-of-squares in clustering. arXiv preprint arXiv:2112.03898.
[59] ZHANG, A. and XIA, D. (2018). Tensor SVD: Statistical and computational limits. *IEEE Trans. Inf. Theory* **64** 7311–7338. MR3876445 https://doi.org/10.1109/TIT.2018.2841377

The Annals of Statistics
2023, Vol. 51, No. 2, 668–690
https://doi.org/10.1214/23-AOS2265
© Institute of Mathematical Statistics, 2023

EFFICIENT FUNCTIONAL ESTIMATION AND THE SUPER-ORACLE PHENOMENON

BY THOMAS B. BERRETT[1,a] AND RICHARD J. SAMWORTH[2,b]

[1]*Department of Statistics, University of Warwick,* [a]*tom.berrett@warwick.ac.uk*
[2]*Statistical Laboratory, University of Cambridge,* [b]*r.samworth@statslab.cam.ac.uk*

We consider the estimation of two-sample integral functionals, of the type that occur naturally, for example, when the object of interest is a divergence between unknown probability densities. Our first main result is that, in wide generality, a weighted nearest neighbour estimator is efficient, in the sense of achieving the local asymptotic minimax lower bound. Moreover, we also prove a corresponding central limit theorem, which facilitates the construction of asymptotically valid confidence intervals for the functional, having asymptotically minimal width. One interesting consequence of our results is the discovery that, for certain functionals, the worst-case performance of our estimator may improve on that of the natural 'oracle' estimator, which itself can be optimal in the related problem where the data consist of the values of the unknown densities at the observations.

REFERENCES

BALDI, P. and RINOTT, Y. (1989). On normal approximations of distributions in terms of dependency graphs. *Ann. Probab.* **17** 1646–1650. MR1048950

BEIRLANT, J., DUDEWICZ, E. J., GYÖRFI, L. and VAN DER MEULEN, E. C. (1997). Nonparametric entropy estimation: An overview. *Int. J. Math. Stat. Sci.* **6** 17–39. MR1471870

BERRETT, T. B. and SAMWORTH, R. J. (2023). Supplement to "Efficient functional estimation and the super-oracle phenomenon." https://doi.org/10.1214/23-AOS2265SUPP

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. 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

BICKEL, P. J. and RITOV, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhyā Ser. A* **50** 381–393. MR1065550

BIRGÉ, L. and MASSART, P. (1995). Estimation of integral functionals of a density. *Ann. Statist.* **23** 11–29. MR1331653 https://doi.org/10.1214/aos/1176324452

DIEUDONNÉ, J. (1969). *Foundations of Modern Analysis. Pure and Applied Mathematics*, *Vol.* 10-*I*. Academic Press, New York. MR0349288

DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24** 508–539. MR1394974 https://doi.org/10.1214/aos/1032894451

GINÉ, E. and NICKL, R. (2008). A simple adaptive estimator of the integrated square of a density. *Bernoulli* **14** 47–61. MR2401653 https://doi.org/10.3150/07-BEJ110

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

HAN, Y., JIAO, J., WEISSMAN, T. and WU, Y. (2020). Optimal rates of entropy estimation over Lipschitz balls. *Ann. Statist.* **48** 3228–3250. MR4185807 https://doi.org/10.1214/19-AOS1927

HERO, A. O., MA, B., MICHEL, O. and GORMAN, J. (2002). Applications of entropic spanning graphs. *IEEE Signal Process. Mag.* **19** 85–95.

IBRAGIMOV, I. A. and KHAS'MINSKIĬ, R. Z. (1991). Asymptotically normal families of distributions and efficient estimation. *Ann. Statist.* **19** 1681–1724. MR1135145 https://doi.org/10.1214/aos/1176348367

JUDITSKY, A. and LAMBERT-LACROIX, S. (2004). On minimax density estimation on $\mathbb{R}$. *Bernoulli* **10** 187–220. MR2046772 https://doi.org/10.3150/bj/1082380217

MSC2020 subject classifications. 62G05, 62G20.
Key words and phrases. Efficiency, functional estimation, weighted nearest neighbours.

KANDASAMY, K., KRISHNAMURTHY, A., PÓCZOS, B., WASSERMAN, L. and ROBINS, J. M. (2015). Nonparametric von Mises estimators for entropies, divergences and mutual informations. *NeurIPS* **28**.

KILBAS, A. A., SRIVASTAVA, H. M. and TRUJILLO, J. J. (2006). *Theory and Applications of Fractional Differential Equations*. *North-Holland Mathematics Studies* **204**. Elsevier, Amsterdam. MR2218073

KOZACHENKO, L. F. and LEONENKO, N. N. (1987). Sample estimate of the entropy of a random vector. *Probl. Inf. Transm.* **23** 95–101.

KRISHNAMURTHY, A., KANDASAMY, K., POĆZOS, B. and WASSERMAN, L. (2014). Nonparametric estimation of Rényi divergence and friends. In *Proc.* 31*st Int. Conf. on Mach. Learn* (*ICML*). **32** 919–927.

LAURENT, B. (1996). Efficient estimation of integral functionals of a density. *Ann. Statist.* **24** 659–681. MR1394981 https://doi.org/10.1214/aos/1032894458

LECAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. *Univ. Calif. Publ. Stat.* **1** 277–329. MR0054913

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

LEONENKO, N., PRONZATO, L. and SAVANI, V. (2008). A class of Rényi information estimators for multidimensional densities. *Ann. Statist.* **36** 2153–2182. MR2458183 https://doi.org/10.1214/07-AOS539

LEONENKO, N. and SELEZNJEV, O. (2010). Statistical inference for the ϵ-entropy and the quadratic Rényi entropy. *J. Multivariate Anal.* **101** 1981–1994. MR2671196 https://doi.org/10.1016/j.jmva.2010.05.009

MOON, K. R., SRICHARAN, K., GREENEWALD, K. and HERO, A. O. III (2018). Ensemble estimation of information divergence. *Entropy* **20** 560. MR3892642 https://doi.org/10.3390/e20080560

NOWOZIN, S., CSEKE, B. and TOMIOKA, R. (2016). F-GAN: Training generative neural samplers using variational divergence minimization. *Adv. Neural Inf. Process. Syst.*

RYU, J., GANGULY, S., KIM, Y., NOH, Y. and LEE, D. D. (2018). Nearest neighbor density functional estimation based on inverse Laplace transform. *IEEE Trans. Inf. Theory* **68** 3511–3551.

SINGH, S. and PÓCZOS, B. (2016). Finite-sample analysis of fixed-k nearest neighbor density functional estimators. In *Annual Conference on Neural Information Processing Systems* (*NIPS*) 1217–1225.

SINGH, S., SRIPERUMBUDUR, B. K. and PÓCZOS, B. (2018). Minimax estimation of quadratic Fourier functionals. Available at https://arxiv.org/abs/1803.11451.

TSYBAKOV, A. B. and VAN DER MEULEN, E. C. (1996). Root-n consistent estimators of entropy for densities with unbounded support. *Scand. J. Stat.* **23** 75–83. MR1380483

VAN DER VAART, A. W. (1997). Superefficiency. In *Festscrift for Lucien Le Cam* (D. Pollard, E. Torgersen and G. Yang, eds.) Springer, Berlin.

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

WORNOWIZKI, M. and FRIED, R. (2016). Two-sample homogeneity tests based on divergence measures. *Comput. Statist.* **31** 291–313. MR3481806 https://doi.org/10.1007/s00180-015-0633-3

The Annals of Statistics
2023, Vol. 51, No. 2, 691–716
https://doi.org/10.1214/23-AOS2266
© Institute of Mathematical Statistics, 2023

DEEP NONPARAMETRIC REGRESSION ON APPROXIMATE MANIFOLDS: NONASYMPTOTIC ERROR BOUNDS WITH POLYNOMIAL PREFACTORS

BY YULING JIAO[1,a], GUOHAO SHEN[2,b], YUANYUAN LIN[3,d] AND JIAN HUANG[2,c]

[1]*School of Mathematics and Statistics, Wuhan University,* [a]*yulingjiaomath@whu.edu.cn*
[2]*Department of Applied Mathematics, Hong Kong Polytechnic University,* [b]*guohao.shen@polyu.edu.hk,* [c]*j.huang@polyu.edu.hk*
[3]*Department of Statistics, Chinese University of Hong Kong,* [d]*ylin@sta.cuhk.edu.hk*

We study the properties of nonparametric least squares regression using deep neural networks. We derive nonasymptotic upper bounds for the excess risk of the empirical risk minimizer of feedforward deep neural regression. Our error bounds achieve minimax optimal rate and improve over the existing ones in the sense that they depend polynomially on the dimension of the predictor, instead of exponentially on dimension. We show that the neural regression estimator can circumvent the curse of dimensionality under the assumption that the predictor is supported on an approximate low-dimensional manifold or a set with low Minkowski dimension. We also establish the optimal convergence rate under the exact manifold support assumption. We investigate how the prediction error of the neural regression estimator depends on the structure of neural networks and propose a notion of network relative efficiency between two types of neural networks, which provides a quantitative measure for evaluating the relative merits of different network structures. To establish these results, we derive a novel approximation error bound for the Hölder smooth functions using ReLU activated neural networks, which may be of independent interest. Our results are derived under weaker assumptions on the data distribution and the neural network structure than those in the existing literature.

REFERENCES

AAMARI, E., KIM, J., CHAZAL, F., MICHEL, B., RINALDO, A. and WASSERMAN, L. (2019). Estimating the reach of a manifold. *Electron. J. Stat.* **13** 1359–1399. MR3938326 https://doi.org/10.1214/19-ejs1551

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

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

ASWANI, A., BICKEL, P. and TOMLIN, C. (2011). Regression on manifolds: Estimation of the exterior derivative. *Ann. Statist.* **39** 48–81. MR2797840 https://doi.org/10.1214/10-AOS823

BARANIUK, R. G. and WAKIN, M. B. (2009). Random projections of smooth manifolds. *Found. Comput. Math.* **9** 51–77. MR2472287 https://doi.org/10.1007/s10208-007-9011-z

BARTLETT, P. L., HARVEY, N., LIAW, C. and MEHRABIAN, A. (2019). Nearly-tight VC-dimension and pseudodimension bounds for piecewise linear neural networks. *J. Mach. Learn. Res.* **20** Paper No. 63, 17. MR3960917

BAUER, B. and KOHLER, M. (2019). On deep learning as a remedy for the curse of dimensionality in nonparametric regression. *Ann. Statist.* **47** 2261–2285. MR3953451 https://doi.org/10.1214/18-AOS1747

BELKIN, M. and NIYOGI, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **15** 1373–1396.

BELKIN, M. and NIYOGI, P. (2004). Semi-supervised learning on Riemannian manifolds. *Mach. Learn.* **56** 209–239.

MSC2020 subject classifications. Primary 62G05, 62G08; secondary 68T07.

Key words and phrases. Approximation error, curse of dimensionality, deep neural network, low-dimensional manifolds, network relative efficiency, nonasymptotic error bound.

BICKEL, P. J. and LI, B. (2007). Local polynomial regression on unknown manifolds. In *Complex Datasets and Inverse Problems. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **54** 177–186. IMS, Beachwood, OH. MR2459188 https://doi.org/10.1214/074921707000000148

BIRGÉ, L. and MASSART, P. (1993). Rates of convergence for minimum contrast estimators. *Probab. Theory Related Fields* **97** 113–150. MR1240719 https://doi.org/10.1007/BF01199316

BIRGÉ, L. and MASSART, P. (1998). Minimum contrast estimators on sieves: Exponential bounds and rates of convergence. *Bernoulli* **4** 329–375. MR1653272 https://doi.org/10.2307/3318720

BISHOP, C. J. and PERES, Y. (2017). *Fractals in Probability and Analysis*. *Cambridge Studies in Advanced Mathematics* **162**. Cambridge Univ. Press, Cambridge. MR3616046 https://doi.org/10.1017/9781316460238

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

BRAND, M. (2002). Charting a manifold. *Adv. Neural Inf. Process. Syst.* **15**.

CANCER GENOME ATLAS RESEARCH NETWORK, WEINSTEIN, J. N., COLLISSON, E. A., MILLS, G. B., SHAW, K. R. M., OZENBERGER, B. A., ELLROTT, K., SHMULEVICH, I., SANDER, C. et al. (2013). The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* **45** 1113–1120. https://doi.org/10.1038/ng.2764

CARLSSON, G. (2009). Topology and data. *Bull. Amer. Math. Soc.* (*N.S.*) **46** 255–308. MR2476414 https://doi.org/10.1090/S0273-0979-09-01249-X

CHEN, M., JIANG, H., LIAO, W. and ZHAO, T. (2022). Nonparametric regression on low-dimensional manifolds using deep ReLU networks: Function approximation and statistical recovery. *Inf. Inference* **11** 1203–1253. MR4526322 https://doi.org/10.1093/imaiai/iaac001

CHEN, M., JIANG, H. and ZHAO, T. (2019). Efficient approximation of deep relu networks for functions on low dimensional manifolds. *Adv. Neural Inf. Process. Syst.*

CHENG, M.-Y. and WU, H.-T. (2013). Local linear regression on manifolds and its geometric interpretation. *J. Amer. Statist. Assoc.* **108** 1421–1434. MR3174718 https://doi.org/10.1080/01621459.2013.827984

CHUI, C. K., LI, X. and MHASKAR, H. N. (1996). Limitations of the approximation capabilities of neural networks with one hidden layer. *Adv. Comput. Math.* **5** 233–243. MR1399382 https://doi.org/10.1007/BF02124745

CLONINGER, A. and KLOCK, T. (2020). ReLU nets adapt to intrinsic dimensionality beyond the target domain. Available at arXiv:2008.02545.

COX, D. D. (1988). Approximation of least squares regression on nested subspaces. *Ann. Statist.* **16** 713–732. MR0947572 https://doi.org/10.1214/aos/1176350830

DENG, J., DONG, W., SOCHER, R., LI, L.-J., LI, K. and FEI-FEI, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 *IEEE Conference on Computer Vision and Pattern Recognition* 248–255. Ieee.

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

DONOHO, D. L. and GRIMES, C. (2003). Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proc. Natl. Acad. Sci. USA* **100** 5591–5596. MR1981019 https://doi.org/10.1073/pnas.1031596100

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

ELDAN, R. and SHAMIR, O. (2016). The power of depth for feedforward neural networks. In *Conference on Learning Theory* 907–940. PMLR.

FALCONER, K. (2003). *Fractal Geometry*: *Mathematical Foundations and Applications*, 2nd ed. Wiley, Hoboken, NJ. MR2118797 https://doi.org/10.1002/0470013850

FARRELL, M. H., LIANG, T. and MISRA, S. (2021). Deep neural networks for estimation and inference. *Econometrica* **89** 181–213. MR4220387 https://doi.org/10.3982/ecta16901

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

FEFFERMAN, C., MITTER, S. and NARAYANAN, H. (2016). Testing the manifold hypothesis. *J. Amer. Math. Soc.* **29** 983–1049. MR3522608 https://doi.org/10.1090/jams/852

FRIEDMAN, J. H. and STUETZLE, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76** 817–823. MR0650892

GEMAN, S. and HWANG, C.-R. (1982). Nonparametric maximum likelihood estimation by the method of sieves. *Ann. Statist.* **10** 401–414. MR0653512

GHORBANI, B., MEI, S., MISIAKIEWICZ, T. and MONTANARI, A. (2020). Discussion of: "Nonparametric regression using deep neural networks with ReLU activation function" [MR4134774]. *Ann. Statist.* **48** 1898–1901. MR4134775 https://doi.org/10.1214/19-AOS1910

GOODFELLOW, I., BENGIO, Y. and COURVILLE, A. (2016). *Deep Learning*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3617773

GYÖRFI, L., KOHLER, M., KRZY˙ZAK, A. and WALK, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. Springer, New York. https://doi.org/10.1007/b97848

HÄRDLE, W., HALL, P. and ICHIMURA, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21** 157–178. MR1212171 https://doi.org/10.1214/aos/1176349020

HENDRIKS, H. (1990). Nonparametric estimation of a probability density on a Riemannian manifold using Fourier expansions. *Ann. Statist.* **18** 832–849. MR1056339 https://doi.org/10.1214/aos/1176347628

HOFFMANN, H., SCHAAL, S. and VIJAYAKUMAR, S. (2009). Local dimensionality reduction for non-parametric regression. *Neural Process. Lett.* **29** 109.

HOROWITZ, J. L. and HÄRDLE, W. (1996). Direct semiparametric estimation of single-index models with discrete covariates. *J. Amer. Statist. Assoc.* **91** 1632–1640. MR1439104 https://doi.org/10.2307/2291590

HUBBARD, J. H. and HUBBARD, B. B. (2015). *Vector Calculus, Linear Algebra, and Differential Forms: A Unified Approach*. Matrix Editions. MR1657732

JIAO, Y., SHEN, G., LIN, Y. and HUANG, J. (2023). Supplement to "Deep nonparametric regression on approximate manifolds: Nonasymptotic error bounds with polynomial prefactors." https://doi.org/10.1214/23-AOS2266SUPP

KOHLER, M., KRZYŻAK, A. and LANGER, S. (2022). Estimation of a function of low local dimensionality by deep neural networks. *IEEE Trans. Inf. Theory* **68** 4032–4042. MR4433267

KOHLER, M. and LANGER, S. (2021). On the rate of convergence of fully connected deep neural network regression estimates. *Ann. Statist.* **49** 2231–2249. MR4319248 https://doi.org/10.1214/20-aos2034

KONG, E. and XIA, Y. (2007). Variable selection for the single-index model. *Biometrika* **94** 217–229. MR2367831 https://doi.org/10.1093/biomet/asm008

KPOTUFE, S. (2011). k-NN regression adapts to local intrinsic dimension. Available at arXiv:1110.4300.

KPOTUFE, S. and GARG, V. K. (2013). Adaptivity to local smoothness and dimension in kernel regression. In *NIPS* 3075–3083.

KRIZHEVSKY, A. (2009). Learning multiple layers of features from tiny images. Technical report, the CIFAR-10 dataset is available at https://www.cs.toronto.edu/~kriz/cifar.html.

LECUN, Y., CORTES, C. and BURGES, C. (2010). MNIST handwritten digit database. AT&T Labs [online]. Available at http://yann.lecun.com/exdb/mnist, 2.

LEE, J. A. and VERLEYSEN, M. (2007). *Nonlinear Dimensionality Reduction*. *Information Science and Statistics*. Springer, New York. MR2373983 https://doi.org/10.1007/978-0-387-39351-3

LEE, J. M. (2003). *Introduction to Smooth Manifolds*. *Graduate Texts in Mathematics* **218**. Springer, New York. MR1930091 https://doi.org/10.1007/978-0-387-21752-9

LEE, J. M. (2006). *Riemannian Manifolds: An Introduction to Curvature* **176**. Springer Science & Business Media. MR1468735 https://doi.org/10.1007/b98852

LEE, W. S., BARTLETT, P. L. and WILLIAMSON, R. C. (1996). Efficient agnostic learning of neural networks with bounded fan-in. *IEEE Trans. Inf. Theory* **42** 2118–2132. MR1447518 https://doi.org/10.1109/18.556601

LIANG, S. and SRIKANT, R. (2016). Why deep neural networks for function approximation? Available at arXiv:1610.04161.

LU, J., SHEN, Z., YANG, H. and ZHANG, S. (2021). Deep network approximation for smooth functions. *SIAM J. Math. Anal.* **53** 5465–5506. MR4319100 https://doi.org/10.1137/20M134695X

LU, Z., PU, H., WANG, F., HU, Z. and WANG, L. (2017). The expressive power of neural networks: A view from the width. In *Proceedings of the* 31*st International Conference on Neural Information Processing Systems* 6232–6240.

MOHRI, M., ROSTAMIZADEH, A. and TALWALKAR, A. (2018). *Foundations of Machine Learning*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. Second edition of [MR3057769]. MR3931734

NAKADA, R. and IMAIZUMI, M. (2020). Adaptive approximation and generalization of deep neural network with intrinsic dimensionality. *J. Mach. Learn. Res.* **21** Paper No. 174, 38. MR4209460

NEMIROVSKIĬ, A. S., POLYAK, B. T. and TSYBAKOV, A. B. (1983). Estimates of the maximum likelihood type for a nonparametric regression. *Dokl. Akad. Nauk SSSR* **273** 1310–1314. MR0731296

NEMIROVSKIĬ, A. S., POLYAK, B. T. and TSYBAKOV, A. B. (1984). Signal processing by the nonparametric maximum likelihood method. *Problemy Peredachi Informatsii* **20** 29–46. MR0791733

NEMIROVSKIĬ, A. S., POLYAK, B. T. and TSYBAKOV, A. B. (1985). The rate of convergence of nonparametric estimates of maximum likelihood type. *Problemy Peredachi Informatsii* **21** 17–33. MR0820705

NGUYEN, P.-M. and PHAM, H. T. (2020). A rigorous framework for the mean field limit of multilayer neural networks. Available at arXiv:2001.11443.

NIYOGI, P., SMALE, S. and WEINBERGER, S. (2008). Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.* **39** 419–441. MR2383768 https://doi.org/10.1007/s00454-008-9053-2

PELLETIER, B. (2005). Kernel density estimation on Riemannian manifolds. *Statist. Probab. Lett.* **73** 297–304. MR2179289 https://doi.org/10.1016/j.spl.2005.04.004

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

POPE, P., ZHU, C., ABDELKADER, A., GOLDBLUM, M. and GOLDSTEIN, T. (2020). The intrinsic dimension of images and its impact on learning. In *International Conference on Learning Representations*.

RAFAJŁOWICZ, E. (1987). Nonparametric orthogonal series estimators of regression: A class attaining the optimal convergence rate in L_2. *Statist. Probab. Lett.* **5** 219–224. MR0881201 https://doi.org/10.1016/0167-7152(87)90044-7

RECANATESI, S., FARRELL, M., ADVANI, M., MOORE, T., LAJOIE, G. and SHEA-BROWN, E. (2019). Dimensionality compression and expansion in deep neural networks. Available at arXiv:1906.00443.

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

SCHMIDT-HIEBER, J. (2019). Deep relu network approximation of functions on a manifold. Available at arXiv:1908.00695.

SCHMIDT-HIEBER, J. (2020). Nonparametric regression using deep neural networks with ReLU activation function. *Ann. Statist.* **48** 1875–1897. MR4134774 https://doi.org/10.1214/19-AOS1875

SHEN, X. and WONG, W. H. (1994). Convergence rate of sieve estimates. *Ann. Statist.* **22** 580–615. MR1292531 https://doi.org/10.1214/aos/1176325486

SHEN, Z., YANG, H. and ZHANG, S. (2019). Nonlinear approximation via compositions. *Neural Netw.* **119** 74–84.

SHEN, Z., YANG, H. and ZHANG, S. (2020). Deep network approximation characterized by number of neurons. *Commun. Comput. Phys.* **28** 1768–1811. MR4188521 https://doi.org/10.4208/cicp.oa-2020-0149

SHEN, Z., YANG, H. and ZHANG, S. (2022). Optimal approximation rate of ReLU networks in terms of width and depth. *J. Math. Pures Appl.* (9) **157** 101–135. MR4351074 https://doi.org/10.1016/j.matpur.2021.07.009

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

STONE, C. J. (1986). The dimensionality reduction principle for generalized additive models. *Ann. Statist.* **14** 590–606. MR0840516 https://doi.org/10.1214/aos/1176349940

STONE, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation. *Ann. Statist.* **22** 118–184. With discussion by Andreas Buja and Trevor Hastie and a rejoinder by the author. MR1272079 https://doi.org/10.1214/aos/1176325361

SUZUKI, T. (2018). Adaptivity of deep ReLU network for learning in Besov and mixed smooth Besov spaces: Optimal rate and curse of dimensionality. Available at arXiv:1810.08033.

TELGARSKY, M. (2016). Benefits of depth in neural networks. In *Conference on Learning Theory* 1517–1539. PMLR.

TENENBAUM, J. B., DE SILVA, V. and LANGFORD, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science* **290** 2319–2323. https://doi.org/10.1126/science.290.5500.2319

VAN DE GEER, S. (1987). A new approach to least-squares estimation, with applications. *Ann. Statist.* **15** 587–602. MR0888427 https://doi.org/10.1214/aos/1176350362

VAN DE GEER, S. (1990). Estimating a regression function. *Ann. Statist.* **18** 907–924. MR1056343 https://doi.org/10.1214/aos/1176347632

VAN DE GEER, S. and WEGKAMP, M. (1996). Consistency for the least squares estimator in nonparametric regression. *Ann. Statist.* **24** 2513–2523. MR1425964 https://doi.org/10.1214/aos/1032181165

VAN DE GEER, S. A. (2000). *Applications of Empirical Process Theory*. *Cambridge Series in Statistical and Probabilistic Mathematics* **6**. Cambridge Univ. Press, Cambridge. MR1739079

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

YANG, Y. and DUNSON, D. B. (2016). Bayesian manifold regression. *Ann. Statist.* **44** 876–905. MR3476620 https://doi.org/10.1214/15-AOS1390

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

YAROTSKY, D. (2017). Error bounds for approximations with deep ReLU networks. *Neural Netw.* **94** 103–114. https://doi.org/10.1016/j.neunet.2017.07.002

YAROTSKY, D. (2018). Optimal approximation of continuous functions by very deep ReLU networks. In *Conference on Learning Theory* 639–649. PMLR.

The Annals of Statistics
2023, Vol. 51, No. 2, 717–737
https://doi.org/10.1214/23-AOS2267
© Institute of Mathematical Statistics, 2023

ON ROBUSTNESS AND LOCAL DIFFERENTIAL PRIVACY

BY MENGCHU LI[a], THOMAS B. BERRETT[b] AND YI YU[c]

Department of Statistics, University of Warwick, [a]*mengchu.li@warwick.ac.uk,* [b]*tom.berrett@warwick.ac.uk,* [c]*yi.yu.2@warwick.ac.uk*

It is of soaring demand to develop statistical analysis tools that are robust against contamination as well as preserving individual data owners' privacy. In spite of the fact that both topics host a rich body of literature, to the best of our knowledge, we are the first to systematically study the connections between the optimality under Huber's contamination model and the local differential privacy (LDP) constraints.

In this paper, we start with a general minimax lower bound result, which disentangles the costs of being robust against Huber contamination and preserving LDP. We further study four concrete examples: a two-point testing problem, a potentially diverging mean estimation problem, a nonparametric density estimation problem and a univariate median estimation problem. For each problem, we demonstrate procedures that are optimal in the presence of both contamination and LDP constraints, comment on the connections with the state-of-the-art methods that are only studied under either contamination or privacy constraints, and unveil the connections between robustness and LDP via partially answering whether LDP procedures are robust and whether robust procedures can be efficiently privatised. Overall, our work showcases a promising prospect of joint study for robustness and local differential privacy.

REFERENCES

ACHARYA, J., CANONNE, C. L., LIU, Y., SUN, Z. and TYAGI, H. (2022). Interactive inference under information constraints. *IEEE Trans. Inf. Theory* **68** 502–516. MR4395425 https://doi.org/10.1109/tit.2021.3123905

ACHARYA, J., SUN, Z. and ZHANG, H. (2021). Robust testing and estimation under manipulation attacks. In *International Conference on Machine Learning* 43–53. PMLR.

ARIDOR, G., CHE, Y.-K. and SALZ, T. (2021). The effect of privacy regulation on the data industry: Empirical evidence from GDPR. In *Proceedings of the* 22*nd ACM Conference on Economics and Computation* 93–94.

AVELLA-MEDINA, M. (2020). The role of robust statistics in private data analysis. *Chance* **33** 37–42.

BERRETT, T. and BUTUCEA, C. (2019). Classification under local differential privacy. *Ann. I.S.U.P.* **63** 191–205.

BERRETT, T. B. and BUTUCEA, C. (2020). Locally private non-asymptotic testing of discrete distributions is faster using interactive mechanisms. *Adv. Neural Inf. Process. Syst.* **33** 3164–3173.

BERRETT, T. B., GYÖRFI, L. and WALK, H. (2021). Strongly universally consistent nonparametric regression and classification with privatised data. *Electron. J. Stat.* **15** 2430–2453. MR4255329 https://doi.org/10.1214/21-ejs1845

BERRETT, T. B. and YU, Y. (2021). Locally private online change point detection. *Adv. Neural Inf. Process. Syst.* **34**. 3425–3437.

BRUNEL, V.-E. and AVELLA-MEDINA, M. (2020). Propose, test, release: Differentially private estimation with high probability. Preprint. Available at arXiv:2002.08774.

BUN, M., KAMATH, G., STEINKE, T. and WU, Z. S. (2021). Private hypothesis selection. *IEEE Trans. Inf. Theory* **67** 1981–2000. MR4282337 https://doi.org/10.1109/TIT.2021.3049802

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

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

MSC2020 subject classifications. 62C20.
Key words and phrases. Huber's contamination model, local differential privacy, minimax optimality.

CANONNE, C. L., KAMATH, G., MCMILLAN, A., SMITH, A. and ULLMAN, J. (2019). The structure of optimal private tests for simple hypotheses. In *STOC'*19*—Proceedings of the* 51*st Annual ACM SIGACT Symposium on Theory of Computing* 310–321. ACM, New York. MR4003341

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

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

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

CHERUKURI, A. and HOTA, A. R. (2021). Consistency of distributionally robust risk- and chance-constrained optimization under Wasserstein ambiguity sets. *IEEE Control Syst. Lett.* **5** 1729–1734. MR4213934

CHEU, A., SMITH, A. and ULLMAN, J. (2021). Manipulation attacks in local differential privacy. In 2021 *IEEE Symposium on Security and Privacy* (*SP*) 883–900. IEEE.

CHHOR, J. and SENTENAC, F. (2022). Robust estimation of discrete distributions under local differential privacy. Preprint. Available at arXiv:2202.06825.

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

DEVROYE, L. and LUGOSI, G. (2001). *Combinatorial Methods in Density Estimation. Springer Series in Statistics*. Springer, New York. MR1843146 https://doi.org/10.1007/978-1-4613-0125-7

DEVROYE, L., MEHRABIAN, A. and REDDAD, T. (2018). The total variation distance between high-dimensional Gaussians. Preprint. Available at arXiv:1810.08693.

DIAKONIKOLAS, I., KAMATH, G., KANE, D., LI, J., MOITRA, A. and STEWART, A. (2019). Robust estimators in high-dimensions without the computational intractability. *SIAM J. Comput.* **48** 742–864. MR3945261 https://doi.org/10.1137/17M1126680

DIAKONIKOLAS, I., KAMATH, G., KANE, D. M., LI, J., MOITRA, A. and STEWART, A. (2017). Being robust (in high dimensions) can be practical. In *International Conference on Machine Learning* 999–1008.

DIMITRAKAKIS, C., NELSON, B., MITROKOTSA, A. and RUBINSTEIN, B. I. P. (2014). Robust and private Bayesian inference. In *Algorithmic Learning Theory. Lecture Notes in Computer Science* **8776** 291–305. Springer, Cham. MR3295543 https://doi.org/10.1007/978-3-319-11662-4_21

DING, B., KULKARNI, J. and YEKHANIN, S. (2017). Collecting telemetry data privately. *Adv. Neural Inf. Process. Syst.* **30**.

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

DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2013). Local privacy, data processing inequalities, and statistical minimax rates. Preprint. Available at arXiv:1302.3203.

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

DWORK, C. and LEI, J. (2009). Differential privacy and robust statistics. In *STOC'*09*—Proceedings of the* 2009 *ACM International Symposium on Theory of Computing* 371–380. ACM, New York. MR2780083

DWORK, C., MCSHERRY, F., NISSIM, K. and SMITH, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography. Lecture Notes in Computer Science* **3876** 265–284. Springer, Berlin. MR2241676 https://doi.org/10.1007/11681878_14

ERLINGSSON, Ú., PIHUR, V. and KOROLOVA, A. (2014). Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the* 2014 *ACM SIGSAC Conference on Computer and Communications Security* 1054–1067.

ESFANDIARI, H., MIRROKNI, V. and NARAYANAN, S. (2021). Tight and robust private mean estimation with few users. Preprint. Available at arXiv:2110.11876.

FORTI, M. (2021). The deployment of artificial intelligence tools in the health sector: Privacy concerns and regulatory answers within the GDPR. *Eur. J. Leg. Stud.* **13** 29.

GHAZI, B., KUMAR, R., MANURANGSI, P. and NGUYEN, T. (2021). Robust and private learning of halfspaces. In *International Conference on Artificial Intelligence and Statistics* 1603–1611.

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

GOPI, S., KAMATH, G., KULKARNI, J., NIKOLOV, A., WU, Z. S. and ZHANG, H. (2020). Locally private hypothesis selection. In *Conference on Learning Theory* 1785–1816.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics*: *The Approach Based on Influence Functions. Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York. MR0829458

HUBER, P. J. (1968). Robust confidence limits. *Z. Wahrsch. Verw. Gebiete* **10** 269–278. MR0242330 https://doi.org/10.1007/BF00531848

HUBER, P. J. (1992). Robust estimation of a location parameter. In *Breakthroughs in Statistics* 492–518.

HUBER, P. J. (2004). *Robust Statistics*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0606374

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

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

JAIN, A., ORLITSKY, A. and RAVINDRAKUMAR, V. (2022). Robust estimation algorithms don't need to know the corruption level. Preprint. Available at arXiv:2202.05453.

JOSEPH, M., MAO, J., NEEL, S. and ROTH, A. (2019). The role of interactivity in local differential privacy. In 2019 *IEEE* 60*th Annual Symposium on Foundations of Computer Science* 94–105. IEEE Comput. Soc. Press, Los Alamitos, CA. MR4228162 https://doi.org/10.1109/FOCS.2019.00015

KAIROUZ, P., OH, S. and VISWANATH, P. (2016). Extremal mechanisms for local differential privacy. *J. Mach. Learn. Res.* **17** Paper No. 17, 51 pp. MR3491111

KAMATH, G., MOUZAKIS, A., SINGHAL, V., STEINKE, T. and ULLMAN, J. (2021). A private and computationally-efficient estimator for unbounded Gaussians. Preprint. Available at arXiv:2111.04609.

KARWA, V. and VADHAN, S. (2017). Finite sample differentially private confidence intervals. Preprint. Available at arXiv:1711.03908.

KOTHARI, P. K., MANURANGSI, P. and VELINGKER, A. (2021). Private robust estimation by stabilizing convex relaxations. Preprint. Available at arXiv:2112.03548.

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

LAM-WEIL, J., LAURENT, B. and LOUBES, J.-M. (2022). Minimax optimal goodness-of-fit testing for densities and multinomials under a local differential privacy constraint. *Bernoulli* **28** 579–600. MR4337717 https://doi.org/10.3150/21-bej1358

LI, M., BERRETT, T. B. and YU, Y. (2022). Network change point localisation under local differential privacy. *Adv. Neural Inf. Process. Syst.* **35**. 15013–15026.

LI, M., BERRETT, T. B. and YU, Y. (2023). Supplement to "On robustness and local differential privacy." https://doi.org/10.1214/23-AOS2267SUPP

LIU, X., KONG, W., KAKADE, S. and OH, S. (2021). Robust and differentially private mean estimation. Preprint. Available at arXiv:2102.09159.

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

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

MARONNA, R. A., MARTIN, R. D., YOHAI, V. J. and SALIBIÁN-BARRERA, M. (2019). *Robust Statistics*: *Theory and Methods* (*with R*). *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR3839299

PENSIA, A., JOG, V. and LOH, P.-L. (2020). Robust regression with covariate filtering: Heavy tails and adversarial contamination. Preprint. Available at arXiv:2009.12976.

PRASAD, A., BALAKRISHNAN, S. and RAVIKUMAR, P. (2019). A unified approach to robust mean estimation. Preprint. Available at arXiv:1907.00927.

ROHDE, A. and STEINBERGER, L. (2020). Geometrizing rates of convergence under local differential privacy constraints. *Ann. Statist.* **48** 2646–2670. MR4152116 https://doi.org/10.1214/19-AOS1901

TANG, J., KOROLOVA, A., BAI, X., WANG, X. and WANG, X. (2017). Privacy loss in Apple's implementation of differential privacy on MacOS 10.12. Preprint. Available at arXiv:1709.02753.

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

UPPAL, A., SINGH, S. and POCZOS, B. (2020). Robust density estimation under Besov IPM losses. *Adv. Neural Inf. Process. Syst.* **33**.

WARNER, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc.* **60** 63–69.

WASSERMAN, L. and ZHOU, S. (2010). A statistical framework for differential privacy. *J. Amer. Statist. Assoc.* **105** 375–389. MR2656057 https://doi.org/10.1198/jasa.2009.tm08651

YATRACOS, Y. G. (1985). Rates of convergence of minimum distance estimators and Kolmogorov's entropy. *Ann. Statist.* **13** 768–774. MR0790571 https://doi.org/10.1214/aos/1176349553

The Annals of Statistics
2023, Vol. 51, No. 2, 738–761
https://doi.org/10.1214/23-AOS2268
© Institute of Mathematical Statistics, 2023

OPTIMALLY TACKLING COVARIATE SHIFT IN RKHS-BASED NONPARAMETRIC REGRESSION

BY CONG MA[1,a], REESE PATHAK[2,b] AND MARTIN J. WAINWRIGHT[3,c]

[1]*Department of Statistics, University of Chicago,* [a]*congm@uchicago.edu*
[2]*Department of Electrical Engineering and Computer Sciences, University of California, Berkeley,* [b]*pathakr@berkeley.edu*
[3]*EECS and Mathematics, Statistics and Data Science Center, Massachusetts Institute of Technology,* [c]*wainwrigwork@gmail.com*

We study the covariate shift problem in the context of nonparametric regression over a reproducing kernel Hilbert space (RKHS). We focus on two natural families of covariate shift problems defined using the likelihood ratios between the source and target distributions. When the likelihood ratios are uniformly bounded, we prove that the kernel ridge regression (KRR) estimator with a carefully chosen regularization parameter is minimax rate-optimal (up to a log factor) for a large family of RKHSs with regular kernel eigenvalues. Interestingly, KRR does not require full knowledge of the likelihood ratio apart from an upper bound on it. In striking contrast to the standard statistical setting without covariate shift, we also demonstrate that a naïve estimator, which minimizes the empirical risk over the function class, is strictly suboptimal under covariate shift as compared to KRR. We then address the larger class of covariate shift problems where likelihood ratio is possibly unbounded yet has a finite second moment. Here, we propose a reweighted KRR estimator that weights samples based on a careful truncation of the likelihood ratios. Again, we are able to show that this estimator is minimax optimal, up to logarithmic factors.

REFERENCES

[1] CAI, T. T. and WEI, H. (2021). Transfer learning for nonparametric classification: Minimax rate and adaptive classifier. *Ann. Statist.* **49** 100–128. MR4206671 https://doi.org/10.1214/20-AOS1949

[2] 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

[3] CHEN, X. and YANG, Y. (2021). Hanson–Wright inequality in Hilbert spaces with application to K-means clustering for non-Euclidean data. *Bernoulli* **27** 586–614. MR4177382 https://doi.org/10.3150/20-BEJ1251

[4] CORTES, C., MANSOUR, Y. and MOHRI, M. (2010). Learning bounds for importance weighting. In *Advances in Neural Information Processing Systems* **23** 442–450.

[5] GUAN, H. and LIU, M. (2021). Domain adaptation for medical image analysis: A survey. *IEEE Trans. Biomed. Eng.* 1–1.

[6] 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

[7] JIANG, J. and ZHAI, C. (2007). Instance weighting for domain adaptation in NLP. In *Proceedings of the* 45*th Annual Meeting of the Association of Computational Linguistics* 264–271. Association for Computational Linguistics, Prague, Czech Republic.

[8] KOH, P. W., SAGAWA, S., MARKLUND, H., XIE, S. M., ZHANG, M., BALSUBRAMANI, A., HU, W., YASUNAGA, M., PHILLIPS, R. L. et al. (2021). Wilds: A benchmark of in-the-wild distribution shifts. In *International Conference on Machine Learning* 5637–5664. PMLR.

[9] KPOTUFE, S. and MARTINET, G. (2021). Marginal singularity and the benefits of labels in covariate-shift. *Ann. Statist.* **49** 3299–3323. MR4352531 https://doi.org/10.1214/21-aos2084

MSC2020 subject classifications. Primary 62C20; secondary 62G08.

Key words and phrases. Covariate shift, nonparametric regression, reproducing kernel Hilbert spaces, kernel ridge regression, transfer learning.

[10] MA, C., PATHAK, R. and WAINWRIGHT, M. J. (2023). Supplement to "Optimally tackling covariate shift in RKHS-based nonparametric regression." https://doi.org/10.1214/23-AOS2268SUPP
[11] MAITY, S., SUN, Y. and BANERJEE, M. (2020). Minimax optimal approaches to the label shift problem. ArXiv preprint. Available at arXiv:2003.10443.
[12] PAN, S. J. and YANG, Q. (2009). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **22** 1345–1359.
[13] PATHAK, R., MA, C. and WAINWRIGHT, M. J. (2022). A new similarity measure for covariate shift with applications to nonparametric regression. In *Proceedings of the* 39*th International Conference on Machine Learning*, *PMLR* **162** 17517–17530.
[14] POLLARD, D. (2012). *Convergence of Stochastic Processes*. Springer, Berlin.
[15] REEVE, H. W. J., CANNINGS, T. I. and SAMWORTH, R. J. (2021). Adaptive transfer learning. *Ann. Statist.* **49** 3618–3649. MR4352543 https://doi.org/10.1214/21-aos2102
[16] SCHMIDT-HIEBER, J. and ZAMOLODTCHIKOV, P. (2022). Local convergence rates of the least squares estimator with applications to transfer learning. ArXiv preprint. Available at arXiv:2204.05003.
[17] SCHÖLKOPF, B., SMOLA, A. J., BACH, F. et al. (2002). *Learning with Kernels*: *Support Vector Machines*, *Regularization*, *Optimization*, *and Beyond*. MIT Press, Cambridge.
[18] SHIMODAIRA, H. (2000). Improving predictive inference under covariate shift by weighting the log-likelihood function. *J. Statist. Plann. Inference* **90** 227–244. MR1795598 https://doi.org/10.1016/S0378-3758(00)00115-4
[19] SUGIYAMA, M. and KAWANABE, M. (2012). *Machine Learning in Non-stationary Environments*: *Introduction to Covariate Shift Adaptation*. MIT Press, Cambridge.
[20] VAPNIK, V. (1999). *The Nature of Statistical Learning Theory*. Springer, Berlin.
[21] 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
[22] WAINWRIGHT, M. J. (2019). *High-Dimensional Statistics*: *A Non-asymptotic Viewpoint*. *Cambridge Series in Statistical and Probabilistic Mathematics* **48**. Cambridge Univ. Press, Cambridge. MR3967104 https://doi.org/10.1017/9781108627771
[23] YANG, Y., PILANCI, M. and WAINWRIGHT, M. J. (2017). Randomized sketches for kernels: Fast and optimal nonparametric regression. *Ann. Statist.* **45** 991–1023. MR3662446 https://doi.org/10.1214/16-AOS1472

The Annals of Statistics
2023, Vol. 51, No. 2, 762–790
https://doi.org/10.1214/23-AOS2270
© Institute of Mathematical Statistics, 2023

MINIMAX RATES FOR CONDITIONAL DENSITY ESTIMATION VIA EMPIRICAL ENTROPY

BY BLAIR BILODEAU[1,a], DYLAN J. FOSTER[2,c] AND DANIEL M. ROY[1,b]

[1]*Department of Statistical Sciences, University of Toronto,* [a]*blair.bilodeau@mail.utoronto.ca,* [b]*daniel.roy@utoronto.ca*
[2]*Microsoft Research, New England,* [c]*dylanfoster@microsoft.com*

We consider the task of estimating a conditional density using i.i.d. samples from a joint distribution, which is a fundamental problem with applications in both classification and uncertainty quantification for regression. For joint density estimation, minimax rates have been characterized for general density classes in terms of uniform (metric) entropy, a well-studied notion of statistical capacity. When applying these results to conditional density estimation, the use of uniform entropy—which is infinite when the covariate space is unbounded and suffers from the curse of dimensionality—can lead to suboptimal rates. Consequently, minimax rates for conditional density estimation cannot be characterized using these classical results.

We resolve this problem for well-specified models, obtaining matching (within logarithmic factors) upper and lower bounds on the minimax Kullback–Leibler risk in terms of the *empirical Hellinger entropy* for the conditional density class. The use of empirical entropy allows us to appeal to concentration arguments based on local Rademacher complexity, which—in contrast to uniform entropy—leads to matching rates for large, potentially nonparametric classes and captures the correct dependence on the complexity of the covariate space. Our results require only that the conditional densities are bounded above, and do not require that they are bounded below or otherwise satisfy any tail conditions.

REFERENCES

[1] AGARWAL, A., KAKADE, S., KRISHNAMURTHY, A. and SUN, W. (2020). FLAMBE: Structural complexity and representation learning of low rank MDPs. In *Advances in Neural Information Processing Systems*, *Vol*. 33.

[2] BARAUD, Y. and CHEN, J. (2020). Robust estimation of a regression function in exponential families. Available at arXiv:2011.01657.

[3] BARRON, A., RISSANEN, J. and YU, B. (1998). The minimum description length principle in coding and modeling **44** 2743–2760. MR1658898 https://doi.org/10.1109/18.720554

[4] BARTLETT, P. L. (1998). The sample complexity of pattern classification with neural networks: The size of the weights is more important than the size of the network. *IEEE Trans. Inf. Theory* **44** 525–536. MR1607706 https://doi.org/10.1109/18.661502

[5] BARTLETT, P. L., BOUSQUET, O. and MENDELSON, S. (2005). Local Rademacher complexities. *Ann. Statist*. **33** 1497–1537. MR2166554 https://doi.org/10.1214/009053605000000282

[6] 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

[7] BEN-DAVID, S. and SHALEV-SHWARTZ, S. (2014). *Understanding Machine Learning*: *From Theory to Algorithms*. Cambridge Univ. Press, Cambridge.

[8] BHATT, A. and KIM, Y.-H. (2021). Sequential prediction under log-loss with side information. In *Algorithmic Learning Theory*. *Proc. Mach. Learn. Res*. **132** 5. PMLR, Online. MR4227325

[9] BILODEAU, B., FOSTER, D. J. and ROY, D. M. (2020). Tight bounds on minimax regret under logarithmic loss via self-concordance. In *Proceedings of the* 37*th International Conference on Machine Learning*.

MSC2020 subject classifications. Primary 62G07, 62G05; secondary 62C20, 94A17.

Key words and phrases. Conditional density estimation, empirical entropy, logarithmic loss, nonparametric estimation, minimax rates.

[10] BILODEAU, B., FOSTER, D. J. and ROY, D. M. (2023). Supplement to "Minimax rates for conditional density estimation via empirical entropy." https://doi.org/10.1214/23-AOS2270SUPP
[11] BIRGÉ, L. (1983). Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsch. Verw. Gebiete* **65** 181–237. MR0722129 https://doi.org/10.1007/BF00532480
[12] BIRGÉ, L. (1986). On estimating a density using Hellinger distance and some other strange facts. *Probab. Theory Related Fields* **71** 271–291. MR0816706 https://doi.org/10.1007/BF00332312
[13] BIRGÉ, L. (1987). Estimating a density under order restrictions: Nonasymptotic minimax risk. *Ann. Statist.* **15** 995–1012. MR0902241 https://doi.org/10.1214/aos/1176350488
[14] BIRGÉ, L. (2013). Robust tests for model selection. In *From Probability to Statistics and Back: High-Dimensional Models and Processes. Inst. Math. Stat. (IMS) Collect.* **9** 47–64. IMS, Beachwood, OH. MR3186748 https://doi.org/10.1214/12-IMSCOLL905
[15] BIRGÉ, L. and MASSART, P. (1993). Rates of convergence for minimum contrast estimators. *Probab. Theory Related Fields* **97** 113–150. MR1240719 https://doi.org/10.1007/BF01199316
[16] BIRGÉ, L. and MASSART, P. (1998). Minimum contrast estimators on sieves: Exponential bounds and rates of convergence. *Bernoulli* **4** 329–375. MR1653272 https://doi.org/10.2307/3318720
[17] BOTT, A.-K. and KOHLER, M. (2017). Nonparametric estimation of a conditional density. *Ann. Inst. Statist. Math.* **69** 189–214. MR3590717 https://doi.org/10.1007/s10463-015-0535-8
[18] BOUSQUET, O. (2002). Concentration inequalities and empirical processes theory applied to the analysis of learning algorithms. Ph.D. thesis, École Polytechnique.
[19] BOUSQUET, O., KOLTCHINSKII, V. and PANCHENKO, D. (2002). Some local measures of complexity of convex hulls and generalization bounds. In *Computational Learning Theory* (*Sydney*, 2002). *Lecture Notes in Computer Science* **2375** 59–73. Springer, Berlin. MR2040405 https://doi.org/10.1007/3-540-45435-7_5
[20] BOYD, D. W. and STEELE, J. M. (1978). Lower bounds for nonparametric density estimation rates. *Ann. Statist.* **6** 932–934. MR0494674
[21] CESA-BIANCHI, N. and LUGOSI, G. (1999). Minimax regret under log loss for general classes of experts. In *Proceedings of the Twelfth Annual Conference on Computational Learning Theory* (*Santa Cruz, CA*, 1999) 12–18. ACM, New York. MR1811597 https://doi.org/10.1145/307400.307407
[22] CESA-BIANCHI, N. and LUGOSI, G. (2006). *Prediction, Learning, and Games*. Cambridge Univ. Press, Cambridge. MR2409394 https://doi.org/10.1017/CBO9780511546921
[23] COHEN, S. and LE PENNEC, E. (2011). Conditional density estimation by penalized likelihood model selection and applications. Available at arXiv:1103.2021.
[24] COVER, T. M. (1991). Universal portfolios. *Math. Finance* **1** 1–29. MR1113417 https://doi.org/10.1111/j.1467-9965.1991.tb00002.x
[25] COVER, T. M. and THOMAS, J. A. (1991). *Elements of Information Theory. Wiley Series in Telecommunications*. Wiley, New York. MR1122806 https://doi.org/10.1002/0471200611
[26] DE HEIDE, R., KIRICHENKO, A., MEHTA, N. A. and GRÜNWALD, P. (2020). Safe-Bayesian generalized linear regression. In *Proceedings of the* 23*rd International Conference on Artificial Intelligence and Statistics*.
[27] DEVROYE, L. (1983). On arbitrarily slow rates of global convergence in density estimation. *Z. Wahrsch. Verw. Gebiete* **62** 475–483. MR0690572 https://doi.org/10.1007/BF00534199
[28] DEVROYE, L. and REDDAD, T. (2019). Discrete minimax estimation with trees. *Electron. J. Stat.* **13** 2595–2623. MR3992499 https://doi.org/10.1214/19-EJS1586
[29] DOSS, C. R. and WELLNER, J. A. (2016). Global rates of convergence of the MLEs of log-concave and s-concave densities. *Ann. Statist.* **44** 954–981. MR3485950 https://doi.org/10.1214/15-AOS1394
[30] DUDLEY, R. M. (1974). Metric entropy of some classes of sets with differentiable boundaries. *J. Approx. Theory* **10** 227–236. MR0358168 https://doi.org/10.1016/0021-9045(74)90120-8
[31] EFROMOVICH, S. (2007). Conditional density estimation in a regression setting. *Ann. Statist.* **35** 2504–2535. MR2382656 https://doi.org/10.1214/009053607000000253
[32] FAHRMEIR, L. and KAUFMANN, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann. Statist.* **13** 342–368. MR0773172 https://doi.org/10.1214/aos/1176346597
[33] FOSTER, D. J., KALE, S., LUO, H., MOHRI, M. and SRIDHARAN, K. (2018). Logistic regression: The importance of being improper. In *Proceedings of the* 31*st Conference on Learning Theory*.
[34] FOSTER, D. J. and KRISHNAMURTHY, A. (2021). Efficient first-order contextual bandits: Prediction, allocation, and triangular discrimination. In *Advances in Neural Information Processing Systems*, *Vol.* 34.
[35] GAO, F. and WELLNER, J. A. (2009). On the rate of convergence of the maximum likelihood estimator of a k-monotone density. *Sci. China Ser. A* **52** 1525–1538. MR2520591 https://doi.org/10.1007/s11425-009-0102-y

[36] GEMAN, S. and HWANG, C.-R. (1982). Nonparametric maximum likelihood estimation by the method of sieves. *Ann. Statist.* **10** 401–414. MR0653512
[37] GRÜNWALD, P. (2011). Safe learning: Bridging the gap between Bayes, MDL, and statistical learning theory via empirical convexity. In *Proceedings of the 24th Conference on Learning Theory*.
[38] 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
[39] GRÜNWALD, P. and MEHTA, N. A. (2019). A tight excess risk bound via a unified PAC-Bayesian–Rademacher–Shtarkov-MDL complexity. In *Algorithmic Learning Theory* 2019. *Proc. Mach. Learn. Res.* **98** 433–465, 33. PMLR, Chicago, IL. MR3932854
[40] GRÜNWALD, P. and MEHTA, N. A. (2020). Fast rates for general unbounded loss functions: From ERM to generalized Bayes. *J. Mach. Learn. Res.* **21** Paper No. 56, 80. MR4095335
[41] GYÖRFI, L. and KOHLER, M. (2007). Nonparametric estimation of conditional distributions. *IEEE Trans. Inf. Theory* **53** 1872–1879. MR2317148 https://doi.org/10.1109/TIT.2007.894631
[42] GYÖRFI, L., KOHLER, M., KRYŻAK, A. and WALK, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. Springer, New York.
[43] HALL, P., RACINE, J. and LI, Q. (2004). Cross-validation and the estimation of conditional probability densities. *J. Amer. Statist. Assoc.* **99** 1015–1026. MR2109491 https://doi.org/10.1198/016214504000000548
[44] HAN, Q. (2021). Set structured global empirical risk minimizers are rate optimal in general dimensions. *Ann. Statist.* **49** 2642–2671. MR4338378 https://doi.org/10.1214/21-aos2049
[45] HAUSSLER, D. (1992). Decision-theoretic generalizations of the PAC model for neural net and other learning applications. *Inform. and Comput.* **100** 78–150. MR1175977 https://doi.org/10.1016/0890-5401(92)90010-D
[46] HOLMES, M. P., GRAY, A. G. and ISBELL, JR., C. L. (2007). Fast nonparametric conditional density estimation. In *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence*.
[47] IBRAGIMOV, I. A. and KHAS'MINSKII, R. Z. (1983). Estimation of distribution density. *J. Sov. Math.* **21** 40–57.
[48] KIM, A. K. H. and SAMWORTH, R. J. (2016). Global rates of convergence in log-concave density estimation. *Ann. Statist.* **44** 2756–2779. MR3576560 https://doi.org/10.1214/16-AOS1480
[49] KOLTCHINSKII, V. (2001). Rademacher penalties and structural risk minimization. *IEEE Trans. Inf. Theory* **47** 1902–1914. MR1842526 https://doi.org/10.1109/18.930926
[50] KOLTCHINSKII, V. (2011). *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems. Lecture Notes in Math.* **2033**. Springer, Heidelberg. MR2829871 https://doi.org/10.1007/978-3-642-22147-7
[51] KOLTCHINSKII, V. and PANCHENKO, D. (2000). Rademacher processes and bounding the risk of function learning. In *High Dimensional Probability, II* (*Seattle, WA*, 1999). *Progress in Probability* **47** 443–457. Birkhäuser, Boston, MA. MR1857339
[52] KUR, G., DAGAN, Y. and RAKHLIN, A. (2019). Optimality of maximum likelihood for log-concave density estimation and bounded convex regression. Available at arXiv:1903.05315.
[53] KUR, G., GAO, F., GUNTUBOYINA, A. and SEN, B. (2020). Convex regression in multidimensions: Suboptimality of least squares estimators. Available at arXiv:2006.02044.
[54] LECAM, L. (1973). Convergence of estimates under dimensionality restrictions. *Ann. Statist.* **1** 38–53. MR0334381
[55] LECUÉ, G. (2006). Lower bounds and aggregation in density estimation. *J. Mach. Learn. Res.* **7** 971–981. MR2274393
[56] LEE, K.-Y. and COURTADE, T. A. (2020). Minimax bounds for generalized linear models. In *Advances in Neural Information Processing Systems*, *Vol.* 33.
[57] LI, M., NEYKOV, M. and BALAKRISHNAN, S. (2022). Minimax optimal conditional density estimation under total variation smoothness. *Electron. J. Stat.* **16** 3937–3972. MR4453593 https://doi.org/10.1214/22-ejs2037
[58] LUGOSI, G. and NOBEL, A. B. (1999). Adaptive model selection using empirical complexities. *Ann. Statist.* **27** 1830–1864. MR1765619 https://doi.org/10.1214/aos/1017939242
[59] 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
[60] MOURTADA, J. and GAÏFFAS, S. (2022). An improper estimator with optimal excess risk in misspecified density estimation and logistic regression. *J. Mach. Learn. Res.* **23** Paper No. 31, 49. MR4420756
[61] OPPER, M. and HAUSSLER, D. (1999). Worst case prediction over sequences under log loss. In *The Mathematics of Information Coding, Extraction and Distribution* (G. Cybenko, D. P. O'Leary and J. Rissanen, eds.) 81–90. Springer, New York.

[62] PARK, J. and MUANDET, K. (2022). Towards empirical process theory for vector-valued functions: metric entropy of smooth function classes. Available at arXiv:2202.04415.
[63] RAKHLIN, A. and SRIDHARAN, K. (2015). Sequential probability assignment with binary alphabets and large classes of experts. Available at arXiv:1501.07340.
[64] RAKHLIN, A., SRIDHARAN, K. and TEWARI, A. (2010). Online learning: Random averages, combinatorial parameters, and learnability. In *Advances in Neural Information Processing Systems*, *Vol*. 23.
[65] RAKHLIN, A., SRIDHARAN, K. and TSYBAKOV, A. B. (2017). Empirical entropy, minimax regret and minimax risk. *Bernoulli* **23** 789–824. MR3606751 https://doi.org/10.3150/14-BEJ679
[66] RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2011). Minimax rates of estimation for high-dimensional linear regression over ℓ_q-balls. *IEEE Trans. Inf. Theory* **57** 6976–6994. MR2882274 https://doi.org/10.1109/TIT.2011.2165799
[67] RIGOLLET, P. (2012). Kullback–Leibler aggregation and misspecified generalized linear models. *Ann. Statist.* **40** 639–665. MR2933661 https://doi.org/10.1214/11-AOS961
[68] RISSANEN, J. (1984). Universal coding, information, prediction, and estimation. *IEEE Trans. Inf. Theory* **30** 629–636. MR0755791 https://doi.org/10.1109/TIT.1984.1056936
[69] SAMWORTH, R. J. (2018). Recent progress in log-concave density estimation. *Statist. Sci.* **33** 493–509. MR3881205 https://doi.org/10.1214/18-STS666
[70] SART, M. (2017). Estimating the conditional density by histogram type estimators and model selection. *ESAIM Probab. Stat.* **21** 34–55. MR3630602 https://doi.org/10.1051/ps/2016026
[71] SHEN, X. (1997). On methods of sieves and penalization. *Ann. Statist.* **25** 2555–2591. MR1604416 https://doi.org/10.1214/aos/1030741085
[72] SHEN, X. and WONG, W. H. (1994). Convergence rate of sieve estimates. *Ann. Statist.* **22** 580–615. MR1292531 https://doi.org/10.1214/aos/1176325486
[73] STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360. MR0594650
[74] 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
[75] VAN DE GEER, S. (1990). Estimating a regression function. *Ann. Statist.* **18** 907–924. MR1056343 https://doi.org/10.1214/aos/1176347632
[76] 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
[77] WAINWRIGHT, M. J. (2019). *High-Dimensional Statistics*: *A Non-asymptotic Viewpoint*. *Cambridge Series in Statistical and Probabilistic Mathematics* **48**. Cambridge Univ. Press, Cambridge. MR3967104 https://doi.org/10.1017/9781108627771
[78] WONG, W. H. and SHEN, X. (1995). Probability inequalities for likelihood ratios and convergence rates of sieve MLEs. *Ann. Statist.* **23** 339–362. MR1332570 https://doi.org/10.1214/aos/1176324524
[79] YANG, Y. (1999). Minimax nonparametric classification. I. Rates of convergence. *IEEE Trans. Inf. Theory* **45** 2271–2284. MR1725115 https://doi.org/10.1109/18.796368
[80] YANG, Y. and BARRON, A. (1999). Information-theoretic determination of minimax rates of convergence. *Ann. Statist.* **27** 1564–1599. MR1742500 https://doi.org/10.1214/aos/1017939142
[81] YANG, Y. and BARRON, A. R. (1998). An asymptotic property of model selection criteria. *IEEE Trans. Inf. Theory* **44** 95–116. MR1486651 https://doi.org/10.1109/18.650993
[82] ZHANG, T. (2006). Information-theoretic upper and lower bounds for statistical estimation. *IEEE Trans. Inf. Theory* **52** 1307–1321. MR2241190 https://doi.org/10.1109/TIT.2005.864439

The Annals of Statistics
2023, Vol. 51, No. 2, 791–815
https://doi.org/10.1214/23-AOS2273
© Institute of Mathematical Statistics, 2023

INFERENCE ON THE MAXIMAL RANK OF TIME-VARYING COVARIANCE MATRICES USING HIGH-FREQUENCY DATA

BY MARKUS REISS[1,a] AND LARS WINKELMANN[2,b]

[1]*Institute of Mathematics, Humboldt-Universität zu Berlin,* [a]*mreiss@math.hu-berlin.de*
[2]*School of Business and Economics, Freie Universität Berlin,* [b]*lars.winkelmann@fu-berlin.de*

We study the rank of the instantaneous or spot covariance matrix $\Sigma_X(t)$ of a multidimensional process $X(t)$. Given high-frequency observations $X(i/n)$, $i = 0, \dots, n$, we test the null hypothesis $\operatorname{rank}(\Sigma_X(t)) \le r$ for all t against local alternatives where the average $(r+1)$st eigenvalue is larger than some signal detection rate v_n.

A major problem is that the inherent averaging in local covariance statistics produces a bias that distorts the rank statistics. We show that the bias depends on the regularity and spectral gap of $\Sigma_X(t)$. We establish explicit matrix perturbation and concentration results that provide nonasymptotic uniform critical values and optimal signal detection rates v_n. This leads to a rank estimation method via sequential testing. For a class of stochastic volatility models, we determine data-driven critical values via normed p-variations of estimated local covariance matrices. The methods are illustrated by simulations and an application to high-frequency data of U.S. government bonds.

REFERENCES

AÏT-SAHALIA, Y. and XIU, D. (2017). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *J. Econometrics* **201** 384–399. MR3717571 https://doi.org/10.1016/j.jeconom.2017.08.015

AÏT-SAHALIA, Y. and XIU, D. (2019). Principal component analysis of high-frequency data. *J. Amer. Statist. Assoc.* **114** 287–303. MR3941255 https://doi.org/10.1080/01621459.2017.1401542

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

BLANCHARD, G., CARPENTIER, A. and GUTZEIT, M. (2018). Minimax Euclidean separation rates for testing convex hypotheses in $\mathbb{R}^d$. *Electron. J. Stat.* **12** 3713–3735. MR3873534 https://doi.org/10.1214/18-ejs1472

BRU, M.-F. (1991). Wishart processes. *J. Theoret. Probab.* **4** 725–751. MR1132135 https://doi.org/10.1007/BF01259552

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

CHRISTENSEN, K., NIELSEN, M. S. and PODOLSKIJ, M. (2022). High-dimensional estimation of quadratic variation based on penalized realized variance. *Stat. Inference Stoch. Process.* https://doi.org/10.1007/s11203-022-09282-8

CIESIELSKI, Z., KERKYACHARIAN, G. and ROYNETTE, B. (1993). Quelques espaces fonctionnels associés à des processus gaussiens. *Studia Math.* **107** 171–204. MR1244574 https://doi.org/10.4064/sm-107-2-171-204

COHEN, A. (2003). *Numerical Analysis of Wavelet Methods. Studies in Mathematics and Its Applications* **32**. North-Holland, Amsterdam. MR1990555

DIEBOLD, F. X. and LI, C. (2006). Forecasting the term structure of government bond yields. *J. Econometrics* **130** 337–364. MR2211798 https://doi.org/10.1016/j.jeconom.2005.03.005

FAN, J. and WANG, Y. (2008). Spot volatility estimation for high-frequency data. *Stat. Interface* **1** 279–288. MR2476744 https://doi.org/10.4310/SII.2008.v1.n2.a5

HOFFMANN, M. (2002). Rate of convergence for parametric estimation in a stochastic volatility model. *Stochastic Process. Appl.* **97** 147–170. MR1870964 https://doi.org/10.1016/S0304-4149(01)00130-2

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

MSC2020 subject classifications. 62G10, 62M07, 62P05, 60B20.

Key words and phrases. Empirical covariance matrix, rank detection, signal detection rate, matrix concentration, eigenvalue perturbation, principal component analysis, factor model, term structure.

JACOD, J., LEJAY, A. and TALAY, D. (2008). Estimation of the Brownian dimension of a continuous Itô process. *Bernoulli* **14** 469–498. MR2544098 https://doi.org/10.3150/07-BEJ6190
JACOD, J. and PODOLSKIJ, M. (2013). A test for the rank of the volatility process: The random perturbation approach. *Ann. Statist.* **41** 2391–2427. MR3127870 https://doi.org/10.1214/13-AOS1153
JACOD, J. and ROSENBAUM, M. (2013). Quarticity and other functionals of volatility: Efficient estimation. *Ann. Statist.* **41** 1462–1484. MR3113818 https://doi.org/10.1214/13-AOS1115
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
JUDITSKY, A. and NEMIROVSKI, A. (2002). On nonparametric tests of positivity/monotonicity/convexity. *Ann. Statist.* **30** 498–527. MR1902897 https://doi.org/10.1214/aos/1021379863
LAM, C. and YAO, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *Ann. Statist.* **40** 694–726. MR2933663 https://doi.org/10.1214/12-AOS970
LI, W., GAO, J., LI, K. and YAO, Q. (2016). Modeling multivariate volatilities via latent common factors. *J. Bus. Econom. Statist.* **34** 564–573. MR3547996 https://doi.org/10.1080/07350015.2015.1092975
NELSON, C. and SIEGEL, A. F. (1987). Parsimonious modeling of yield curves. *J. Bus.* **60** 473–489.
ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2014). Signal detection in high dimension: The multispiked case. *Ann. Statist.* **42** 225–254. MR3189485 https://doi.org/10.1214/13-AOS1181
REISS, M. and WAHL, M. (2020). Nonasymptotic upper bounds for the reconstruction error of PCA. *Ann. Statist.* **48** 1098–1123. MR4102689 https://doi.org/10.1214/19-AOS1839
REISS, M. and WINKELMANN, L. (2023). Supplement to "Inference on the maximal rank of time-varying covariance matrices using high-frequency data." https://doi.org/10.1214/23-AOS2273SUPP
SU, L. and WANG, X. (2017). On time-varying factor models: Estimation and testing. *J. Econometrics* **198** 84–101. MR3628100 https://doi.org/10.1016/j.jeconom.2016.12.004
TAO, T. (2012). *Topics in Random Matrix Theory*. *Graduate Studies in Mathematics* **132**. Amer. Math. Soc., Providence, RI. MR2906465 https://doi.org/10.1090/gsm/132
TROPP, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.* **12** 389–434. MR2946459 https://doi.org/10.1007/s10208-011-9099-z
VETTER, M. (2015). Estimation of integrated volatility of volatility with applications to goodness-of-fit testing. *Bernoulli* **21** 2393–2418. MR3378471 https://doi.org/10.3150/14-BEJ648

The Annals of Statistics
2023, Vol. 51, No. 2, 816–845
https://doi.org/10.1214/23-AOS2276
© Institute of Mathematical Statistics, 2023

CONFORMAL PREDICTION BEYOND EXCHANGEABILITY

BY RINA FOYGEL BARBER[1,a], EMMANUEL J. CANDÈS[2,b], ADITYA RAMDAS[3,c] AND RYAN J. TIBSHIRANI[3,d]

[1]*Department of Statistics, University of Chicago,* [a]*rina@uchicago.edu*
[2]*Departments of Statistics and Mathematics, Stanford University,* [b]*candes@stanford.edu*
[3]*Departments of Statistics and Machine Learning, Carnegie Mellon University,* [c]*aramdas@cmu.edu,* [d]*ryantibs@cmu.edu*

Conformal prediction is a popular, modern technique for providing valid predictive inference for arbitrary machine learning models. Its validity relies on the assumptions of exchangeability of the data, and symmetry of the given model fitting algorithm as a function of the data. However, exchangeability is often violated when predictive models are deployed in practice. For example, if the data distribution drifts over time, then the data points are no longer exchangeable; moreover, in such settings, we might want to use a nonsymmetric algorithm that treats recent observations as more relevant. This paper generalizes conformal prediction to deal with both aspects: we employ weighted quantiles to introduce robustness against distribution drift, and design a new randomization technique to allow for algorithms that do not treat data points symmetrically. Our new methods are provably robust, with substantially less loss of coverage when exchangeability is violated due to distribution drift or other challenging features of real data, while also achieving the same coverage guarantees as existing conformal prediction methods if the data points are in fact exchangeable. We demonstrate the practical utility of these new tools with simulations and real-data experiments on electricity and election forecasting.

REFERENCES

ANGELOPOULOS, A. and BATES, S. (2023). Conformal prediction: A gentle introduction. *Found. Trends Mach. Learn.* **16** 495–591.

BARBER, R. F., CANDÈS, E. J., RAMDAS, A. and TIBSHIRANI, R. J. (2021). Predictive inference with the jackknife+. *Ann. Statist.* **49** 486–507. MR4206687 https://doi.org/10.1214/20-AOS1965

BARBER, R. F., CANDÈS, E. J., RAMDAS, A. and TIBSHIRANI, R. J. (2023). Supplement to "Conformal prediction beyond exchangeability." https://doi.org/10.1214/23-AOS2276SUPP

BATES, S., CANDÈS, E., LEI, L., ROMANO, Y. and SESIA, M. (2023). Testing for outliers with conformal p-values. *Ann. Statist.* **51** 149–178. MR4564852 https://doi.org/10.1214/22-aos2244

BURNAEV, E. and VOVK, V. (2014). Efficiency of conformalized ridge regression. In *Conference on Learning Theory* 605–622.

CANDÈS, E. J., LEI, L. and REN, Z. (2023). Conformalized survival analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **85** 24–45. https://doi.org/10.1093/jrsssb/qkac004

CAUCHOIS, M., GUPTA, S. and ALI, A. (2020). Robust validation: Confident predictions even when distributions shift. arXiv preprint, arXiv:2008.04267.

CHERIAN, J. and BRONNER, L. (2020). How the Washington Post estimates outstanding votes for the 2020 presidential election. Available at https://s3.us-east-1.amazonaws.com/elex-models-prod/2020-general/write-up/election_model_writeup.pdf.

CHERNOZHUKOV, V., WÜTHRICH, K. and YINCHU, Z. (2018). Exact and robust conformal inference methods for predictive machine learning with dependent data. In *Conference on Learning Theory* 732–749. PMLR.

DUNN, R., WASSERMAN, L. and RAMDAS, A. (2022). Distribution-free prediction sets for two-layer hierarchical models. *J. Amer. Statist. Assoc.* To appear.

MSC2020 subject classifications. Primary 62G35; secondary 62F40, 62G86.

Key words and phrases. Conformal prediction, distribution-free inference, robust statistics, jackknife, exchangeability.

FANNJIANG, C., BATES, S., ANGELOPOULOS, A. N., LISTGARTEN, J. and JORDAN, M. I. (2022). Conformal prediction for the design problem. arXiv preprint arXiv:2202.03613.

GIBBS, I. and CANDÈS, E. J. (2021). Adaptive conformal inference under distribution shift. *Adv. Neural Inf. Process. Syst.* **34**.

GUAN, L. (2023). Localized conformal prediction: A generalized inference framework for conformal prediction. *Biometrika* **110** 33–50. MR4565442 https://doi.org/10.1093/biomet/asac040

HARRIES, M. (1999). Splice-2 comparative evaluation: Electricity pricing. Tech. rept., Univ.New South Wales.

HARRISON, M. T. (2012). Conservative hypothesis tests and confidence intervals using importance sampling. *Biometrika* **99** 57–69. MR2899663 https://doi.org/10.1093/biomet/asr079

KIVARANOVIC, D., JOHNSON, K. D. and LEEB, H. (2020). Adaptive, distribution-free prediction intervals for deep networks. In *International Conference on Artificial Intelligence and Statistics*. PMLR.

LEI, J. (2019). Fast exact conformalization of the lasso using piecewise linear homotopy. *Biometrika* **106** 749–764. MR4031197 https://doi.org/10.1093/biomet/asz046

LEI, J., G'SELL, M., RINALDO, A., TIBSHIRANI, R. J. and WASSERMAN, L. (2018). Distribution-free predictive inference for regression. *J. Amer. Statist. Assoc.* **113** 1094–1111. MR3862342 https://doi.org/10.1080/01621459.2017.1307116

LEI, J., ROBINS, J. and WASSERMAN, L. (2013). Distribution-free prediction sets. *J. Amer. Statist. Assoc.* **108** 278–287. MR3174619 https://doi.org/10.1080/01621459.2012.751873

LEI, J. and WASSERMAN, L. (2014). Distribution-free prediction bands for non-parametric regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 71–96. MR3153934 https://doi.org/10.1111/rssb.12021

LEI, L. and CANDÈS, E. J. (2021). Conformal inference of counterfactuals and individual treatment effects. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 911–938. MR4349122

MAO, H., MARTIN, R. and REICH, B. (2023). Valid model-free spatial prediction. *J. Amer. Statist. Assoc.* To appear.

PODKOPAEV, A. and RAMDAS, A. (2021). Distribution-free uncertainty quantification for classification under label shift. In *Uncertainty in Artificial Intelligence*. PMLR.

ROMANO, Y., PATTERSON, E. and CANDÈS, E. J. (2019). Conformalized quantile regression. *Adv. Neural Inf. Process. Syst.* **32**.

SHAFER, G. and VOVK, V. (2008). A tutorial on conformal prediction. *J. Mach. Learn. Res.* **9** 371–421. MR2417240

STANKEVICIUTE, K., ALAA, A. M. and VAN DER SCHAAR, M. (2021). Conformal time-series forecasting. *Adv. Neural Inf. Process. Syst.* **34**.

TIBSHIRANI, R. J., BARBER, R. F., CANDÈS, E. J. and RAMDAS, A. (2019). Conformal prediction under covariate shift. *Adv. Neural Inf. Process. Syst.* **32**.

VOLKHONSKIY, D., BURNAEV, E., NOURETDINOV, I., GAMMERMAN, A. and VOVK, V. (2017). Inductive conformal martingales for change-point detection. In *Conformal and Probabilistic Prediction and Applications* 132–153. PMLR.

VOVK, V. (2015). Cross-conformal predictors. *Ann. Math. Artif. Intell.* **74** 9–28. MR3353894 https://doi.org/10.1007/s10472-013-9368-4

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

VOVK, V., GAMMERMAN, A. and SHAFER, G. (2005). *Algorithmic Learning in a Random World*. Springer, New York. MR2161220

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

VOVK, V., PETEJ, I. and GAMMERMAN, A. (2021). Protected probabilistic classification. In *Conformal and Probabilistic Prediction and Applications* 297–299. PMLR.

XU, C. and XIE, Y. (2021). Conformal prediction interval for dynamic time-series. In *International Conference on Machine Learning*. PMLR.

ZAFFRAN, M., FÉRON, O., GOUDE, Y. and JOSSE, J. (2022). Adaptive conformal predictions for time series. In *International Conference on Machine Learning*. PMLR.

The Annals of Statistics
2023, Vol. 51, No. 2, 846–877
https://doi.org/10.1214/23-AOS2278
© Institute of Mathematical Statistics, 2023

RATE-OPTIMAL ROBUST ESTIMATION OF HIGH-DIMENSIONAL VECTOR AUTOREGRESSIVE MODELS

BY DI WANG[1,a] AND RUEY S. TSAY[2,b]

[1]*School of Mathematical Sciences, Shanghai Jiao Tong University,* [a]*di.wang@sjtu.edu.cn*
[2]*Booth School of Business, University of Chicago,* [b]*ruey.tsay@chicagobooth.edu*

High-dimensional time series data appear in many scientific areas in the current data-rich environment. Analysis of such data poses new challenges to data analysts because of not only the complicated dynamic dependence between the series, but also the existence of aberrant observations, such as missing values, contaminated observations, and heavy-tailed distributions. For high-dimensional vector autoregressive (VAR) models, we introduce a unified estimation procedure that is robust to model misspecification, heavy-tailed noise contamination, and conditional heteroscedasticity. The proposed methodology enjoys both statistical optimality and computational efficiency, and can handle many popular high-dimensional models, such as sparse, reduced-rank, banded, and network-structured VAR models. With proper regularization and data truncation, the estimation convergence rates are shown to be almost optimal in the minimax sense under a bounded $(2 + 2\epsilon)$th moment condition. When $\epsilon \geq 1$, the rates of convergence match those obtained under the sub-Gaussian assumption. Consistency of the proposed estimators is also established for some $\epsilon \in (0, 1)$, with minimax optimal convergence rates associated with ϵ. The efficacy of the proposed estimation methods is demonstrated by simulation and a U.S. macroeconomic example.

REFERENCES

AVELLA-MEDINA, M., BATTEY, H. S., FAN, J. and LI, Q. (2018). Robust estimation of high-dimensional covariance and precision matrices. *Biometrika* **105** 271–284. MR3804402 https://doi.org/10.1093/biomet/asy011

BANNA, M., MERLEVÈDE, F. and YOUSSEF, P. (2016). Bernstein-type inequality for a class of dependent random matrices. *Random Matrices Theory Appl.* **5** 1650006, 28. MR3493552 https://doi.org/10.1142/S2010326316500064

BASU, S., LI, X. and MICHAILIDIS, G. (2019). Low rank and structured modeling of high-dimensional vector autoregressions. *IEEE Trans. Signal Process.* **67** 1207–1222. MR3925874 https://doi.org/10.1109/TSP.2018.2887401

BASU, S. and MICHAILIDIS, G. (2015). Regularized estimation in sparse high-dimensional time series models. *Ann. Statist.* **43** 1535–1567. MR3357870 https://doi.org/10.1214/15-AOS1315

BELLONI, A. and CHERNOZHUKOV, V. (2011). ℓ_1-penalized quantile regression in high-dimensional sparse models. *Ann. Statist.* **39** 82–130. MR2797841 https://doi.org/10.1214/10-AOS827

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

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

BUBECK, S., CESA-BIANCHI, N. and LUGOSI, G. (2013). Bandits with heavy tail. *IEEE Trans. Inf. Theory* **59** 7711–7717. MR3124669 https://doi.org/10.1109/TIT.2013.2277869

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

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

MSC2020 subject classifications. Primary 62M10, 62F35; secondary 62J07.

Key words and phrases. Autocovariance, high-dimensional time series, minimax optimal, robust statistics, truncation.

CANDÈS, E. J. and PLAN, Y. (2011). Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Trans. Inf. Theory* **57** 2342–2359. MR2809094 https://doi.org/10.1109/TIT.2011.2111771

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

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

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

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., WANG, W. and ZHU, Z. (2021). A shrinkage principle for heavy-tailed data: High-dimensional robust low-rank matrix recovery. *Ann. Statist.* **49** 1239–1266. MR4298863 https://doi.org/10.1214/20-aos1980

GORROSTIETA, C., OMBAO, H., BÉDARD, P. and SANES, J. N. (2012). Investigating brain connectivity using mixed effects vector autoregressive models. *NeuroImage* **59** 3347–3355. https://doi.org/10.1016/j.neuroimage.2011.08.115

GUO, S., WANG, Y. and YAO, Q. (2016). High-dimensional and banded vector autoregressions. *Biometrika* **103** 889–903. MR3620446 https://doi.org/10.1093/biomet/asw046

HAMPEL, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Stat.* **42** 1887–1896. MR0301858 https://doi.org/10.1214/aoms/1177693054

HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393. MR0362657

HAMPEL, F. R. (2001). Robust statistics: A brief introduction and overview. In *Research Report/Seminar Für Statistik, Eidgenössische Technische Hochschule (ETH)* **94**. Seminar für Statistik, Eidgenössische Technische Hochschule, Zurich.

HAN, F., LU, H. and LIU, H. (2015). A direct estimation of high dimensional stationary vector autoregressions. *J. Mach. Learn. Res.* **16** 3115–3150. MR3450535

HAN, Y., TSAY, R. S. and WU, W. B. (2023). High dimensional generalized linear models for temporal dependent data. *Bernoulli* **29** 105–131. MR4497241 https://doi.org/10.3150/21-bej1451

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

KE, Y., MINSKER, S., REN, Z., SUN, Q. and ZHOU, W.-X. (2019). User-friendly covariance estimation for heavy-tailed distributions. *Statist. Sci.* **34** 454–471. MR4017523 https://doi.org/10.1214/19-STS711

KOCK, A. B. and CALLOT, L. (2015). Oracle inequalities for high dimensional vector autoregressions. *J. Econometrics* **186** 325–344. MR3343790 https://doi.org/10.1016/j.jeconom.2015.02.013

KOOP, G. M. (2013). Forecasting with medium and large Bayesian VARs. *J. Appl. Econometrics* **28** 177–203. MR3045863 https://doi.org/10.1002/jae.1270

LIEBSCHER, E. (2005). Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes. *J. Time Series Anal.* **26** 669–689. MR2188304 https://doi.org/10.1111/j.1467-9892.2005.00412.x

LING, S. (2004). Estimation and testing stationarity for double-autoregressive models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 63–78. MR2035759 https://doi.org/10.1111/j.1467-9868.2004.00432.x

LOH, P.-L. (2017). Statistical consistency and asymptotic normality for high-dimensional robust M-estimators. *Ann. Statist.* **45** 866–896. MR3650403 https://doi.org/10.1214/16-AOS1471

LÜTKEPOHL, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer, Berlin. MR2172368 https://doi.org/10.1007/978-3-540-27752-1

MARTIN, R. D. (1981). Robust methods for time series. In *Applied Time Series Analysis, II (Tulsa, Okla., 1980)* 683–759. Academic Press, New York. MR0651950

MICHAILIDIS, G. and D'ALCHÉ-BUC, F. (2013). Autoregressive models for gene regulatory network inference: Sparsity, stability and causality issues. *Math. Biosci.* **246** 326–334. MR3132054 https://doi.org/10.1016/j.mbs.2013.10.003

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

MULER, N., PEÑA, D. and YOHAI, V. J. (2009). Robust estimation for ARMA models. *Ann. Statist.* **37** 816–840. MR2502652 https://doi.org/10.1214/07-AOS570

MULER, N. and YOHAI, V. J. (2013). Robust estimation for vector autoregressive models. *Comput. Statist. Data Anal.* **65** 68–79. MR3064944 https://doi.org/10.1016/j.csda.2012.02.011

NEGAHBAN, S. and WAINWRIGHT, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *Ann. Statist.* **39** 1069–1097. MR2816348 https://doi.org/10.1214/10-AOS850

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

QIU, H., XU, S., HAN, F., LIU, H. and CAFFO, B. (2015). Robust estimation of transition matrices in high dimensional heavy-tailed vector autoregressive processes. In *International Conference on Machine Learning* 1843–1851. PMLR, Lille.

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

REINSEL, G. C. and VELU, R. P. (1998). *Multivariate Reduced-Rank Regression*: *Theory and Applications*. *Lecture Notes in Statistics* **136**. Springer, New York. MR1719704 https://doi.org/10.1007/978-1-4757-2853-8

STOCK, J. H. and WATSON, M. W. (2009). Forecasting to dynamic factor models subject to structural instability. In *The Methodology and Practice of Econometrics* 173–205. Oxford Univ. Press, Oxford. MR3204305 https://doi.org/10.1093/acprof:oso/9780199237197.003.0007

SUN, Q., ZHOU, W.-X. and FAN, J. (2020). Adaptive Huber regression. *J. Amer. Statist. Assoc.* **115** 254–265. MR4078461 https://doi.org/10.1080/01621459.2018.1543124

TAN, K. M., SUN, Q. and WITTEN, D. (2022). Sparse reduced rank Huber regression in high dimensions. *J. Amer. Statist. Assoc*. To appear.

TSAY, R. S. (2014). *Multivariate Time Series Analysis*: *With R and Financial Applications*. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR3236787

WANG, D. and TSAY, R. S. (2023). Supplement to "Rate-optimal robust estimation of high-dimensional vector autoregressive models." https://doi.org/10.1214/23-AOS2278SUPP

WANG, D., ZHENG, Y. and LI, G. (2021). High-dimensional low-rank tensor autoregressive time series modeling. arXiv preprint. Available at arXiv:2101.04276.

WANG, D., ZHENG, Y., LIAN, H. and LI, G. (2022). High-dimensional vector autoregressive time series modeling via tensor decomposition. *J. Amer. Statist. Assoc.* **117** 1338–1356. MR4480716 https://doi.org/10.1080/01621459.2020.1855183

WANG, H. and BANERJEE, A. (2014). Bregman alternating direction method of multipliers. *Adv. Neural Inf. Process. Syst.* **4** 2816–2824.

WANG, H., LI, G. and JIANG, G. (2007). Robust regression shrinkage and consistent variable selection through the LAD-Lasso. *J. Bus. Econom. Statist.* **25** 347–355. MR2380753 https://doi.org/10.1198/073500106000000251

WANG, L., PENG, B., BRADIC, J., LI, R. and WU, Y. (2020). A tuning-free robust and efficient approach to high-dimensional regression. *J. Amer. Statist. Assoc.* **115** 1700–1714. MR4189748 https://doi.org/10.1080/01621459.2020.1840989

WANG, L., WU, Y. and LI, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *J. Amer. Statist. Assoc.* **107** 214–222. MR2949353 https://doi.org/10.1080/01621459.2012.656014

WANG, X. and YUAN, X. (2012). The linearized alternating direction method of multipliers for Dantzig selector. *SIAM J. Sci. Comput.* **34** A2792–A2811. MR3023726 https://doi.org/10.1137/110833543

WONG, K. C., LI, Z. and TEWARI, A. (2020). Lasso guarantees for β-mixing heavy-tailed time series. *Ann. Statist.* **48** 1124–1142. MR4102690 https://doi.org/10.1214/19-AOS1840

WU, J. C. and XIA, F. D. (2016). Measuring the macroeconomic impact of monetary policy at the zero lower bound. *J. Money Credit Bank.* **48** 253–291.

WU, W.-B. and WU, Y. N. (2016). Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electron. J. Stat.* **10** 352–379. MR3466186 https://doi.org/10.1214/16-EJS1108

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

ZHANG, D. (2021). Robust estimation of the mean and covariance matrix for high dimensional time series. *Statist. Sinica* **31** 797–820. MR4286195 https://doi.org/10.5705/ss.20

ZHENG, L. and RASKUTTI, G. (2019). Testing for high-dimensional network parameters in auto-regressive models. *Electron. J. Stat.* **13** 4977–5043. MR4041701 https://doi.org/10.1214/19-EJS1646

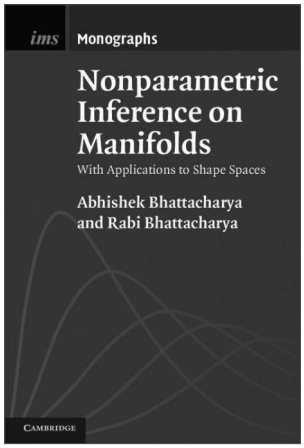
ZHENG, Y. and CHENG, G. (2021). Finite-time analysis of vector autoregressive models under linear restrictions. *Biometrika* **108** 469–489. MR4259144 https://doi.org/10.1093/biomet/asaa065

ZHU, Q., ZHENG, Y. and LI, G. (2018). Linear double autoregression. *J. Econometrics* **207** 162–174. MR3856765 https://doi.org/10.1016/j.jeconom.2018.05.006

ZHU, X., PAN, R., LI, G., LIU, Y. and WANG, H. (2017). Network vector autoregression. *Ann. Statist.* **45** 1096–1123. MR3662449 https://doi.org/10.1214/16-AOS1476

The Institute of Mathematical Statistics presents

IMS MONOGRAPHS



Nonparametric Inference on Manifolds

With Applications to Shape Spaces

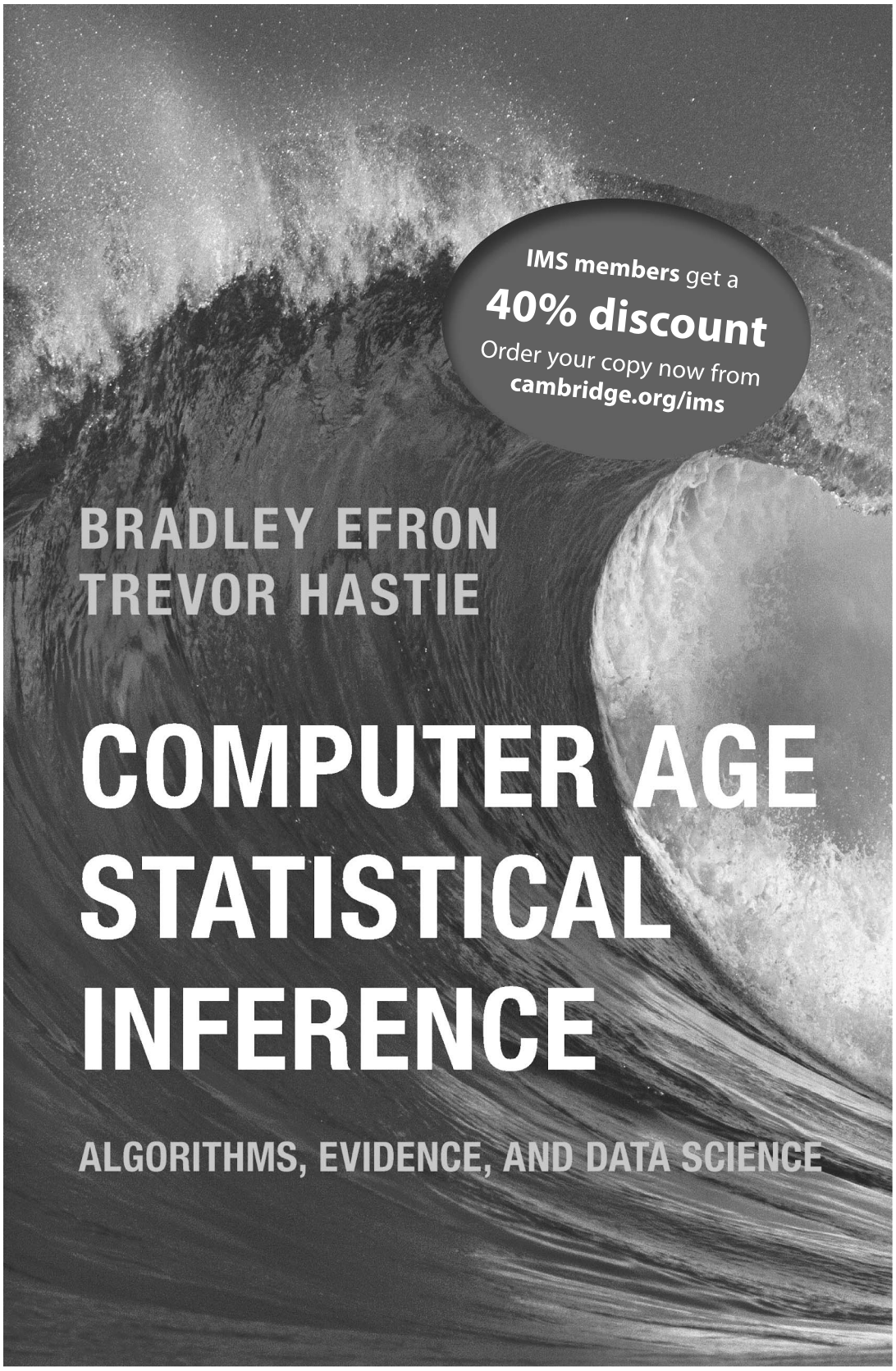
Abhishek Bhattacharya, Rabi Bhattacharya

This book introduces in a systematic manner a general nonparametric theory of statistics on manifolds, with emphasis on manifolds of shapes. The theory has important and varied applications in medical diagnostics, image analysis, and machine vision. An early chapter of examples establishes the effectiveness of the new methods and demonstrates how they outperform their parametric counterparts. Inference is developed for both intrinsic and extrinsic Fréchet means of probability distributions on manifolds, then applied to shape spaces defined as orbits of landmarks under a Lie group of transformations—in particular, similarity, reflection similarity, affine and projective transformations. In addition, nonparametric Bayesian theory is adapted and extended to manifolds for the purposes of density estimation, regression and classification. Ideal for statisticians who analyze manifold data and wish to develop their own methodology, this book is also of interest to probabilists, mathematicians, computer scientists and morphometricians with mathematical training.

IMS member? Claim your 40% discount: www.cambridge.org/ims

Hardback price US$51.00 (non-member price $85.00)

Cambridge University Press, in conjunction with the Institute of Mathematical Statistics, established the IMS Monographs and IMS Textbooks series of high-quality books. The Series Editors are Xiao-Li Meng, Susan Holmes, Ben Hambly, D. R. Cox and Alan Agresti.


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