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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

Statistical inference for rough volatility: Minimax theory .... CARSTEN H. CHONG, MARC HOFFMANN, YANGHUI LIU, MATHIEU ROSENBAUM AND GRÉGOIRE SZYMANSKY 1277
Optimal parameter estimation for linear SPDEs from multiple measurements RANDOLF ALTMEYER, ANTON TIEPNER AND MARTIN WAHL 1307
Optimal estimation of Schatten norms of a rectangular matrix SOLÈNE THÉPAUT AND NICOLAS VERZELEN 1334
Higher-order coverage errors of batching methods via Edgeworth expansions on t-statistics .... SHENGYI HE AND HENRY LAM 1360
On the approximation accuracy of Gaussian variational inference ANYA KATSEVICH AND PHILIPPE RIGOLLET 1384
E-statistics, group invariance and anytime-valid testing MURIEL FELIPE PÉREZ-ORTIZ, TYRON LARDY, RIANNE DE HEIDE AND PETER D. GRÜNWALD 1410
Heavy-tailed Bayesian nonparametric adaptation SERGIOS AGAPIOU AND ISMAËL CASTILLO 1433
Fundamental limits of low-rank matrix estimation with diverging aspect ratios ANDREA MONTANARI AND YUCHEN WU 1460
Asymptotic normality and optimality in nonsmooth stochastic approximation DAMEK DAVIS, DMITRIY DRUSVYATSKIY AND LIWEI JIANG 1485
Bootstrap-assisted inference for generalized Grenander-type estimators MATIAS D. CATTANEO, MICHAEL JANSSON AND KENICHI NAGASAWA 1509
One-step estimation of differentiable Hilbert-valued parameters ALEX LUEDTKE AND INCHEOUL CHUNG 1534
Sharp multiple testing boundary for sparse sequences KWEKU ABRAHAM, ISMAËL CASTILLO AND ÉTIENNE ROQUAIN 1564
A nonparametric doubly robust test for a continuous treatment effect CHARLES R. DOSS, GUANGWEI WENG, LAN WANG, IRA MOSCOVICE AND TONGTAN CHANTARAT 1592
Gromov–Wasserstein distances: Entropic regularization, duality and sample complexity ZHENGXIN ZHANG, ZIV GOLDFELD, YOUSSEF MROUEH AND BHARATH K. SRIPERUMBUDUR 1616
Online change-point detection for matrix-valued time series with latent two-way factor structure .... YONG HE, XINBING KONG, LORENZO TRAPANI AND LONG YU 1646
Majority vote for distributed differentially private sign selection WEIDONG LIU, JIYUAN TU, XIAOJUN MAO AND XI CHEN 1671
Wasserstein convergence in Bayesian and frequentist deconvolution models JUDITH ROUSSEAU AND CATIA SCRICCIOLO 1691
Detection and estimation of structural breaks in high-dimensional functional time series DEGUI LI, RUNZE LI AND HAN LIN SHANG 1716
Efficient functional Lasso kernel smoothing for high-dimensional additive regression EUN RYUNG LEE, SEYOUNG PARK, ENNO MAMMEN AND BYEONG U. PARK 1741
Learning Gaussian mixtures using the Wasserstein–Fisher–Rao gradient flow YULING YAN, KAIZHENG WANG AND PHILIPPE RIGOLLET 1774

continued

Vol. 52, No. 4—August 2024

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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles—*Continued from front cover*

Efficient and multiply robust risk estimation under general forms of dataset shift
HONGXIANG QIU, ERIC TCHETGEN TCHETGEN AND EDGAR DOBRIBAN 1796
On posterior consistency of data assimilation with Gaussian process priors: The 2D-Navier–Stokes equations . . . . RICHARD NICKL AND EDRISS S. TITI 1825
How do noise tails impact on deep ReLU networks?
JIANQING FAN, YIHONG GU AND WEN-XIN ZHOU 1845

Vol. 52, No. 4—August 2024

THE ANNALS OF STATISTICS Vol. 52, No. 4, pp. 1277–1871 August 2024

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

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

The Annals of Statistics **[ISSN 0090-5364 (print); ISSN 2168-8966 (online)],** Volume 52, Number 4, August 2024. Published bimonthly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, OH 44094, USA. Periodicals postage paid at Cleveland, Ohio, and at additional mailing offices.

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

The Annals of Statistics
2024, Vol. 52, No. 4, 1277–1306
https://doi.org/10.1214/23-AOS2343

STATISTICAL INFERENCE FOR ROUGH VOLATILITY: MINIMAX THEORY

BY CARSTEN H. CHONG[1,a], MARC HOFFMANN[2,b], YANGHUI LIU[3,c], MATHIEU ROSENBAUM[4,d] AND GRÉGOIRE SZYMANSKY[4,e]

[1]*Department of Information Systems, Business Statistics and Operations Management, Hong Kong University of Science and Technology,* [a]*carstenchong@ust.hk*

[2]*Ceremade, University Paris Dauphine-PSL,* [b]*hoffmann@ceremade.dauphine.fr*

[3]*Department of Mathematics, Baruch College CUNY,* [c]*yanghui.liu@baruch.cuny.edu*

[4]*CMAP, École Polytechnique,* [d]*mathieu.rosenbaum@polytechnique.edu,* [e]*gregoire.szymanski@polytechnique.edu*

In recent years, rough volatility models have gained considerable attention in quantitative finance. In this paradigm, the stochastic volatility of the price of an asset has quantitative properties similar to that of a fractional Brownian motion with small Hurst index $H < 1/2$. In this work, we provide the first rigorous statistical analysis of the problem of estimating H from historical observations of the underlying asset. We establish minimax lower bounds and design optimal procedures based on adaptive estimation of quadratic functionals based on wavelets. We prove in particular that the optimal rate of convergence for estimating H based on price observations at n time points is of order $n^{-1/(4H+2)}$ as n grows to infinity, extending results that were known only for $H > 1/2$. Our study positively answers the question whether H can be inferred, although it is the regularity of a latent process (the volatility); in rough models, when H is close to 0, we even obtain an accuracy comparable to usual $\sqrt{n}$-consistent regular statistical models.

REFERENCES

[1] ABI JABER, E. (2022). The characteristic function of Gaussian stochastic volatility models: An analytic expression. *Finance Stoch.* **26** 733–769. MR4487683 https://doi.org/10.1007/s00780-022-00489-4

[2] ABI JABER, E. and EL EUCH, O. (2019). Multifactor approximation of rough volatility models. *SIAM J. Financial Math.* **10** 309–349. MR3934104 https://doi.org/10.1137/18M1170236

[3] ABI JABER, E., LARSSON, M. and PULIDO, S. (2019). Affine Volterra processes. *Ann. Appl. Probab.* **29** 3155–3200. MR4019885 https://doi.org/10.1214/19-AAP1477

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

[5] ALÒS, E., LEÓN, J. A. and VIVES, J. (2007). On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility. *Finance Stoch.* **11** 571–589. MR2335834 https://doi.org/10.1007/s00780-007-0049-1

[6] ALÒS, E., ROLLOOS, F. and SHIRAYA, K. (2022). Forward start volatility swaps in rough volatility models. arXiv preprint, available at arXiv:2207.10370.

[7] BARNDORFF-NIELSEN, O. E., CORCUERA, J. M. and PODOLSKIJ, M. (2013). Limit theorems for functionals of higher order differences of Brownian semi-stationary processes. In *Prokhorov and Contemporary Probability Theory*. *Springer Proc. Math. Stat.* **33** 69–96. Springer, Heidelberg. MR3070467 https://doi.org/10.1007/978-3-642-33549-5_4

[8] BAYER, C., FRIZ, P. and GATHERAL, J. (2016). Pricing under rough volatility. *Quant. Finance* **16** 887–904. MR3494612 https://doi.org/10.1080/14697688.2015.1099717

[9] BAYER, C., FRIZ, P. K., GASSIAT, P., MARTIN, J. and STEMPER, B. (2020). A regularity structure for rough volatility. *Math. Finance* **30** 782–832. MR4116451 https://doi.org/10.1111/mafi.12233

[10] BENNEDSEN, M., LUNDE, A. and PAKKANEN, M. S. (2017). Hybrid scheme for Brownian semistationary processes. *Finance Stoch.* **21** 931–965. MR3723378 https://doi.org/10.1007/s00780-017-0335-5

MSC2020 subject classifications. Primary 60G22, 62C20; secondary 62F12, 62M09, 62P20.
Key words and phrases. Rough volatility, fractional Brownian motion, wavelets, minimax optimality, pre-averaging.

[11] BENNEDSEN, M., LUNDE, A. and PAKKANEN, M. S. (2022). Decoupling the short-and long-term behavior of stochastic volatility. *J. Financ. Econom.* **25** 961–1006. https://doi.org/10.1093/jjfinec/nbaa049

[12] BIBINGER, M., JIRAK, M. and VETTER, M. (2017). Nonparametric change-point analysis of volatility. *Ann. Statist.* **45** 1542–1578. MR3670188 https://doi.org/10.1214/16-AOS1499

[13] BOLKO, A. E., CHRISTENSEN, K., PAKKANEN, M. S. and VELIYEV, B. (2023). A GMM approach to estimate the roughness of stochastic volatility. *J. Econometrics* **235** 745–778. MR4602890 https://doi.org/10.1016/j.jeconom.2022.06.009

[14] BROUSTE, A. and FUKASAWA, M. (2018). Local asymptotic normality property for fractional Gaussian noise under high-frequency observations. *Ann. Statist.* **46** 2045–2061. MR3845010 https://doi.org/10.1214/17-AOS1611

[15] CALLEGARO, G., GRASSELLI, M. and PAGÈS, G. (2021). Fast hybrid schemes for fractional Riccati equations (rough is not so tough). *Math. Oper. Res.* **46** 221–254. MR4224426 https://doi.org/10.1287/moor.2020.1054

[16] CHONG, C. H., HOFFMANN, M., LIU, Y., ROSENBAUM, M. and SZYMANSKI, G. (2024). Statistical inference for rough volatility: Central limit theorems. *Ann. Appl. Probab.* **34** 2600–2649. MR4757488 https://doi.org/10.1214/23-aap2002

[17] CHONG, C. H. and TODOROV, V. (2022). Short-time expansion of characteristic functions in a rough volatility setting with applications. *Bernoulli*. To appear.

[18] CHONG, C. H., HOFFMANN, M., LIU, Y., ROSENBAUM, M. and SZYMANSKY, G. (2024). Supplement to "Statistical inference for rough volatility: Minimax theory." https://doi.org/10.1214/23-AOS2343SUPP

[19] 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

[20] COEURJOLLY, J.-F. (2001). Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths. *Stat. Inference Stoch. Process.* **4** 199–227. MR1856174 https://doi.org/10.1023/A:1017507306245

[21] COMTE, F. and RENAULT, E. (1998). Long memory in continuous-time stochastic volatility models. *Math. Finance* **8** 291–323. MR1645101 https://doi.org/10.1111/1467-9965.00057

[22] CUCHIERO, C. and TEICHMANN, J. (2019). Markovian lifts of positive semidefinite affine Volterra-type processes. *Decis. Econ. Finance* **42** 407–448. MR4031333 https://doi.org/10.1007/s10203-019-00268-5

[23] EL EUCH, O., FUKASAWA, M., GATHERAL, J. and ROSENBAUM, M. (2019). Short-term at-the-money asymptotics under stochastic volatility models. *SIAM J. Financial Math.* **10** 491–511. MR3945237 https://doi.org/10.1137/18M1167565

[24] EL EUCH, O., FUKASAWA, M. and ROSENBAUM, M. (2018). The microstructural foundations of leverage effect and rough volatility. *Finance Stoch.* **22** 241–280. MR3778355 https://doi.org/10.1007/s00780-018-0360-z

[25] EL EUCH, O. and ROSENBAUM, M. (2018). Perfect hedging in rough Heston models. *Ann. Appl. Probab.* **28** 3813–3856. MR3861827 https://doi.org/10.1214/18-AAP1408

[26] EL EUCH, O. and ROSENBAUM, M. (2019). The characteristic function of rough Heston models. *Math. Finance* **29** 3–38. MR3905737 https://doi.org/10.1111/mafi.12173

[27] FORDE, M., FUKASAWA, M., GERHOLD, S. and SMITH, B. (2020). The Rough Bergomi model as $H \to 0$ - skew flattening/blow up and non-Gaussian rough volatility. Preprint.

[28] FORDE, M. and ZHANG, H. (2017). Asymptotics for rough stochastic volatility models. *SIAM J. Financial Math.* **8** 114–145. MR3608743 https://doi.org/10.1137/15M1009330

[29] FRIZ, P. K., GASSIAT, P. and PIGATO, P. (2022). Short-dated smile under rough volatility: Asymptotics and numerics. *Quant. Finance* **22** 463–480. MR4401811 https://doi.org/10.1080/14697688.2021.1999486

[30] FRIZ, P. K., GATHERAL, J. and RADOIČIĆ, R. (2022). Forests, cumulants, martingales. *Ann. Probab.* **50** 1418–1445. MR4420423 https://doi.org/10.1214/21-aop1560

[31] FUKASAWA, M. (2021). Volatility has to be rough. *Quant. Finance* **21** 1–8. MR4188876 https://doi.org/10.1080/14697688.2020.1825781

[32] FUKASAWA, M., HORVATH, B. and TANKOV, P. (2024). Hedging under rough volatility. In *Rough Volatility. Financ. Math.* 115–125. Society for Industrial and Applied Mathematics, Philadelphia, PA. https://doi.org/10.1137/1.9781611977783.ch6

[33] FUKASAWA, M. and TAKABATAKE, T. (2019). Asymptotically efficient estimators for self-similar stationary Gaussian noises under high frequency observations. *Bernoulli* **25** 1870–1900. MR3961234 https://doi.org/10.3150/18-BEJ1039

[34] GASSIAT, P. (2019). On the martingale property in the rough Bergomi model. *Electron. Commun. Probab.* **24** Paper No. 33, 9. MR3962483 https://doi.org/10.1214/19-ECP239

[35] GASSIAT, P. (2023). Weak error rates of numerical schemes for rough volatility. *SIAM J. Financial Math.* **14** 475–496. MR4589567 https://doi.org/10.1137/22M1485760

[36] GATHERAL, J., JAISSON, T. and ROSENBAUM, M. (2018). Volatility is rough. *Quant. Finance* **18** 933–949. MR3805308 https://doi.org/10.1080/14697688.2017.1393551

[37] GATHERAL, J., JUSSELIN, P. and ROSENBAUM, M. (2020). The quadratic rough Heston model and the joint S&P 500/VIX smile calibration problem. Risk.

[38] GATHERAL, J. and KELLER-RESSEL, M. (2019). Affine forward variance models. *Finance Stoch.* **23** 501–533. MR3968277 https://doi.org/10.1007/s00780-019-00392-5

[39] GERHOLD, S., GERSTENECKER, C. and PINTER, A. (2019). Moment explosions in the rough Heston model. *Decis. Econ. Finance* **42** 575–608. MR4031338 https://doi.org/10.1007/s10203-019-00267-6

[40] GLOTER, A. and HOFFMANN, M. (2004). Stochastic volatility and fractional Brownian motion. *Stochastic Process. Appl.* **113** 143–172. MR2078541 https://doi.org/10.1016/j.spa.2004.03.008

[41] GLOTER, A. and HOFFMANN, M. (2007). Estimation of the Hurst parameter from discrete noisy data. *Ann. Statist.* **35** 1947–1974. MR2363959 https://doi.org/10.1214/009053607000000316

[42] 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

[43] HORVATH, B., TEICHMANN, J. and ŽURIČ, Ž. (2021). Deep hedging under rough volatility. *Risks* **9** 138.

[44] ISTAS, J. and LANG, G. (1997). Quadratic variations and estimation of the local Hölder index of a Gaussian process. *Ann. Inst. Henri Poincaré Probab. Stat.* **33** 407–436. MR1465796 https://doi.org/10.1016/S0246-0203(97)80099-4

[45] JACQUIER, A., MUGURUZA, A. and PANNIER, A. (2021). Rough multifactor volatility for SPX and VIX options. arXiv preprint, available at arXiv:2112.14310.

[46] JACQUIER, A. and PANNIER, A. (2022). Large and moderate deviations for stochastic Volterra systems. *Stochastic Process. Appl.* **149** 142–187. MR4410038 https://doi.org/10.1016/j.spa.2022.03.017

[47] JAISSON, T. and ROSENBAUM, M. (2015). Limit theorems for nearly unstable Hawkes processes. *Ann. Appl. Probab.* **25** 600–631. MR3313750 https://doi.org/10.1214/14-AAP1005

[48] JUSSELIN, P. and ROSENBAUM, M. (2020). No-arbitrage implies power-law market impact and rough volatility. *Math. Finance* **30** 1309–1336. MR4154771 https://doi.org/10.1111/mafi.12254

[49] KAWAI, R. (2013). Fisher information for fractional Brownian motion under high-frequency discrete sampling. *Comm. Statist. Theory Methods* **42** 1628–1636. MR3041490 https://doi.org/10.1080/03610926.2011.594540

[50] LIVIERI, G., MOUTI, S., PALLAVICINI, A. and ROSENBAUM, M. (2018). Rough volatility: Evidence from option prices. *IISE Trans.* **50** 767–776.

[51] MCCRICKERD, R. and PAKKANEN, M. S. (2018). Turbocharging Monte Carlo pricing for the rough Bergomi model. *Quant. Finance* **18** 1877–1886. MR3867719 https://doi.org/10.1080/14697688.2018.1459812

[52] MEYER, Y., SELLAN, F. and TAQQU, M. S. (1999). Wavelets, generalized white noise and fractional integration: The synthesis of fractional Brownian motion. *J. Fourier Anal. Appl.* **5** 465–494. MR1755100 https://doi.org/10.1007/BF01261639

[53] ROSENBAUM, M. (2008). Estimation of the volatility persistence in a discretely observed diffusion model. *Stochastic Process. Appl.* **118** 1434–1462. MR2427046 https://doi.org/10.1016/j.spa.2007.09.004

[54] ROSENBAUM, M. (2009). First order p-variations and Besov spaces. *Statist. Probab. Lett.* **79** 55–62. MR2483397 https://doi.org/10.1016/j.spl.2008.07.019

[55] ROSENBAUM, M. and ZHANG, J. (2021). Deep calibration of the quadratic rough Heston model. arXiv preprint, available at arXiv:2107.01611.

[56] SZYMANSKI, G. (2024). Optimal estimation of the rough Hurst parameter in additive noise. *Stochastic Process. Appl.* **170** Paper no. 104302. MR4700236 https://doi.org/10.1016/j.spa.2024.104302

The Annals of Statistics
2024, Vol. 52, No. 4, 1307–1333
https://doi.org/10.1214/24-AOS2364

OPTIMAL PARAMETER ESTIMATION FOR LINEAR SPDES FROM MULTIPLE MEASUREMENTS

BY RANDOLF ALTMEYER[1,a], ANTON TIEPNER[2,b] AND MARTIN WAHL[3,c]

[1]*Department of Pure Mathematics and Mathematical Statistics, University of Cambridge*, [a]*ra591@cam.ac.uk*

[2]*Department of Mathematics, Aarhus University*, [b]*tiepner@math.au.dk*

[3]*Fakultät für Mathematik, Universität Bielefeld*, [c]*martin.wahl@math.uni-bielefeld.de*

The coefficients in a second order parabolic linear stochastic partial differential equation (SPDE) are estimated from multiple spatially localised measurements. Assuming that the spatial resolution tends to zero and the number of measurements is nondecreasing, the rate of convergence for each coefficient depends on its differential order and is faster for higher order coefficients. Based on an explicit analysis of the reproducing kernel Hilbert space of a general stochastic evolution equation, a Gaussian lower bound scheme is introduced. As a result, minimax optimality of the rates as well as sufficient and necessary conditions for consistent estimation are established.

REFERENCES

[1] ALONSO, S., STANGE, M. and BETA, C. (2018). Modeling random crawling, membrane deformation and intracellular polarity of motile amoeboid cells. *PLoS ONE* **13** e0201977. https://doi.org/10.1371/journal.pone.0201977

[2] ALTMEYER, R., BRETSCHNEIDER, T., JANÁK, J. and REISS, M. (2022). Parameter estimation in an SPDE model for cell repolarization. *SIAM/ASA J. Uncertain. Quantificat.* **10** 179–199. MR4376300 https://doi.org/10.1137/20M1373347

[3] ALTMEYER, R., CIALENCO, I. and PASEMANN, G. (2023). Parameter estimation for semilinear SPDEs from local measurements. *Bernoulli* **29** 2035–2061. MR4580906 https://doi.org/10.3150/22-bej1531

[4] ALTMEYER, R. and REISS, M. (2021). Nonparametric estimation for linear SPDEs from local measurements. *Ann. Appl. Probab.* **31** 1–38. MR4254472 https://doi.org/10.1214/20-aap1581

[5] ALTMEYER, R., TIEPNER, A. and WAHL, M. (2024). Supplement to "Optimal parameter estimation for linear SPDEs from multiple measurements." https://doi.org/10.1214/24-AOS2364SUPP

[6] AMANN, H. (1995). *Linear and Quasilinear Parabolic Problems. Vol. I. Abstract Linear Theory. Monographs in Mathematics* **89**. Birkhäuser, Boston, MA. MR1345385 https://doi.org/10.1007/978-3-0348-9221-6

[7] ASPELMEIER, T., EGNER, A. and MUNK, A. (2015). Modern statistical challenges in high-resolution fluorescence microscopy. *Annu. Rev. Stat. Appl.* **2** 163–202.

[8] BACKER, A. S. and MOERNER, W. E. Extending single-molecule microscopy using optical Fourier processing. *J. Phys. Chem., B* **118** 8313–8329. https://doi.org/10.1021/jp501778z

[9] BENTH, F. E., SCHROERS, D. and VERAART, A. E. D. (2022). A weak law of large numbers for realised covariation in a Hilbert space setting. *Stochastic Process. Appl.* **145** 241–268. MR4360403 https://doi.org/10.1016/j.spa.2021.12.011

[10] BIBINGER, M. and TRABS, M. (2020). Volatility estimation for stochastic PDEs using high-frequency observations. *Stochastic Process. Appl.* **130** 3005–3052. MR4080737 https://doi.org/10.1016/j.spa.2019.09.002

[11] CATANIA, F., MASSABÒ, M. and PALADINO, O. (2006). Estimation of transport and kinetic parameters using analytical solutions of the 2D advection-dispersion-reaction model. *Environmetrics* **17** 199–216. MR2225538 https://doi.org/10.1002/env.770

[12] CHONG, C. (2020). High-frequency analysis of parabolic stochastic PDEs. *Ann. Statist.* **48** 1143–1167. MR4102691 https://doi.org/10.1214/19-AOS1841

MSC2020 subject classifications. Primary 60H15, 60F05; secondary 62F12, 62F35.

Key words and phrases. Linear SPDEs, parameter estimation, central limit theorem, minimax lower bound, reproducing kernel Hilbert space, local measurements.

[13] CIALENCO, I., DELGADO-VENCES, F. and KIM, H.-J. (2020). Drift estimation for discretely sampled SPDEs. *Stoch. Partial Differ. Equ. Anal. Comput.* **8** 895–920. MR4174073 https://doi.org/10.1007/s40072-019-00164-4

[14] CIALENCO, I., KIM, H.-J. and PASEMANN, G. (2024). Statistical analysis of discretely sampled semilinear SPDEs: A power variation approach. *Stoch. Partial Differ. Equ. Anal. Comput.* **12** 326–351. MR4709545 https://doi.org/10.1007/s40072-022-00285-3

[15] DA PRATO, G. and ZABCZYK, J. (2014). *Stochastic Equations in Infinite Dimensions*, 2nd ed. *Encyclopedia of Mathematics and Its Applications* **152**. Cambridge Univ. Press, Cambridge. MR3236753 https://doi.org/10.1017/CBO9781107295513

[16] DEBUSSCHE, A., DE MOOR, S. and HOFMANOVÁ, M. (2015). A regularity result for quasilinear stochastic partial differential equations of parabolic type. *SIAM J. Math. Anal.* **47** 1590–1614. MR3340199 https://doi.org/10.1137/130950549

[17] EGNER, A., GEISLER, C. and SIEGMUND, R. (2020). STED nanoscopy. In *Nanoscale Photonic Imaging* (T. Salditt, A. Egner and D. R. Luke, eds.). *Topics in Applied Physics* 3–34. Springer, Cham.

[18] ENGEL, K.-J. and NAGEL, R. (2000). *One-Parameter Semigroups for Linear Evolution Equations*. *Graduate Texts in Mathematics* **194**. Springer, New York. MR1721989

[19] EVANS, L. C. (2010). *Partial Differential Equations*, 2nd ed. *Graduate Studies in Mathematics* **19**. Amer. Math. Soc., Providence, RI. MR2597943 https://doi.org/10.1090/gsm/019

[20] EVANS, L. C. and GARIEPY, R. F. (2015). *Measure Theory and Fine Properties of Functions*, Revised ed. *Textbooks in Mathematics*. CRC Press, Boca Raton, FL. MR3409135

[21] GAUDLITZ, S. and REISS, M. (2023). Estimation for the reaction term in semi-linear SPDEs under small diffusivity. *Bernoulli* **29** 3033–3058. MR4632130 https://doi.org/10.3150/22-bej1573

[22] 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

[23] GUGUSHVILI, S., VAN DER VAART, A. and YAN, D. (2020). Bayesian linear inverse problems in regularity scales. *Ann. Inst. Henri Poincaré Probab. Stat.* **56** 2081–2107. MR4116718 https://doi.org/10.1214/19-AIHP1029

[24] HAIRER, M. (2009). An Introduction to Stochastic PDEs. arXiv preprint. Available at arXiv:0907.4178.

[25] HILDEBRANDT, F. and TRABS, M. (2021). Parameter estimation for SPDEs based on discrete observations in time and space. *Electron. J. Stat.* **15** 2716–2776. MR4280160 https://doi.org/10.1214/21-ejs1848

[26] HILDEBRANDT, F. and TRABS, M. (2023). Nonparametric calibration for stochastic reaction-diffusion equations based on discrete observations. *Stochastic Process. Appl.* **162** 171–217. MR4591370 https://doi.org/10.1016/j.spa.2023.04.019

[27] HUEBNER, M. and ROZOVSKIĬ, B. L. (1995). On asymptotic properties of maximum likelihood estimators for parabolic stochastic PDE's. *Probab. Theory Related Fields* **103** 143–163. MR1355054 https://doi.org/10.1007/BF01204212

[28] JANSON, S. (1997). *Gaussian Hilbert Spaces*. *Cambridge Tracts in Mathematics* **129**. Cambridge Univ. Press, Cambridge. MR1474726 https://doi.org/10.1017/CBO9780511526169

[29] KARALASHVILI, M., GROSS, S., MARQUARDT, W., MHAMDI, A. and REUSKEN, A. (2011). Identification of transport coefficient models in convection-diffusion equations. *SIAM J. Sci. Comput.* **33** 303–327. MR2783196 https://doi.org/10.1137/09077360X

[30] KARATZAS, I. and SHREVE, S. E. (1998). *Brownian Motion and Stochastic Calculus*. Springer, New York.

[31] KATO, T. (1995). *Perturbation Theory for Linear Operators*. *Classics in Mathematics*. Springer, Berlin. MR1335452

[32] KŘÍŽ, P. and MASLOWSKI, B. (2019). Central limit theorems and minimum-contrast estimators for linear stochastic evolution equations. *Stochastics* **91** 1109–1140. MR4023510 https://doi.org/10.1080/17442508.2019.1576688

[33] KULAITIS, G., MUNK, A. and WERNER, F. (2021). What is resolution? A statistical minimax testing perspective on superresolution microscopy. *Ann. Statist.* **49** 2292–2312. MR4319251 https://doi.org/10.1214/20-aos2037

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

[35] KWAŚNICKI, M. (2017). Ten equivalent definitions of the fractional Laplace operator. *Fract. Calc. Appl. Anal.* **20** 7–51. MR3613319 https://doi.org/10.1515/fca-2017-0002

[36] LIFSHITS, M. (2012). *Lectures on Gaussian Processes*. *SpringerBriefs in Mathematics*. Springer, Heidelberg. MR3024389 https://doi.org/10.1007/978-3-642-24939-6

[37] LIPTSER, R. S. and SHIRYAEV, A. N. (2001). *Statistics of Random Processes*. *I. General Theory*, expanded ed. *Applications of Mathematics* (*New York*) **5**. Springer, Berlin. MR1800857

[38] LIU, W. and RÖCKNER, M. (2015). *Stochastic Partial Differential Equations*: *An Introduction*. *Universitext*. Springer, Cham. MR3410409 https://doi.org/10.1007/978-3-319-22354-4

[39] LIU, X., YEO, K. and LU, S. (2022). Statistical modeling for spatio-temporal data from stochastic convection-diffusion processes. *J. Amer. Statist. Assoc.* **117** 1482–1499. MR4480726 https://doi.org/10.1080/01621459.2020.1863223

[40] LLOPIS, F. P., KANTAS, N., BESKOS, A. and JASRA, A. (2018). Particle filtering for stochastic Navier-Stokes signal observed with linear additive noise. *SIAM J. Sci. Comput.* **40** A1544–A1565. MR3807951 https://doi.org/10.1137/17M1151900

[41] LOCKLEY, R. (2017). Image-based Modelling of Cell Reorientation PhD Thesis Univ. Warwick.

[42] LOTOTSKY, S. (2003). Parameter estimation for stochastic parabolic equations: Asymptotic properties of a two-dimensional projection-based estimator. *Stat. Inference Stoch. Process.* **6** 65–87. MR1965185 https://doi.org/10.1023/A:1022699622088

[43] LUCE, C. H., TONINA, D., GARIGLIO, F. and APPLEBEE, R. (2013). Solutions for the diurnally forced advection-diffusion equation to estimate bulk fluid velocity and diffusivity in streambeds from temperature time series. *Water Resour. Res.* **49** 488–506.

[44] LUNARDI, A. (1995). *Analytic Semigroups and Optimal Regularity in Parabolic Problems*. *Modern Birkhäuser Classics*. Birkhäuser/Springer Basel AG, Basel. MR3012216

[45] PAZY, A. (1983). *Semigroups of Linear Operators and Applications to Partial Differential Equations*. *Applied Mathematical Sciences* **44**. Springer, New York. MR0710486 https://doi.org/10.1007/978-1-4612-5561-1

[46] REDDY, S. C. and TREFETHEN, L. N. (1994). Pseudospectra of the convection-diffusion operator. *SIAM J. Appl. Math.* **54** 1634–1649. MR1301275 https://doi.org/10.1137/S0036139993246982

[47] SAUER, M. and STANNAT, W. (2016). Analysis and approximation of stochastic nerve axon equations. *Math. Comp.* **85** 2457–2481. MR3511288 https://doi.org/10.1090/mcom/3068

[48] SHEU, S. J. (1991). Some estimates of the transition density of a nondegenerate diffusion Markov process. *Ann. Probab.* **19** 538–561. MR1106275

[49] SHIMAKURA, N. (1992). *Partial Differential Operators of Elliptic Type*. *Translations of Mathematical Monographs* **99**. Amer. Math. Soc., Providence, RI. MR1168472 https://doi.org/10.1090/mmono/099

[50] SIGRIST, F., KÜNSCH, H. R. and STAHEL, W. A. (2015). Stochastic partial differential equation based modelling of large space-time data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 3–33. MR3299397 https://doi.org/10.1111/rssb.12061

[51] TONAKI, Y., KAINO, Y. and UCHIDA, M. (2023). Parameter estimation for linear parabolic SPDEs in two space dimensions based on high frequency data. *Scand. J. Stat.* **50** 1568–1589. MR4677368

[52] TRIEBEL, H. (1983). *Theory of Function Spaces*. *Monographs in Mathematics* **78**. Birkhäuser, Basel. MR0781540 https://doi.org/10.1007/978-3-0346-0416-1

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

[54] TUCKWELL, H. C. (2013). Stochastic partial differential equations in neurobiology: Linear and nonlinear models for spiking neurons. In *Stochastic Biomathematical Models*. *Lecture Notes in Math.* **2058** 149–173. Springer, Heidelberg. MR3051032 https://doi.org/10.1007/978-3-642-32157-3_6

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

[56] 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

[57] WALSH, J. B. (1981). A stochastic model of neural response. *Adv. in Appl. Probab.* **13** 231–281. MR0612203 https://doi.org/10.2307/1426683

[58] YAGI, A. (2010). *Abstract Parabolic Evolution Equations and Their Applications*. *Springer Monographs in Mathematics*. Springer, Berlin. MR2573296 https://doi.org/10.1007/978-3-642-04631-5

The Annals of Statistics
2024, Vol. 52, No. 4, 1334–1359
https://doi.org/10.1214/24-AOS2374

OPTIMAL ESTIMATION OF SCHATTEN NORMS OF A RECTANGULAR MATRIX

BY SOLÈNE THÉPAUT[1,a] AND NICOLAS VERZELEN[2,b]

[1]*Safran,* [a]*solenethepautv@gmail.com*
[2]*MISTEA, INRAE, Université de Montpellier,* [b]*nicolas.verzelen@inrae.fr*

We consider the twin problems of estimating the effective rank and the Schatten norms $\|\mathbf{A}\|_s$ of a rectangular $p \times q$ matrix $\mathbf{A}$ from noisy observations. When s is an even integer, we introduce a polynomial-time estimator of $\|\mathbf{A}\|_s$ that achieves the minimax rate $(pq)^{1/4}$. Interestingly, this optimal rate does not depend on the underlying rank of the matrix $\mathbf{A}$. When s is not an even integer, the optimal rate is much slower. A simple thresholding estimator of the singular values achieves the rate $(q \wedge p)(pq)^{1/4}$, which turns out to be optimal up to a logarithmic multiplicative term. The tight minimax rate is achieved by a more involved polynomial approximation method. This allows us to build estimators for a class of effective rank indices. As a byproduct, we also characterize the minimax rate for estimating the sequence of singular values of a matrix.

REFERENCES

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

[2] BALCAN, M.-F., LI, Y., WOODRUFF, D. P. and ZHANG, H. (2019). Testing matrix rank, optimally. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms* 727–746. SIAM, Philadelphia, PA. MR3909515 https://doi.org/10.1137/1.9781611975482.46

[3] BERNSTEIN, S. (1913). Sur la valeur asymptotique de la meilleure approximation des fonctions analytiques. *Comm*. *Soc*. *Math*. *Kharkov* **13** 263–273.

[4] BERNSTEIN, S. (1938). Sur la meilleure approximation de $|x|^p$ par des polynômes de degrés très élevés. *Izv. Ross*. *Akad*. *Nauk Ser. Mat*. **2** 169–190.

[5] BHATIA, R. (2013). *Matrix Analysis*. *Graduate Texts in Mathematics* **169**. Springer, New York. MR1477662 https://doi.org/10.1007/978-1-4612-0653-8

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

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

[8] 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

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

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

[11] 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

[12] 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

[13] HAN, Y., JIAO, J. and MUKHERJEE, R. (2020). On estimation of L_r-norms in Gaussian white noise models. *Probab*. *Theory Related Fields* **177** 1243–1294. MR4126939 https://doi.org/10.1007/s00440-020-00982-x

MSC2020 subject classifications. Primary 62H12; secondary 62C20.
Key words and phrases. Schatten norm, effective rank, matrix estimation, random matrix, polynomial approximation.

[14] HAN, Y., JIAO, J. and WEISSMAN, T. (2018). Local moment matching: A unified methodology for symmetric functional estimation and distribution estimation under Wasserstein distance. In *Conference on Learning Theory* 3189–3221. PMLR.

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

[16] JOST, L. (2006). Entropy and diversity. *Oikos* **113** 363–375.

[17] KHETAN, A. and OH, S. (2019). Spectrum estimation from a few entries. *J. Mach. Learn. Res.* **20** Paper No. 21, 55. MR3911428

[18] KOLTCHINSKII, V. and LOUNICI, K. (2017). Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli* **23** 110–133. MR3556768 https://doi.org/10.3150/15-BEJ730

[19] KONG, W. and VALIANT, G. (2017). Spectrum estimation from samples. *Ann. Statist.* **45** 2218–2247. MR3718167 https://doi.org/10.1214/16-AOS1525

[20] LEPSKI, O., NEMIROVSKI, A. and SPOKOINY, V. (1999). On estimation of the L_r norm of a regression function. *Probab. Theory Related Fields* **113** 221–253. MR1670867 https://doi.org/10.1007/s004409970006

[21] LI, Y., NGUYEN, H. L. and WOODRUFF, D. P. (2014). On sketching matrix norms and the top singular vector. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms* 1562–1581. ACM, New York. MR3376474 https://doi.org/10.1137/1.9781611973402.114

[22] NADAKUDITI, R. R. (2014). OptShrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage. *IEEE Trans. Inf. Theory* **60** 3002–3018. MR3200641 https://doi.org/10.1109/TIT.2014.2311661

[23] PACHÓN, R. and TREFETHEN, L. N. (2009). Barycentric–Remez algorithms for best polynomial approximation in the chebfun system. *BIT* **49** 721–741. MR2564378 https://doi.org/10.1007/s10543-009-0240-1

[24] POLYANSKIY, Y. and WU, Y. (2019). Dualizing Le Cam's method for functional estimation, with applications to estimating the unseens. arXiv preprint arXiv:1902.05616.

[25] REMEZ, E. Y. (1934). Sur la détermination des polynômes d'approximation de degré donnée. *Comm. Soc. Math. Kharkov* **10** 41–63.

[26] RIVLIN, T. J. (1990). *Chebyshev Polynomials: From Approximation Theory to Algebra and Number Theory*, 2nd ed. *Pure and Applied Mathematics (New York)*. Wiley, New York. MR1060735

[27] ROY, O. and VETTERLI, M. (2007). The effective rank: A measure of effective dimensionality. In 2007 15*th European Signal Processing Conference* 606–610. IEEE.

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

[29] THÉPAUT, S. and VERZELEN, N. (2024). Supplement to "Optimal estimation of Schatten norms of a rectangular matrix." https://doi.org/10.1214/24-AOS2374SUPP

[30] TIMAN, A. F. (1963). *Theory of Approximation of Functions of a Real Variable*. *International Series of Monographs in Pure and Applied Mathematics* **34**. The Macmillan Company, New York. Translated from the Russian by J. Berry, English translation edited and editorial preface by J. Cossar. MR0192238

[31] VALLET, P., LOUBATON, P. and MESTRE, X. (2012). Improved subspace estimation for multivariate observations of high dimension: The deterministic signals case. *IEEE Trans. Inf. Theory* **58** 1043–1068. MR2918009 https://doi.org/10.1109/TIT.2011.2173718

[32] 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

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

[34] WU, Y. and YANG, P. (2021). Polynomial methods in statistical inference: Theory and practice. arXiv preprint arXiv:2104.07317.

The Annals of Statistics
2024, Vol. 52, No. 4, 1360–1383
https://doi.org/10.1214/24-AOS2377

HIGHER-ORDER COVERAGE ERRORS OF BATCHING METHODS VIA EDGEWORTH EXPANSIONS ON t-STATISTICS

BY SHENGYI HE[a] AND HENRY LAM[b]

Industrial Engineering and Operations Research, Columbia University, [a]*sh3972@columbia.edu,* [b]*khl2114@columbia.edu*

While batching methods have been widely used in simulation and statistics, their higher-order coverage behaviors and relative advantages in this regard remain open. We develop techniques to obtain higher-order coverage errors for batching methods by building Edgeworth-type expansions on t-statistics. The coefficients in these expansions are intricate analytically, but we provide algorithms to estimate the coefficients of the n^{-1} error terms via Monte Carlo simulation. We provide insights on the effect of the number of batches on the coverage error, where we demonstrate generally nonmonotonic relations. We also compare different batching methods both theoretically and numerically, and argue that none of the methods is uniformly better than others in terms of coverage. However, when the number of batches is large, sectioned jackknife has the best coverage among all.

REFERENCES

ALEXOPOULOS, C., ARGON, N. T., GOLDSMAN, D., STEIGER, N. M., TOKOL, G. and WILSON, J. R. (2007a). Efficient computation of overlapping variance estimators for simulation. *INFORMS J. Comput*. **19** 314–327. MR2344067 https://doi.org/10.1287/ijoc.1060.0198

ALEXOPOULOS, C., ARGON, N. T., GOLDSMAN, D., TOKOL, G. and WILSON, J. R. (2007b). Overlapping variance estimators for simulation. *Oper. Res*. **55** 1090–1103. MR2372279 https://doi.org/10.1287/opre.1070.0475

ALEXOPOULOS, C., BOONE, J. H., GOLDSMAN, D., LOLOS, A., DINGEÇ, K. D. and WILSON, J. R. (2020). Steady-state quantile estimation using standardized time series. In *Proceedings of the* 2020 *Winter Simulation Conference* 289–300.

ALEXOPOULOS, C., GOLDSMAN, D., MOKASHI, A., NIE, R., SUN, Q., TIEN, K.-W. and WILSON, J. R. (2014). Sequest: A sequential procedure for estimating steady-state quantiles. In *Proceedings of the* 2014 *Winter Simulation Conference* 662–673.

ALEXOPOULOS, C., GOLDSMAN, D., MOKASHI, A. C. and WILSON, J. R. (2019). Sequential estimation of steady-state quantiles: Some new developments in methods and software. In *Proceedings of the* 2019 *Winter Simulation Conference* 3774–3785.

ALEXOPOULOS, C., GOLDSMAN, D., TANG, P. and WILSON, J. R. (2013). A sequential procedure for estimating the steady-state mean using standardized time series. In *Proceedings of the* 2013 *Winter Simulation Conference* 613–622.

ALEXOPOULOS, C., GOLDSMAN, D., TANG, P. and WILSON, J. R. (2016). SPSTS: A sequential procedure for estimating the steady-state mean using standardized time series. *IIE Trans*. **48** 864–880.

ALEXOPOULOS, C. and SEILA, A. F. (1996). Implementing the batch means method in simulation experiments. In *Proceedings of the* 1996 *Winter Simulation Conference* 214–221.

ANTONINI, C., ALEXOPOULOS, C., GOLDSMAN, D. and WILSON, J. R. (2009). Area variance estimators for simulation using folded standardized time series. *IIE Trans*. **41** 134–144.

ASMUSSEN, S. and GLYNN, P. W. (2007). *Stochastic Simulation*: *Algorithms and Analysis*. *Stochastic Modelling and Applied Probability* **57**. Springer, New York. MR2331321

BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist*. **6** 434–451. MR0471142

BHATTACHARYA, R. N. and PURI, M. L. (1983). On the order of magnitude of cumulants of von Mises functionals and related statistics. *Ann. Probab.* **11** 346–354. MR0690132

BICKEL, P. J., GÖTZE, F. and VAN ZWET, W. R. (1997). Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statist. Sinica* **7** 1–31.

MSC2020 subject classifications. Primary 62E20; secondary 62F40.
Key words and phrases. Batching methods, Edgeworth expansion, coverage errors, simulation analysis.

CALLAERT, H., JANSSEN, P. and VERAVERBEKE, N. (1980). An Edgeworth expansion for U-statistics. *Ann. Statist.* **8** 299–312. MR0560731

CALVIN, J. M. and NAKAYAMA, M. K. (2006). Permuted standardized time series for steady-state simulations. *Math. Oper. Res.* **31** 351–368. MR2234001 https://doi.org/10.1287/moor.1050.0183

CALVIN, J. M. and NAKAYAMA, M. K. (2013). Confidence intervals for quantiles with standardized time series. In *Proceedings of the* 2013 *Winter Simulations Conference* 601–612.

CHIEN, C. (1994). Batch size selection for the batch means method. In *Proceedings of the* 1994 *Winter Simulation Conference* 345–352.

DONG, J. and GLYNN, P. (2019a). A new approach to sequential stopping for stochastic simulation.

DONG, J. and GLYNN, P. W. (2019b). The asymptotic validity of sequential stopping rules for confidence interval construction using standardized time series. In *Proceedings of the* 2019 *Winter Simulation Conference* 332–343.

DUCHI, J. C., GLYNN, P. W. and NAMKOONG, H. (2021). Statistics of robust optimization: A generalized empirical likelihood approach. *Math. Oper. Res.* **46** 946–969. MR4312583 https://doi.org/10.1287/moor.2020.1085

FLEGAL, J. M. and JONES, G. L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *Ann. Statist.* **38** 1034–1070. MR2604704 https://doi.org/10.1214/09-AOS735

GEYER, C. J. (1992). Practical Markov chain Monte Carlo. *Statist. Sci.* **7** 473–483.

GLYNN, P. W. and IGLEHART, D. L. (1990). Simulation output analysis using standardized time series. *Math. Oper. Res.* **15** 1–16. MR1038232 https://doi.org/10.1287/moor.15.1.1

GLYNN, P. W. and LAM, H. (2018). Constructing simulation output intervals under input uncertainty via data sectioning. In *Proceedings of the* 2018 *Winter Simulation Conference* 1551–1562.

GOLDSMAN, D. and SCHRUBEN, L. (1990). New confidence interval estimators using standardized time series. *Manage. Sci.* **36** 393–397.

GÖTZE, F. (1984). Expansions for von Mises functionals. *Z. Wahrsch. Verw. Gebiete* **65** 599–625. MR0736148 https://doi.org/10.1007/BF00531841

GÖTZE, F. and HIPP, C. (1983). Asymptotic expansions for sums of weakly dependent random vectors. *Z. Wahrsch. Verw. Gebiete* **64** 211–239. MR0714144 https://doi.org/10.1007/BF01844607

HAAS, P. J. (2002). Regenerative Simulation. In *Stochastic Petri Nets: Modelling, Stability, Simulation* 189–273. Springer, New York, NY.

HALL, P. (1992). *The Bootstrap and Edgeworth Expansion. Springer Series in Statistics*. Springer, New York. MR1145237 https://doi.org/10.1007/978-1-4612-4384-7

HE, S. and LAM, H. (2021). Higher-order coverage error analysis for batching and sectioning. In *Proceedings of* 2021 *Winter Simulation Conference* 1–12.

HE, S. and LAM, H. (2024a). Supplement to "Higher-order coverage errors of batching methods via Edgeworth expansions on t-statistics." https://doi.org/10.1214/24-AOS2377SUPPA

HE, S. and LAM, H. (2024b). Codes for "Higher-order coverage errors of batching methods via Edgeworth expansions on t-statistics." https://doi.org/10.1214/24-AOS2377SUPPB

JENSEN, J. L. (1989). Asymptotic expansions for strongly mixing Harris recurrent Markov chains. *Scand. J. Stat.* **16** 47–63. MR1003968

JONES, G. L., HARAN, M., CAFFO, B. S. and NEATH, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **101** 1537–1547. MR2279478 https://doi.org/10.1198/016214506000000492

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

LADA, E. K. and WILSON, J. R. (2006). A wavelet-based spectral procedure for steady-state simulation analysis. *European J. Oper. Res.* **174** 1769–1801.

LEE, S. M. S. and YOUNG, G. A. (1995). Asymptotic iterated bootstrap confidence intervals. *Ann. Statist.* **23** 1301–1330. MR1353507 https://doi.org/10.1214/aos/1176324710

LEWIS, P. A. W. and ORAV, E. J. (1989). *Simulation Methodology for Statisticians, Operations Analysts, and Engineers. Vol. I. The Wadsworth & Brooks/Cole Statistics/Probability Series*. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA. MR0969125

MALINOVSKII, V. K. (1987). Limit theorems for Harris Markov chains, I. *Theory Probab. Appl.* **31** 269–317.

MEKETON, M. S. and SCHMEISER, B. (1984). Overlapping batch means: Something for nothing? In *Proceedings of the* 16*th Conference on Winter Simulation. WSC* '84 226–230. IEEE Press, New York.

MUÑOZ, D. F. and GLYNN, P. W. (1997). A batch means methodology for estimation of a nonlinear function of a steady-state mean. *Manage. Sci.* **43** 1121–1135.

NAKAYAMA, M. K. (2007). Fixed-width multiple-comparison procedures using common random numbers for steady-state simulations. *European J. Oper. Res.* **182** 1330–1349.

NAKAYAMA, M. K. (2014). Confidence intervals for quantiles using sectioning when applying variance-reduction techniques. *ACM Trans. Model. Comput. Simul.* **24** 19. MR3287088 https://doi.org/10.1145/2558328

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

POPE, A. (1995). Improving confidence intervals obtained by sectioning in Monte Carlo analysis. In *MODSIM* 1995 *International Congress on Modelling and Simulation. November* 27*th*-30*th* 150–154, Newcastle, NSW, Australia.

SCHMEISER, B. (1982). Batch size effects in the analysis of simulation output. *Oper. Res.* **30** 556–568. MR0659055 https://doi.org/10.1287/opre.30.3.556

SCHRUBEN, L. (1983). Confidence interval estimation using standardized time series. *Oper. Res.* **31** 1090–1108.

SKOVGAARD, I. M. (1986). On multivariate Edgeworth expansions. *Int. Stat. Rev.* **54** 169–186. MR0962933 https://doi.org/10.2307/1403142

SONG, W. T. and SCHMEISER, B. W. (1995). Optimal mean-squared-error batch sizes. *Manage. Sci.* **41** 110–123.

STEIGER, N. M., LADA, E. K., WILSON, J. R., ALEXOPOULOS, C., GOLDSMAN, D. and ZOUAOUI, F. (2002). ASAP2: An improved batch means procedure for simulation output analysis. In *Proceedings of the* 2002 *Winter Simulation Conference* 336–344.

STEIGER, N. M., LADA, E. K., WILSON, J. R., JOINES, J. A., ALEXOPOULOS, C. and GOLDSMAN, D. (2005). ASAP3: A batch means procedure for steady-state simulation analysis. *ACM Trans. Model. Comput. Simul.* **15** 39–73.

STEIGER, N. M. and WILSON, J. R. (2002). An improved batch means procedure for simulation output analysis. *Manage. Sci.* **48** 1569–1586.

SU, Z., PASUPATHY, R., YEH, Y. and GLYNN, P. W. (2024). Overlapping batch confidence intervals on statistical functionals constructed from time series: Application to quantiles, optimization, and estimation. *ACM Trans. Model. Comput. Simul.* **34** 10. MR4751429

TAFAZZOLI, A., STEIGER, N. M. and WILSON, J. R. (2011). N-Skart: A nonsequential skewness- and autoregression-adjusted batch-means procedure for simulation analysis. *IEEE Trans. Automat. Control* **56** 254–264. MR2761085 https://doi.org/10.1109/TAC.2010.2052137

TAFAZZOLI, A., WILSON, J. R., LADA, E. K. and STEIGER, N. M. (2008). Skart: A skewness- and autoregression-adjusted batch-means procedure for simulation analysis. In 2008 *Winter Simulation Conference* 387–395.

TAKAHASHI, H. (1988). A note on Edgeworth expansion for the von Mises functionals. *J. Multivariate Anal.* **24** 56–65. MR0925130 https://doi.org/10.1016/0047-259X(88)90102-9

WITHERS, C. S. (1983). Expansions for the distribution and quantiles of a regular functional of the empirical distribution with applications to nonparametric confidence intervals. *Ann. Statist.* **11** 577–587.

The Annals of Statistics
2024, Vol. 52, No. 4, 1384–1409
https://doi.org/10.1214/24-AOS2393

ON THE APPROXIMATION ACCURACY OF GAUSSIAN VARIATIONAL INFERENCE

BY ANYA KATSEVICH[a] AND PHILIPPE RIGOLLET[b]

Department of Mathematics, Massachusetts Institute of Technology, [a]*akatsevi@mit.edu,* [b]*rigollet@math.mit.edu*

The main computational challenge in Bayesian inference is to compute integrals against a high-dimensional posterior distribution. In the past decades, variational inference (VI) has emerged as a tractable approximation to these integrals, and a viable alternative to the more established paradigm of Markov chain Monte Carlo. However, little is known about the approximation accuracy of VI. In this work, we bound the TV error and the mean and covariance approximation error of Gaussian VI in terms of dimension and sample size. Our error analysis relies on a Hermite series expansion of the log posterior whose first terms are precisely cancelled out by the first order optimality conditions associated to the Gaussian VI optimization problem.

REFERENCES

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

[2] ALBANESE, L. and ALESSANDRELLI, A. (2022). On Gaussian spin glass with P-wise interactions. *J. Math. Phys.* **63** Paper No. 043302, 28. MR4403546 https://doi.org/10.1063/5.0079776

[3] 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

[4] BARRA, A., GENOVESE, G., GUERRA, F. and TANTARI, D. (2014). About a solvable mean field model of a Gaussian spin glass. *J. Phys. A* **47** 155002, 16. MR3191671 https://doi.org/10.1088/1751-8113/47/15/155002

[5] BEN AROUS, G., DEMBO, A. and GUIONNET, A. (2001). Aging of spherical spin glasses. *Probab. Theory Related Fields* **120** 1–67. MR1856194 https://doi.org/10.1007/PL00008774

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

[7] 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

[8] BURT, D. R., RASMUSSEN, C. E. and VAN DER WILK, M. (2020). Convergence of sparse variational inference in Gaussian processes regression. *J. Mach. Learn. Res.* **21** Paper No. 131, 63. MR4138115

[9] CHALLIS, E. and BARBER, D. (2013). Gaussian Kullback–Leibler approximate inference. *J. Mach. Learn. Res.* **14** 2239–2286. MR3111362

[10] CHÉRIEF-ABDELLATIF, B.-E. and ALQUIER, P. (2018). Consistency of variational Bayes inference for estimation and model selection in mixtures. *Electron. J. Stat.* **12** 2995–3035. MR3855643 https://doi.org/10.1214/18-EJS1475

[11] COWLES, M. K. and CARLIN, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *J. Amer. Statist. Assoc.* **91** 883–904. MR1395755 https://doi.org/10.2307/2291683

[12] DAUDEL, K. and DOUC, R. (2021). Mixture weights optimisation for alpha-divergence variational inference. *Adv. Neural Inf. Process. Syst.* **34** 4397–4408.

[13] DAUDEL, K., DOUC, R. and PORTIER, F. (2021). Infinite-dimensional gradient-based descent for alpha-divergence minimisation. *Ann. Statist.* **49** 2250–2270. MR4319249 https://doi.org/10.1214/20-aos2035

[14] DEHAENE, G. and BARTHELMÉ, S. (2018). Expectation propagation in the large data limit. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 199–217. MR3744718 https://doi.org/10.1111/rssb.12241

MSC2020 subject classifications. 62E17, 62F15.
Key words and phrases. Gaussian variational inference, posterior integrals, Hermite polynomials.

[15] DEHAENE, G. P. (2019). A deterministic and computable Bernstein–von Mises theorem. arXiv preprint. Available at arXiv:1904.02505.

[16] DEHAENE, G. P. and BARTHELMÉ, S. (2015). Bounding errors of expectation-propagation. *Adv. Neural Inf. Process. Syst.* **28**.

[17] DIAO, M., BALASUBRAMANIAN, K., CHEWI, S. and SALIM, A. (2023). Forward-backward Gaussian variational inference via JKO in the Bures–Wasserstein space. arXiv preprint. Available at arXiv:2304.05398.

[18] DURANTE, D., POZZA, F. and SZABO, B. (2023). Skewed Bernstein–von Mises theorem and skew-modal approximations. arXiv preprint. Available at arXiv:2301.03038.

[19] FASANO, A., DURANTE, D. and ZANELLA, G. (2022). Scalable and accurate variational Bayes for high-dimensional binary regression models. *Biometrika* **109** 901–919. MR4519107 https://doi.org/10.1093/biomet/asac026

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

[21] GHOSH, J. K. and RAMAMOORTHI, R. V. (2003). *Bayesian Nonparametrics*. *Springer Series in Statistics*. Springer, New York. MR1992245

[22] 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

[23] HAN, W. and YANG, Y. (2019). Statistical inference in mean-field variational Bayes. arXiv preprint. Available at arXiv:1911.01525.

[24] HELIN, T. and KRETSCHMANN, R. (2022). Non-asymptotic error estimates for the Laplace approximation in Bayesian inverse problems. *Numer. Math.* **150** 521–549. MR4382587 https://doi.org/10.1007/s00211-021-01266-9

[25] JOHNSON, R. A. (1970). Asymptotic expansions associated with posterior distributions. *Ann. Math. Stat.* **41** 851–864. MR0263198 https://doi.org/10.1214/aoms/1177696963

[26] KASPRZAK, M. J., GIORDANO, R. and BRODERICK, T. (2022). How good is your Gaussian approximation of the posterior? Finite-sample computable error bounds for a variety of useful divergences. arXiv preprint. Available at arXiv:2209.14992.

[27] KATSEVICH, A. (2023). The Laplace approximation accuracy in high dimensions: A refined analysis and new skew adjustment. arXiv preprint. Available at arXiv:2306.07262.

[28] KATSEVICH, A. (2023). Improved dimension dependence in the Bernstein von Mises Theorem via a new Laplace approximation bound. arXiv preprint. Available at arXiv:2308.06899.

[29] KATSEVICH, A. and RIGOLLET, P. (2024). Supplement to "On the approximation accuracy of Gaussian variational inference." https://doi.org/10.1214/24-AOS2393SUPP

[30] KLEIJN, B. J. K. and VAN DER VAART, A. W. (2012). The Bernstein–Von-Mises theorem under misspecification. *Electron. J. Stat.* **6** 354–381. MR2988412 https://doi.org/10.1214/12-EJS675

[31] LAMBERT, M., CHEWI, S., BACH, F., BONNABEL, S. and RIGOLLET, P. (2022). Variational inference via Wasserstein gradient flows. arXiv preprint. Available at arXiv:2205.15902.

[32] LANG, S. (1993). *Real and Functional Analysis*, 3rd ed. *Graduate Texts in Mathematics* **142**. Springer, New York. MR1216137 https://doi.org/10.1007/978-1-4612-0897-6

[33] LEBEDEV, N. N. (1972). *Special Functions and Their Applications*. Dover, New York. Revised edition, translated from the Russian and edited by Richard A. Silverman, Unabridged and corrected republication. MR0350075

[34] LU, Y. (2017). On the Bernstein–von Mises theorem for high dimensional nonlinear Bayesian inverse problems. arXiv preprint. Available at arXiv:1706.00289.

[35] MILLER, J. W. (2021). Asymptotic normality, concentration, and coverage of generalized posteriors. *J. Mach. Learn. Res.* **22** Paper No. 168, 53. MR4318524

[36] MUKHERJEE, S. and SEN, S. (2022). Variational inference in high-dimensionalAccurate approximations for posterior linear regression. *J. Mach. Learn. Res.* **23** Paper No. [304], 56. MR4577743

[37] SPOKOINY, V. (2013). Bernstein–von Mises Theorem for growing parameter dimension. arXiv preprint. Available at arXiv:1302.3430.

[38] SPOKOINY, V. (2023). Dimension free nonasymptotic bounds on the accuracy of high-dimensional Laplace approximation. *SIAM/ASA J. Uncertain. Quantificat.* **11** 1044–1068. MR4646387 https://doi.org/10.1137/22M1495688

[39] SUR, P. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. Ph.D. thesis, Stanford Univ. MR4197622

[40] TIERNEY, L. and KADANE, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc.* **81** 82–86. MR0830567

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

[42] 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

[43] WENG, R. C. (2010). A Bayesian Edgeworth expansion by Stein's identity. *Bayesian Anal.* **5** 741–763. MR2740155 https://doi.org/10.1214/10-BA526

[44] 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

[45] 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

[46] ZHANG, X., LING, C. and QI, L. (2012). The best rank-1 approximation of a symmetric tensor and related spherical optimization problems. *SIAM J. Matrix Anal. Appl.* **33** 806–821. MR3023452 https://doi.org/10.1137/110835335

The Annals of Statistics
2024, Vol. 52, No. 4, 1410–1432
https://doi.org/10.1214/24-AOS2394

E-STATISTICS, GROUP INVARIANCE AND ANYTIME-VALID TESTING

BY MURIEL FELIPE PÉREZ-ORTIZ[1,a], TYRON LARDY[2,b], RIANNE DE HEIDE[3,c] AND PETER D. GRÜNWALD[4,d]

[1]*Statistics, Probability and Operations Research Cluster, Eindhoven University of Technology,* [a]*m.f.perez.ortiz@tue.nl*
[2]*Mathematical Institute, Leiden University,* [b]*t.d.lardy@math.leidenuniv.nl*
[3]*Department of Mathematics, Vrije Universiteit,* [c]*r.de.heide@vu.nl*
[4]*Centrum Wiskunde & Informatica,* [d]*pdg@cwi.nl*

We study worst-case-growth-rate-optimal (GROW) e-statistics for hypothesis testing between two group models. It is known that under a mild condition on the action of the underlying group G on the data, there exists a maximally invariant statistic. We show that among all e-statistics, invariant or not, the likelihood ratio of the maximally invariant statistic is GROW, both in the absolute and in the relative sense, and that an anytime-valid test can be based on it. The GROW e-statistic is equal to a Bayes factor with a right Haar prior on G. Our treatment avoids nonuniqueness issues that sometimes arise for such priors in Bayesian contexts. A crucial assumption on the group G is its amenability, a well-known group-theoretical condition, which holds, for instance, in scale-location families. Our results also apply to finite-dimensional linear regression.

REFERENCES

ANDERSSON, S. (1982). Distributions of maximal invariants using quotient measures. *Ann. Statist.* **10** 955–961. MR0663446

BERGER, J. O., PERICCHI, L. R. and VARSHAVSKY, J. A. (1998). Bayes factors and marginal distributions in invariant situations. *Sankhyā Ser. A* **60** 307–321. MR1718789

BERGER, J. O. and SUN, D. (2008). Objective priors for the bivariate normal model. *Ann. Statist.* **36** 963–982. MR2396821 https://doi.org/10.1214/07-AOS501

BHOWMIK, J. L. and KING, M. L. (2007). Maximal invariant likelihood based testing of semi-linear models. *Statist. Papers* **48** 357–383. MR2395928 https://doi.org/10.1007/s00362-006-0342-7

BONDAR, J. V. (1976). Borel cross-sections and maximal invariants. *Ann. Statist.* **4** 866–877. MR0474589

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

BOURBAKI, N. (2004). *Integration. II. Chapters* 7–9. *Elements of Mathematics* (*Berlin*). Springer, Berlin. Translated from the 1963 and 1969 French originals by Sterling K. Berberian. MR2098271

CHANG, J. T. and POLLARD, D. (1997). Conditioning as disintegration. *Stat. Neerl.* **51** 287–317. MR1484954 https://doi.org/10.1111/1467-9574.00056

COVER, T. M. and THOMAS, J. A. (2006). *Elements of Information Theory. Wiley Series in Telecommunications and Signal Processing*. Wiley, New York.

COX, D. R. (1952). Sequential tests for composite hypotheses. *Proc. Camb. Philos. Soc.* **48** 290–299. MR0047292 https://doi.org/10.1017/s030500410002764x

DARLING, D. A. and ROBBINS, H. (1968). Some nonparametric sequential tests with power one. *Proc. Natl. Acad. Sci. USA* **61** 804–809. MR0238437 https://doi.org/10.1073/pnas.61.3.804

DAWID, A. P., STONE, M. and ZIDEK, J. V. (1973). Marginalization paradoxes in Bayesian and structural inference. *J. Roy. Statist. Soc. Ser. B* **35** 189–233. MR0365805

EATON, M. L. (1989). *Group Invariance Applications in Statistics. NSF-CBMS Regional Conference Series in Probability and Statistics* **1**. IMS, Hayward. MR1089423

EATON, M. L. and SUDDERTH, W. D. (1999). Consistency and strong inconsistency of group-invariant predictive inferences. *Bernoulli* **5** 833–854. MR1715441 https://doi.org/10.2307/3318446

MSC2020 subject classifications. Primary 62L10; secondary 62F03, 62B99.
Key words and phrases. E-values, anytime-valid methods, group invariance, Bayes factors, right Haar priors.

EATON, M. L. and SUDDERTH, W. D. (2002). Group invariant inference and right Haar measure. *J. Statist. Plann. Inference* **103** 87–99. MR1896985 https://doi.org/10.1016/S0378-3758(01)00199-9

GIRI, N., KIEFER, J. and STEIN, C. (1963). Minimax character of Hotelling's T^2 test in the simplest case. *Ann. Math. Stat.* **34** 1524–1535. MR0156408 https://doi.org/10.1214/aoms/1177703884

GRÜNWALD, P., DE HEIDE, R. and KOOLEN, W. (2023). Safe testing. Available at: arXiv:1906.07801 [cs, math, stat]. First version on arXiv 2019. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* To appear.

GRÜNWALD, P. D. (2023). The e-posterior. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **381** Paper No. 20220146, 21. MR4590499

HALL, W. J., WIJSMAN, R. A. and GHOSH, J. K. (1965). The relationship between sufficiency and invariance with applications in sequential analysis. *Ann. Math. Stat.* **36** 575–614. MR0178552 https://doi.org/10.1214/aoms/1177700169

HENZI, A., PUKE, M., DIMITRIADIS, T. and ZIEGEL, J. (2024). A safe Hosmer–Lemeshow test. *New Engl. J. Statist. Data Sci.* **2** 175–189.

JEFFREYS, H. (1961). *Theory of Probability*, 3rd ed. Clarendon Press, Oxford. MR0187257

KARIYA, T. (1980). Locally robust tests for serial correlation in least squares regression. *Ann. Statist.* **8** 1065–1070. MR0585704

KULLBACK, S. and LEIBLER, R. A. (1951). On information and sufficiency. *Ann. Math. Stat.* **22** 79–86. MR0039968 https://doi.org/10.1214/aoms/1177729694

LAI, T. L. (1976). On confidence sequences. *Ann. Statist.* **4** 265–280. MR0395103

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

LIANG, F. and BARRON, A. (2004). Exact minimax strategies for predictive density estimation, data compression, and model selection. *IEEE Trans. Inf. Theory* **50** 2708–2726. MR2096988 https://doi.org/10.1109/TIT.2004.836922

PATERSON, A. L. T. (1988). *Amenability*. *Mathematical Surveys and Monographs* **29**. Amer. Math. Soc., Providence, RI. MR0961261 https://doi.org/10.1090/surv/029

PÉREZ-ORTIZ, M. F., LARDY, T., DE HEIDE, R. and GRÜNWALD, P. D. (2024). Supplement to "E-statistics, group invariance and anytime-valid testing." https://doi.org/10.1214/24-AOS2394SUPP

RAMDAS, A., GRÜNWALD, P., VOVK, V. and SHAFER, G. (2023). Game-theoretic statistics and safe anytime-valid inference. *Statist. Sci.* **38** 576–601. MR4665027 https://doi.org/10.1214/23-sts894

RAMDAS, A., RUF, J., LARSSON, M. and KOOLEN, W. (2020). Admissible anytime-valid sequential inference must rely on nonnegative martingales. arXiv:2009.03167 [math, stat].

REITER, H. and STEGEMAN, J. D. (2000). *Classical Harmonic Analysis and Locally Compact Groups*, 2nd ed. *London Mathematical Society Monographs*. *New Series* **22**. Clarendon, New York. MR1802924

REN, Z. and BARBER, R. F. (2024). Derandomised knockoffs: Leveraging e-values for false discovery rate control. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **86** 122–154. MR4754084 https://doi.org/10.1093/jrsssb/qkad085

ROBBINS, H. (1970). Statistical methods related to the law of the iterated logarithm. *Ann. Math. Stat.* **41** 1397–1409. MR0277063 https://doi.org/10.1214/aoms/1177696786

ROUDER, J. N., SPECKMAN, P. L., SUN, D., MOREY, R. D. and IVERSON, G. (2009). Bayesian t-tests for accepting and rejecting the null hypothesis. *Psychon. Bull. Rev.* **16** 225–237. https://doi.org/10.3758/PBR.16.2.225

ROY, S. N. and BARGMANN, R. E. (1958). Tests of multiple independence and the associated confidence bounds. *Ann. Math. Stat.* **29** 491–503. MR0093862 https://doi.org/10.1214/aoms/1177706624

RUSHTON, S. (1950). On a sequential t-test. *Biometrika* **37** 326–333. MR0044080 https://doi.org/10.1093/biomet/37.3-4.326

SHAFER, G. (2021). Testing by betting: A strategy for statistical and scientific communication. *J. Roy. Statist. Soc. Ser. A* **184** 407–478. MR4255905 https://doi.org/10.1111/rssa.12647

SHALAEVSKII, O. V. (1971). Minimax character of Hotelling's T2 test. I. In *Investigations in Classical Problems of Probability Theory and Mathematical Statistics*: *Part I* (V. M. Kalinin and O. V. Shalaevskii, eds.) 74–101. Springer, Boston. https://doi.org/10.1007/978-1-4684-8211-9_2

SUBBAIAH, P. and MUDHOLKAR, G. S. (1978). A comparison of two tests for the significance of a mean vector. *J. Amer. Statist. Assoc.* **73** 414–418. MR0497422

SUN, D. and BERGER, J. O. (2007). Objective Bayesian analysis for the multivariate normal model. In *Bayesian Statistics* 8. *Oxford Sci. Publ.* 525–562. Oxford Univ. Press, Oxford. MR2433206

TURNER, R. J., LY, A. and GRÜNWALD, P. D. (2024). Generic E-variables for exact sequential k-sample tests that allow for optional stopping. *J. Statist. Plann. Inference* **230** Paper No. 106116, 15. MR4661178 https://doi.org/10.1016/j.jspi.2023.106116

VOVK, V. and WANG, R. (2021). E-values: Calibration, combination and applications. *Ann. Statist.* **49** 1736–1754. MR4298879 https://doi.org/10.1214/20-aos2020

WALD, A. (1945). Sequential tests of statistical hypotheses. *Ann. Math. Stat.* **16** 117–186. MR0013275 https://doi.org/10.1214/aoms/1177731118

WANG, R. and RAMDAS, A. (2022). False discovery rate control with e-values. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 822–852. MR4460577

WAUDBY-SMITH, I. and RAMDAS, A. (2024). Estimating means of bounded random variables by betting. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **86** 1–27. MR4716192 https://doi.org/10.1093/jrsssb/qkad009

ZHANG, Z., RAMDAS, A. and WANG, R. (2023). When do exact and powerful p-values and e-values exist? arXiv preprint. Available at: arXiv:2305.16539.

The Annals of Statistics
2024, Vol. 52, No. 4, 1433–1459
https://doi.org/10.1214/24-AOS2397

HEAVY-TAILED BAYESIAN NONPARAMETRIC ADAPTATION

BY SERGIOS AGAPIOU[1,a] AND ISMAËL CASTILLO[2,b]

[1]*Department of Mathematics and Statistics, University of Cyprus,* [a]*agapiou.sergios@ucy.ac.cy*
[2]*Laboratoire de Probabilités, Statistique et Modélisation, Sorbonne University,* [b]*ismael.castillo@upmc.fr*

We propose a new Bayesian strategy for adaptation to smoothness in nonparametric models based on heavy-tailed series priors. We illustrate it in a variety of settings, showing in particular that the corresponding Bayesian posterior distributions achieve adaptive rates of contraction in the minimax sense (up to logarithmic factors) without the need to sample hyperparameters. Unlike many existing procedures, where a form of direct model (or estimator) selection is performed, the method can be seen as performing a *soft* selection through the prior tail. In Gaussian regression, such heavy-tailed priors are shown to lead to (near-)optimal simultaneous adaptation both in the L^2- and L^∞-sense. Results are also derived for linear inverse problems, for anisotropic Besov classes, and for certain losses in more general models through the use of tempered posterior distributions. We present numerical simulations corroborating the theory.

REFERENCES

[1] AGAPIOU, S., BARDSLEY, J. M., PAPASPILIOPOULOS, O. and STUART, A. M. (2014). Analysis of the Gibbs sampler for hierarchical inverse problems. *SIAM/ASA J. Uncertain. Quantificat.* **2** 511–544. MR3283919 https://doi.org/10.1137/130944229

[2] AGAPIOU, S., BURGER, M., DASHTI, M. and HELIN, T. (2018). Sparsity-promoting and edge-preserving maximum *a posteriori* estimators in non-parametric Bayesian inverse problems. *Inverse Probl.* **34** 045002, 37. MR3774703 https://doi.org/10.1088/1361-6420/aaacac

[3] AGAPIOU, S. and CASTILLO, I. (2024). Supplement to "Heavy-tailed Bayesian nonparametric adaptation." https://doi.org/10.1214/24-AOS2397SUPP

[4] AGAPIOU, S., DASHTI, M. and HELIN, T. (2021). Rates of contraction of posterior distributions based on p-exponential priors. *Bernoulli* **27** 1616–1642. MR4278794 https://doi.org/10.3150/20-bej1285

[5] AGAPIOU, S. and SAVVA, A. (2024). Adaptive inference over Besov spaces in the white noise model using p-exponential priors. *Bernoulli* **30** 2275–2300. MR4746608 https://doi.org/10.3150/23-bej1673

[6] AGAPIOU, S. and WANG, S. (2024). Laplace priors and spatial inhomogeneity in Bayesian inverse problems. *Bernoulli* **30** 878–910. MR4699538 https://doi.org/10.3150/22-bej1563

[7] 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

[8] ARBEL, J., GAYRAUD, G. and ROUSSEAU, J. (2013). Bayesian optimal adaptive estimation using a sieve prior. *Scand. J. Stat.* **40** 549–570. MR3091697 https://doi.org/10.1002/sjos.12002

[9] BANDEIRA, A. S., MAILLARD, A., NICKL, R. and WANG, S. (2023). On free energy barriers in Gaussian priors and failure of cold start MCMC for high-dimensional unimodal distributions. *Philos. Trans. R. Soc. A* **381** Paper No. 20220150, 29. MR4590496 https://doi.org/10.1098/rsta.2022.0150

[10] BARRON, A., BIRGÉ, L. and MASSART, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113** 301–413. MR1679028 https://doi.org/10.1007/s004400050210

[11] BELITSER, E. and GHOSAL, S. (2003). Adaptive Bayesian inference on the mean of an infinite-dimensional normal distribution. *Ann. Statist.* **31** 536–559. Dedicated to the memory of Herbert E. Robbins. MR1983541 https://doi.org/10.1214/aos/1051027880

[12] BUI-THANH, T. and GHATTAS, O. (2015). A scalable algorithm for MAP estimators in Bayesian inverse problems with Besov priors. *Inverse Probl. Imaging* **9** 27–53. MR3305885 https://doi.org/10.3934/ipi.2015.9.27

MSC2020 subject classifications. 62G05, 62G20.
Key words and phrases. Bayesian nonparametrics, frequentist analysis of posterior distributions, adaptation to smoothness, heavy tails, fractional posteriors.

[13] CAI, T. T. (2008). On information pooling, adaptability and superefficiency in nonparametric function estimation. *J. Multivariate Anal.* **99** 421–436. MR2396972 https://doi.org/10.1016/j.jmva.2006.11.010

[14] CARVALHO, C. M., POLSON, N. G. and SCOTT, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika* **97** 465–480. MR2650751 https://doi.org/10.1093/biomet/asq017

[15] CASTILLO, I. (2008). Lower bounds for posterior rates with Gaussian process priors. *Electron. J. Stat.* **2** 1281–1299. MR2471287 https://doi.org/10.1214/08-EJS273

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

[17] CASTILLO, I., KERKYACHARIAN, G. and PICARD, D. (2014). Thomas Bayes' walk on manifolds. *Probab. Theory Related Fields* **158** 665–710. MR3176362 https://doi.org/10.1007/s00440-013-0493-0

[18] CASTILLO, I. and MISMER, R. (2021). Spike and slab Pólya tree posterior densities: Adaptive inference. *Ann. Inst. Henri Poincaré Probab. Stat.* **57** 1521–1548. MR4291462 https://doi.org/10.1214/20-aihp1132

[19] CASTILLO, I. and NICKL, R. (2013). Nonparametric Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.* **41** 1999–2028. MR3127856 https://doi.org/10.1214/13-AOS1133

[20] CASTILLO, I. and NICKL, R. (2014). On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. *Ann. Statist.* **42** 1941–1969. MR3262473 https://doi.org/10.1214/14-AOS1246

[21] CASTILLO, I. and ROČKOVÁ, V. (2021). Uncertainty quantification for Bayesian CART. *Ann. Statist.* **49** 3482–3509. MR4352538 https://doi.org/10.1214/21-aos2093

[22] CAVALIER, L. (2011). Inverse problems in statistics. In *Inverse Problems and High-Dimensional Estimation. Lect. Notes Stat. Proc.* **203** 3–96. Springer, Heidelberg. MR2868199 https://doi.org/10.1007/978-3-642-19989-9_1

[23] DASHTI, M., HARRIS, S. and STUART, A. (2012). Besov priors for Bayesian inverse problems. *Inverse Probl. Imaging* **6** 183–200. MR2942737 https://doi.org/10.3934/ipi.2012.6.183

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

[25] DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1995). Wavelet shrinkage: Asymptopia? *J. Roy. Statist. Soc. Ser. B* **57** 301–369. With discussion and a reply by the authors. MR1323344

[26] GAO, C. and ZHOU, H. H. (2016). Rate exact Bayesian adaptation with modified block priors. *Ann. Statist.* **44** 318–345. MR3449770 https://doi.org/10.1214/15-AOS1368

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

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

[29] GINÉ, E. and NICKL, R. (2011). Rates of contraction for posterior distributions in L^r-metrics, $1 \le r \le \infty$. *Ann. Statist.* **39** 2883–2911. MR3012395 https://doi.org/10.1214/11-AOS924

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

[31] GIORDANO, M. (2023). Besov–Laplace priors in density estimation: Optimal posterior contraction rates and adaptation. *Electron. J. Stat.* **17** 2210–2249. MR4649387 https://doi.org/10.1214/23-ejs2161

[32] HOFFMANN, M., ROUSSEAU, J. and SCHMIDT-HIEBER, J. (2015). On adaptive posterior concentration rates. *Ann. Statist.* **43** 2259–2295. MR3396985 https://doi.org/10.1214/15-AOS1341

[33] IBRAGIMOV, I. A. and HAS'MINSKIĬ, R. Z. (1980). An estimate of the density of a distribution. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **98** 61–85, 161–162, 166. Studies in mathematical statistics, IV. MR0591862

[34] JOHNSTONE, I. M. (2019). Gaussian estimation: Sequence and wavelet models. Unpublished manuscript.

[35] KNAPIK, B. T., SZABÓ, B. T., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2016). Bayes procedures for adaptive inference in inverse problems for the white noise model. *Probab. Theory Related Fields* **164** 771–813. MR3477780 https://doi.org/10.1007/s00440-015-0619-7

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

[37] KOLEHMAINEN, V., LASSAS, M., NIINIMÄKI, K. and SILTANEN, S. (2012). Sparsity-promoting Bayesian inversion. *Inverse Probl.* **28** 025005, 28. MR2876856 https://doi.org/10.1088/0266-5611/28/2/025005

[38] L'HUILLIER, A., TRAVIS, L., CASTILLO, I. and RAY, K. (2023). Semiparametric inference using fractional posteriors. *J. Mach. Learn. Res.* **24** Paper No. [389], 61. MR4720845 https://doi.org/10.4995/agt.2023.18504

[39] LASSAS, M., SAKSMAN, E. and SILTANEN, S. (2009). Discretization-invariant Bayesian inversion and Besov space priors. *Inverse Probl. Imaging* **3** 87–122. MR2558305 https://doi.org/10.3934/ipi.2009.3.87

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

[41] LEPSKIĬ, O. V. (1991). Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Teor. Veroyatn. Primen.* **36** 645–659. MR1147167 https://doi.org/10.1137/1136085

[42] NAULET, Z. (2022). Adaptive Bayesian density estimation in sup-norm. *Bernoulli* **28** 1284–1308. MR4388939 https://doi.org/10.3150/21-bej1387

[43] NICKL, R.and WANG, S. (2024). On polynomial-time computation of high-dimensional posterior measures by Langevin-type algorithms *J. Eur. Math. Soc.* (*JEMS*) **26** 1031–1112. MR4721029 https://doi.org/10.4171/jems/1304

[44] RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning. Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR2514435

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

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

[47] SCRICCIOLO, C. (2006). Convergence rates for Bayesian density estimation of infinite-dimensional exponential families. *Ann. Statist.* **34** 2897–2920. MR2329472 https://doi.org/10.1214/009053606000000911

[48] SHAH, A., WILSON, A. and GHAHRAMANI, Z. (2014). Student-t processes as alternatives to Gaussian processes. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics* (S. Kaski and J. Corander, eds.). *Proceedings of Machine Learning Research* **33** 877–885. PMLR, Reykjavik , Iceland.

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

[50] STAN DEVELOPMENT TEAM (2024). Stan Modelling Language Users Guide and Reference Manual v. 2.34.

[51] SULLIVAN, T. J. (2017). Well-posed Bayesian inverse problems and heavy-tailed stable quasi-Banach space priors. *Inverse Probl. Imaging* **11** 857–874. MR3681971 https://doi.org/10.3934/ipi.2017040

[52] SUURONEN, J., CHADA, N. K. and ROININEN, L. (2022). Cauchy Markov random field priors for Bayesian inversion. *Stat. Comput.* **32** Paper No. 33, 26. MR4402180 https://doi.org/10.1007/s11222-022-10089-z

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

[54] SZABÓ, B. T., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2013). Empirical Bayes scaling of Gaussian priors in the white noise model. *Electron. J. Stat.* **7** 991–1018. MR3044507 https://doi.org/10.1214/13-EJS798

[55] TRIEBEL, H. (1983). *Theory of Function Spaces. Monographs in Mathematics* **78**. Birkhäuser, Basel. MR0781540 https://doi.org/10.1007/978-3-0346-0416-1

[56] 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

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

[58] VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2009). Adaptive Bayesian estimation using a Gaussian random field with inverse gamma bandwidth. *Ann. Statist.* **37** 2655–2675. MR2541442 https://doi.org/10.1214/08-AOS678

[59] WALKER, S. and HJORT, N. L. (2001). On Bayesian consistency. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 811–821. MR1872068 https://doi.org/10.1111/1467-9868.00314

[60] 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

[61] 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

[62] ZHANG, T. (2006). From ϵ-entropy to KL-entropy: Analysis of minimum information complexity density estimation. *Ann. Statist.* **34** 2180–2210. MR2291497 https://doi.org/10.1214/009053606000000704

[63] ZHAO, L. H. (2000). Bayesian aspects of some nonparametric problems. *Ann. Statist.* **28** 532–552. MR1790008 https://doi.org/10.1214/aos/1016218229

The Annals of Statistics
2024, Vol. 52, No. 4, 1460–1484
https://doi.org/10.1214/24-AOS2400

FUNDAMENTAL LIMITS OF LOW-RANK MATRIX ESTIMATION WITH DIVERGING ASPECT RATIOS

BY ANDREA MONTANARI[1,a] AND YUCHEN WU[2,b]

[1]*Department of Statistics and Department of Mathematics, Stanford University,* [a]*montanari@stanford.edu*
[2]*Department of Statistics and Data Science, the Wharton School, University of Pennsylvania,* [b]*wuyc14@wharton@upenn.edu*

We consider the problem of estimating the factors of a low-rank $n \times d$ matrix, when this is corrupted by additive Gaussian noise. A special example of our setting corresponds to clustering mixtures of Gaussians with equal (known) covariances. Simple spectral methods do not take into account the distribution of the entries of these factors and are therefore often suboptimal. Here, we characterize the asymptotics of the minimum estimation error under the assumption that the distribution of the entries is known to the statistician.

Our results apply to the high-dimensional regime $n, d \to \infty$ and $d/n \to \infty$ (or $d/n \to 0$) and generalize earlier work that focused on the proportional asymptotics $n, d \to \infty$, $d/n \to \delta \in (0, \infty)$. We outline an interesting signal strength regime in which $d/n \to \infty$ and partial recovery is possible for the left singular vectors while impossible for the right singular vectors.

We illustrate the general theory by deriving consequences for Gaussian mixture clustering and carrying out a numerical study on genomics data.

REFERENCES

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

[2] ACHLIOPTAS, D. and MCSHERRY, F. (2005). On spectral learning of mixtures of distributions. In *Learning Theory. Lecture Notes in Computer Science* **3559** 458–469. Springer, Berlin. MR2203280 https://doi.org/10.1007/11503415_31

[3] AMINI, A. A. and WAINWRIGHT, M. J. (2009). High-dimensional analysis of semidefinite relaxations for sparse principal components. *Ann. Statist.* **37** 2877–2921. MR2541450 https://doi.org/10.1214/08-AOS664

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

[5] AWASTHI, P. and SHEFFET, O. (2012). Improved spectral-norm bounds for clustering. In *Approximation, Randomization, and Combinatorial Optimization. Lecture Notes in Computer Science* **7408** 37–49. Springer, Heidelberg. MR3003539 https://doi.org/10.1007/978-3-642-32512-0_4

[6] 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

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

[8] 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

[9] BANKS, J., MOORE, C., VERSHYNIN, R., VERZELEN, N. and XU, J. (2018). Information-theoretic bounds and phase transitions in clustering, sparse PCA, and submatrix localization. *IEEE Trans. Inf. Theory* **64** 4872–4994. MR3819345 https://doi.org/10.1109/tit.2018.2810020

[10] BARBIER, J. and MACRIS, N. (2019). The adaptive interpolation method: A simple scheme to prove replica formulas in Bayesian inference. *Probab. Theory Related Fields* **174** 1133–1185. MR3980313 https://doi.org/10.1007/s00440-018-0879-0

MSC2020 subject classifications. Primary 62B10, 60B20; secondary 94A17.
Key words and phrases. Matrix factorization, Gaussian mixture model.

[11] 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

[12] BELKIN, M. and SINHA, K. (2015). Polynomial learning of distribution families. *SIAM J. Comput.* **44** 889–911. MR3366914 https://doi.org/10.1137/13090818X

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

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

[15] BERTHET, Q. and RIGOLLET, P. (2013). Optimal detection of sparse principal components in high dimension. *Ann. Statist.* **41** 1780–1815. MR3127849 https://doi.org/10.1214/13-AOS1127

[16] BERTHET, Q. and RIGOLLET, P. (2013). Computational lower bounds for sparse PCA. ArXiv preprint. Available at arXiv:1304.0828.

[17] BRUBAKER, S. C. and VEMPALA, S. S. (2008). Isotropic PCA and affine-invariant clustering. In *Building Bridges*. *Bolyai Soc. Math. Stud.* **19** 241–281. Springer, Berlin. MR2484643 https://doi.org/10.1007/978-3-540-85221-6_8

[18] BUBECK, S., DING, J., ELDAN, R. and RÁCZ, M. Z. (2016). Testing for high-dimensional geometry in random graphs. *Random Structures Algorithms* **49** 503–532. MR3545825 https://doi.org/10.1002/rsa.20633

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

[20] 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

[21] DASGUPTA, S. (1999). Learning mixtures of Gaussians. In 40*th Annual Symposium on Foundations of Computer Science* (*New York*, 1999) 634–644. IEEE Computer Soc., Los Alamitos, CA. MR1917603 https://doi.org/10.1109/SFFCS.1999.814639

[22] DASGUPTA, S. and SCHULMAN, L. (2007). A probabilistic analysis of EM for mixtures of separated, spherical Gaussians. *J. Mach. Learn. Res.* **8** 203–226. MR2320668

[23] 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

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

[25] DESHPANDE, Y. and MONTANARI, A. (2016). Sparse PCA via covariance thresholding. *J. Mach. Learn. Res.* **17** Paper No. 141. MR3555032

[26] DIA, M., MACRIS, N., KRZAKALA, F., LESIEUR, T., ZDEBOROVÁ, L. et al. (2016). Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. *Adv. Neural Inf. Process. Syst.* **29**.

[27] DOMINGUEZ, T. and MOURRAT, J.-C. (2024). Mutual information for the sparse stochastic block model. *Ann. Probab.* **52** 434–501. MR4718399 https://doi.org/10.1214/23-aop1665

[28] DONOHO, D. L. and FELDMAN, M. J. (2022). Optimal Eigenvalue Shrinkage in the Semicircle Limit. Available at arXiv:2210.04488.

[29] DUA, D. and GRAFF, C. (2017). UCI machine learning repository.

[30] EL ALAOUI, A. and KRZAKALA, F. (2018). Estimation in the spiked Wigner model: A short proof of the replica formula. In 2018 *IEEE International Symposium on Information Theory* (*ISIT*) 1874–1878. IEEE, New York.

[31] FEI, Y. and CHEN, Y. (2018). Hidden integrality of SDP relaxations for sub-Gaussian mixture models. In *Conference on Learning Theory* 1931–1965. PMLR.

[32] FELDMAN, J., SERVEDIO, R. A. and O'DONNELL, R. (2006). PAC learning axis-aligned mixtures of Gaussians with no separation assumption. In *Learning Theory*. *Lecture Notes in Computer Science* **4005** 20–34. Springer, Berlin. MR2277916 https://doi.org/10.1007/11776420_5

[33] FELDMAN, M. J. (2023). Spiked singular values and vectors under extreme aspect ratios. *J. Multivariate Anal.* **196** Paper No. 105187. MR4575691 https://doi.org/10.1016/j.jmva.2023.105187

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

[35] GUIONNET, A., KO, J., KRZAKALA, F. and ZDEBOROVÁ, L. (2023). Estimating rank-one matrices with mismatched prior and noise: Universality and large deviations. Available at arXiv:2306.09283.

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

[37] HAJEK, B., WU, Y. and XU, J. (2017). Submatrix localization via message passing. *J. Mach. Learn. Res.* **18** Paper No. 186. MR3827074

[38] HARDT, M. and PRICE, E. (2015). Tight bounds for learning a mixture of two Gaussians [extended abstract]. In *STOC'*15—*Proceedings of the* 2015 *ACM Symposium on Theory of Computing* 753–760. ACM, New York. MR3388255

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

[40] HSU, D. and KAKADE, S. M. (2013). Learning mixtures of spherical Gaussians: Moment methods and spectral decompositions. In *ITCS'*13—*Proceedings of the* 2013 *ACM Conference on Innovations in Theoretical Computer Science* 11–19. ACM, New York. MR3385380

[41] IGUCHI, T., MIXON, D. G., PETERSON, J. and VILLAR, S. (2015). On the tightness of an SDP relaxation of k-means. ArXiv preprint. Available at arXiv:1505.04778.

[42] JIN, C., ZHANG, Y., BALAKRISHNAN, S., WAINWRIGHT, M. J. and JORDAN, M. I. (2016). Local maxima in the likelihood of Gaussian mixture models: Structural results and algorithmic consequences. *Adv. Neural Inf. Process. Syst.* **29**.

[43] 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

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

[45] JOHNSTONE, I. M. and LU, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* **104** 682–693. MR2751448 https://doi.org/10.1198/jasa.2009.0121

[46] JOHNSTONE, I. M. and ONATSKI, A. (2020). Testing in high-dimensional spiked models. *Ann. Statist.* **48** 1231–1254. MR4124321 https://doi.org/10.1214/18-AOS1697

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

[48] KALAI, A. T., MOITRA, A. and VALIANT, G. (2010). Efficiently learning mixtures of two Gaussians. In *STOC'*10—*Proceedings of the* 2010 *ACM International Symposium on Theory of Computing* 553–562. ACM, New York. MR2743304

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

[50] KOLAR, M., BALAKRISHNAN, S., RINALDO, A. and SINGH, A. (2011). Minimax localization of structural information in large noisy matrices. *Adv. Neural Inf. Process. Syst.* **24**.

[51] KUMAR, A. and KANNAN, R. (2010). Clustering with spectral norm and the k-means algorithm. In 2010 *IEEE* 51*st Annual Symposium on Foundations of Computer Science—FOCS* 2010 299–308. IEEE Computer Soc., Los Alamitos, CA. MR3025203

[52] 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

[53] LESIEUR, T., DE BACCO, C., BANKS, J., KRZAKALA, F., MOORE, C. and ZDEBOROVÁ, L. (2016). Phase transitions and optimal algorithms in high-dimensional Gaussian mixture clustering. In 2016 54*th Annual Allerton Conference on Communication, Control, and Computing* (*Allerton*) 601–608. IEEE, New York.

[54] LESIEUR, T., KRZAKALA, F. and ZDEBOROVÁ, L. (2015). MMSE of probabilistic low-rank matrix estimation: Universality with respect to the output channel. In 2015 53*rd Annual Allerton Conference on Communication, Control, and Computing* (*Allerton*) 680–687. IEEE.

[55] LI, X., LI, Y., LING, S., STROHMER, T. and WEI, K. (2020). When do birds of a feather flock together? k-means, proximity, and conic programming. *Math. Program.* **179** 295–341. MR4050142 https://doi.org/10.1007/s10107-018-1333-x

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

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

[58] MACRIS, N., RUSH, C. et al. (2020). All-or-nothing statistical and computational phase transitions in sparse spiked matrix estimation. *Adv. Neural Inf. Process. Syst.* **33** 14915–14926.

[59] MÉZARD, M., PARISI, G. and VIRASORO, M. A. (1987). *Spin Glass Theory and Beyond*. *World Scientific Lecture Notes in Physics* **9**. World Scientific Co., Inc., Teaneck, NJ. MR1026102

[60] MIOLANE, L. (2017). Fundamental limits of low-rank matrix estimation: the non-symmetric case. ArXiv preprint. Available at arXiv:1702.00473.

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

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

[63] MONDELLI, M. and VENKATARAMANAN, R. (2021). PCA initialization for approximate message passing in rotationally invariant models. *Adv. Neural Inf. Process. Syst.* **34** 29616–29629.

[64] MONTANARI, A. and SEN, S. (2022). A Short Tutorial on Mean-Field Spin Glass Techniques for Non-Physicists. Available at arXiv:2204.02909.

[65] 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

[66] MONTANARI, A. and WU, Y. (2024). Supplement to "Fundamental limits of low-rank matrix estimation with diverging aspect ratios." https://doi.org/10.1214/24-AOS2400SUPP

[67] NDAOUD, M. (2022). Sharp optimal recovery in the two component Gaussian mixture model. *Ann. Statist.* **50** 2096–2126. MR4474484 https://doi.org/10.1214/22-aos2178

[68] NOVEMBRE, J., JOHNSON, T., BRYC, K., KUTALIK, Z., BOYKO, A. R., AUTON, A., INDAP, A., KING, K. S., BERGMANN, S. et al. (2008). Genes mirror geography within Europe. *Nature* **456** 98–101.

[69] NOVEMBRE, J. and STEPHENS, M. (2008). Interpreting principal component analyses of spatial population genetic variation. *Nat. Genet.* **40** 646–649. https://doi.org/10.1038/ng.139

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

[71] PENG, J. and WEI, Y. (2007). Approximating K-means-type clustering via semidefinite programming. *SIAM J. Optim.* **18** 186–205. MR2299680 https://doi.org/10.1137/050641983

[72] PERRY, A., WEIN, A. S. and BANDEIRA, A. S. (2020). Statistical limits of spiked tensor models. *Ann. Inst. Henri Poincaré Probab. Stat.* **56** 230–264. MR4058987 https://doi.org/10.1214/19-AIHP960

[73] PERRY, A., WEIN, A. S., BANDEIRA, A. S. and MOITRA, A. (2018). Optimality and sub-optimality of PCA I: Spiked random matrix models. *Ann. Statist.* **46** 2416–2451. MR3845022 https://doi.org/10.1214/17-AOS1625

[74] REGEV, O. and VIJAYARAGHAVAN, A. (2017). On learning mixtures of well-separated Gaussians. In 58*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 85–96. IEEE Computer Soc., Los Alamitos, CA. MR3734220 https://doi.org/10.1109/FOCS.2017.17

[75] ROYER, M. (2017). Adaptive clustering through semidefinite programming. *Adv. Neural Inf. Process. Syst.* **30**.

[76] STAM, A. J. (1959). Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Inf. Control* **2** 101–112. MR0109101

[77] THE 1000 GENOMES PROJECT CONSORTIUM (2015). A global reference for human genetic variation. *Nature* **526** 68.

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

[79] ZHONG, X., SU, C. and FAN, Z. (2022). Empirical Bayes PCA in high dimensions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 853–878. MR4460578

The Annals of Statistics
2024, Vol. 52, No. 4, 1485–1508
https://doi.org/10.1214/24-AOS2401

ASYMPTOTIC NORMALITY AND OPTIMALITY IN NONSMOOTH STOCHASTIC APPROXIMATION

BY DAMEK DAVIS[1,a], DMITRIY DRUSVYATSKIY[2,c] AND LIWEI JIANG[1,b]

[1]*School of ORIE, Cornell University,* [a]*dsd95@cornell.edu,* [b]*lj282@cornell.edu*
[2]*Department of Mathematics, University of Washington,* [c]*ddrusv@uw.edu*

In their seminal work, Polyak and Juditsky showed that stochastic approximation algorithms for solving smooth equations enjoy a central limit theorem. Moreover, it has since been argued that the asymptotic covariance of the method is best possible among any estimation procedure in a local minimax sense of Hájek and Le Cam. A long-standing open question in this line of work is whether similar guarantees hold for important nonsmooth problems, such as stochastic nonlinear programming or stochastic variational inequalities. In this work, we show that this is indeed the case.

REFERENCES

[1] BORWEIN, J. and LEWIS, A. S. (2017). *Convex Analysis and Nonlinear Optimization*: *Theory and Examples*. Springer, Berlin.
[2] BOUMAL, N. (2023). *An Introduction to Optimization on Smooth Manifolds*. Cambridge Univ. Press, Cambridge. MR4533407
[3] CLARKE, F. H., LEDYAEV, Y. S., STERN, R. J. and WOLENSKI, P. R. (1998). *Nonsmooth Analysis and Control Theory*. *Graduate Texts in Mathematics* **178**. Springer, New York. MR1488695
[4] CUTLER, J., DÍAZ, M. and DRUSVYATSKIY, D. (2024). Stochastic approximation with decision-dependent distributions: Asymptotic normality and optimality. *J. Mach. Learn. Res.* **25** Paper No. 90, 49. MR4749126
[5] DAVIS, D., DRUSVYATSKIY, D. and CHARISOPOULOS, V. (2019). Stochastic algorithms with geometric step decay converge linearly on sharp functions. arXiv preprint. Available at arXiv:1907.09547.
[6] DAVIS, D., DRUSVYATSKIY, D. and JIANG, L. (2021). Active manifolds, stratifications, and convergence to local minima in nonsmooth optimization. arXiv preprint. Available at arXiv:2108.11832v2.
[7] DAVIS, D., DRUSVYATSKIY, D. and JIANG, L. (2024). Supplement to "Asymptotic normality and optimality in nonsmooth stochastic approximation." https://doi.org/10.1214/24-AOS2401SUPP
[8] DAVIS, D., DRUSVYATSKIY, D., KAKADE, S. and LEE, J. D. (2020). Stochastic subgradient method converges on tame functions. *Found. Comput. Math.* **20** 119–154. MR4056927 https://doi.org/10.1007/s10208-018-09409-5
[9] DONTCHEV, A. L. and ROCKAFELLAR, R. T. (2009). *Implicit Functions and Solution Mappings*: *A view from variational analysis*. *Springer Monographs in Mathematics*. Springer, Dordrecht. MR2515104 https://doi.org/10.1007/978-0-387-87821-8
[10] DRUSVYATSKIY, D., IOFFE, A. D. and LEWIS, A. S. (2016). Generic minimizing behavior in semialgebraic optimization. *SIAM J. Optim.* **26** 513–534. MR3461323 https://doi.org/10.1137/15M1020770
[11] DRUSVYATSKIY, D. and LEWIS, A. S. (2014). Optimality, identifiability, and sensitivity. *Math. Program.* **147** 467–498. MR3258532 https://doi.org/10.1007/s10107-013-0730-4
[12] DUCHI, J. C. and RUAN, F. (2021). Asymptotic optimality in stochastic optimization. *Ann. Statist.* **49** 21–48. MR4206668 https://doi.org/10.1214/19-AOS1831
[13] DUPAČOVÁ, J. and WETS, R. (1988). Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *Ann. Statist.* **16** 1517–1549. MR0964937 https://doi.org/10.1214/aos/1176351052
[14] KING, A. J. and ROCKAFELLAR, R. T. (1993). Asymptotic theory for solutions in statistical estimation and stochastic programming. *Math. Oper. Res.* **18** 148–162. MR1250111 https://doi.org/10.1287/moor.18.1.148

MSC2020 subject classifications. Primary 90C15; secondary 65K05, 65K10.
Key words and phrases. Stochastic gradient, asymptotic normality, local asymptotic minimax theory, variational inequality, active manifold.

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

[16] LEE, J. M. (2013). Smooth manifolds. In *Introduction to Smooth Manifolds*, 2nd ed. *Graduate Texts in Mathematics* **218** 1–31. Springer, New York. MR2954043

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

[18] LEMARÉCHAL, C., OUSTRY, F. and SAGASTIZÁBAL, C. (2000). The $\mathcal{U}$-Lagrangian of a convex function. *Trans. Amer. Math. Soc.* **352** 711–729. MR1487623 https://doi.org/10.1090/S0002-9947-99-02243-6

[19] LEWIS, A. S. (2002). Active sets, nonsmoothness, and sensitivity. *SIAM J. Optim.* **13** 702–725. MR1972212 https://doi.org/10.1137/S1052623401387623

[20] LEWIS, A. S. and MALICK, J. (2008). Alternating projections on manifolds. *Math. Oper. Res.* **33** 216–234. MR2393548 https://doi.org/10.1287/moor.1070.0291

[21] MIFFLIN, R. and SAGASTIZÁBAL, C. (2005). A *VU*-algorithm for convex minimization. *Math. Program.* **104** 583–608. MR2179252 https://doi.org/10.1007/s10107-005-0630-3

[22] MILLER, S. A. and MALICK, J. (2005). Newton methods for nonsmooth convex minimization: Connections among $\mathcal{U}$-Lagrangian, Riemannian Newton and SQP methods. *Math. Program.* **104** 609–633. MR2179253 https://doi.org/10.1007/s10107-005-0631-2

[23] MORDUKHOVICH, B. S. (2006). *Variational Analysis and Generalized Differentiation. I*: *Basic Theory. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **330**. Springer, Berlin. MR2191744

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

[25] PENOT, J.-P. (2012). *Calculus Without Derivatives*. *Graduate Texts in Mathematics* **266**. Springer, New York. MR2986672 https://doi.org/10.1007/978-1-4614-4538-8

[26] POLIQUIN, R. A. and ROCKAFELLAR, R. T. (1996). Prox-regular functions in variational analysis. *Trans. Amer. Math. Soc.* **348** 1805–1838. MR1333397 https://doi.org/10.1090/S0002-9947-96-01544-9

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

[28] ROCKAFELLAR, R. T. and WETS, R. J.-B. (2009). *Variational Analysis* **317**. Springer, Berlin.

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

[30] SHAPIRO, A. (2003). On a class of nonsmooth composite functions. *Math. Oper. Res.* **28** 677–692. MR2015908 https://doi.org/10.1287/moor.28.4.677.20512

[31] VAN DER VAART, A. W. (2000). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

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

[33] XIAO, L. (2009). Dual averaging method for regularized stochastic learning and online optimization. In *Advances in Neural Information Processing Systems* **22**.

The Annals of Statistics
2024, Vol. 52, No. 4, 1509–1533
https://doi.org/10.1214/24-AOS2402

BOOTSTRAP-ASSISTED INFERENCE FOR GENERALIZED GRENANDER-TYPE ESTIMATORS

BY MATIAS D. CATTANEO[1,a], MICHAEL JANSSON[2,b] AND KENICHI NAGASAWA[3,c]

[1]*Department of Operations Research and Financial Engineering, Princeton University,* [a]*cattaneo@princeton.edu*
[2]*Department of Economics, University of California, Berkeley,* [b]*mjansson@econ.berkeley.edu*
[3]*Department of Economics, University of Warwick,* [c]*kenichi.nagasawa@warwick.ac.uk*

Westling and Carone (*Ann. Statist.* **48** (2020) 1001–1024) proposed a framework for studying the large sample distributional properties of generalized Grenander-type estimators, a versatile class of nonparametric estimators of monotone functions. The limiting distribution of those estimators is representable as the left derivative of the greatest convex minorant of a Gaussian process whose monomial mean can be of unknown order (when the degree of flatness of the function of interest is unknown). The standard nonparametric bootstrap is unable to consistently approximate the large sample distribution of the generalized Grenander-type estimators even if the monomial order of the mean is known, making statistical inference a challenging endeavour in applications. To address this inferential problem, we present a bootstrap-assisted inference procedure for generalized Grenander-type estimators. The procedure relies on a carefully crafted, yet automatic, transformation of the estimator. Moreover, our proposed method can be made "flatness robust" in the sense that it can be made adaptive to the (possibly unknown) degree of flatness of the function of interest. The method requires only the consistent estimation of a single scalar quantity, for which we propose an automatic procedure based on numerical derivative estimation and the generalized jackknife. Under random sampling, our inference method can be implemented using a computationally attractive exchangeable bootstrap procedure. We illustrate our methods with examples and we also provide a small simulation study. The development of formal results is made possible by some technical results that may be of independent interest.

REFERENCES

BRUNK, H. D. (1958). On the estimation of parameters restricted by inequalities. *Ann. Math. Stat.* **29** 437–454. MR0132632 https://doi.org/10.1214/aoms/1177706621

CATTANEO, M. D., JANSSON, M. and NAGASAWA, K. (2020). Bootstrap-based inference for cube root asymptotics. *Econometrica* **88** 2203–2219. MR4162697 https://doi.org/10.3982/ecta17950

CATTANEO, M. D., JANSSON, M. and NAGASAWA, K. (2024). Supplement to "Bootstrap-Assisted Inference for Generalized Grenander-type Estimators." https://doi.org/10.1214/24-AOS2402SUPP

CAVALIERE, G. and GEORGIEV, I. (2020). Inference under random limit bootstrap measures. *Econometrica* **88** 2547–2574. MR4192549 https://doi.org/10.3982/ECTA16557

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

COX, G. (2022). A generalized argmax theorem with applications. Preprint. Available at arXiv:2209.08793.

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

GROENEBOOM, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II* (*Berkeley, Calif.*, 1983). *Wadsworth Statist./Probab. Ser.* 539–555. Wadsworth, Belmont, CA. MR0822052

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

MSC2020 subject classifications. Primary 62G09, 62G20; secondary 62G07, 62G08.
Key words and phrases. Monotone estimation, bootstrapping, robust inference.

GROENEBOOM, P. and JONGBLOED, G. (2018). Some developments in the theory of shape constrained inference. *Statist. Sci.* **33** 473–492. MR3881204 https://doi.org/10.1214/18-STS657

HAN, Q. and KATO, K. (2022). Berry-Esseen bounds for Chernoff-type nonstandard asymptotics in isotonic regression. *Ann. Appl. Probab.* **32** 1459–1498. MR4414710 https://doi.org/10.1214/21-aap1716

HONG, H. and LI, J. (2020). The numerical bootstrap. *Ann. Statist.* **48** 397–412. MR4065167 https://doi.org/10.1214/19-AOS1812

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

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

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

KUCHIBHOTLA, A. K., BALAKRISHNAN, S. and WASSERMAN, L.. (2024). The HulC: Confidence regions from convex hulls. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*

LEE, S. M. S. and YANG, P. (2020). Bootstrap confidence regions based on M-estimators under nonstandard conditions. *Ann. Statist.* **48** 274–299. MR4065162 https://doi.org/10.1214/18-AOS1803

MALLICK, S., SARKAR, S. and KUCHIBHOTLA, A. K. (2024). New asymptotic limit theory and inference for monotone regression. Preprint. Available at arXiv:2310.20058.

POLITIS, D. N. and ROMANO, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22** 2031–2050. MR1329181 https://doi.org/10.1214/aos/1176325770

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

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

SEO, M. H. and OTSU, T. (2018). Local M-estimation with discontinuous criterion for dependent and limited observations. *Ann. Statist.* **46** 344–369. MR3766955 https://doi.org/10.1214/17-AOS1552

VAN DER LAAN, M. J. and ROBINS, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality. Springer Series in Statistics*. Springer, New York. MR1958123 https://doi.org/10.1007/978-0-387-21700-0

VAN DER VAART, A. and VAN DER LAAN, M. J. (2006). Estimating a survival distribution with current status data and high-dimensional covariates. *Int. J. Biostat.* **2** Art. 9, 42 pp. MR2306498 https://doi.org/10.2202/1557-4679.1014

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

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

WESTLING, T. and CARONE, M. (2020). A unified study of nonparametric inference for monotone functions. *Ann. Statist.* **48** 1001–1024. MR4102685 https://doi.org/10.1214/19-AOS1835

WESTLING, T., GILBERT, P. and CARONE, M. (2020). Causal isotonic regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 719–747. MR4112782 https://doi.org/10.1111/rssb.12372

The Annals of Statistics
2024, Vol. 52, No. 4, 1534–1563
https://doi.org/10.1214/24-AOS2403

ONE-STEP ESTIMATION OF DIFFERENTIABLE HILBERT-VALUED PARAMETERS

BY ALEX LUEDTKE[a] AND INCHEOUL CHUNG[b]

Department of Statistics, University of Washington, [a]*aluedtke@uw.edu,* [b]*ic247@uw.edu*

We present estimators for smooth Hilbert-valued parameters, where smoothness is characterized by a pathwise differentiability condition. When the parameter space is a reproducing kernel Hilbert space, we provide a means to obtain efficient, root-n rate estimators and corresponding confidence sets. These estimators correspond to generalizations of cross-fitted one-step estimators based on Hilbert-valued efficient influence functions. We give theoretical guarantees even when arbitrary estimators of nuisance functions are used, including those based on machine learning techniques. We show that these results naturally extend to Hilbert spaces that lack a reproducing kernel, as long as the parameter has an efficient influence function. However, we also uncover the unfortunate fact that, when there is no reproducing kernel, many interesting parameters fail to have an efficient influence function, even though they are pathwise differentiable. To handle these cases, we propose a regularized one-step estimator and associated confidence sets. We also show that pathwise differentiability, which is a central requirement of our approach, holds in many cases. Specifically, we provide multiple examples of pathwise differentiable parameters and develop corresponding estimators and confidence sets. Among these examples, four are particularly relevant to ongoing research by the causal inference community: the counterfactual density function, dose-response function, conditional average treatment effect function, and counterfactual kernel mean embedding.

REFERENCES

AGARWAL, R., CHEN, Z. and SARMA, S. V. (2015). Nonparametric estimation of band-limited probability density functions. arXiv preprint. Available at arXiv:1503.06236.

BAIARDI, A. and NAGHI, A. A. (2024). The value added of machine learning to causal inference: Evidence from revisited studies. *Econom. J.* **27** 213–234. MR4761726 https://doi.org/10.1093/ectj/utae004

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

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

BONGIOANNI, B. and TORREA, J. L. (2006). Sobolev spaces associated to the harmonic oscillator. *Proc. Indian Acad. Sci. Math. Sci.* **116** 337–360. MR2256010 https://doi.org/10.1007/BF02829750

BONGIOANNI, B. and TORREA, J. L. (2009). What is a Sobolev space for the Laguerre function systems? *Studia Math.* **192** 147–172. MR2496241 https://doi.org/10.4064/sm192-2-4

CHEN, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Handb. Econom.* **6** 5549–5632.

CHERNOZHUKOV, V., CHETVERIKOV, D., DEMIRER, M., DUFLO, E., HANSEN, C. and NEWEY, W. (2017). Double/debiased/Neyman machine learning of treatment effects. *Amer. Econ. Rev.* **107** 261–265.

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

CHERNOZHUKOV, V., NEWEY, W. K. and SINGH, R. (2022). Debiased machine learning of global and local parameters using regularized Riesz representers. *Econom. J.* **25** 576–601. MR4565439 https://doi.org/10.1093/ectj/utac002

MSC2020 subject classifications. Primary 62G20, 62G15; secondary 62G05.
Key words and phrases. One-step estimation, Hilbert-valued parameter, confidence set, nonparametric.

CHERNOZHUKOV, V., NEWEY, W. K. and SINGH, R. (2023). A simple and general debiased machine learning theorem with finite-sample guarantees. *Biometrika* **110** 257–264. MR4565455 https://doi.org/10.1093/biomet/asac033

COLANGELO, K. and LEE, Y.-Y. (2020). Double debiased machine learning nonparametric inference with continuous treatments. arXiv preprint. Available at arXiv:2004.03036.

DA PRATO, G. (2006). *An Introduction to Infinite-Dimensional Analysis*. *Universitext*. Springer, Berlin. Revised and extended from the 2001 original by Da Prato. MR2244975 https://doi.org/10.1007/3-540-29021-4

DÍAZ, I. and VAN DER LAAN, M. J. (2013). Targeted data adaptive estimation of the causal dose-response curve. *J. Causal Inference* **1** 171–192. MR4289406 https://doi.org/10.1515/jci-2012-0005

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26. MR0515681

FAWKES, J., HU, R., EVANS, R. J. and SEJDINOVIC, D. (2022). Doubly robust kernel statistics for testing distributional treatment effects even under one sided overlap. arXiv preprint. Available at arXiv:2212.04922.

FOSTER, D. J. and SYRGKANIS, V. (2023). Orthogonal statistical learning. *Ann. Statist.* **51** 879–908. MR4630373 https://doi.org/10.1214/23-AOS2258

FRIEDMAN, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Statist.* **29** 1189–1232. MR1873328 https://doi.org/10.1214/aos/1013203451

GARREAU, D., JITKRITTUM, W. and KANAGAWA, M. (2017). Large sample analysis of the median heuristic. arXiv preprint. Available at arXiv:1707.07269.

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

GRENANDER, U. (1963). *Probabilities on Algebraic Structures*. Wiley, New York. MR0206994

GRETTON, A., BORGWARDT, K. M., RASCH, M. J., SCHÖLKOPF, B. and SMOLA, A. (2012). A kernel two-sample test. *J. Mach. Learn. Res.* **13** 723–773. MR2913716

HINES, O., DUKES, O., DIAZ-ORDAZ, K. and VANSTEELANDT, S. (2022). Demystifying statistical learning based on efficient influence functions. *Amer. Statist.* **76** 292–304. MR4453533 https://doi.org/10.1080/00031305.2021.2021984

HO, T. K. (1995). Random decision forests. In *Proceedings of* 3rd *International Conference on Document Analysis and Recognition* 1 278–282. IEEE Press, New York.

HUDSON, A., CARONE, M. and SHOJAIE, A. (2021). Inference on function-valued parameters using a restricted score test. arXiv preprint. Available at arXiv:2105.06646.

IBRAGIMOV, I. and KHAS'MINSKII, R. (1983). Estimation of distribution density belonging to a class of entire functions. *Theory Probab. Appl.* **27** 551–562.

JUNG, Y., TIAN, J. and BAREINBOIM, E. (2021). Double machine learning density estimation for local treatment effects with instruments. *Adv. Neural Inf. Process. Syst.* **34** 21821–21833.

KENNEDY, E. H. (2022). Semiparametric doubly robust targeted double machine learning: a review. arXiv preprint. Available at arXiv:2203.06469.

KENNEDY, E. H. (2023). Towards optimal doubly robust estimation of heterogeneous causal effects. *Electron. J. Stat.* **17** 3008–3049. MR4667730 https://doi.org/10.1214/23-ejs2157

KENNEDY, E. H., BALAKRISHNAN, S., ROBINS, J. M. and WASSERMAN, L. (2024). Minimax rates for heterogeneous causal effect estimation. *Ann. Statist.* **52** 793–816. MR4744196 https://doi.org/10.1214/24-aos2369

KENNEDY, E. H., BALAKRISHNAN, S. and WASSERMAN, L. A. (2023). Semiparametric counterfactual density estimation. *Biometrika* **110** 875–896. MR4667430 https://doi.org/10.1093/biomet/asad017

KENNEDY, E. H., MA, Z., MCHUGH, M. D. and SMALL, D. S. (2017). Non-parametric methods for doubly robust estimation of continuous treatment effects. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1229–1245. MR3689316 https://doi.org/10.1111/rssb.12212

KLAASSEN, C. A. J. (1987). Consistent estimation of the influence function of locally asymptotically linear estimators. *Ann. Statist.* **15** 1548–1562. MR0913573 https://doi.org/10.1214/aos/1176350609

LUEDTKE, A. (2023). HilbertOneStep R Package. Available at www.github.com/alexluedtke12/HilbertOneStep.

LUEDTKE, A., CARONE, M. and VAN DER LAAN, M. J. (2019). An omnibus non-parametric test of equality in distribution for unknown functions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 75–99. MR3904780 https://doi.org/10.1111/rssb.12299

LUEDTKE, A. and CHUNG, I. (2024). Supplement to "One-step estimation of differentiable Hilbert-valued parameters." https://doi.org/10.1214/24-AOS2403SUPP

LUEDTKE, A. and WU, J. (2020). Efficient principally stratified treatment effect estimation in crossover studies with absorbent binary endpoints. *J. SFdS* **161** 176–200. MR4125254

LUEDTKE, A. R. and VAN DER LAAN, M. J. (2016). Super-learning of an optimal dynamic treatment rule. *Int. J. Biostat.* **12** 305–332. MR3505699 https://doi.org/10.1515/ijb-2015-0052

MARCELLÁN, F. and XU, Y. (2015). On Sobolev orthogonal polynomials. *Expo. Math.* **33** 308–352. MR3360352 https://doi.org/10.1016/j.exmath.2014.10.002

MUANDET, K., KANAGAWA, M., SAENGKYONGAM, S. and MARUKATAT, S. (2021). Counterfactual mean embeddings. *J. Mach. Learn. Res.* **22** Paper No. 162, 71. MR4318518 https://doi.org/10.22405/2226-8383-2021-22-1-152-162

NEWEY, W. K. and MCFADDEN, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics, Vol. IV. Handbooks in Econom.* **2** 2111–2245. North-Holland, Amsterdam. MR1315971

NIE, X. and WAGER, S. (2021). Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika* **108** 299–319. MR4259133 https://doi.org/10.1093/biomet/asaa076

PFANZAGL, J. (1982). *Contributions to a General Asymptotic Statistical Theory. Lecture Notes in Statistics* **13**. Springer, New York-Berlin. With the assistance of W. Wefelmeyer. MR0675954

PFANZAGL, J. (1990). *Estimation in Semiparametric Models: Some Recent Developments. Lecture Notes in Statistics* **63**. Springer, New York. MR1048589 https://doi.org/10.1007/978-1-4612-3396-1

ROBINS, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period application to control of the healthy worker survivor effect. *Math. Model.* **7**. MR0877758 https://doi.org/10.1016/0270-0255(86)90088-6

ROBINS, J., LI, L., TCHETGEN, E. and VAN DER VAART, A. (2008). Higher order influence functions and minimax estimation of nonlinear functionals. In *Probability and Statistics: Essays in Honor of David A. Freedman. Inst. Math. Stat. (IMS) Collect.* **2** 335–421. IMS, Beachwood, OH. MR2459958 https://doi.org/10.1214/193940307000000527

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

RUDIN, W. (1987). *Real and Complex Analysis*, 3rd ed. McGraw-Hill, New York. MR0924157

SCHICK, A. (1986). On asymptotically efficient estimation in semiparametric models. *Ann. Statist.* **14** 1139–1151. MR0856811 https://doi.org/10.1214/aos/1176350055

SCHWARTZ, S. C. (1967). Estimation of probability density by an orthogonal series. *Ann. Math. Stat.* **38** 1261–1265. MR0221638 https://doi.org/10.1214/aoms/1177698795

SZÉKELY, G. J. and BAKIROV, N. K. (2003). Extremal probabilities for Gaussian quadratic forms. *Probab. Theory Related Fields* **126** 184–202. MR1990053 https://doi.org/10.1007/s00440-003-0262-6

TAKATSU, K. and WESTLING, T. (2022). Debiased inference for a covariate-adjusted regression function. arXiv preprint. Available at arXiv:2210.06448.

TIKHONOV, A. N., GONCHARSKY, A. V., STEPANOV, V. V. and YAGOLA, A. G. (1995). *Numerical Methods for the Solution of Ill-Posed Problems. Mathematics and Its Applications* **328**. Kluwer Academic, Dordrecht. Translated from the 1990 Russian original by R. A. M. Hoksbergen and revised by the authors. MR1350538 https://doi.org/10.1007/978-94-015-8480-7

TSIATIS, A. A. (2006). *Semiparametric Theory and Missing Data. Springer Series in Statistics*. Springer, New York. MR2233926

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

VON MISES, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Stat.* **18** 309–348. MR0022330 https://doi.org/10.1214/aoms/1177730385

VAN DER LAAN, M. J. (2006). Statistical inference for variable importance. *Int. J. Biostat.* **2** Art. 2, 33. MR2275897 https://doi.org/10.2202/1557-4679.1008

VAN DER LAAN, M. J., BIBAUT, A. and LUEDTKE, A. R. (2018). CV-TMLE for nonpathwise differentiable target parameters. In *Targeted Learning in Data Science. Springer Ser. Statist.* 455–481. Springer, Cham. MR3820740

VAN DER LAAN, M. J. and DUDOIT, S. (2003). Unified cross-validation methodology for selection among estimators and a general cross-validated adaptive epsilon-net estimator: Finite sample oracle inequalities and examples.

VAN DER LAAN, M. J., LAAN, M. and ROBINS, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality. Springer Series in Statistics*. Springer, New York. MR1958123 https://doi.org/10.1007/978-0-387-21700-0

VAN DER LAAN, M. J., ROSE, S. et al. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data. Springer Series in Statistics*. Springer, New York. MR2867111 https://doi.org/10.1007/978-1-4419-9782-1

VAN DER LAAN, M. J. and RUBIN, D. (2006). Targeted maximum likelihood learning. *Int. J. Biostat.* **2** Art. 11, 40. MR2306500 https://doi.org/10.2202/1557-4679.1043

VAN DER VAART, A. (1991). On differentiable functionals. *Ann. Statist.* **19** 178–204. MR1091845 https://doi.org/10.1214/aos/1176347976

VAN DER VAART, A. (2014). Higher order tangent spaces and influence functions. *Statist. Sci.* **29** 679–686. MR3300365 https://doi.org/10.1214/14-STS478

VAN DER VAART, A. and WELLNER, J. A. (1989). Prohorov and continuous mapping theorems in the Hoffmann–Jørgensen weak convergence theory, with application to convolution and asymptotic minimax theorems. *Tech. Rep.* **157**.

VAN DER VAART, A. W., DUDOIT, S. and VAN DER LAAN, M. J. (2006). Oracle inequalities for multi-fold cross validation. *Statist. Decisions* **24** 351–371. MR2305112 https://doi.org/10.1524/stnd.2006.24.3.351

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

WILLIAMSON, B. D., GILBERT, P. B., SIMON, N. R. and CARONE, M. (2023). A general framework for inference on algorithm-agnostic variable importance. *J. Amer. Statist. Assoc.* **118** 1645–1658. MR4646595 https://doi.org/10.1080/01621459.2021.2003200

WRIGHT, M. N. and ZIEGLER, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Softw.* **77** 1–17. https://doi.org/10.18637/jss.v077.i01

YAO, K. (1967). Applications of reproducing kernel Hilbert spaces–bandlimited signal models. *Inf. Control* **11** 429–444.

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

The Annals of Statistics
2024, Vol. 52, No. 4, 1564–1591
https://doi.org/10.1214/24-AOS2404

SHARP MULTIPLE TESTING BOUNDARY FOR SPARSE SEQUENCES

BY KWEKU ABRAHAM[1,a], ISMAËL CASTILLO[2,b] AND ÉTIENNE ROQUAIN[2,c]

[1]*Department of Pure Mathematics and Mathematical Statistics, University of Cambridge,* [a]*lkwabraham@statslab.cam.ac.uk*
[2]*Laboratoire de Probabilités, Statistique et Modélisation (LPSM), Sorbonne Université,* [b]*ismael.castillo@upmc.fr,* [c]*etienne.roquain@upmc.fr*

This work investigates multiple testing by considering minimax separation rates in the sparse sequence model, when the testing risk is measured as the sum FDR+FNR (False Discovery Rate plus False Negative Rate). First, using the popular beta-min separation condition, with all nonzero signals separated from 0 by at least some amount, we determine the sharp minimax testing risk asymptotically and thereby explicitly describe the transition from "achievable multiple testing with vanishing risk" to "impossible multiple testing." Adaptive multiple testing procedures achieving the corresponding optimal boundary are provided: the Benjamini–Hochberg procedure with a properly tuned level, and an empirical Bayes ℓ-value ('local FDR') procedure. We prove that the FDR and FNR make nonsymmetric contributions to the testing risk for most optimal procedures, the FNR part being dominant at the boundary. The multiple testing hardness is then investigated for classes of *arbitrary* sparse signals. A number of extensions, including results for classification losses and convergence rates in the case of large signals, are also investigated.

REFERENCES

[1] ABRAHAM, K., CASTILLO, I. and ROQUAIN, É. (2022). Empirical Bayes cumulative ℓ-value multiple testing procedure for sparse sequences. *Electron. J. Stat.* **16** 2033–2081. MR4415394 https://doi.org/10.1214/22-ejs1979

[2] ABRAHAM, K., CASTILLO, I. and ROQUAIN, E. (2024). Supplement to "Sharp multiple testing boundary for sparse sequences." https://doi.org/10.1214/24-AOS2404SUPP

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

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

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

[6] BANERJEE, S., CASTILLO, I. and GHOSAL, S. (2021). Bayesian inference in high-dimensional models. Book chapter to appear in Springer volume on data science. Preprint. Available at arXiv:2101.04491.

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

[8] BELITSER, E. and NURUSHEV, N. (2022). Uncertainty quantification for robust variable selection and multiple testing. *Electron. J. Stat.* **16** 5955–5979. MR4515719 https://doi.org/10.1214/22-ejs2088

[9] BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. MR1325392

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

[11] BOGDAN, M., CHAKRABARTI, A., FROMMLET, F. and GHOSH, J. K. (2011). Asymptotic Bayes-optimality under sparsity of some multiple testing procedures. *Ann. Statist.* **39** 1551–1579. MR2850212 https://doi.org/10.1214/10-AOS869

MSC2020 subject classifications. 62G20, 62G15.
Key words and phrases. Multiple testing, sharp asymptotic minimaxity, False Discovery Rate, Benjamini–Hochberg procedure, frequentist analysis of Bayesian procedures.

[12] BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data*: *Methods*, *Theory and Applications*. *Springer Series in Statistics*. Springer, Heidelberg. MR2807761 https://doi.org/10.1007/978-3-642-20192-9
[13] BUTUCEA, C., MAMMEN, E., NDAOUD, M. and TSYBAKOV, A. B. (2023). Variable selection, monotone likelihood ratio and group sparsity. *Ann*. *Statist*. **51** 312–333. MR4564858 https://doi.org/10.1214/22-aos2251
[14] BUTUCEA, C., NDAOUD, M., STEPANOVA, N. A. and TSYBAKOV, A. B. (2018). Variable selection with Hamming loss. *Ann*. *Statist*. **46** 1837–1875. MR3845003 https://doi.org/10.1214/17-AOS1572
[15] BUTUCEA, C. and STEPANOVA, N. (2017). Adaptive variable selection in nonparametric sparse additive models. *Electron*. *J*. *Stat*. **11** 2321–2357. MR3656494 https://doi.org/10.1214/17-EJS1275
[16] CAI, T. T. and SUN, W. (2009). Simultaneous testing of grouped hypotheses: Finding needles in multiple haystacks. *J*. *Amer*. *Statist*. *Assoc*. **104** 1467–1481. MR2597000 https://doi.org/10.1198/jasa.2009.tm08415
[17] CAI, T. T., SUN, W. and WANG, W. (2019). Covariate-assisted ranking and screening for large-scale two-sample inference. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **81** 187–234. Includes discussions and a rejoinder. MR3928141
[18] CASTILLO, I. and MISMER, R. (2018). Empirical Bayes analysis of spike and slab posterior distributions. *Electron*. *J*. *Stat*. **12** 3953–4001. MR3885271 https://doi.org/10.1214/18-EJS1494
[19] CASTILLO, I. and ROQUAIN, É. (2020). On spike and slab empirical Bayes multiple testing. *Ann*. *Statist*. **48** 2548–2574. MR4152112 https://doi.org/10.1214/19-AOS1897
[20] CASTILLO, I. and SZABÓ, B. (2020). Spike and slab empirical Bayes sparse credible sets. *Bernoulli* **26** 127–158. MR4036030 https://doi.org/10.3150/19-BEJ1119
[21] DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann*. *Statist*. **32** 962–994. MR2065195 https://doi.org/10.1214/009053604000000265
[22] DURAND, G. (2019). Adaptive p-value weighting with power optimality. *Electron*. *J*. *Stat*. **13** 3336–3385. MR4010982 https://doi.org/10.1214/19-ejs1578
[23] EFRON, B., TIBSHIRANI, R., STOREY, J. D. and TUSHER, V. (2001). Empirical Bayes analysis of a microarray experiment. *J*. *Amer*. *Statist*. *Assoc*. **96** 1151–1160. MR1946571 https://doi.org/10.1198/016214501753382129
[24] FROMONT, M., LERASLE, M. and REYNAUD-BOURET, P. (2016). Family-wise separation rates for multiple testing. *Ann*. *Statist*. **44** 2533–2563. MR3576553 https://doi.org/10.1214/15-AOS1418
[25] GENOVESE, C. and WASSERMAN, L. (2002). Operating characteristics and extensions of the false discovery rate procedure. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **64** 499–517. MR1924303 https://doi.org/10.1111/1467-9868.00347
[26] GINÉ, E. and NICKL, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge Univ. Press, New York. MR3588285 https://doi.org/10.1017/CBO9781107337862
[27] HALL, P. and JIN, J. (2008). Properties of higher criticism under strong dependence. *Ann*. *Statist*. **36** 381–402. MR2387976 https://doi.org/10.1214/009053607000000767
[28] HALL, P. and JIN, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann*. *Statist*. **38** 1686–1732. MR2662357 https://doi.org/10.1214/09-AOS764
[29] HELLER, R. and ROSSET, S. (2021). Optimal control of false discovery criteria in the two-group model. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **83** 133–155. MR4220987 https://doi.org/10.1111/rssb.12403
[30] IGNATIADIS, N. and HUBER, W. (2021). Covariate powered cross-weighted multiple testing. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **83** 720–751. MR4319999 https://doi.org/10.1111/rssb.12411
[31] INGSTER, Y. I., TSYBAKOV, A. B. and VERZELEN, N. (2010). Detection boundary in sparse regression. *Electron*. *J*. *Stat*. **4** 1476–1526. MR2747131 https://doi.org/10.1214/10-EJS589
[32] 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
[33] JIN, J. and KE, Z. T. (2016). Rare and weak effects in large-scale inference: Methods and phase diagrams. *Statist*. *Sinica* **26** 1–34. MR3468343
[34] JOHNSTONE, I. M. and SILVERMAN, B. W. (2004). Needles and straw in haystacks: Empirical Bayes estimates of possibly sparse sequences. *Ann*. *Statist*. **32** 1594–1649. MR2089135 https://doi.org/10.1214/009053604000000030
[35] JOHNSTONE, I. M. and SILVERMAN, B. W. (2005). EbayesThresh: R programs for empirical Bayes thresholding. *J*. *Stat*. *Softw*. **12**.
[36] LI, X. and FITHIAN, W. (2020). Optimality of the max test for detecting sparse signals with Gaussian or heavier tail. Preprint. Available at arXiv:2006.12489.

[37] MILLER, J. C. and STEPANOVA, N. A. (2023). Adaptive signal recovery with Subbotin noise. *Statist. Probab. Lett.* **196** Paper No. 109791, 6. MR4549624 https://doi.org/10.1016/j.spl.2023.109791

[38] NEUVIAL, P. and ROQUAIN, E. (2012). On false discovery rate thresholding for classification under sparsity. *Ann. Statist.* **40** 2572–2600. MR3097613 https://doi.org/10.1214/12-AOS1042

[39] NICKL, R. and VAN DE GEER, S. (2013). Confidence sets in sparse regression. *Ann. Statist.* **41** 2852–2876. MR3161450 https://doi.org/10.1214/13-AOS1170

[40] RABINOVICH, M., JORDAN, M. I. and WAINWRIGHT, M. J. (2020). Lower bounds in multiple testing: A framework based on derandomized proxies. Arxiv eprint. Available at arXiv:2005.03725.

[41] RABINOVICH, M., RAMDAS, A., JORDAN, M. I. and WAINWRIGHT, M. J. (2020). Optimal rates and trade-offs in multiple testing. *Statist. Sinica* **30** 741–762. MR4214160

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

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

[44] SUN, W. and CAI, T. T. (2007). Oracle and adaptive compound decision rules for false discovery rate control. *J. Amer. Statist. Assoc.* **102** 901–912. MR2411657 https://doi.org/10.1198/016214507000000545

[45] SUN, W. and CAI, T. T. (2009). Large-scale multiple testing under dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 393–424. MR2649603 https://doi.org/10.1111/j.1467-9868.2008.00694.x

The Annals of Statistics
2024, Vol. 52, No. 4, 1592–1615
https://doi.org/10.1214/24-AOS2405

A NONPARAMETRIC DOUBLY ROBUST TEST FOR A CONTINUOUS TREATMENT EFFECT

BY CHARLES R. DOSS[1,a], GUANGWEI WENG[1,b], LAN WANG[2,c], IRA MOSCOVICE[3,d] AND TONGTAN CHANTARAT[3,e]

[1]*School of Statistics, University of Minnesota,* [a]*cdoss@stat.umn.edu,* [b]*wengx076@umn.edu*
[2]*Miami Herbert Business School, University of Miami,* [c]*lxw611@miami.edu*
[3]*School of Public Health, University of Minnesota,* [d]*mosco001@umn.edu,* [e]*chant083@umn.edu*

The vast majority of literature on evaluating the significance of a treatment effect based on observational data has been confined to discrete treatments. These methods are not applicable to drawing inference for a continuous treatment, which arises in many important applications. To adjust for confounders when evaluating a continuous treatment, existing inference methods often rely on discretizing the treatment or using (possibly misspecified) parametric models for the effect curve. Recently, Kennedy et al. (*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** (2017) 1229–1245) proposed nonparametric doubly robust estimation for a continuous treatment effect in observational studies. However, inference for the continuous treatment effect is a harder problem. To the best of our knowledge, a completely nonparametric doubly robust approach for inference in this setting is not yet available. We develop such a nonparametric doubly robust procedure in this paper for making inference on the continuous treatment effect curve. Using empirical process techniques for local U- and V-processes, we establish the test statistic's asymptotic distribution. Furthermore, we propose a wild bootstrap procedure for implementing the test in practice. In addition, we define a version of the test procedure based on sample splitting. We illustrate the new method(s) via simulations and a study of a constructed dataset relating the effect of nurse staffing hours on hospital performance. We implement our doubly robust dose response test in the R package DRDRtest on CRAN.

REFERENCES

ALCALÁ, J. T., CRISTÓBAL, J. A. and GONZÁLEZ-MANTEIGA, W. (1999). Goodness-of-fit test for linear models based on local polynomials. *Statist. Probab. Lett.* **42** 39–46. MR1671753 https://doi.org/10.1016/S0167-7152(98)00184-9

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

BENKESER, D., CARONE, M., VAN DER LAAN, M. J. and GILBERT, P. B. (2017). Doubly robust nonparametric inference on the average treatment effect. *Biometrika* **104** 863–880. MR3737309 https://doi.org/10.1093/biomet/asx053

CHEN, G., LI, X. and YU, M. (2022). Policy learning for optimal individualized dose intervals. In *Proceedings of the* 25*th International Conference on Artificial Intelligence and Statistics* (G. Camps-Valls, F. J. R. Ruiz and I. Valera, eds.). *Proceedings of Machine Learning Research* **151** 1671–1693. PMLR.

CHEN, G., ZENG, D. and KOSOROK, M. R. (2016). Personalized dose finding using outcome weighted learning. *J. Amer. Statist. Assoc.* **111** 1509–1521. MR3601705 https://doi.org/10.1080/01621459.2016.1148611

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

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

MSC2020 subject classifications. Primary 62D20; secondary 62G10.
Key words and phrases. Causal inference, double robustness, nonparametric, hypothesis test, U-statistics.

CHERNOZHUKOV, V., NEWEY, W. K. and SINGH, R. (2022). Debiased machine learning of global and local parameters using regularized Riesz representers. *Econom. J.* **25** 576–601. MR4565439 https://doi.org/10.1093/ectj/utac002

COLANGELO, K. and LEE, Y.-Y. (2020). Double debiased machine learning nonparametric inference with continuous treatments. arXiv.org.

COULOMBE, J., MOODIE, E. E. M. and PLATT, R. W. (2021). Estimating the marginal effect of a continuous exposure on an ordinal outcome using data subject to covariate-driven treatment and visit processes. *Stat. Med.* **40** 5746–5764. MR4330577 https://doi.org/10.1002/sim.9151

DAVIDSON, R. and FLACHAIRE, E. (2008). The wild bootstrap, tamed at last. *J. Econometrics* **146** 162–169. MR2459651 https://doi.org/10.1016/j.jeconom.2008.08.003

DETTE, H. and NEUMEYER, N. (2001). Nonparametric analysis of covariance. *Ann. Statist.* **29** 1361–1400. MR1873335 https://doi.org/10.1214/aos/1013203458

DOSS, C. R., WENG, G., WANG, L., MOSCOVICE, I. and CHANTARAT, T. (2024). Supplement to "A nonparametric doubly robust test for a continuous treatment effect." https://doi.org/10.1214/24-AOS2405SUPP

DUDLEY, M. (1999). *Uniform Central Limit Theorems*. *Cambridge Studies in Advanced Mathematics* **63**. Cambridge University Press, Cambridge. MR1720712 https://doi.org/10.1017/CBO9780511665622

EFRON, B. and TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap*. *Monographs on Statistics and Applied Probability* **57**. CRC Press, New York. MR1270903 https://doi.org/10.1007/978-1-4899-4541-9

EUBANK, R. L. and LARICCIA, V. N. (1992). Asymptotic comparison of Cramér–von Mises and nonparametric function estimation techniques for testing goodness-of-fit. *Ann. Statist.* **20** 2071–2086. MR1193326 https://doi.org/10.1214/aos/1176348903

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. *Monographs on Statistics and Applied Probability* **66**. CRC Press, London. MR1383587

FOSTER, D. J. and SYRGKANIS, V. (2023). Orthogonal statistical learning. *Ann. Statist.* **51** 879–908. MR4630373 https://doi.org/10.1214/23-AOS2258

GALVAO, A. F. and WANG, L. (2015). Uniformly semiparametric efficient estimation of treatment effects with a continuous treatment. *J. Amer. Statist. Assoc.* **110** 1528–1542. MR3449052 https://doi.org/10.1080/01621459.2014.978005

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

GRUBER, S. and VAN DER LAAN, M. (2012). tmle: An R package for targeted maximum likelihood estimation. *J. Stat. Softw.* **51** 1–35. https://doi.org/10.18637/jss.v051.i13

HÄRDLE, W. and MAMMEN, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21** 1926–1947. MR1245774 https://doi.org/10.1214/aos/1176349403

HASTIE, T. and TIBSHIRANI, R. (1987). Generalized additive models: some applications. *J. Amer. Statist. Assoc.* **82** 371–386.

HASTIE, T. J. (2017). Generalized additive models. In *Statistical Models in S* 249–307. Routledge, London.

HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models*. *Monographs on Statistics and Applied Probability* **43**. CRC Press, London. MR1082147

HILL, J. L. (2011). Bayesian nonparametric modeling for causal inference. *J. Comput. Graph. Statist.* **20** 217–240. MR2816546 https://doi.org/10.1198/jcgs.2010.08162

HIRANO, K. and IMBENS, G. W. (2004). The propensity score with continuous treatments. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*. *Wiley Ser. Probab. Stat.* 73–84. Wiley, Chichester. MR2134803 https://doi.org/10.1002/0470090456.ch7

HOROWITZ, J. L. and SPOKOINY, V. G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* **69** 599–631. MR1828537 https://doi.org/10.1111/1468-0262.00207

IMAI, K. and VAN DYK, D. A. (2004). Causal inference with general treatment regimes: Generalizing the propensity score. *J. Amer. Statist. Assoc.* **99** 854–866. MR2090918 https://doi.org/10.1198/016214504000001187

IMBENS, G. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Rev. Econ. Stat.* **86** 4–29.

INGSTER, Y. I. (1993a). Asymptotically minimax hypothesis testing for nonparametric alternatives. I. *Math. Methods Statist.* **2** 85–114. MR1257978

INGSTER, Y. I. (1993b). Asymptotically minimax hypothesis testing for nonparametric alternatives. II. *Math. Methods Statist.* **2** 171–189. MR1257983

INGSTER, Y. I. (1993c). Asymptotically minimax hypothesis testing for nonparametric alternatives. III. *Math. Methods Statist.* **2** 249–268. MR1259685

KALLUS, N. and ZHOU, A. (2018). Policy evaluation and optimization with continuous treatments. In *International Conference on Artificial Intelligence and Statistics* 1243–1251. PMLR.

KANG, J. D. Y. and SCHAFER, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statist. Sci.* **22** 523–539. MR2420458 https://doi.org/10.1214/07-STS227

KENNEDY, E. H. (2023). Towards optimal doubly robust estimation of heterogeneous causal effects. *Electron. J. Stat.* **17** 3008–3049. MR4667730 https://doi.org/10.1214/23-ejs2157

KENNEDY, E. H., BALAKRISHNAN, S. and WASSERMAN, L. (2022). Minimax rates for heterogeneous causal effect estimation. arXiv.

KENNEDY, E. H., MA, Z., MCHUGH, M. D. and SMALL, D. S. (2017). Non-parametric methods for doubly robust estimation of continuous treatment effects. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1229–1245. MR3689316 https://doi.org/10.1111/rssb.12212

KREIF, N., GRIEVE, R., DÍAZ, I. and HARRISON, D. (2015). Evaluation of the effect of a continuous treatment: A machine learning approach with an application to treatment for traumatic brain injury. *Health Econ.* **24** 1213–1228. https://doi.org/10.1002/hec.3189

LEE, S., OKUI, R. and WHANG, Y.-J. (2017). Doubly robust uniform confidence band for the conditional average treatment effect function. *J. Appl. Econometrics* **32** 1207–1225. MR3734484 https://doi.org/10.1002/jae.2574

LUEDTKE, A., CARONE, M. and VAN DER LAAN, M. J. (2019). An omnibus non-parametric test of equality in distribution for unknown functions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 75–99. MR3904780 https://doi.org/10.1111/rssb.12299

LUEDTKE, A. R. and VAN DER LAAN, M. J. (2016). Super-learning of an optimal dynamic treatment rule. *Int. J. Biostat.* **12** 305–332. MR3505699 https://doi.org/10.1515/ijb-2015-0052

MCHUGH, M. D., BEREZ, J. and SMALL, D. S. (2013). Hospitals with higher nurse staffing had lower odds of readmissions penalties than hospitals with lower staffing. *Health Aff.* **32** 1740–1747. https://doi.org/10.1377/hlthaff.2013.0613

NEUGEBAUER, R. and VAN DER LAAN, M. (2007). Nonparametric causal effects based on marginal structural models. *J. Statist. Plann. Inference* **137** 419–434. MR2298947 https://doi.org/10.1016/j.jspi.2005.12.008

NEWEY, W. K. and ROBINS, J. R. (2018). Cross-fitting and fast remainder rates for semiparametric estimation. arXiv.

NIE, X. and WAGER, S. (2021). Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika* **108** 299–319. MR4259133 https://doi.org/10.1093/biomet/asaa076

POUET, C. (2001). An asymptotically optimal test for a parametric set of regression functions against a non-parametric alternative. *J. Statist. Plann. Inference* **98** 177–189. MR1860233 https://doi.org/10.1016/S0378-3758(00)00300-1

ROBINS, J., SUED, M., LEI-GOMEZ, Q. and ROTNITZKY, A. (2007). Comment: Performance of double-robust estimators when "inverse probability" weights are highly variable [MR2420458]. *Statist. Sci.* **22** 544–559. MR2420460 https://doi.org/10.1214/07-STS227D

ROBINS, J. M. (2000). Marginal structural models versus structural nested models as tools for causal inference. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials* (*Minneapolis, MN*, 1997). *IMA Vol. Math. Appl.* **116** 95–133. Springer, New York. MR1731682 https://doi.org/10.1007/978-1-4612-1284-3_2

ROBINSON, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica* **56** 931–954. MR0951762 https://doi.org/10.2307/1912705

RUBIN, D. B. (1975). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66** 688–701.

SCHULZ, J. and MOODIE, E. E. M. (2021). Doubly robust estimation of optimal dosing strategies. *J. Amer. Statist. Assoc.* **116** 256–268. MR4227692 https://doi.org/10.1080/01621459.2020.1753521

SEMENOVA, V. and CHERNOZHUKOV, V. (2021). Debiased machine learning of conditional average treatment effects and other causal functions. *Econom. J.* **24** 264–289. MR4281225 https://doi.org/10.1093/ectj/utaa027

SHORACK, G. R. (2000). *Probability for Statisticians. Springer Texts in Statistics*. Springer, New York. MR1762415

SU, L., URA, T. and ZHANG, Y. (2019). Non-separable models with high-dimensional data. *J. Econometrics* **212** 646–677. MR4001678 https://doi.org/10.1016/j.jeconom.2019.06.004

VAN DER LAAN, M. J. (2014). Targeted estimation of nuisance parameters to obtain valid statistical inference. *Int. J. Biostat.* **10** 29–57. MR3208072 https://doi.org/10.1515/ijb-2012-0038

VAN DER LAAN, M. J. and DUDOIT, S. (2003). Unified cross-validation methodology for selection among estimators and a general cross-validated adaptive epsilon-net estimator: Finite sample oracle inequalities and examples. Technical Report 130, Division of Biostatistics, Univ. California Berkeley.

VAN DER LAAN, M. J., POLLEY, E. C. and HUBBARD, A. E. (2007). Super learner. *Stat. Appl. Genet. Mol. Biol.* **6** Art. 25, 23. MR2349918 https://doi.org/10.2202/1544-6115.1309

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

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

WANG, J., WONG, R. K. W., YANG, S. and CHAN, K. C. G. (2022). Estimation of partially conditional average treatment effect by double kernel-covariate balancing. *Electron*. *J*. *Stat*. **16** 4332–4378. MR4474576 https://doi.org/10.1214/22-ejs2000

WESTLING, T. (2022). Nonparametric tests of the causal null with nondiscrete exposures. *J*. *Amer*. *Statist*. *Assoc*. **117** 1551–1562. MR4480731 https://doi.org/10.1080/01621459.2020.1865168

WESTLING, T., GILBERT, P. and CARONE, M. (2020). Causal isotonic regression. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **82** 719–747. MR4112782 https://doi.org/10.1111/rssb.12372

WOOD, S. N. (2017). *Generalized Additive Models*: *An Introduction with R*: *An Introduction with R*. *Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. Second edition of [MR2206355]. MR3726911

ZIMMERT, M. and LECHNER, M. (2019). Nonparametric estimation of causal heterogeneity under high-dimensional confounding. arXiv preprint. Available at arXiv:1908.08779.

The Annals of Statistics
2024, Vol. 52, No. 4, 1616–1645
https://doi.org/10.1214/24-AOS2406

GROMOV–WASSERSTEIN DISTANCES: ENTROPIC REGULARIZATION, DUALITY AND SAMPLE COMPLEXITY

BY ZHENGXIN ZHANG[1,a], ZIV GOLDFELD[2,b], YOUSSEF MROUEH[3,c] AND BHARATH K. SRIPERUMBUDUR[4,d]

[1]*Center for Applied Mathematics, Cornell University,* [a]*zz658@cornell.edu*
[2]*School of Electrical and Computer Engineering, Cornell University,* [b]*goldfeld@cornell.edu*
[3]*IBM Research, IBM,* [c]*mroueh@us.ibm.com*
[4]*Department of Statistics, Pennsylvania State University,* [d]*bks18@psu.edu*

The Gromov–Wasserstein (GW) distance, rooted in optimal transport (OT) theory, quantifies dissimilarity between metric measure spaces and provides a framework for aligning heterogeneous datasets. While computational aspects of the GW problem have been widely studied, a duality theory and fundamental statistical questions concerning empirical convergence rates remained obscure. This work closes these gaps for the quadratic GW distance over Euclidean spaces of different dimensions d_x and d_y. We treat both the standard and the entropically regularized GW distance, and derive dual forms that represent them in terms of the well-understood OT and entropic OT (EOT) problems, respectively. This enables employing proof techniques from statistical OT based on regularity analysis of dual potentials and empirical process theory, using which we establish the first GW empirical convergence rates. The derived two-sample rates are $n^{-2/\max\{\min\{d_x,d_y\},4\}}$ (up to a log factor when $\min\{d_x,d_y\}=4$) for standard GW and $n^{-1/2}$ for entropic GW (EGW), which matches the corresponding rates for standard and entropic OT. The parametric rate for EGW is evidently optimal, while for standard GW we provide matching lower bounds, which establish sharpness of the derived rates. We also study stability of EGW in the entropic regularization parameter and prove approximation and continuity results for the cost and optimal couplings. Lastly, the duality is leveraged to shed new light on the open problem of the one-dimensional GW distance between uniform distributions on n points, illuminating why the identity and anti-identity permutations may not be optimal. Our results serve as a first step towards a comprehensive statistical theory as well as computational advancements for GW distances, based on the discovered dual formulations.

REFERENCES

[1] ALVAREZ-MELIS, D. and JAAKKOLA, T. (2018). Gromov–Wasserstein alignment of word embedding spaces. In *Proceedings of the* 2018 *Conference on Empirical Methods in Natural Language Processing* (E. Riloff, D. Chiang, J. Hockenmaier and J. Tsujii, eds.) 1881–1890. Association for Computational Linguistics, Brussels, Belgium. https://doi.org/10.18653/v1/D18-1214

[2] BEINERT, R., HEISS, C. and STEIDL, G. (2023). On assignment problems related to Gromov–Wasserstein distances on the real line. *SIAM J. Imaging Sci*. **16** 1028–1032. MR4605212 https://doi.org/10.1137/22M1497808

[3] BENAMOU, J.-D. and BRENIER, Y. (2000). A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numer. Math*. **84** 375–393. MR1738163 https://doi.org/10.1007/s002110050002

[4] BIGOT, J., CAZELLES, E. and PAPADAKIS, N. (2019). Central limit theorems for entropy-regularized optimal transport on finite spaces and statistical applications. *Electron. J. Stat*. **13** 5120–5150. MR4041704 https://doi.org/10.1214/19-EJS1637

MSC2020 subject classifications. Primary 62G99; secondary 62H99.
Key words and phrases. Empirical convergence, entropic regularization, Gromov–Wasserstein distance, sample complexity, strong duality.

[5] BLUMBERG, A. J., CARRIERE, M., MANDELL, M. A., RABADAN, R. and VILLAR, S. (2020). MREC: a fast and versatile framework for aligning and matching point clouds with applications to single cell molecular data. arXiv preprint. Available at arXiv:2001.01666.

[6] BOISSARD, E. and LE GOUIC, T. (2014). On the mean speed of convergence of empirical and occupation measures in Wasserstein distance. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** 539–563. MR3189084 https://doi.org/10.1214/12-AIHP517

[7] BUNNE, C., ALVAREZ-MELIS, D., KRAUSE, A. and JEGELKA, S. (2019). Learning generative models across incomparable spaces. In *Proceedings of the* 36*th International Conference on Machine Learning* (K. Chaudhuri and R. Salakhutdinov, eds.). *Proceedings of Machine Learning Research* **97** 851–861. PMLR, Long Beach, CA, USA.

[8] CARLIER, G., DUVAL, V., PEYRÉ, G. and SCHMITZER, B. (2017). Convergence of entropic schemes for optimal transport and gradient flows. *SIAM J. Math. Anal.* **49** 1385–1418. MR3635459 https://doi.org/10.1137/15M1050264

[9] CHEN, Y., GEORGIOU, T. T. and PAVON, M. (2016). On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint. *J. Optim. Theory Appl.* **169** 671–691. MR3489825 https://doi.org/10.1007/s10957-015-0803-z

[10] CHIZAT, L., ROUSSILLON, P., LÉGER, F., VIALARD, F.-X. and PEYRÉ, G. (2020). Faster Wasserstein distance estimation with the Sinkhorn divergence. In *Advances in Neural Information Processing Systems* (H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan and H. Lin, eds.) **33** 2257–2269. Curran Associates, Red Hook.

[11] COMMANDER, C. W. (2005). A survey of the quadratic assignment problem, with applications. *Morehead Electron. J. of Appl. Math.* **4**. MATH-2005-01.

[12] CONFORTI, G. and TAMANINI, L. (2021). A formula for the time derivative of the entropic cost and applications. *J. Funct. Anal.* **280** Paper No. 108964, 48. MR4232667 https://doi.org/10.1016/j.jfa.2021.108964

[13] DEL BARRIO, E., GONZÁLEZ SANZ, A., LOUBES, J.-M. and NILES-WEED, J. (2023). An improved central limit theorem and fast convergence rates for entropic transportation costs. *SIAM J. Math. Data Sci.* **5** 639–669. MR4616887 https://doi.org/10.1137/22M149260X

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

[15] DELON, J., DESOLNEUX, A. and SALMONA, A. (2022). Gromov–Wasserstein distances between Gaussian distributions. *J. Appl. Probab.* **59** 1178–1198. MR4507687 https://doi.org/10.1017/jpr.2022.16

[16] DEMETCI, P., SANTORELLA, R., SANDSTEDE, B., NOBLE, W. S. and SINGH, R. (2022). SCOT: Single-cell multi-omics alignment with optimal transport. *J. Comput. Biol.* **29** 3–18. MR4370741 https://doi.org/10.1089/cmb.2021.0446

[17] DEREICH, S., SCHEUTZOW, M. and SCHOTTSTEDT, R. (2013). Constructive quantization: Approximation by empirical measures. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 1183–1203. MR3127919 https://doi.org/10.1214/12-AIHP489

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

[19] DUMONT, T., LACOMBE, T. and VIALARD, F.-X. (2024). On the existence of Monge maps for the Gromov–Wasserstein problem. *Found. Comput. Math.* 1–48.

[20] FOURNIER, N. and GUILLIN, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* **162** 707–738. MR3383341 https://doi.org/10.1007/s00440-014-0583-7

[21] GENEVAY, A., CHIZAT, L., BACH, F., CUTURI, M. and PEYRÉ, G. (2019). Sample complexity of Sinkhorn divergences. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics* (K. Chaudhuri and M. Sugiyama, eds.). *Proceedings of Machine Learning Research* **89** 1574–1583. PMLR, Naha, Okinawa, Japan.

[22] GENTIL, I., LÉONARD, C. and RIPANI, L. (2017). About the analogy between optimal transport and minimal entropy. *Ann. Fac. Sci. Toulouse Math.* (6) **26** 569–601. MR3669966 https://doi.org/10.5802/afst.1546

[23] GIGLI, N. and TAMANINI, L. (2020). Benamou–Brenier and duality formulas for the entropic cost on $\mathsf{RCD}^*(K,N)$ spaces. *Probab. Theory Related Fields* **176** 1–34. MR4055185 https://doi.org/10.1007/s00440-019-00909-1

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

[25] GOLDFELD, Z. and GREENEWALD, K. (2020). Gaussian-smoothed optimal transport: Metric structure and statistical efficiency. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics* (S. Chiappa and R. Calandra, eds.). *Proceedings of Machine Learning Research* **108** 3327–3337. PMLR, Online.

[26] GOLDFELD, Z., KATO, K., NIETERT, S. and RIOUX, G. (2024). Limit distribution theory for smooth p-Wasserstein distances. *Ann. Appl. Probab.* **34** 2447–2487. MR4728174 https://doi.org/10.1214/23-aap2028

[27] GOLDFELD, Z., KATO, K., RIOUX, G. and SADHU, R. (2024). Statistical inference with regularized optimal transport. *Inf. Inference* **13** Paper No. 13, 68. MR4701828 https://doi.org/10.1093/imaiai/iaad056

[28] GOLDFELD, Z., KATO, K., RIOUX, G. and SADHU, R. (2024). Limit theorems for entropic optimal transport maps and Sinkhorn divergence. *Electron. J. Stat.* **18** 980–1041. MR4718466 https://doi.org/10.1214/24-ejs2217

[29] GONZÁLEZ-SANZ, A. and HUNDRIESER, S. (2023). Weak limits for empirical entropic optimal transport: Beyond smooth costs. arXiv preprint. Available at arXiv:2305.09745.

[30] GONZÁLEZ-SANZ, A., LOUBES, J.-M. and NILES-WEED, J. (2022). Weak limits of entropy regularized optimal transport; potentials, plans and divergences. arXiv preprint. Available at arXiv:2207.07427.

[31] GOODALL, C. (1991). Procrustes methods in the statistical analysis of shape. *J. Roy. Statist. Soc. Ser. B* **53** 285–339. MR1108330

[32] GRAVE, E., JOULIN, A. and BERTHET, Q. (2019). Unsupervised alignment of embeddings with Wasserstein procrustes. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics* (K. Chaudhuri and M. Sugiyama, eds.). *Proceedings of Machine Learning Research* **89** 1880–1890. PMLR, Naha, Okinawa, Japan.

[33] GROPPE, M. and HUNDRIESER, S. (2023). Lower complexity adaptation for empirical entropic optimal transport. arXiv preprint. Available at arXiv:2306.13580.

[34] GUNSILIUS, F. and XU, Y. (2021). Matching for causal effects via multimarginal unbalanced optimal transport. arXiv preprint, updated on July 9, 2022. Available at arXiv:2112.04398.

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

[36] HUNDRIESER, S., KLATT, M., STAUDT, T. and MUNK, A. (2022). A unifying approach to distributional limits for empirical optimal transport. arXiv preprint. Available at arXiv:2202.12790.

[37] HUNDRIESER, S., STAUDT, T. and MUNK, A. (2024). Empirical optimal transport between different measures adapts to lower complexity. *Ann. Inst. Henri Poincaré Probab. Stat.* **60** 824–846. MR4757509 https://doi.org/10.1214/23-aihp1369

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

[39] KOEHL, P., DELARUE, M. and ORLAND, H. (2023). Computing the Gromov–Wasserstein distance between two surface meshes using optimal transport. *Algorithms* **16** 131.

[40] LE, K., LE, D. Q., NGUYEN, H., DO, D., PHAM, T. and HO, N. (2022). Entropic Gromov–Wasserstein between Gaussian distributions. In *Proceedings of the* 39*th International Conference on Machine Learning* (K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu and S. Sabato, eds.). *Proceedings of Machine Learning Research* **162** 12164–12203. PMLR, Baltimore, MD, USA.

[41] MANOLE, T., BALAKRISHNAN, S., NILES-WEED, J. and WASSERMAN, L. (2021). Plugin estimation of smooth optimal transport maps. arXiv preprint. Available at arXiv:2107.12364.

[42] MANOLE, T. and NILES-WEED, J. (2024). Sharp convergence rates for empirical optimal transport with smooth costs. *Ann. Appl. Probab.* **34** 1108–1135. MR4700254 https://doi.org/10.1214/23-aap1986

[43] MARON, H. and LIPMAN, Y. (2018). (Probably) concave graph matching. In *Advances in Neural Information Processing Systems* (S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi and R. Garnett, eds.). **31**. Curran Associates, Red Hook.

[44] MÉMOLI, F. (2009). Spectral Gromov–Wasserstein distances for shape matching. In 2009 *IEEE* 12*th International Conference on Computer Vision Workshops*. *ICCV Workshops* 256–263. IEEE Press, New York.

[45] MÉMOLI, F. (2011). Gromov–Wasserstein distances and the metric approach to object matching. *Found. Comput. Math.* **11** 417–487. MR2811584 https://doi.org/10.1007/s10208-011-9093-5

[46] MENA, G. and NILES-WEED, J. (2019). Statistical bounds for entropic optimal transport: Sample complexity and the central limit theorem. In *Advances in Neural Information Processing Systems* (H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox and R. Garnett, eds.). **32**. Curran Associates, Red Hook.

[47] NUTZ, M. and WIESEL, J. (2022). Entropic optimal transport: Convergence of potentials. *Probab. Theory Related Fields* **184** 401–424. MR4498514 https://doi.org/10.1007/s00440-021-01096-8

[48] PEYRÉ, G., CUTURI, M. et al. (2019). Computational optimal transport: With applications to data science. Foundations and trends® in machine learning, Vol. 11. 355–607.

[49] PEYRÉ, G., CUTURI, M. and SOLOMON, J. (2016). Gromov–Wasserstein averaging of kernel and distance matrices. 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** 2664–2672. PMLR, New York.

[50] RIGOLLET, P. and STROMME, A. J. (2022). On the sample complexity of entropic optimal transport. arXiv preprint. Available at arXiv:2206.13472.

[51] RIOUX, G., GOLDFELD, Z. and KATO, K. (2023). Entropic Gromov–Wasserstein distances: Stability, algorithms, and distributional limits. arXiv preprint. Available at arXiv:2306.00182.

[52] RÖMISCH, W. (2004). *Delta Method*, *Infinite Dimensional*. *Encyclopedia of Statistical Sciences*. Wiley, New York.

[53] SANTAMBROGIO, F. (2015). *Optimal Transport for Applied Mathematicians*: *Calculus of Variations*, *PDEs*, *and Modeling*. *Progress in Nonlinear Differential Equations and Their Applications* **87**. Birkhäuser, Cham. MR3409718 https://doi.org/10.1007/978-3-319-20828-2

[54] SCETBON, M., PEYRÉ, G. and CUTURI, M. (2022). Linear-time Gromov Wasserstein distances using low rank couplings and costs. In *Proceedings of the* 39*th International Conference on Machine Learning* (K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu and S. Sabato, eds.). *Proceedings of Machine Learning Research* **162** 19347–19365. PMLR, Baltimore, MD, USA.

[55] SCHÖNEMANN, P. H. (1966). A generalized solution of the orthogonal Procrustes problem. *Psychometrika* **31** 1–10. MR0215870 https://doi.org/10.1007/BF02289451

[56] SEJOURNE, T., VIALARD, F.-X. and PEYRÉ, G. (2021). The unbalanced Gromov–Wasserstein distance: Conic formulation and relaxation. In *Advances in Neural Information Processing Systems* (M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang and J. W. Vaughan, eds.) **34** 8766–8779. Curran Associates, Red Hook.

[57] SHAPIRO, A. (1990). On concepts of directional differentiability. *J*. *Optim*. *Theory Appl*. **66** 477–487. MR1080259 https://doi.org/10.1007/BF00940933

[58] SOLOMON, J., PEYRÉ, G., KIM, V. G. and SRA, S. (2016). Entropic metric alignment for correspondence problems. *ACM Trans*. *Graph*. **35** 1–13.

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

[60] STROMME, A. J. (2023). Minimum intrinsic dimension scaling for entropic optimal transport. arXiv preprint. Available at arXiv:2306.03398.

[61] STURM, K.-T. (2023). The space of spaces: Curvature bounds and gradient flows on the space of metric measure spaces. *Mem*. *Amer*. *Math*. *Soc*. **290** v+111. MR4659681 https://doi.org/10.1090/memo/1443

[62] TAMELING, C., SOMMERFELD, M. and MUNK, A. (2019). Empirical optimal transport on countable metric spaces: Distributional limits and statistical applications. *Ann*. *Appl*. *Probab*. **29** 2744–2781. MR4019874 https://doi.org/10.1214/19-AAP1463

[63] TITOUAN, V., FLAMARY, R., COURTY, N., TAVENARD, R. and CHAPEL, L. (2019). Sliced Gromov–Wasserstein. In *Advances in Neural Information Processing Systems* (H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox and R. Garnett, eds.). **32**. Curran Associates, Red Hook.

[64] VAN DER VAART, A. W. and WELLNER, J. A. (2023). *Weak Convergence and Empirical Processes—with Applications to Statistics*. *Springer Series in Statistics*. Springer, Cham. Second edition [of 1385671]. MR4628026 https://doi.org/10.1007/978-3-031-29040-4

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

[66] WEED, J. and BACH, F. (2019). Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli* **25** 2620–2648. MR4003560 https://doi.org/10.3150/18-BEJ1065

[67] XU, H., LUO, D. and CARIN, L. (2019). Scalable Gromov–Wasserstein learning for graph partitioning and matching. In *Advances in Neural Information Processing Systems* (H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox and R. Garnett, eds.). **32**. Curran Associates, Red Hook.

[68] XU, H., LUO, D., ZHA, H. and DUKE, L. C. (2019). Gromov–Wasserstein learning for graph matching and node embedding. In *Proceedings of the* 36*th International Conference on Machine Learning* (K. Chaudhuri and R. Salakhutdinov, eds.). *Proceedings of Machine Learning Research* **97** 6932–6941. PMLR, Long Beach, CA, USA.

[69] YAN, Y., LI, W., WU, H., MIN, H., TAN, M. and WU, Q. (2018). Semi-supervised optimal transport for heterogeneous domain adaptation. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, *IJCAI*-18 2969–2975. International Joint Conferences on Artificial Intelligence Organization. https://doi.org/10.24963/ijcai.2018/412

[70] ZHANG, Z., GOLDFELD, Z., MROUEH, Y. and SRIPERUMBUDUR, B. K. (2024). Supplement to "Gromov–Wasserstein Distances: Entropic Regularization, Duality and Sample Complexity." https://doi.org/10.1214/24-AOS2406SUPP

[71] ZHANG, Z., MROUEH, Y., GOLDFELD, Z. and SRIPERUMBUDUR, B. (2022). Cycle consistent probability divergences across different spaces. In *Proceedings of the* 25*th International Conference on Artificial Intelligence and Statistics* (G. Camps-Valls, F. J. R. Ruiz and I. Valera, eds.). *Proceedings of Machine Learning Research* **151** 7257–7285. PMLR, Online.

The Annals of Statistics
2024, Vol. 52, No. 4, 1646–1670
https://doi.org/10.1214/24-AOS2410

ONLINE CHANGE-POINT DETECTION FOR MATRIX-VALUED TIME SERIES WITH LATENT TWO-WAY FACTOR STRUCTURE

BY YONG HE[1,a], XINBING KONG[2,b], LORENZO TRAPANI[3,c] AND LONG YU[4,d]

[1]*Institute for Financial Studies, Shandong University,* [a]*heyong@sdu.edu.cn*
[2]*School of Statistics and Data Science, Nanjing Audit University,* [b]*xinbingkong@126.com*
[3]*Department of Economics and Management, Universita' di Pavia,* [c]*lorenzo.trapani@unipv.it*
[4]*School of Statistics and Management, Shanghai University of Finance and Economics,* [d]*yulong@mail.shufe.edu.cn*

This paper proposes a novel methodology for the online detection of changepoints in the factor structure of large matrix time series. Our approach is based on the well-known fact that, in the presence of a changepoint, the number of spiked eigenvalues in the second moment matrix of the data increases (e.g., in the presence of a change in the loadings, or if a new factor emerges). Based on this, we propose two families of procedures—one based on the fluctuations of partial sums, and one based on extreme value theory—to monitor whether the first nonspiked eigenvalue diverges after a point in time in the monitoring horizon, thereby indicating the presence of a changepoint. Our procedure is based only on rates; at each point in time, we randomise the estimated eigenvalue, thus obtaining a normally distributed sequence which is i.i.d. with mean zero under the null of no break, whereas it diverges to positive infinity in the presence of a changepoint. We base our monitoring procedures on such sequence. Extensive simulation studies and empirical analysis justify the theory. An R package implementing the procedure is available on CRAN.[1]

REFERENCES

AUE, A. and HORVÁTH, L. (2004). Delay time in sequential detection of change. *Statist. Probab. Lett.* **67** 221–231. MR2053524 https://doi.org/10.1016/j.spl.2004.01.002

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

BAI, J., DUAN, J. and HAN, X. (2024). The likelihood ratio test for structural changes in factor models. *J. Econometrics* **238** Paper No. 105631, 23 pp. MR4683916 https://doi.org/10.1016/j.jeconom.2023.105631

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

BALTAGI, B. H., KAO, C. and WANG, F. (2017). Identification and estimation of a large factor model with structural instability. *J. Econometrics* **197** 87–100. MR3598648 https://doi.org/10.1016/j.jeconom.2016.10.007

BALTAGI, B. H., KAO, C. and WANG, F. (2021). Estimating and testing high dimensional factor models with multiple structural changes. *J. Econometrics* **220** 349–365. MR4201494 https://doi.org/10.1016/j.jeconom.2020.04.005

BANDI, F. M. and CORRADI, V. (2014). Nonparametric nonstationarity tests. *Econometric Theory* **30** 127–149. MR3177794 https://doi.org/10.1017/S0266466613000145

BARIGOZZI, M., CAVALIERE, G. and TRAPANI, L. (2024). Inference in heavy-tailed nonstationary multivariate time series. *J. Amer. Statist. Assoc.* **119** 565–581. MR4713914 https://doi.org/10.1080/01621459.2022.2128807

BARIGOZZI, M., CHO, H. and FRYZLEWICZ, P. (2018). Simultaneous multiple change-point and factor analysis for high-dimensional time series. *J. Econometrics* **206** 187–225. MR3840788 https://doi.org/10.1016/j.jeconom.2018.05.003

MSC2020 subject classifications. Primary 62L10, 62H25; secondary 62R07.
Key words and phrases. Matrix factor model, factor space, online changepoint detection, projection estimation, randomisation.

[1]https://cran.r-project.org/web/packages/OLCPM/index.html.

BARIGOZZI, M. and TRAPANI, L. (2020). Sequential testing for structural stability in approximate factor models. *Stochastic Process. Appl.* **130** 5149–5187. MR4108485 https://doi.org/10.1016/j.spa.2020.03.003

BARIGOZZI, M. and TRAPANI, L. (2022). Testing for common trends in nonstationary large datasets. *J. Bus. Econom. Statist.* **40** 1107–1122. MR4439276 https://doi.org/10.1080/07350015.2021.1901719

BREITUNG, J. and EICKMEIER, S. (2011). Testing for structural breaks in dynamic factor models. *J. Econometrics* **163** 71–84. MR2803667 https://doi.org/10.1016/j.jeconom.2010.11.008

CHEN, E. Y. and CHEN, R. (2020). Modeling dynamic transport network with matrix factor models: With an application to international trade flow. Preprint. Available at arXiv:1901.00769.

CHEN, E. Y. and FAN, J. (2023). Statistical inference for high-dimensional matrix-variate factor models. *J. Amer. Statist. Assoc.* **118** 1038–1055. MR4595475 https://doi.org/10.1080/01621459.2021.1970569

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

CHEN, E. Y., XIA, D., CAI, C. and FAN, J. (2020). Semiparametric tensor factor analysis by iteratively projected SVD. Preprint. Available at arXiv:2007.02404.

CHEN, L., DOLADO, J. J. and GONZALO, J. (2014). Detecting big structural breaks in large factor models. *J. Econometrics* **180** 30–48. MR3188910 https://doi.org/10.1016/j.jeconom.2014.01.006

CHEN, R., YANG, D. and ZHANG, C.-H. (2022). Factor models for high-dimensional tensor time series. *J. Amer. Statist. Assoc.* **117** 94–116. MR4399070 https://doi.org/10.1080/01621459.2021.1912757

CHEN, X., YANG, D., XU, Y., XIA, Y., WANG, D. and SHEN, H. (2023). Testing and support recovery of correlation structures for matrix-valued observations with an application to stock market data. *J. Econometrics* **232** 544–564. MR4539499 https://doi.org/10.1016/j.jeconom.2021.09.014

CHEN, Y., WANG, T. and SAMWORTH, R. J. (2022). High-dimensional, multiscale online changepoint detection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 234–266. MR4400396 https://doi.org/10.1111/rssb.12447

CHEN, Y., WANG, T. and SAMWORTH, R. J. (2024). Inference in high-dimensional online changepoint detection. *J. Amer. Statist. Assoc.* **119** 1461–1472. MR4766002 https://doi.org/10.1080/01621459.2023.2199962

CHENG, X., LIAO, Z. and SCHORFHEIDE, F. (2016). Shrinkage estimation of high-dimensional factor models with structural instabilities. *Rev. Econ. Stud.* **83** 1511–1543. MR3556505 https://doi.org/10.1093/restud/rdw005

CHU, C. SJ., STINCHCOMBE, M. and WHITE, H. (1996). Monitoring structural change. *Econometrica* **64** 1045–1066.

CORRADI, V. and SWANSON, N. R. (2006). The effect of data transformation on common cycle, cointegration, and unit root tests: Monte Carlo results and a simple test. *J. Econometrics* **132** 195–229. MR2271396 https://doi.org/10.1016/j.jeconom.2005.01.028

CORRADI, V. and SWANSON, N. R. (2014). Testing for structural stability of factor augmented forecasting models. *J. Econometrics* **182** 100–118. MR3212764 https://doi.org/10.1016/j.jeconom.2014.04.011

CSÖRGŐ, M. and HORVÁTH, L. (1997). *Limit Theorems in Change-Point Analysis*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR2743035

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

DETTE, H. and GÖSMANN, J. (2020). A likelihood ratio approach to sequential change point detection for a general class of parameters. *J. Amer. Statist. Assoc.* **115** 1361–1377. MR4143471 https://doi.org/10.1080/01621459.2019.1630562

DUAN, J., BAI, J. and HAN, X. (2023). Quasi-maximum likelihood estimation of break point in high-dimensional factor models. *J. Econometrics* **233** 209–236. MR4554733 https://doi.org/10.1016/j.jeconom.2021.12.011

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

FEDOTENKOV, I. (2013). A bootstrap method to test for the existence of finite moments. *J. Nonparametr. Stat.* **25** 315–322. MR3056087 https://doi.org/10.1080/10485252.2012.752487

GASULL, A., JOLIS, M. and UTZET, F. (2015). On the norming constants for normal maxima. *J. Math. Anal. Appl.* **422** 376–396. MR3263464 https://doi.org/10.1016/j.jmaa.2014.08.025

GOMBAY, E. and HORVÁTH, L. (1996). On the rate of approximations for maximum likelihood tests in change-point models. *J. Multivariate Anal.* **56** 120–152. MR1380184 https://doi.org/10.1006/jmva.1996.0007

HALL, P. (1979). On the rate of convergence of normal extremes. *J. Appl. Probab.* **16** 433–439. MR0531778 https://doi.org/10.1017/s0021900200046647

HALLIN, M. and LIŠKA, R. (2007). Determining the number of factors in the general dynamic factor model. *J. Amer. Statist. Assoc.* **102** 603–617. MR2325115 https://doi.org/10.1198/016214506000001275

HAN, X. and INOUE, A. (2015). Tests for parameter instability in dynamic factor models. *Econometric Theory* **31** 1117–1152. MR3396242 https://doi.org/10.1017/S0266466614000486

HAN, Y., CHEN, R. and ZHANG, C.-H. (2022). Rank determination in tensor factor model. *Electron. J. Stat.* **16** 1726–1803. MR4393792 https://doi.org/10.1214/22-ejs1991

HE, Y., KONG, X., TRAPANI, L. and YU, L. (2023). One-way or two-way factor model for matrix sequences? *J. Econometrics* **235** 1981–2004. MR4602941 https://doi.org/10.1016/j.jeconom.2023.02.008

HE, Y., KONG, X., TRAPANI, L. and YU, L. (2024). Supplement to "Online change-point detection for matrix-valued time series with latent two-way factor structure." https://doi.org/10.1214/24-AOS2410SUPPA, https://doi.org/10.1214/24-AOS2410SUPPB

HOPKE, P. K., LEUNG, M., LI, N. and NAVASCA, C. (2011). Block tensor decomposition for source apportionment of air pollution. Preprint. Available at arXiv:1110.4133.

HORVÁTH, L., HUŠKOVÁ, M., KOKOSZKA, P. and STEINEBACH, J. (2004). Monitoring changes in linear models. *J. Statist. Plann. Inference* **126** 225–251. MR2090695 https://doi.org/10.1016/j.jspi.2003.07.014

HORVÁTH, L., KOKOSZKA, P. and STEINEBACH, J. (2007). On sequential detection of parameter changes in linear regression. *Statist. Probab. Lett.* **77** 885–895. MR2363438 https://doi.org/10.1016/j.spl.2006.12.014

HORVÁTH, L., MILLER, C. and RICE, G. (2020). A new class of change point test statistics of Rényi type. *J. Bus. Econom. Statist.* **38** 570–579. MR4115418 https://doi.org/10.1080/07350015.2018.1537923

HORVÁTH, L. and TRAPANI, L. (2019). Testing for randomness in a random coefficient autoregression model. *J. Econometrics* **209** 338–352. MR3944754 https://doi.org/10.1016/j.jeconom.2019.01.005

HORVÁTH, L. and TRAPANI, L. (2023). Changepoint detection in heteroscedastic random coefficient autoregressive models. *J. Bus. Econom. Statist.* **41** 1300–1314. MR4650463 https://doi.org/10.1080/07350015.2022.2120485

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131. MR0375412 https://doi.org/10.1007/BF00533093

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrsch. Verw. Gebiete* **34** 33–58. MR0402883 https://doi.org/10.1007/BF00532688

KOREN, Y., BELL, R. and VOLINSKY, C. (2009). Matrix factorization techniques for recommender systems. *Computer* **42** 30–37.

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, Q., CHENG, G., FAN, J. and WANG, Y. (2018). Embracing the blessing of dimensionality in factor models. *J. Amer. Statist. Assoc.* **113** 380–389. MR3803472 https://doi.org/10.1080/01621459.2016.1256815

LIU, X. and CHEN, E. Y. (2022). Identification and estimation of threshold matrix-variate factor models. *Scand. J. Stat.* **49** 1383–1417. MR4471290

MASSACCI, D. (2023). Testing for regime changes in portfolios with a large number of assets: A robust approach to factor heteroskedasticity. *J. Financ. Econom.* **21** 316–367.

MASSACCI, D., SARNO, L. and TRAPANI, L. (2021). Factor models with downside risk. Available at SSRN.

TRAPANI, L. (2016). Testing for (in)finite moments. *J. Econometrics* **191** 57–68. MR3434435 https://doi.org/10.1016/j.jeconom.2015.08.006

TRAPANI, L. (2018). A randomized sequential procedure to determine the number of factors. *J. Amer. Statist. Assoc.* **113** 1341–1349. MR3862361 https://doi.org/10.1080/01621459.2017.1328359

UEMATSU, Y. and YAMAGATA, T. (2023). Inference in sparsity-induced weak factor models. *J. Bus. Econom. Statist.* **41** 126–139. MR4522157 https://doi.org/10.1080/07350015.2021.2003203

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

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

WANG, Y. (2023). Identification and estimation of parameter instability in a high dimensional approximate factor model. Technical report.

WOODRUFF, D. P. (2014). Sketching as a tool for numerical linear algebra. *Found. Trends Theor. Comput. Sci.* **10** iv+157. MR3285427 https://doi.org/10.1561/0400000060

YAMAMOTO, Y. and TANAKA, S. (2015). Testing for factor loading structural change under common breaks. *J. Econometrics* **189** 187–206. MR3397360 https://doi.org/10.1016/j.jeconom.2015.06.018

YU, L., HE, Y., KONG, X. and ZHANG, X. (2022). Projected estimation for large-dimensional matrix factor models. *J. Econometrics* **229** 201–217. MR4414019 https://doi.org/10.1016/j.jeconom.2021.04.001

The Annals of Statistics
2024, Vol. 52, No. 4, 1671–1690
https://doi.org/10.1214/24-AOS2411

MAJORITY VOTE FOR DISTRIBUTED DIFFERENTIALLY PRIVATE SIGN SELECTION

BY WEIDONG LIU[1,a], JIYUAN TU[2,c], XIAOJUN MAO[1,b] AND XI CHEN[3,d]

[1]*School of Mathematical Sciences, Shanghai Jiao Tong University,* [a]*weidongl@sjtu.edu.cn,* [b]*maoxj@sjtu.edu.cn*
[2]*School of Mathematics, Shanghai University of Finance and Economics,* [c]*tujy.19@gmail.com*
[3]*Stern School of Business, New York University,* [d]*xc13@stern.nyu.edu*

Privacy-preserving data analysis has become more prevalent in recent years. In this study, we propose a distributed group differentially private Majority Vote mechanism, for the sign selection problem in a distributed setup. To achieve this, we apply the iterative peeling to the stability function and use the exponential mechanism to recover the signs. For enhanced applicability, we study the private sign selection for mean estimation and linear regression problems, in distributed systems. Our method recovers the support and signs with the optimal signal-to-noise ratio as in the nonprivate scenario, which is better than contemporary works of private variable selections. Moreover, the sign selection consistency is justified by theoretical guarantees. Simulation studies are conducted to demonstrate the effectiveness of the proposed method.

REFERENCES

[1] BAFNA, M. and ULLMAN, J. (2017). The price of selection in differential privacy. In *Proceedings of the* 2017 *Conference on Learning Theory*. *Proceedings of Machine Learning Research* **65** 151–168. PMLR.

[2] BASSILY, R., SMITH, A. and THAKURTA, A. (2014). Private empirical risk minimization: Efficient algorithms and tight error bounds. In 55*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2014 464–473. IEEE Comput. Soc., Los Alamitos, CA. MR3344896 https://doi.org/10.1109/FOCS.2014.56

[3] BATTEY, H., FAN, J., LIU, H., LU, J. and ZHU, Z. (2018). Distributed estimation and inference with statistical guarantees. *Ann. Statist.* **46** 1352–1382. MR4337068

[4] BELLET, A., LIANG, Y., GARAKANI, A. B., BALCAN, M.-F. and SHA, F. A distributed frank-Wolfe algorithm for communication-efficient sparse learning. In *Proceedings of the* 2015 *SIAM International Conference on Data Mining* (*SDM*) 478–486.

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

[6] 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.

[7] BU, Z., DONG, J., LONG, Q. and SU, W. J. (2020). Deep learning with Gaussian differential privacy. Harvard data science review 2020.

[8] 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

[9] BUTUCEA, C., DUBOIS, A. and SAUMARD, A. (2023). Phase transitions for support recovery under local differential privacy. *Math. Stat. Learn.* **6** 1–50. MR4619820 https://doi.org/10.4171/msl/37

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

[11] CAI, T. and LIU, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *J. Amer. Statist. Assoc.* **106** 672–684. MR2847949 https://doi.org/10.1198/jasa.2011.tm10560

MSC2020 subject classifications. Primary 62F30; secondary 62F12, 62J05.
Key words and phrases. Differential privacy, distributed learning, high-dimensional regression, majority voting, sign selection.

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

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

[14] 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

[15] CHEN, X., MIAO, S. and WANG, Y. (2023). Differential privacy in personalized pricing with nonparametric demand models. *Oper. Res.* **71** 581–602. MR4585721

[16] CHEN, X., SIMCHI-LEVI, D. and WANG, Y. (2022). Privacy-preserving dynamic personalized pricing with demand learning. *Manage. Sci.* **68** 4878–4898.

[17] DONG, J., ROTH, A. and SU, W. J. (2022). Gaussian differential privacy. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 3–54. With discussions and a reply by the authors. MR4400389

[18] DU, W. and ATALLAH, M. J. (2001). Secure multi-party computation problems and their applications: A review and open problems. In *Proceedings of the* 2001 *Workshop on New Security Paradigms*. *NSPW'*01 13–22. New York Assoc. Comput. Mach.

[19] 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

[20] 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

[21] 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

[22] DWORK, C. and ROTH, A. (2013). The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.* **9** 211–487. MR3254020 https://doi.org/10.1561/0400000042

[23] DWORK, C., SU, W. and ZHANG, L. (2021). Differentially private false discovery rate control. *J. Priv. Confid.* **11**.

[24] DWORK, C., TALWAR, K., THAKURTA, A. and ZHANG, L. (2014). Analyze Gauss: Optimal bounds for privacy-preserving principal component analysis. In *STOC'*14*—Proceedings of the* 2014 *ACM Symposium on Theory of Computing* 11–20. ACM, New York. MR3238926 https://doi.org/10.1145/2591796.2591883

[25] EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499. With discussion, and a rejoinder by the authors. MR2060166 https://doi.org/10.1214/009053604000000067

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

[27] EVANS, D., KOLESNIKOV, V. and ROSULEK, M. (2018). *A Pragmatic Introduction to Secure Multi-Party Computation*. *Foundations and Trends in Privacy and Security* **2** 70–246.

[28] 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

[29] GOLDREICH, O. (1998). Secure multi-party computation. Manuscript. Preliminary version 78.

[30] HAMM, J., CAO, Y. and BELKIN, M. (2016). Learning privately from multiparty data. In *Proceedings of the* 33*rd International Conference on Machine Learning*. *Proceedings of Machine Learning Research* **48** 555–563. PMLR.

[31] JAIN, P. and THAKURTA, A. G. (2014). (Near) dimension independent risk bounds for differentially private learning. In *Proceedings of the* 31*st International Conference on Machine Learning* **32** 476–484.

[32] JAVANMARD, A. and MONTANARI, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909. MR3277152

[33] 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

[34] JAYARAMAN, B., WANG, L., EVANS, D. and GU, Q. (2018). Distributed learning without distress: Privacy-preserving empirical risk minimization. In *Advances in Neural Information Processing Systems* **31** 6343–6354. Curran Associates, Red Hook.

[35] KASIVISWANATHAN, S. P. and JIN, H. (2016). Efficient private empirical risk minimization for high-dimensional learning. In *Proceedings of the* 33*rd International Conference on Machine Learning* **48** 488–497.

[36] KIFER, D., SMITH, A. and THAKURTA, A. (2012). Private convex empirical risk minimization and high-dimensional regression. In *Proceedings of the* 25*th Annual Conference on Learning Theory*. *Proceedings of Machine Learning Research* **23** 25.1–25.40. PMLR.

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

[38] LEI, J. (2011). Differentially private M-estimators. In *Advances in Neural Information Processing Systems* **24** 361–369. Curran Associates, Red Hook.

[39] LEI, J., CHAREST, A.-S., SLAVKOVIC, A., SMITH, A. and FIENBERG, S. (2018). Differentially private model selection with penalized and constrained likelihood. *J. Roy. Statist. Soc. Ser. A* **181** 609–633. MR3807500 https://doi.org/10.1111/rssa.12324

[40] LIU, W., TU, J., MAO, X. and CHEN, X. (2024). Supplement to "Majority Vote for Distributed Differentially Private Sign Selection." https://doi.org/10.1214/24-AOS2411SUPP

[41] MCSHERRY, F. and TALWAR, K. (2007). Mechanism design via differential privacy. In 48*th Annual IEEE Symposium on Foundations of Computer Science* (*FOCS*'07) 94–103.

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

[43] MILLER, A. (2002). *Subset Selection in Regression*, 2nd ed. *Monographs on Statistics and Applied Probability* **95**. CRC Press, Boca Raton, FL. MR2001193 https://doi.org/10.1201/9781420035933

[44] MIRONOV, I. (2017). Rényi differential privacy. In 2017 *IEEE* 30*th Computer Security Foundations Symposium* (*CSF*) 263–275.

[45] NISSIM, K., RASKHODNIKOVA, S. and SMITH, A. (2007). Smooth sensitivity and sampling in private data analysis. In *STOC*'07—*Proceedings of the* 39*th Annual ACM Symposium on Theory of Computing* 75–84. ACM, New York. MR2402430 https://doi.org/10.1145/1250790.1250803

[46] PATHAK, M., RANE, S. and RAJ, B. (2010). Multiparty differential privacy via aggregation of locally trained classifiers. In *Advances in Neural Information Processing Systems* **23** 1876–1884. Curran Associates, Red Hook.

[47] QIAO, G., SU, W. and ZHANG, L. (2021). Oneshot differentially private top-k selection. In *Proceedings of the* 38*th International Conference on Machine Learning*. *Proceedings of Machine Learning Research* **139** 8672–8681. PMLR.

[48] SHI, W., LING, Q., WU, G. and YIN, W. (2015). A proximal gradient algorithm for decentralized composite optimization. *IEEE Trans. Signal Process.* **63** 6013–6023. MR3411374 https://doi.org/10.1109/TSP.2015.2461520

[49] SMITH, A. (2011). Privacy-preserving statistical estimation with optimal convergence rates [extended abstract]. 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

[50] STEINKE, T. and ULLMAN, J. (2017). Tight lower bounds for differentially private selection. In 58*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 552–563. IEEE Comput. Soc., Los Alamitos, CA. MR3734260 https://doi.org/10.1109/FOCS.2017.57

[51] TALWAR, K., GUHA THAKURTA, A. and ZHANG, L. (2015). Nearly optimal private LASSO. In *Advances in Neural Information Processing Systems*, *Vol*. 28. Curran Associates, Red Hook.

[52] THAKURTA, A. G. and SMITH, A. (2013). Differentially private feature selection via stability arguments, and the robustness of the Lasso. In *Proceedings of the* 26*th Annual Conference on Learning Theory*. *Proceedings of Machine Learning Research* **30** 819–850. PMLR.

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

[54] VADHAN, S. (2017). The complexity of differential privacy. In *Tutorials on the Foundations of Cryptography*. *Inf. Secur. Cryptography* 347–450. Springer, Cham. MR3837668

[55] WAINWRIGHT, M. J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso). *IEEE Trans. Inf. Theory* **55** 2183–2202. MR2729873 https://doi.org/10.1109/TIT.2009.2016018

[56] WANG, D. and XU, J. (2021). On sparse linear regression in the local differential privacy model. *IEEE Trans. Inf. Theory* **67** 1182–1200. MR4232008 https://doi.org/10.1109/TIT.2020.3040406

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

[58] WANG, Y.-X., LEI, J. and FIENBERG, S. E. (2016). On-average KL-privacy and its equivalence to generalization for max-entropy mechanisms. In *Privacy in Statistical Databases* 121–134 Springer, Cham.

[59] 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

[60] YUAN, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *J. Mach. Learn. Res.* **11** 2261–2286. MR2719856

[61] ZHU, X., LI, F. and WANG, H. (2021). Least-square approximation for a distributed system. *J. Comput. Graph. Statist.* **30** 1004–1018. MR4356601 https://doi.org/10.1080/10618600.2021.1923517

The Annals of Statistics
2024, Vol. 52, No. 4, 1691–1715
https://doi.org/10.1214/24-AOS2413

WASSERSTEIN CONVERGENCE IN BAYESIAN AND FREQUENTIST DECONVOLUTION MODELS

BY JUDITH ROUSSEAU[1,a] AND CATIA SCRICCIOLO[2,b]

[1]*Department of Statistics, University of Oxford,* [a]*judith.rousseau@stats.ox.ac.uk*
[2]*Dipartimento di Scienze Economiche, Università degli Study di Verona,* [b]*catia.scricciolo@univr.it*

We study the multivariate deconvolution problem of recovering the distribution of a signal from independent and identically distributed observations additively contaminated with random errors having known distribution. For errors with ordinary smooth distribution, we recast the multidimensional problem as a one-dimensional problem leveraging the equivalence between the L^1-Wasserstein and the max-sliced L^1-Wasserstein metrics and derive an inversion inequality relating the L^1-Wasserstein distance between two distributions of the signal to the L^1-distance between the corresponding mixture densities of the observations. This smoothing inequality outperforms existing inversion inequalities. We apply it to derive L^1-Wasserstein rates of convergence for the distribution of the signal. As an application to the Bayesian framework, we consider L^1-Wasserstein deconvolution with the Laplace noise in dimension one using a Dirichlet process mixture of normal densities as a prior measure for the mixing distribution. We construct an adaptive approximation of the sampling density by convolving the Laplace density with a well-chosen mixture of Gaussian densities and show that the posterior measure contracts at a nearly minimax-optimal rate, up to a log-factor, in the L^1-distance. The rate automatically adapts to the unknown Sobolev regularity of the mixing density, thus leading to a new Bayesian adaptive estimation procedure over the full scale of regularity levels. We illustrate the utility of the inversion inequality also in a frequentist setting by showing that a minimum distance estimator attains the minimax convergence rates for L^1-Wasserstein deconvolution in any dimension $d \geq 1$, lower bounds being derived here.

REFERENCES

[1] BALABDAOUI, F. and WELLNER, J. A. (2007). Estimation of a k-monotone density: Limit distribution theory and the spline connection. *Ann. Statist.* **35** 2536–2564. MR2382657 https://doi.org/10.1214/009053607000000262

[2] BAYRAKTAR, E. and GUOÏ, G. (2021). Strong equivalence between metrics of Wasserstein type. *Electron. Commun. Probab.* **26** Paper No. 13, 13. MR4236683 https://doi.org/10.3390/mca26010013

[3] BISSANTZ, N., DÜMBGEN, L., HOLZMANN, H. and MUNK, A. (2007). Non-parametric confidence bands in deconvolution density estimation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 483–506. MR2323764 https://doi.org/10.1111/j.1467-9868.2007.599.x

[4] BOURGAIN, J. (1999). On triples in arithmetic progression. *Geom. Funct. Anal.* **9** 968–984. MR1726234 https://doi.org/10.1007/s000390050105

[5] BUTUCEA, C. and COMTE, F. (2009). Adaptive estimation of linear functionals in the convolution model and applications. *Bernoulli* **15** 69–98. MR2546799 https://doi.org/10.3150/08-BEJ146

[6] 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

[7] BUTUCEA, C. and TSYBAKOV, A. B. (2007). Sharp optimality in density deconvolution with dominating bias. II. *Teor. Veroyatn. Primen.* **52** 336–349. MR2742504 https://doi.org/10.1137/S0040585X97982992

MSC2020 subject classifications. 62G07, 62G20.
Key words and phrases. Adaptation, density estimation, Dirichlet process mixtures, inversion inequalities, max-sliced Wasserstein metrics, minimax rates, mixtures of Laplace densities, multivariate deconvolution, rates of convergence, Sobolev classes, Wasserstein metrics.

[8] BUTUCEA, C. and TSYBAKOV, A. B. (2007). Sharp optimality in density deconvolution with dominating bias. I. *Teor. Veroyatn. Primen.* **52** 111–128. MR2354572 https://doi.org/10.1137/S0040585X97982840

[9] CAILLERIE, C., CHAZAL, F., DEDECKER, J. and MICHEL, B. (2011). Deconvolution for the Wasserstein metric and geometric inference. *Electron. J. Stat.* **5** 1394–1423. MR2851684 https://doi.org/10.1214/11-EJS646

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

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

[12] COMTE, F. and REBAFKA, T. (2012). Adaptive density estimation in the pile-up model involving measurement errors. *Electron. J. Stat.* **6** 2002–2037. MR3020255 https://doi.org/10.1214/12-EJS737

[13] COMTE, F., ROZENHOLC, Y. and TAUPIN, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Canad. J. Statist.* **34** 431–452. MR2328553 https://doi.org/10.1002/cjs.5550340305

[14] DATTNER, I., REISS, M. and TRABS, M. (2016). Adaptive quantile estimation in deconvolution with unknown error distribution. *Bernoulli* **22** 143–192. MR3449779 https://doi.org/10.3150/14-BEJ626

[15] DE JONGE, R. and VAN ZANTEN, J. H. (2010). Adaptive nonparametric Bayesian inference using location-scale mixture priors. *Ann. Statist.* **38** 3300–3320. MR2766853 https://doi.org/10.1214/10-AOS811

[16] DEDECKER, J., FISCHER, A. and MICHEL, B. (2015). Improved rates for Wasserstein deconvolution with ordinary smooth error in dimension one. *Electron. J. Stat.* **9** 234–265. MR3314482 https://doi.org/10.1214/15-EJS997

[17] DEDECKER, J. and MICHEL, B. (2013). Minimax rates of convergence for Wasserstein deconvolution with supersmooth errors in any dimension. *J. Multivariate Anal.* **122** 278–291. MR3189324 https://doi.org/10.1016/j.jmva.2013.08.009

[18] DELAIGLE, A. and GIJBELS, I. (2004). Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Ann. Inst. Statist. Math.* **56** 19–47. MR2053727 https://doi.org/10.1007/BF02530523

[19] DIGGLE, P. J. and HALL, P. (1993). A Fourier approach to nonparametric deconvolution of a density estimate. *J. Roy. Statist. Soc. Ser. B* **55** 523–531. MR1224414

[20] DINUR, I. and NISSIM, K. (2003). Revealing information while preserving privacy. In *Twenty Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, *PODS* 2003 202–210. Conference Date: 09-06-2003. Through 11-06-2003. https://doi.org/10.1145/773153.773173

[21] DIVOL, V. (2022). Measure estimation on manifolds: An optimal transport approach. *Probab. Theory Related Fields* **183** 581–647. MR4421180 https://doi.org/10.1007/s00440-022-01118-z

[22] 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

[23] DWORK, C. (2008). Differential privacy: A survey of results. In *Theory and Applications of Models of Computation*. *Lecture Notes in Computer Science* **4978** 1–19. Springer, Berlin. MR2472670 https://doi.org/10.1007/978-3-540-79228-4_1

[24] DWORK, C. and NISSIM, K. (2004). Privacy-preserving datamining on vertically partitioned databases. In *Advances in Cryptology—CRYPTO* 2004. *Lecture Notes in Computer Science* **3152** 528–544. Springer, Berlin. MR2147523 https://doi.org/10.1007/978-3-540-28628-8_32

[25] EVFIMIEVSKI, A. V., GEHRKE, J. and SRIKANT, R. (2003). Limiting privacy breaches in privacy preserving data mining. In *PODS* ’03.

[26] FAN, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.* **21** 196–216. MR1212173 https://doi.org/10.1214/aos/1176349022

[27] FRY, R. and MCMANUS, S. (2002). Smooth bump functions and the geometry of Banach spaces: A brief survey. *Expo. Math.* **20** 143–183. MR1904712 https://doi.org/10.1016/S0723-0869(02)80017-2

[28] GAO, F. and VAN DER VAART, A. (2016). Posterior contraction rates for deconvolution of Dirichlet–Laplace mixtures. *Electron. J. Stat.* **10** 608–627. MR3471990 https://doi.org/10.1214/16-EJS1119

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

[30] GHOSAL, S. and VAN DER VAART, A. (2007). Convergence rates of posterior distributions for non-i.i.d. observations. *Ann. Statist.* **35** 192–223. MR2332274 https://doi.org/10.1214/009053606000001172

[31] 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

[32] GINÉ, E. and NICKL, R. (2011). Rates of contraction for posterior distributions in L^r-metrics, $1 \le r \le \infty$. *Ann. Statist.* **39** 2883–2911. MR3012395 https://doi.org/10.1214/11-AOS924

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

[34] KOTZ, S., KOZUBOWSKI, T. J. and PODGÓRSKI, K. (2001). *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*. Birkhäuser, Boston, MA. MR1935481 https://doi.org/10.1007/978-1-4612-0173-1

[35] KRUIJER, W., ROUSSEAU, J. and VAN DER VAART, A. (2010). Adaptive Bayesian density estimation with location-scale mixtures. *Electron. J. Stat.* **4** 1225–1257. MR2735885 https://doi.org/10.1214/10-EJS584

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

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

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

[39] PENSKY, M. and VIDAKOVIC, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.* **27** 2033–2053. MR1765627 https://doi.org/10.1214/aos/1017939249

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

[41] 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

[42] ROUSSEAU, J. (2010). Rates of convergence for the posterior distributions of mixtures of betas and adaptive nonparametric estimation of the density. *Ann. Statist.* **38** 146–180. MR2589319 https://doi.org/10.1214/09-AOS703

[43] ROUSSEAU, J. and SCRICCIOLO, C. (2024). Supplement to "Wasserstein convergence in Bayesian and frequentist deconvolution models." https://doi.org/10.1214/24-AOS2413SUPP

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

[45] SCRICCIOLO, C. (2011). Posterior rates of convergence for Dirichlet mixtures of exponential power densities. *Electron. J. Stat.* **5** 270–308. MR2802044 https://doi.org/10.1214/11-EJS604

[46] SCRICCIOLO, C. (2014). Adaptive Bayesian density estimation in L^p-metrics with Pitman–Yor or normalized inverse-Gaussian process kernel mixtures. *Bayesian Anal.* **9** 475–520. MR3217004 https://doi.org/10.1214/14-BA863

[47] SCRICCIOLO, C. (2018). Bayes and maximum likelihood for L^1-Wasserstein deconvolution of Laplace mixtures. *Stat. Methods Appl.* **27** 333–362. MR3807373 https://doi.org/10.1007/s10260-017-0400-4

[48] SHEN, W., TOKDAR, S. T. and GHOSAL, S. (2013). Adaptive Bayesian multivariate density estimation with Dirichlet mixtures. *Biometrika* **100** 623–640. MR3094441 https://doi.org/10.1093/biomet/ast015

[49] SU, Y., BHATTACHARYA, A., ZHANG, Y., CHATTERJEE, N. and CARROLL, R. J. (2020). Nonparametric Bayesian deconvolution of a symmetric unimodal density. Preprint. Available at arXiv:2002.07255.

[50] VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2009). Adaptive Bayesian estimation using a Gaussian random field with inverse gamma bandwidth. *Ann. Statist.* **37** 2655–2675. MR2541442 https://doi.org/10.1214/08-AOS678

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

[52] WILLIAMSON, R. E. (1956). Multiply monotone functions and their Laplace transforms. *Duke Math. J.* **23** 189–207. MR0077581

The Annals of Statistics
2024, Vol. 52, No. 4, 1716–1740
https://doi.org/10.1214/24-AOS2414

DETECTION AND ESTIMATION OF STRUCTURAL BREAKS IN HIGH-DIMENSIONAL FUNCTIONAL TIME SERIES

BY DEGUI LI[1,a], RUNZE LI[2,b] AND HAN LIN SHANG[3,c]

[1]*Faculty of Business Administration, University of Macau,* [a]*deguili@um.edu.mo*
[2]*Department of Statistics, Pennsylvania State University,* [b]*rzli@psu.edu*
[3]*Department of Actuarial Studies and Business Analytics, Macquarie University,* [c]*hanlin.shang@mq.edu.au*

We consider detecting and estimating breaks in heterogenous mean functions of high-dimensional functional time series which are allowed to be cross-sectionally correlated. A new test statistic combining the functional CUSUM statistic and power enhancement component is proposed with asymptotic null distribution comparable to the conventional CUSUM theory derived for a single functional time series. In particular, the extra power enhancement component enlarges the region where the proposed test has power, and results in stable power performance when breaks are sparse in the alternative hypothesis. Furthermore, we impose a latent group structure on the subjects with heterogenous break points and introduce an easy-to-implement clustering algorithm with an information criterion to consistently estimate the unknown group number and membership. The estimated group structure improves the convergence property of the break point estimate. Monte Carlo simulation studies and empirical applications show that the proposed estimation and testing techniques have satisfactory performance in finite samples.

REFERENCES

ASTON, J. A. D. and KIRCH, C. (2012). Detecting and estimating changes in dependent functional data. *J. Multivariate Anal.* **109** 204–220. MR2922864 https://doi.org/10.1016/j.jmva.2012.03.006

AUE, A., GABRYS, R., HORVÁTH, L. and KOKOSZKA, P. (2009). Estimation of a change-point in the mean function of functional data. *J. Multivariate Anal.* **100** 2254–2269. MR2560367 https://doi.org/10.1016/j.jmva.2009.04.001

AUE, A., RICE, G. and SÖNMEZ, O. (2018). Detecting and dating structural breaks in functional data without dimension reduction. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 509–529. MR3798876 https://doi.org/10.1111/rssb.12257

BAI, J. (1997). Estimating multiple breaks one at a time. *Econometric Theory* **13** 315–352. MR1455175 https://doi.org/10.1017/S0266466600005831

BAI, J. (2010). Common breaks in means and variances for panel data. *J. Econometrics* **157** 78–92. MR2652280 https://doi.org/10.1016/j.jeconom.2009.10.020

BARIGOZZI, M., CHO, H. and FRYZLEWICZ, P. (2018). Simultaneous multiple change-point and factor analysis for high-dimensional time series. *J. Econometrics* **206** 187–225. MR3840788 https://doi.org/10.1016/j.jeconom.2018.05.003

BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. MR1325392

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

BERKES, I., GABRYS, R., HORVÁTH, L. and KOKOSZKA, P. (2009). Detecting changes in the mean of functional observations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 927–946. MR2750251 https://doi.org/10.1111/j.1467-9868.2009.00713.x

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York. MR0233396

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

MSC2020 subject classifications. Primary 62F03, 62F12; secondary 62M10.
Key words and phrases. Clustering, CUSUM, functional time series, power enhancement, structural breaks.

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

CHIOU, J.-M. and LI, P.-L. (2007). Functional clustering and identifying substructures of longitudinal data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 679–699. MR2370075 https://doi.org/10.1111/j.1467-9868.2007.00605.x

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

DELAIGLE, A. and HALL, P. (2012). Achieving near perfect classification for functional data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 267–286. MR2899863 https://doi.org/10.1111/j.1467-9868.2011.01003.x

DELAIGLE, A., HALL, P. and PHAM, T. (2019). Clustering functional data into groups by using projections. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 271–304. MR3928143 https://doi.org/10.1111/rssb.12310

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

FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis*: *Theory and Practice*. *Springer Series in Statistics*. Springer, New York. MR2229687

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

HÖRMANN, S., HORVÁTH, L. and REEDER, R. (2013). A functional version of the ARCH model. *Econometric Theory* **29** 267–288. MR3042756 https://doi.org/10.1017/S0266466612000345

HÖRMANN, S. and KOKOSZKA, P. (2010). Weakly dependent functional data. *Ann. Statist.* **38** 1845–1884. MR2662361 https://doi.org/10.1214/09-AOS768

HORVÁTH, L. and HUŠKOVÁ, M. (2012). Change-point detection in panel data. *J. Time Series Anal.* **33** 631–648. MR2944843 https://doi.org/10.1111/j.1467-9892.2012.00796.x

HORVÁTH, L. and KOKOSZKA, P. (2012). *Inference for Functional Data with Applications*. *Springer Series in Statistics*. Springer, New York. MR2920735 https://doi.org/10.1007/978-1-4614-3655-3

HORVÁTH, L., KOKOSZKA, P. and RICE, G. (2014). Testing stationarity of functional time series. *J. Econometrics* **179** 66–82. MR3153649 https://doi.org/10.1016/j.jeconom.2013.11.002

HUMAN MORTALITY DATABASE (2023). Max Planck Institute for Demographic Research (Germany), Univ. California, Berkeley USA, and French Institute for Demographic Studies (France). Available at http://www.mortality.org (data downloaded on 2023-3-3).

HYNDMAN, R., ATHANASOPOULOS, G., BERGMEIR, C., CACERES, G., CHHAY, L., O'HARA-WILD, M., PETROPOULOS, F., RAZBASH, S., WANG, E. et al. (2019). forecast: Forecasting functions for time series and linear models. R package version 8.4. Available at https://CRAN.R-project.org/package=forecast.

HYNDMAN, R. J. and ULLAH, M. S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Comput. Statist. Data Anal.* **51** 4942–4956. MR2364551 https://doi.org/10.1016/j.csda.2006.07.028

JIRAK, M. (2015). Uniform change point tests in high dimension. *Ann. Statist.* **43** 2451–2483. MR3405600 https://doi.org/10.1214/15-AOS1347

KOCK, A. B. and PREINERSTORFER, D. (2019). Power in high-dimensional testing problems. *Econometrica* **87** 1055–1069. MR3957338 https://doi.org/10.3982/ECTA15844

LI, D., LI, R. and SHANG, H. L. (2024). Supplement to "Detection and Estimation of Structural Breaks in High-Dimensional Functional Time Series." https://doi.org/10.1214/24-AOS2414SUPP

LI, D., ROBINSON, P. M. and SHANG, H. L. (2020). Long-range dependent curve time series. *J. Amer. Statist. Assoc.* **115** 957–971. MR4107692 https://doi.org/10.1080/01621459.2019.1604362

LI, D., ROBINSON, P. M. and SHANG, H. L. (2023). Nonstationary fractionally integrated functional time series. *Bernoulli* **29** 1505–1526. MR4550233 https://doi.org/10.3150/22-bej1508

LI, Y.-N., LI, D. and FRYZLEWICZ, P. (2023). Detection of multiple structural breaks in large covariance matrices. *J. Bus. Econom. Statist.* **41** 846–861. MR4600853 https://doi.org/10.1080/07350015.2022.2076686

PENG, J. and MÜLLER, H.-G. (2008). Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions. *Ann. Appl. Stat.* **2** 1056–1077. MR2516804 https://doi.org/10.1214/08-AOAS172

RICE, G., WIRJANTO, T. and ZHAO, Y. (2020). Tests for conditional heteroscedasticity of functional data. *J. Time Series Anal.* **41** 733–758. MR4176358 https://doi.org/10.1111/jtsa.12532

SAFIKHANI, A. and SHOJAIE, A. (2022). Joint structural break detection and parameter estimation in high-dimensional nonstationary VAR models. *J. Amer. Statist. Assoc.* **117** 251–264. MR4399083 https://doi.org/10.1080/01621459.2020.1770097

SHARIPOV, O., TEWES, J. and WENDLER, M. (2016). Sequential block bootstrap in a Hilbert space with application to change point analysis. *Canad. J. Statist.* **44** 300–322. MR3536199 https://doi.org/10.1002/cjs.11293

TANG, C., SHANG, H. L. and YANG, Y. (2022). Clustering and forecasting multiple functional time series. *Ann. Appl. Stat.* **16** 2523–2553. MR4489222 https://doi.org/10.1214/22-aoas1602

TARPEY, T. and KINATEDER, K. K. J. (2003). Clustering functional data. *J. Classification* **20** 93–114. MR1983123 https://doi.org/10.1007/s00357-003-0007-3

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

WOOD, S. N. (1994). Monotonic smoothing splines fitted by cross validation. *SIAM J. Sci. Comput.* **15** 1126–1133. MR1289157 https://doi.org/10.1137/0915069

ZHANG, X., SHAO, X., HAYHOE, K. and WUEBBLES, D. J. (2011). Testing the structural stability of temporally dependent functional observations and application to climate projections. *Electron. J. Stat.* **5** 1765–1796. MR2870150 https://doi.org/10.1214/11-EJS655

The Annals of Statistics
2024, Vol. 52, No. 4, 1741–1773
https://doi.org/10.1214/24-AOS2415

EFFICIENT FUNCTIONAL LASSO KERNEL SMOOTHING FOR HIGH-DIMENSIONAL ADDITIVE REGRESSION

BY EUN RYUNG LEE[1,a], SEYOUNG PARK[2,b], ENNO MAMMEN[3,c] AND BYEONG U. PARK[4,d]

[1]*Department of Statistics, Sungkyunkwan University,* [a]*erlee@skku.edu*
[2]*Department of Applied Statistics, Yonsei University,* [b]*ishspsy@yonsei.ac.kr*
[3]*Institute for Applied Mathematics, Heidelberg University,* [c]*mammen@math.uni-heidelberg.de*
[4]*Department of Statistics, Seoul National University,* [d]*bupark@snu.ac.kr*

Smooth backfitting has been proposed and proved as a powerful nonparametric estimation technique for additive regression models in various settings. Existing studies are restricted to cases with a moderate number of covariates and are not directly applicable to high dimensional settings. In this paper, we develop new kernel estimators based on the idea of smooth backfitting for high dimensional additive models. We introduce a novel penalization scheme, combining the idea of functional Lasso with the smooth backfitting technique. We investigate the theoretical properties of the functional Lasso smooth backfitting estimation. For the implementation of the proposed method, we devise a simple iterative algorithm where the iteration is defined by a truncated projection operator. The algorithm has only an additional thresholding operator over the projection-based iteration of the smooth backfitting algorithm. We further present a debiased version of the proposed estimator with implementation details, and investigate its theoretical properties for statistical inference. We demonstrate the finite sample performance of the methods via simulation and real data analysis.

REFERENCES

[1] ATKINSON, A. C. (1987). *Plots, Transformations, and Regression*. Oxford.
[2] BACH, F. R. (2008). Consistency of the group lasso and multiple kernel learning. *J. Mach. Learn. Res.* **9** 1179–1225. MR2417268
[3] BARRETINA, J., CAPONIGRO, G., STRANSKY, N., VENKATESAN, K., MARGOLIN, A. A., KIM, S. et al. (2012). The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483** 603–607.
[4] BECK, A. and TETRUASHVILI, L. (2013). On the convergence of block coordinate descent type methods. *SIAM J. Optim.* **23** 2037–2060. MR3116649 https://doi.org/10.1137/120887679
[5] 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
[6] BLANCO, T. J., FRIGOLA, M. D. and ALOY, P. (2018). Rationalizing drug response in cancer cell lines. *J. Mol. Biol.* **430** 3016–3027.
[7] BREIMAN, L. and FRIEDMAN, J. H. (1985). Estimating optimal transformations for multiple regression and correlation (with discussion). *J. Amer. Statist. Assoc.* **80** 580–619. MR0803258
[8] BUJA, A., HASTIE, T. and TIBSHIRANI, R. (1989). Linear smoothers and additive models (with discussion). *Ann. Statist.* **17** 453–555. MR0994249 https://doi.org/10.1214/aos/1176347115
[9] CAI, T., LIU, W. and LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.* **106** 594–607. MR2847973 https://doi.org/10.1198/jasa.2011.tm10155
[10] CHEN, Z., FAN, J. and LI, R. (2018). Error variance estimation in ultrahigh-dimensional additive models. *J. Amer. Statist. Assoc.* **113** 315–327. MR3803467 https://doi.org/10.1080/01621459.2016.1251440

MSC2020 subject classifications. Primary 62G08; secondary 62G05, 62G20.
Key words and phrases. Nonparametric regression, additive models, kernel smoothing, smooth backfitting, sparse estimation, penalization, functional Lasso, debiasing.

[11] CIUFFREDA, L., BUFALO, D. D., DESIDERI, M., SANZA, C. D., STOPPACCIARO, A., RICCIARDI, M. R., CHIARETTI, S., TAVOLARO, S., BENASSI, B. et al. (2009). Growth-inhibitory and antiangiogenic activity of the MEK inhibitor PD0325901 in malignant melanoma with or without BRAF mutations. *Neoplasia* **11** 720–731. https://doi.org/10.1593/neo.09398

[12] EUBANK, R. L. and SPECKMAN, P. L. (1991). A bias reduction theorem with applications in nonparametric regression. *Scand. J. Stat.* **18** 211–222. MR1146178

[13] FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications. Monographs on Statistics and Applied Probability* **66**. CRC Press, London. MR1383587

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

[15] FERREIRA, D., ADEGA, F. and CHAVES, R. (2013). *The Importance of Cancer Cell Lines as in Vitro Models in Cancer Methylome Analysis and Anticancer Drugs Testing* (L.-C. Cesar, ed.).

[16] GARNETT, M. J. et al. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483** 570–575.

[17] GRANT, M., BOYD, S. and YE, Y. (2006). Disciplined convex programming. In *Global Optimization* (L. Liberti and N. Maculan, eds.) *Nonconvex Optim. Appl.* **84** 155–210. Springer, New York. MR2206954 https://doi.org/10.1007/0-387-30528-9_7

[18] GREGORY, K., MAMMEN, E. and WAHL, M. (2021). Statistical inference in sparse high-dimensional additive models. *Ann. Statist.* **49** 1514–1536. MR4302573 https://doi.org/10.1214/20-aos2011

[19] HALL, P. and WEHRLY, T. E. (1991). A geometrical method for removing edge effects from kernel-type nonparametric regression estimators. *J. Amer. Statist. Assoc.* **86** 665–672. MR1147090

[20] HAN, K., MÜLLER, H.-G. and PARK, B. U. (2020). Additive functional regression for densities as responses. *J. Amer. Statist. Assoc.* **115** 997–1010. MR4107695 https://doi.org/10.1080/01621459.2019.1604365

[21] HAN, K. and PARK, B. U. (2018). Smooth backfitting for errors-in-variables additive models. *Ann. Statist.* **46** 2216–2250. MR3845016 https://doi.org/10.1214/17-AOS1617

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

[23] HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models. Monographs on Statistics and Applied Probability* **43**. CRC Press, London. MR1082147

[24] HONG, M., WANG, X., RAZAVIYAYN, M. and LUO, Z.-Q. (2017). Iteration complexity analysis of block coordinate descent methods. *Math. Program.* **163** 85–114. MR3632975 https://doi.org/10.1007/s10107-016-1057-8

[25] HUANG, J., HOROWITZ, J. L. and WEI, F. (2010). Variable selection in nonparametric additive models. *Ann. Statist.* **38** 2282–2313. MR2676890 https://doi.org/10.1214/09-AOS781

[26] JEON, J. M., LEE, Y. K., MAMMEN, E. and PARK, B. U. (2022). Locally polynomial Hilbertian additive regression. *Bernoulli* **28** 2034–2066. MR4411521 https://doi.org/10.3150/21-bej1410

[27] JEON, J. M. and PARK, B. U. (2020). Additive regression with Hilbertian responses. *Ann. Statist.* **48** 2671–2697. MR4152117 https://doi.org/10.1214/19-AOS1902

[28] JEON, J. M. and PARK, B. U. (2024). Supplement to "Additive regression with Hilbertian responses." https://doi.org/10.1214/24-AOS2415SUPP

[29] JEON, J. M., PARK, B. U. and VAN KEILEGOM, I. (2021). Additive regression for non-Euclidean responses and predictors. *Ann. Statist.* **49** 2611–2641. MR4338377 https://doi.org/10.1214/21-aos2048

[30] JONES, M. C. (1991). Simple boundary correction for kernel density estimation. *Stat. Comput.* **3** 133–146. MR1125651

[31] JONES, M. C. and HENDERSON, D. A. (2007). Kernel-type density estimation on the unit interval. *Biometrika* **94** 977–984. MR2416803 https://doi.org/10.1093/biomet/asm068

[32] KLEJEWSKI, A., SWIERCZEWSKA, M., ZAORSKA, K., BRAZERT, M., NOWICKI, M., ZABEL, M. and JANUCHOWSKI, R. (2017). New and old genes associated with Topotecan resistance development in ovarian cancer cell lines. *Anticancer Res.* **37** 625–636.

[33] LEE, E. R., HAN, K. and PARK, B. U. (2018). Estimation of errors-in-variables partially linear additive models. *Statist. Sinica* **28** 2353–2373. MR3839864

[34] LEE, E. R., NOH, H. and PARK, B. U. (2014). Model selection via Bayesian information criterion for quantile regression models. *J. Amer. Statist. Assoc.* **109** 216–229. MR3180558 https://doi.org/10.1080/01621459.2013.836975

[35] LEE, E. R., PARK, S., MAMMEN, E. and PARK, B. U. (2024). Supplement to "Efficient Functional Lasso Kernel Smoothing for High-dimensional Additive Regression." https://doi.org/10.1214/24-AOS2415SUPP

[36] LEE, Y. K., MAMMEN, E. and PARK, B. U. (2010). Backfitting and smooth backfitting for additive quantile models. *Ann. Statist.* **38** 2857–2883. MR2722458 https://doi.org/10.1214/10-AOS808

[37] LEE, Y. K., MAMMEN, E. and PARK, B. U. (2012). Flexible generalized varying coefficient regression models. *Ann*. *Statist*. **40** 1906–1933. MR3015048 https://doi.org/10.1214/12-AOS1026

[38] LIN, Y. and ZHANG, H. H. (2006). Component selection and smoothing in multivariate nonparametric regression. *Ann*. *Statist*. **34** 2272–2297. MR2291500 https://doi.org/10.1214/009053606000000722

[39] LIN, Z., MÜLLER, H.-G. and PARK, B. U. (2023). Additive models for symmetric positive-definite matrices and Lie groups. *Biometrika* **110** 361–379. MR4588358 https://doi.org/10.1093/biomet/asac055

[40] LINTON, O. and NIELSEN, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82** 93–100. MR1332841 https://doi.org/10.1093/biomet/82.1.93

[41] LIU, H., WASSERMAN, L., LAFFERTY, J. and RAVIKUMAR, P. (2007). SpAM: Sparse additive models. In *NIPS* 20.

[42] LU, J., KOLAR, M. and LIU, H. (2020). Kernel meets sieve: Post-regularization confidence bands for sparse additive model. *J*. *Amer*. *Statist*. *Assoc*. **115** 2084–2099. MR4189778 https://doi.org/10.1080/01621459.2019.1689984

[43] MAKONDI, P. T., CHU, C.-M., WEI, P.-L. and CHANG, Y.-J. (2017). Prediction of novel target genes and pathways involved in irinotecan-resistant colorectal cancer. *PLoS ONE* **12** e0180616.

[44] MAMMEN, E., LINTON, O. and NIELSEN, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Ann*. *Statist*. **27** 1443–1490. MR1742496 https://doi.org/10.1214/aos/1017939137

[45] MEIER, L., VAN DE GEER, S. and BÜHLMANN, P. (2009). High-dimensional additive modeling. *Ann*. *Statist*. **37** 3779–3821. MR2572443 https://doi.org/10.1214/09-AOS692

[46] MESSER, K. and GOLDSTEIN, L. (1993). A new class of kernels for nonparametric curve estimation. *Ann*. *Statist*. **21** 179–195. MR1212172 https://doi.org/10.1214/aos/1176349021

[47] MÜLLER, H.-G. (1993). On the boundary kernel method for nonparametric curve estimation near endpoints. *Scand*. *J*. *Stat*. **20** 313–328. MR1276696

[48] MÜLLER, H. G. and STADTMÜLLER, U. (1999). Multivariate boundary kernels and a continuous least squares principle. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **61** 439–458. MR1680306 https://doi.org/10.1111/1467-9868.00186

[49] NIELSEN, J. P. and LINTON, O. B. (1998). An optimization interpretation of integration and back-fitting estimators for separable nonparametric models. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **60** 217–222. MR1625624 https://doi.org/10.1111/1467-9868.00120

[50] NISHII, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression. *Ann*. *Statist*. **12** 758–765. MR0740928 https://doi.org/10.1214/aos/1176346522

[51] OPSOMER, J. D. and RUPPERT, D. (1997). Fitting a bivariate additive model by local polynomial regression. *Ann*. *Statist*. **25** 186–211. MR1429922 https://doi.org/10.1214/aos/1034276626

[52] PARK, J. and SEIFERT, B. (2010). Local additive estimation. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **72** 171–191. MR2830763 https://doi.org/10.1111/j.1467-9868.2009.00731.x

[53] QIU, K., LEE, J., KIM, H., YOON, S. and KANG, K. (2021). Machine learning based anti-cancer drug response prediction and search for predictor genes using cancer cell line gene expression. *Genomics Inform*. **19** e10.

[54] RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2012). Minimax-optimal rates for sparse additive models over kernel classes via convex programming. *J*. *Mach*. *Learn*. *Res*. **13** 389–427. MR2913704

[55] RAVIKUMAR, P., LAFFERTY, J., LIU, H. and WASSERMAN, L. (2009). Sparse additive models. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **71** 1009–1030. MR2750255 https://doi.org/10.1111/j.1467-9868.2009.00718.x

[56] SHARMA, S. V., HABER, D. A. and SETTLEMAN, J. (2010). Cell line-based platforms to evaluate the therapeutic efficacy of candidate anticancer agents. *Nat*. *Rev*. *Cancer* **10** 241–253. https://doi.org/10.1038/nrc2820

[57] STAUNTON, J. E. et al. (2001). Chemosensitivity prediction by transcriptional profiling. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **98** 10787–92.

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

[59] TENTLER, J. J., NALLAPAREDDY, S., TAN, A. C. et al. (2010). Identification of predictive markers of response to the MEK1/2 inhibitor selumetinib (AZD6244) in K-ras–mutated colorectal cancer. *Mol*. *Cancer Ther*. **9** 3351–3362.

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

[61] VAN DE GEER, S. A. and BÜHLMANN, P. (2009). On the conditions used to prove oracle results for the Lasso. *Electron*. *J*. *Stat*. **3** 1360–1392. MR2576316 https://doi.org/10.1214/09-EJS506

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

[63] YU, K., PARK, B. U. and MAMMEN, E. (2008). Smooth backfitting in generalized additive models. *Ann. Statist.* **36** 228–260. MR2387970 https://doi.org/10.1214/009053607000000596

[64] ZHENG, Q., PENG, L. and HE, X. (2015). Globally adaptive quantile regression with ultra-high dimensional data. *Ann. Statist.* **43** 2225–2258. MR3396984 https://doi.org/10.1214/15-AOS1340

[65] ZUO, Y., XU, H., CHEN, Z. et al. (2020). 17-AAG synergizes with belinostat to exhibit a negative effect on the proliferation and invasion of MDA-MB-231 breast cancer cells. *Oncol. Rep.* **43** 1928–1944.

The Annals of Statistics
2024, Vol. 52, No. 4, 1774–1795
https://doi.org/10.1214/24-AOS2416

LEARNING GAUSSIAN MIXTURES USING THE WASSERSTEIN–FISHER–RAO GRADIENT FLOW

BY YULING YAN[1,a], KAIZHENG WANG[2,b] AND PHILIPPE RIGOLLET[3,c]

[1]*Department of Statistics, University of Wisconsin-Madison,* [a]*yuling.yan@wisc.edu*
[2]*Department of IEOR and Data Science Institute, Columbia University,* [b]*kaizheng.wang@columbia.edu*
[3]*Department of Mathematics, MIT,* [c]*rigollet@math.mit.edu*

Gaussian mixture models form a flexible and expressive parametric family of distributions that has found a variety of applications. Unfortunately, fitting these models to data is a notoriously hard problem from a computational perspective. Currently, only moment-based methods enjoy theoretical guarantees while likelihood-based methods are dominated by heuristics such as Expectation-Maximization that are known to fail in simple examples. In this work, we propose a new algorithm to compute the nonparametric maximum likelihood estimator (NPMLE) in a Gaussian mixture model. Our method is based on gradient descent over the space of probability measures equipped with the Wasserstein–Fisher–Rao geometry for which we establish convergence guarantees. In practice, it can be approximated using an interacting particle system where the weight and location of particles are updated alternately. We conduct extensive numerical experiments to confirm the effectiveness of the proposed algorithm compared not only to classical benchmarks but also to similar gradient descent algorithms with respect to simpler geometries. In particular, these simulations illustrate the benefit of updating both weight and location of the interacting particles.

REFERENCES

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.

AMBROSIO, L., GIGLI, N. and SAVARÉ, G. (2008). *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed. *Lectures in Mathematics ETH Zürich*. Birkhäuser, Basel. MR2401600

BAUER, M., BRUVERIS, M. and MICHOR, P. W. (2016). Uniqueness of the Fisher–Rao metric on the space of smooth densities. *Bull. Lond. Math. Soc.* **48** 499–506. MR3509909 https://doi.org/10.1112/blms/bdw020

CHEWI, S. (2022). Log-concave sampling. Unpublished manuscript.

CHEWI, S., MAUNU, T., RIGOLLET, P. and STROMME, A. J. (2020). Gradient descent algorithms for Bures–Wasserstein barycenters. In *Conference on Learning Theory* 1276–1304. PMLR.

CHIZAT, L. and BACH, F. (2018). On the global convergence of gradient descent for over-parameterized models using optimal transport. *Adv. Neural Inf. Process. Syst.* **31**.

CHIZAT, L., PEYRÉ, G., SCHMITZER, B. and VIALARD, F.-X. (2018). An interpolating distance between optimal transport and Fisher–Rao metrics. *Found. Comput. Math.* **18** 1–44. MR3749413 https://doi.org/10.1007/s10208-016-9331-y

DICKER, L. H. and ZHAO, S. D. (2016). High-dimensional classification via nonparametric empirical Bayes and maximum likelihood inference. *Biometrika* **103** 21–34. MR3465819 https://doi.org/10.1093/biomet/asv067

DU, S. S., JIN, C., LEE, J. D., JORDAN, M. I., SINGH, A. and POCZOS, B. (2017). Gradient descent can take exponential time to escape saddle points. *Adv. Neural Inf. Process. Syst.* **30**.

GALLOUËT, T. O. and MONSAINGEON, L. (2017). A JKO splitting scheme for Kantorovich–Fisher–Rao gradient flows. *SIAM J. Math. Anal.* **49** 1100–1130. MR3628313 https://doi.org/10.1137/16M106666X

GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2008). The support reduction algorithm for computing non-parametric function estimates in mixture models. *Scand. J. Stat.* **35** 385–399. MR2446726 https://doi.org/10.1111/j.1467-9469.2007.00588.x

MSC2020 subject classifications. Primary 62G05; secondary 62H30.

Key words and phrases. Gaussian mixture model, nonparametric MLE, Wasserstein–Fisher–Rao geometry, optimal transport, Wasserstein gradient flows, overparameterization.

JEWELL, N. P. (1982). Mixtures of exponential distributions. *Ann. Statist.* **10** 479–484. MR0653523

JIANG, W. and ZHANG, C.-H. (2009). General maximum likelihood empirical Bayes estimation of normal means. *Ann. Statist.* **37** 1647–1684. MR2533467 https://doi.org/10.1214/08-AOS638

JIN, C., ZHANG, Y., BALAKRISHNAN, S., WAINWRIGHT, M. J. and JORDAN, M. I. (2016). Local maxima in the likelihood of Gaussian mixture models: Structural results and algorithmic consequences. *Adv. Neural Inf. Process. Syst.* **29** 4116–4124.

JORDAN, R., KINDERLEHRER, D. and OTTO, F. (1998). The variational formulation of the Fokker–Planck equation. *SIAM J. Math. Anal.* **29** 1–17. MR1617171 https://doi.org/10.1137/S0036141096303359

KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Stat.* **27** 887–906. MR0086464 https://doi.org/10.1214/aoms/1177728066

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

KONDRATYEV, S., MONSAINGEON, L. and VOROTNIKOV, D. (2016). A new optimal transport distance on the space of finite Radon measures. *Adv. Differential Equations* **21** 1117–1164. MR3556762

LAMBERT, M., CHEWI, S., BACH, F., BONNABEL, S. and RIGOLLET, P. (2022). Variational inference via Wasserstein gradient flows. *Adv. Neural Inf. Process. Syst.* **35** 14434–14447.

LEE, J. D., SIMCHOWITZ, M., JORDAN, M. I. and RECHT, B. (2016). Gradient descent only converges to minimizers. In *Conference on Learning Theory* 1246–1257. PMLR.

LIERO, M., MIELKE, A. and SAVARÉ, G. (2018). Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures. *Invent. Math.* **211** 969–1117. MR3763404 https://doi.org/10.1007/s00222-017-0759-8

LINDSAY, B. G. (1983). The geometry of mixture likelihoods: A general theory. *Ann. Statist.* **11** 86–94. MR0684866 https://doi.org/10.1214/aos/1176346059

LINDSAY, B. G. and ROEDER, K. (1993). Uniqueness of estimation and identifiability in mixture models. *Canad. J. Statist.* **21** 139–147. MR1234757 https://doi.org/10.2307/3315807

LU, Y., LU, J. and NOLEN, J. (2019). Accelerating Langevin sampling with birth-death. arXiv preprint. Available at arXiv:1905.09863.

LU, Y., SLEPČEV, D. and WANG, L. (2023). Birth-death dynamics for sampling: Global convergence, approximations and their asymptotics. *Nonlinearity* **36** 5731–5772. MR4646041 https://doi.org/10.1088/1361-6544/acf988

MEI, S., MONTANARI, A. and NGUYEN, P.-M. (2018). A mean field view of the landscape of two-layer neural networks. *Proc. Natl. Acad. Sci. USA* **115** E7665–E7671. MR3845070 https://doi.org/10.1073/pnas.1806579115

OTTO, F. (2001). The geometry of dissipative evolution equations: The porous medium equation. *Comm. Partial Differential Equations* **26** 101–174. MR1842429 https://doi.org/10.1081/PDE-100002243

POLYANSKIY, Y. and WU, Y. (2020). Self-regularizing property of nonparametric maximum likelihood estimator in mixture models. arXiv preprint. Available at arXiv:2008.08244.

SAHA, S. and GUNTUBOYINA, A. (2020). On the nonparametric maximum likelihood estimator for Gaussian location mixture densities with application to Gaussian denoising. *Ann. Statist.* **48** 738–762. MR4102674 https://doi.org/10.1214/19-AOS1817

SANDER, M. E., ABLIN, P., BLONDEL, M. and PEYRÉ, G. (2022). Sinkformers: Transformers with doubly stochastic attention. In *International Conference on Artificial Intelligence and Statistics* 3515–3530. PMLR.

SANTAMBROGIO, F. (2017). {Euclidean, metric, and Wasserstein} gradient flows: An overview. *Bull. Math. Sci.* **7** 87–154. MR3625852 https://doi.org/10.1007/s13373-017-0101-1

SOLOFF, J. A., GUNTUBOYINA, A. and SEN, B. (2024). Multivariate, heteroscedastic empirical Bayes via nonparametric maximum likelihood. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* qkae040.

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

YAN, Y., WANG, K. and RIGOLLET, P. (2024). Supplement to "Learning Gaussian mixtures using the Wasserstein–Fisher–Rao gradient flow." https://doi.org/10.1214/24-AOS2416SUPP

ZHANG, C.-H. (2009). Generalized maximum likelihood estimation of normal mixture densities. *Statist. Sinica* **19** 1297–1318. MR2536157

ZHANG, Y., CUI, Y., SEN, B. and TOH, K.-C. (2024). On efficient and scalable computation of the nonparametric maximum likelihood estimator in mixture models. *J. Mach. Learn. Res.* **25** Paper No. [8], 46. MR4723858

The Annals of Statistics
2024, Vol. 52, No. 4, 1796–1824
https://doi.org/10.1214/24-AOS2422

EFFICIENT AND MULTIPLY ROBUST RISK ESTIMATION UNDER GENERAL FORMS OF DATASET SHIFT

BY HONGXIANG QIU[1,a], ERIC TCHETGEN TCHETGEN[2,b] AND EDGAR DOBRIBAN[2,c]

[1]*Department of Epidemiology and Biostatistics, Michigan State University,* [a]*qiuhongx@msu.edu*
[2]*Department of Statistics and Data Science, the Wharton School, University of Pennsylvania,* [b]*ett@wharton.upenn.edu,* [c]*dobriban@wharton.upenn.edu*

Statistical machine learning methods often face the challenge of limited data available from the population of interest. One remedy is to leverage data from auxiliary source populations, which share some conditional distributions or are linked in other ways with the target domain. Techniques leveraging such *dataset shift* conditions are known as *domain adaptation* or *transfer learning*. Despite extensive literature on dataset shift, limited works address how to efficiently use the auxiliary populations to improve the accuracy of risk evaluation for a given machine learning task in the target population.

In this paper, we study the general problem of efficiently estimating target population risk under various dataset shift conditions, leveraging semiparametric efficiency theory. We consider a general class of dataset shift conditions, which includes three popular conditions—covariate, label and concept shift—as special cases. We allow for partially nonoverlapping support between the source and target populations. We develop efficient and multiply robust estimators along with a straightforward specification test of these dataset shift conditions. We also derive efficiency bounds for two other dataset shift conditions, posterior drift and location-scale shift. Simulation studies support the efficiency gains due to leveraging plausible dataset shift conditions.

REFERENCES

ANGELOPOULOS, A. N., BATES, S., CANDÈS, E. J., JORDAN, M. I. and LEI, L. (2021). Learn then test: Calibrating predictive algorithms to achieve risk control. Preprint. Available at arXiv:2110.01052v5.

ANGELOPOULOS, A. N., BATES, S., FANNJIANG, C., JORDAN, M. I. and ZRNIC, T. (2023). Prediction-powered inference. Preprint. Available at arXiv:2301.09633v1.

ANGRIST, J. D. and KRUEGER, A. B. (1992). The effect of age at school entry on educational attainment: An application of instrumental variables with moments from two samples. *J. Amer. Statist. Assoc.* **87** 328–336.

AZRIEL, D., BROWN, L. D., SKLAR, M., BERK, R., BUJA, A. and ZHAO, L. (2021). Semi-supervised linear regression. *J. Amer. Statist. Assoc.* **117** 2238–2251.

BALAKRISHNAN, S., KENNEDY, E. H. and WASSERMAN, L. (2023). The fundamental limits of structure-agnostic functional estimation. Preprint. Available at arXiv:2305.04116v1.

BASTANI, H. (2021). Predicting with proxies: Transfer learning in high dimension. *Manage. Sci.* **67** 2964–2984.

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

BHATTACHARYA, P. K., CHERNOFF, H. and YANG, S. S. (2007). Nonparametric estimation of the slope of a truncated regression. *Ann. Statist.* **11** 505–514. https://doi.org/10.1214/aos/1176346157

BICKEL, P., KLAASSEN, C. A., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins Univ. Press, Baltimore, MD.

BICKEL, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10** 647–671.

BOLTHAUSEN, E., PERKINS, E. and VAN DER VAART, A. (2002). *Lectures on Probability Theory and Statistics*: *Ecole D'Eté de Probabilités de Saint-Flour XXIX*-1999. *Lecture Notes in Math*. **1781**. Springer, Berlin, Heidelberg.

BROOKHART, M. A. and VAN DER LAAN, M. J. (2006). A semiparametric model selection criterion with applications to the marginal structural model. *Comput. Statist. Data Anal.* **50** 475–498.

MSC2020 subject classifications. 68Q32, 62G20.
Key words and phrases. Dataset shift, domain adaptation, efficiency, multiple robustness, semiparametric model, transfer learning.

CAI, T., LI, M. and LIU, M. (2022). Semi-supervised triply robust inductive transfer learning. Preprint. Available at arXiv:2209.04977.

CAI, T. T. and WEI, H. (2021). Transfer learning for nonparametric classification: Minimax rate and adaptive classifier. *Ann. Statist.* **49** 100–128.

CHAKRABORTTY, A. and CAI, T. (2018). Efficient and adaptive linear regression in semi-supervised settings. *Ann. Statist.* **46** 1541–1572.

CHATTERJEE, N., CHEN, Y. H., MAAS, P. and CARROLL, R. J. (2016). Constrained maximum likelihood estimation for model calibration using summary-level information from external big data sources. *J. Amer. Statist. Assoc.* **111** 107–117.

CHEN, T. and GUESTRIN, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 13–17 *August* 2016 785–794.

CHEN, X. and POUZO, D. (2015). Sieve Wald and QLR inferences on semi/nonparametric conditional moment models. *Econometrica* **83** 1013–1079.

CHEN, Y. H. and CHEN, H. (2000). A unified approach to regression analysis under double-sampling designs. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 449–460.

CHERNOZHUKOV, V., CHETVERIKOV, D., DEMIRER, M., DUFLO, E., HANSEN, C. and NEWEY, W. (2017). Double/debiased/Neyman machine learning of treatment effects. *Amer. Econ. Rev.* **107** 261–265.

CHERNOZHUKOV, V., CHETVERIKOV, D., DEMIRER, M., DUFLO, E., HANSEN, C., NEWEY, W. and ROBINS, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econom. J.* **21** C1–C68.

CHRISTODOULIDIS, S., ANTHIMOPOULOS, M., EBNER, L., CHRISTE, A. and MOUGIAKAKOU, S. (2017). Multisource transfer learning with convolutional neural networks for lung pattern analysis. *IEEE J. Biomed. Health Inform.* **21** 76–84. https://doi.org/10.1109/JBHI.2016.2636929

D'ORAZIO, M., DI ZIO, M. and SCANU, M. (2006). *Statistical Matching: Theory and Practice*. Wiley, Hoboken, NJ.

D'ORAZIO, M., DI ZIO, M. and SCANU, M. (2010). Old and new approaches in statistical matching when samples are drawn with complex survey designs. In *Proceedings of the* 45*th "Riunione Scientifica della Societa'Italiana di Statistica", Padova* 16–18.

DORAN, G., MUANDET, K., ZHANG, K. and SCHÖLKOPF, B. (2014). A permutation-based kernel conditional independence test. In *Uncertainty in Artificial Intelligence - Proceedings of the* 30*th Conference, UAI* 2014 132–141.

EVANS, K., SUN, B. L., ROBINS, J. and TCHETGEN TCHETGEN, E. J. (2021). Doubly robust regression analysis for data fusion. *Statist. Sinica* **31** 1285–1307.

FOSTER, D. J. and SYRGKANIS, V. (2023). Orthogonal statistical learning. *Ann. Statist.* **51** 879–908.

FRIEDMAN, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Statist.* **29** 1189–1232.

FRIEDMAN, J. H. (2002). Stochastic gradient boosting. *Comput. Statist. Data Anal.* **38** 367–378.

GRONSBELL, J., LIU, M., TIAN, L. and CAI, T. (2022). Efficient evaluation of prediction rules in semi-supervised settings under stratified sampling. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 1353–1391.

GU, T., HAN, Y. and DUAN, R. (2022). Robust angle-based transfer learning in high dimensions. Preprint. Available at arXiv:2210.12759.

GYÖRFI, L., KOHLER, M., KRZYŻAK, A. and WALK, H. (2002). *A Distribution-Free Theory of Nonparametric Regression. Springer Series in Statistics*. Springer New York, New York, NY. https://doi.org/10.1007/B97848

HAJEK, J. (1962). Asymptotically most powerful rank-order tests. *Ann. Math. Stat.* **33** 1124–1147.

HASTIE, T., BUJA, A. and TIBSHIRANI, R. (1995). Penalized discriminant analysis. *Ann. Statist.* **23** 73–102.

HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models*. CRC Press/CRC, Boca Raton, FL.

HAUSMAN, J. A. (1978). Specification tests in econometrics. *Econometrica* **46** 1251–1271.

HE, Y., LI, Q., HU, Q. and LIU, L. (2022). Transfer learning in high-dimensional semiparametric graphical models with application to brain connectivity analysis. *Stat. Med.* **41** 4112–4129.

HU, X. and LEI, J. (2023). A two-sample conditional distribution test using conformal prediction and weighted rank sum. *J. Amer. Statist. Assoc.* 1–19.

JEWELL, N. P. (1985). Least squares regression with data arising from stratified samples of the dependent variable. *Biometrika* **72** 11–21. https://doi.org/10.1093/biomet/72.1.11

KENNEDY, E. H. (2020). Towards optimal doubly robust estimation of heterogeneous causal effects. Preprint. Available at arXiv:2004.14497v3.

KOUW, W. M. and LOOG, M. (2018). An introduction to domain adaptation and transfer learning. Preprint. Available at arXiv:1812.11806.

LI, S., CAI, T. T. and LI, H. (2022). Transfer learning in large-scale Gaussian graphical models with false discovery rate control. *J. Amer. Statist. Assoc.* 1–13.

LI, S., GILBERT, P. B. and LUEDTKE, A. (2023). Data fusion using weakly aligned sources. Preprint. Available at arXiv:2308.14836v1.

LI, S. and LUEDTKE, A. (2023). Efficient estimation under data fusion. *Biometrika* **110** 1041–1054. https://doi.org/10.1093/biomet/asad007

LIPTON, Z., WANG, Y.-X. and SMOLA, A. (2018). Detecting and correcting for label shift with black box predictors. In *International Conference on Machine Learning* 3122–3130. PMLR.

LIU, L., MUKHERJEE, R. and ROBINS, J. M. (2020). On nearly assumption-free tests of nominal confidence interval coverage for causal parameters estimated by machine learning. *Statist. Sci.* **35** 518–539.

LIU, L., MUKHERJEE, R. and ROBINS, J. M. (2023). Can we falsify the justification of the validity of Wald confidence intervals of doubly robust functionals, without assumptions? Preprint. Available at arXiv:2306.10590v1.

LIU, M., ZHANG, Y. and CAI, T. (2020). Augmented transfer regression learning with semi-non-parametric nuisance models. Preprint. Available at arXiv:2010.02521.

LIU, Y., LIU, M., GUO, Z. and CAI, T. (2023). Surrogate-assisted federated learning of high dimensional electronic health record data. Preprint. Available at arXiv:2302.04970v1.

MA, C., PATHAK, R. and WAINWRIGHT, M. J. (2023). Optimally tackling covariate shift in Rkhs-based nonparametric regression. *Ann. Statist.* **51** 738–761. https://doi.org/10.1214/23-AOS2268

MANSKI, C. F. and LERMAN, S. R. (1977). The estimation of choice probabilities from choice based samples. *Econometrica* **45** 1977.

MASON, L., BAXTER, J., BARTLETT, P. L. and FREAN, M. (1999). Boosting algorithms as gradient descent. In *Advances in Neural Information Processing Systems* **12**.

MCGRATH, S. and MUKHERJEE, R. (2022). On undersmoothing and sample splitting for estimating a doubly robust functional. Preprint. Available at arXiv:2212.14857v1.

MORENO-TORRES, J. G., RAEDER, T., ALAIZ-RODRÍGUEZ, R., CHAWLA, N. V. and HERRERA, F. (2012). A unifying view on dataset shift in classification. *Pattern Recognit.* **45** 521–530.

NEWEY, W. K. (1985). Generalized method of moments specification testing. *J. Econometrics* **29** 229–256.

NEWEY, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica* **62** 1349.

NIE, X. and WAGER, S. (2021). Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika* **108** 299–319.

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

PARK, S., DOBRIBAN, E., LEE, I. and BASTANI, O. (2022). PAC prediction sets under covariate shift. In *International Conference on Learning Representations*.

PARK, S., LI, S., LEE, I. and BASTANI, O. (2020). PAC confidence predictions for deep neural network classifiers. Preprint. Available at arXiv:2011.00716.

PATHAK, R., MA, C. and WAINWRIGHT, M. J. (2022). A new similarity measure for covariate shift with applications to nonparametric regression. In *Proceedings of Machine Learning Research* **162** 17517–17530. PMLR.

PFANZAGL, J. (1985). *Contributions to a General Asymptotic Statistical Theory*. *Lecture Notes in Statistics* **3**. Springer New York, New York, NY.

PFANZAGL, J. (1990). *Estimation in Semiparametric Models*. *Lecture Notes in Statistics* **63**. Springer, New York, NY.

POLO, F. M., IZBICKI, R., LACERDA, E. G., IBIETA-JIMENEZ, J. P. and VICENTE, R. (2022). A unified framework for dataset shift diagnostics. Preprint. Available at arXiv:2205.08340.

QIU, H., DOBRIBAN, E. and TCHETGEN TCHETGEN, E. (2022). Prediction sets adaptive to unknown covariate shift. Preprint. Available at arXiv:2203.06126v5.

QIU, H., TCHETGEN TCHETGEN, E. and DOBRIBAN, E. (2024). Supplement to "Efficient and multiply robust risk estimation under general forms of dataset shift." https://doi.org/10.1214/24-AOS2422SUPP

RÄSSLER, S. (2012). *Statistical Matching*: *A Frequentist Theory*, *Practical Applications*, *and Alternative Bayesian Approaches*. *Lecture Notes in Statistics* **168**. Springer, New York, NY.

ROBINS, J. M., HSIEH, F. and NEWEY, W. (1995). Semiparametric efficient estimation of a conditional density with missing or mismeasured covariates. *J. Roy. Statist. Soc. Ser. B* **57** 409–424.

ROTNITZKY, A., FARAGGI, D. and SCHISTERMAN, E. (2006). Doubly robust estimation of the area under the receiver-operating characteristic curve in the presence of verification bias. *J. Amer. Statist. Assoc.* **101** 1276–1288.

RUBIN, D. and VAN DER LAAN, M. J. (2007). A doubly robust censoring unbiased transformation. *Int. J. Biostat.* **3** Article 4. https://doi.org/10.2202/1557-4679.1052

SCHICK, A. (1986). On asymptotically efficient estimation in semiparametric models. *Ann. Statist.* **14** 1139–1151.

SCHÖLKOPF, B., JANZING, D., PETERS, J., SGOURITSA, E., ZHANG, K. and MOOIJ, J. M. (2012). On causal and anticausal learning. In *ICML*.

SCOTT, C. (2019). A generalized Neyman-Pearson criterion for optimal domain adaptation. In *Proceedings of the 30th International Conference on Algorithmic Learning Theory* **98** 738–761. PMLR.

SHAH, R. D. and PETERS, J. (2020). The hardness of conditional independence testing and the generalised covariance measure. *Ann. Statist.* **48** 1514–1538.

SHIMODAIRA, H. (2000). Improving predictive inference under covariate shift by weighting the log-likelihood function. *J. Statist. Plann. Inference* **90** 227–244.

STORKEY, A. (2013). When training and test sets are different: Characterizing learning transfer. In *Dataset Shift in Machine Learning* 2–28. MIT Press, Cambridge.

SUGIYAMA, M. and KAWANABE, M. (2012). *Machine Learning in Non-stationary Environments*: *Introduction to Covariate Shift Adaptation*. MIT Press, Cambridge, MA.

SUGIYAMA, M., KRAULEDAT, M. and MULLER, K.-R. (2007). Covariate shift adaptation by importance weighted cross validation. *J. Mach. Learn. Res.* **8** 985–1005.

TASCHE, D. (2017). Fisher consistency for prior probability shift. *J. Mach. Learn. Res.* **18** 1–32.

TCHETGEN TCHETGEN, E. J. (2009). A commentary on G. Molenberghs's review of missing data methods. *Drug Inf. J.* **43** 433–435.

TIAN, Y. and FENG, Y. (2022). Transfer learning under high-dimensional generalized linear models. *J. Amer. Statist. Assoc.* 1–14.

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

VANSTEELANDT, S. and DUKES, O. (2022). Assumption-lean inference for generalised linear model parameters. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 657–685.

VANSTEELANDT, S., ROTNITZKY, A. and ROBINS, J. (2007). Estimation of regression models for the mean of repeated outcomes under nonignorable nonmonotone nonresponse. *Biometrika* **94** 841–860. https://doi.org/10.1093/biomet/asm070

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

VAN DER LAAN, M. J., POLLEY, E. C. and HUBBARD, A. E. (2007). Super learner. *Stat. Appl. Genet. Mol. Biol.* **6** Article 25. https://doi.org/10.2202/1544-6115.1309

VAN DER LAAN, M. J. and ROSE, S. (2018). *Targeted Learning in Data Science*: *Causal Inference for Complex Longitudinal Studies*. Springer, New York, NY.

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge Univ. Press, Cambridge, UK.

VAN DER VAART, A. W., DUDOIT, S. and VAN DER LAAN, M. J. (2006). Oracle inequalities for multi-fold cross validation. *Statist. Decisions* **24** 351–371.

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

VAPNIK, V. (1992). Principles of risk minimization for learning theory. *Adv. Neural Inf. Process. Syst.* **4** 831–838.

VOVK, V. (2013). Conditional validity of inductive conformal predictors. *Mach. Learn.* **92** 349–376.

YANG, Y., KUCHIBHOTLA, A. K. and TCHETGEN TCHETGEN, E. (2024). Doubly robust calibration of prediction sets under covariate shift. *Journal of the Royal Statistical Society Series B*: *Statistical Methodology* **86** 943–965. https://doi.org/10.1093/jrsssb/qkae009

YUVAL, O. and ROSSET, S. (2023). Mixed semi-supervised generalized-linear-regression with applications to deep learning. Preprint. Available at arXiv:2302.09526v1.

ZHANG, A., BROWN, L. D. and CAI, T. T. (2019). Semi-supervised inference: General theory and estimation of means. *Ann. Statist.* **47** 2538–2566.

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

ZHANG, K., SCHÖLKOPF, B., MUANDET, K. and WANG, Z. (2013). Domain adaptation under target and conditional shift. In *International Conference on Machine Learning* 819–827. PMLR.

ZHANG, X., BLANCHET, J., GHOSH, S. and SQUILLANTE, M. S. (2022). A class of geometric structures in transfer learning: Minimax bounds and optimality. In *International Conference on Artificial Intelligence and Statistics* 3794–3820. PMLR.

ZHANG, Y. and BRADIC, J. (2022). High-dimensional semi-supervised learning: In search of optimal inference of the mean. *Biometrika* **109** 387–403.

ZHANG, Y., CHAKRABORTTY, A. and BRADIC, J. (2021). Double robust semi-supervised inference for the mean: Selection bias under mar labeling with decaying overlap. Preprint. Available at arXiv:2104.06667.

ZHOU, D., LIU, M., LI, M. and CAI, T. (2022). Doubly robust augmented model accuracy transfer inference with high dimensional features. Preprint. Available at arXiv:2208.05134.

The Annals of Statistics
2024, Vol. 52, No. 4, 1825–1844
https://doi.org/10.1214/24-AOS2427
This research was funded, in whole or in part, by ERC Advanced Grant (Horizon Europe UKRI G116786), Qatar National Fund NPRP grant #S-0207-200290 and KAUST Office of Sponsored Research Award No. OSR-2020-CRG9-4336.

ON POSTERIOR CONSISTENCY OF DATA ASSIMILATION WITH GAUSSIAN PROCESS PRIORS: THE 2D-NAVIER–STOKES EQUATIONS

BY RICHARD NICKL[1,a] AND EDRISS S. TITI[2,b]

[1]*Department of Pure Mathematics and Mathematical Statistics, University of Cambridge,* [a]*nickl@maths.cam.ac.uk*
[2]*Department of Applied Mathematics and Theoretical Physics, University of Cambridge,* [b]*Edriss.Titi@maths.cam.ac.uk*

We consider a nonlinear Bayesian data assimilation model for the periodic two-dimensional Navier–Stokes equations with initial condition modelled by a Gaussian process prior. We show that if the system is updated with sufficiently many discrete noisy measurements of the velocity field, then the posterior distribution eventually concentrates near the ground truth solution of the time evolution equation, and in particular that the initial condition is recovered consistently by the posterior mean vector field. We further show that the convergence rate can in general not be faster than inverse logarithmic in sample size, but describe specific conditions on the initial conditions when faster rates are possible. In the proofs, we provide an explicit quantitative estimate for backward uniqueness of solutions of the two-dimensional Navier–Stokes equations.

REFERENCES

[1] ABRAHAM, K. and NICKL, R. (2019). On statistical Calderón problems. *Math. Stat. Learn.* **2** 165–216. MR4130599

[2] BARDOS, C. and TARTAR, L. (1973). Sur l'unicité rétrograde des équations paraboliques et quelques questions voisines. *Arch. Ration. Mech. Anal.* **50** 10–25. MR0338517 https://doi.org/10.1007/BF00251291

[3] BENNETT, A. F. (2002). *Inverse Modeling of the Ocean and Atmosphere*. Cambridge Univ. Press, Cambridge. MR1920432 https://doi.org/10.1017/CBO9780511535895

[4] BESKOS, A., GIROLAMI, M., LAN, S., FARRELL, P. E. and STUART, A. M. (2017). Geometric MCMC for infinite-dimensional inverse problems. *J. Comput. Phys.* **335** 327–351. MR3612501 https://doi.org/10.1016/j.jcp.2016.12.041

[5] BOHR, J. (2021). Stability of the non-Abelian X-ray transform in dimension ≥ 3. *J. Geom. Anal.* **31** 11226–11269. MR4310169 https://doi.org/10.1007/s12220-021-00679-0

[6] BOHR, J. and NICKL, R. (2023). On log-concave approximations of high-dimensional posterior measures and stability properties in non-linear inverse problems. *Ann. Inst. Henri Poincaré B, Probab. Stat.*. (to appear).

[7] BORGGAARD, J., GLATT-HOLTZ, N. and KROMETIS, J. (2020). On Bayesian consistency for flows observed through a passive scalar. *Ann. Appl. Probab.* **30** 1762–1783. MR4133382 https://doi.org/10.1214/19-AAP1542

[8] CONSTANTIN, P. and FOIAS, C. (1988). *Navier-Stokes Equations*. *Chicago Lectures in Mathematics*. Univ. Chicago Press, Chicago, IL. MR0972259

[9] COTTER, S. L., DASHTI, M., ROBINSON, J. C. and STUART, A. M. (2009). Bayesian inverse problems for functions and applications to fluid mechanics. *Inverse Probl.* **25** 115008, 43. MR2558668 https://doi.org/10.1088/0266-5611/25/11/115008

[10] COTTER, S. L., ROBERTS, G. O., STUART, A. M. and WHITE, D. (2013). MCMC methods for functions: Modifying old algorithms to make them faster. *Statist. Sci.* **28** 424–446. MR3135540 https://doi.org/10.1214/13-STS421

[11] CUI, T., LAW, K. J. H. and MARZOUK, Y. M. (2016). Dimension-independent likelihood-informed MCMC. *J. Comput. Phys.* **304** 109–137. MR3422405 https://doi.org/10.1016/j.jcp.2015.10.008

[12] EVANS, L. C. (2010). *Partial Differential Equations*, 2nd ed. *Graduate Studies in Mathematics* **19**. Amer. Math. Soc., Providence, RI. MR2597943 https://doi.org/10.1090/gsm/019

[13] EVENSEN, G., VOSSEPOEL, F. C. and VAN LEEUWEN, J. (2022). *Data Assimilation Fundamentals*. *Springer Textbooks in Earth Sciences*, *Geography and Environment*. Springer, Cham.

MSC2020 subject classifications. Primary 35Q30; secondary 62G20.
Key words and phrases. Bayesian inference for the initial condition, nonlinear dynamical system.

[14] Foias, C. and Saut, J.-C. (1984). Asymptotic behavior, as $t \to +\infty$, of solutions of Navier–Stokes equations and nonlinear spectral manifolds. *Indiana Univ. Math. J.* **33** 459–477. MR0740960 https://doi.org/10.1512/iumj.1984.33.33025

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

[16] Giné, E. and Nickl, R. (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

[17] Giordano, M. and Nickl, R. (2020). Consistency of Bayesian inference with Gaussian process priors in an elliptic inverse problem. *Inverse Probl.* **36** 085001, 35. MR4151406 https://doi.org/10.1088/1361-6420/ab7d2a

[18] Giordano, M. and Ray, K. (2022). Nonparametric Bayesian inference for reversible multidimensional diffusions. *Ann. Statist.* **50** 2872–2898. MR4500628 https://doi.org/10.1214/22-aos2213

[19] Hairer, M., Stuart, A. M. and Vollmer, S. J. (2014). Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions. *Ann. Appl. Probab.* **24** 2455–2490. MR3262508 https://doi.org/10.1214/13-AAP982

[20] Hoffmann, M. and Ray, K. (2022). Bayesian estimation in a multidimensional diffusion model with high frequency data. arXiv.

[21] Kalnay, E. (2003). *Atmospheric Modelling*, *Data Assimilation*, *and Predictability*. Cambridge Univ. Press, Cambridge, UK.

[22] Kekkonen, H. (2022). Consistency of Bayesian inference with Gaussian process priors for a parabolic inverse problem. *Inverse Probl.* **38** Paper No. 035002, 29. MR4385425 https://doi.org/10.1088/1361-6420/ac4839

[23] Knapik, B. T., van der Vaart, A. W. and van Zanten, J. H. (2011). Bayesian inverse problems with Gaussian priors. *Ann. Statist.* **39** 2626–2657. MR2906881 https://doi.org/10.1214/11-AOS920

[24] Knapik, B. T., van der Vaart, A. W. and van Zanten, J. H. (2013). Bayesian recovery of the initial condition for the heat equation. *Comm. Statist. Theory Methods* **42** 1294–1313. MR3031282 https://doi.org/10.1080/03610926.2012.681417

[25] Law, K., Stuart, A. and Zygalakis, K. (2015). *Data Assimilation*: *A Mathematical Introduction*. *Texts in Applied Mathematics* **62**. Springer, Cham. MR3363508 https://doi.org/10.1007/978-3-319-20325-6

[26] Lorenz, E. N. (1963). Deterministic nonperiodic flow. *J. Atmos. Sci.* **20** 130–141. MR4021434 https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2

[27] Majda, A. J. and Harlim, J. (2012). *Filtering Complex Turbulent Systems*. Cambridge Univ. Press, Cambridge. MR2934167 https://doi.org/10.1017/CBO9781139061308

[28] Monard, F., Nickl, R. and Paternain, G. P. (2021). Consistent inversion of noisy non-Abelian X-ray transforms. *Comm. Pure Appl. Math.* **74** 1045–1099. MR4230066 https://doi.org/10.1002/cpa.21942

[29] Monard, F., Nickl, R. and Paternain, G. P. (2021). Statistical guarantees for Bayesian uncertainty quantification in nonlinear inverse problems with Gaussian process priors. *Ann. Statist.* **49** 3255–3298. MR4352530 https://doi.org/10.1214/21-aos2082

[30] Nickl, R. (2020). Bernstein–von Mises theorems for statistical inverse problems I: Schrödinger equation. *J. Eur. Math. Soc.* (*JEMS*) **22** 2697–2750. MR4118619 https://doi.org/10.4171/JEMS/975

[31] Nickl, R. (2024). Inference for diffusions from low frequency measurements. *Ann. Statist.* **52** 519–549. MR4744186

[32] Nickl, R. (2023). *Bayesian Non-linear Statistical Inverse Problems*. *Zurich Lectures in Advanced Mathematics*. EMS Press, Berlin. MR4604099 https://doi.org/10.4171/zlam/30

[33] Nickl, R. and Ray, K. (2020). Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions. *Ann. Statist.* **48** 1383–1408. MR4124327 https://doi.org/10.1214/19-AOS1851

[34] Nickl, R., van de Geer, S. and Wang, S. (2020). Convergence rates for penalized least squares estimators in PDE constrained regression problems. *SIAM/ASA J. Uncertain. Quantificat.* **8** 374–413. MR4074017 https://doi.org/10.1137/18M1236137

[35] Nickl, R. and Wang, S. (2024). On polynomial-time computation of high-dimensional posterior measures by Langevin-type algorithms. *J. Eur. Math. Soc.* (*JEMS*) **26** 1031–1112. MR4721029 https://doi.org/10.4171/jems/1304

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

[37] Reich, S. and Cotter, C. (2015). *Probabilistic Forecasting and Bayesian Data Assimilation*. Cambridge Univ. Press, New York. MR3242790 https://doi.org/10.1017/CBO9781107706804

[38] ROBINSON, J. C. (2001). *Infinite-Dimensional Dynamical Systems*: *An Introduction to Dissipative Parabolic PDEs and the Theory of Global Attractors*. *Cambridge Texts in Applied Mathematics*. Cambridge Univ. Press, Cambridge. MR1881888 https://doi.org/10.1007/978-94-010-0732-0

[39] SCHWAB, C. and STUART, A. M. (2012). Sparse deterministic approximation of Bayesian inverse problems. *Inverse Probl*. **28** 045003, 32. MR2903278 https://doi.org/10.1088/0266-5611/28/4/045003

[40] ST-AMANT, S. (2022). Stability estimate for the broken non-Abelian X-ray transform in Minkowski space. *Inverse Probl*. **38** Paper No. 105007, 36. MR4482473

[41] STUART, A. M. (2010). Inverse problems: A Bayesian perspective. *Acta Numer.* **19** 451–559. MR2652785 https://doi.org/10.1017/S0962492910000061

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

The Annals of Statistics
2024, Vol. 52, No. 4, 1845–1871
https://doi.org/10.1214/24-AOS2428

HOW DO NOISE TAILS IMPACT ON DEEP RELU NETWORKS?

BY JIANQING FAN[1,a], YIHONG GU[1,b] AND WEN-XIN ZHOU[2,c]

[1]*Department of Operations Research and Financial Engineering, Princeton University,* [a]*jqfan@princeton.edu,* [b]*yihongg@princeton.edu*

[2]*Department of Information and Decision Sciences, University of Illinois at Chicago,* [c]*wenxinz@uic.edu*

This paper investigates the stability of deep ReLU neural networks for nonparametric regression under the assumption that the noise has only a finite pth moment. We unveil how the optimal rate of convergence depends on p, the degree of smoothness and the intrinsic dimension in a class of nonparametric regression functions with hierarchical composition structure when both the adaptive Huber loss and deep ReLU neural networks are used. This optimal rate of convergence cannot be obtained by the ordinary least squares but can be achieved by the Huber loss with a properly chosen parameter that adapts to the sample size, smoothness, and moment parameters. A concentration inequality for the adaptive Huber ReLU neural network estimators with allowable optimization errors is also derived. To establish a matching lower bound within the class of neural network estimators using the Huber loss, we employ a different strategy from the traditional route: constructing a deep ReLU network estimator that has a better empirical loss than the true function and the difference between these two functions furnishes a low bound. This step is related to the Huberization bias, yet more critically to the approximability of deep ReLU networks. As a result, we also contribute some new results on the approximation theory of deep ReLU neural networks.

REFERENCES

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

[2] BARRON, A. R. (1994). Approximation and estimation bounds for artificial neural networks. *Mach. Learn.* **14** 115–133.

[3] 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

[4] BARTLETT, P. L. and LONG, P. M. (2021). Failures of model-dependent generalization bounds for least-norm interpolation. *J. Mach. Learn. Res.* **22** Paper No. 204, 15. MR4329783

[5] 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

[6] BEATON, A. E. and TUKEY, J. W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics* **16** 147–185.

[7] BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. and STONE, C. J. (2017). *Classification and Regression Trees*. *Wadsworth Statistics/Probability Series*. Wadsworth, Belmont, CA.

[8] 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

[9] CLEVELAND, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* **74** 829–836. MR0556476

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

[11] 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

MSC2020 subject classifications. Primary 62G08; secondary 62G35.

Key words and phrases. Robustness, truncation, heavy tails, optimal rates, approximablility of ReLU networks, composition of functions.

[12] DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence. III. *Ann. Statist.* **19** 668–701.

[13] FAN, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.* **87** 998–1004. MR1209561

[14] FAN, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.* **21** 196–216. MR1212173 https://doi.org/10.1214/aos/1176349022

[15] FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications. Monographs on Statistics and Applied Probability* **66**. CRC Press, London. MR1383587

[16] FAN, J. and GU, Y. (2023). Factor augmented sparse throughput deep ReLU neural networks for high dimensional regression. *J. Amer. Statist. Assoc.* 1–28. https://doi.org/10.1080/01621459.2023.2271605

[17] FAN, J., GU, Y. and ZHOU, W.-X. (2024). Supplement to "How do noise tails impact on deep ReLU networks?." https://doi.org/10.1214/24-AOS2428SUPP

[18] 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

[19] 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

[20] 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

[21] FRIEDMAN, J. H. (1991). Multivariate adaptive regression splines. *Ann. Statist.* **19** 1–67. MR1091842 https://doi.org/10.1214/aos/1176347963

[22] GYÖRFI, L., KOHLER, M., KRZYŻAK, A. and WALK, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. Springer, Berlin.

[23] HAAS, M., HOLZMÜLLER, D., LUXBURG, U. and STEINWART, I. (2023). Mind the spikes: Benign overfitting of kernels and neural networks in fixed dimension. *Adv. Neural Inf. Process. Syst.* **36** 20763–20826.

[24] HABERMAN, S. J. (1977). Maximum likelihood estimates in exponential response models. *Ann. Statist.* **5** 815–841. MR0501540

[25] 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

[26] HAN, Q. and WELLNER, J. A. (2018). Robustness of shape-restricted regression estimators: An envelope perspective. arXiv preprint. Available at arXiv:1805.02542.

[27] HAN, Q. and WELLNER, J. A. (2019). Convergence rates of least squares regression estimators with heavy-tailed errors. *Ann. Statist.* **47** 2286–2319. MR3953452 https://doi.org/10.1214/18-AOS1748

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

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

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

[31] 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

[32] 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

[33] KUCHIBHOTLA, A. K. and PATRA, R. K. (2022). On least squares estimation under heteroscedastic and heavy-tailed errors. *Ann. Statist.* **50** 277–302. MR4382017 https://doi.org/10.1214/21-aos2105

[34] KUZBORSKIJ, I. and SZEPESVÁRI, C. (2022). Learning Lipschitz functions by GD-trained shallow overparameterized ReLU neural networks. ArXiv preprint. Available at arXiv:2212.13848.

[35] LECUN, Y., BENGIO, Y. and HINTON, G. (2015). Deep learning. *Nature* **521** 436–444. https://doi.org/10.1038/nature14539

[36] LEDERER, J. (2020). Risk bounds for robust deep learning. ArXiv preprint. Available at arXiv:2009.06202.

[37] 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

[38] 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

[39] MADRID PADILLA, O. H., TANSEY, W. and CHEN, Y. (2022). Quantile regression with ReLU networks: Estimators and minimax rates. *J. Mach. Learn. Res.* **23** Paper No. [247], 42. MR4577686

[40] MCCAFFREY, D. F. and GALLANT, A. R. (1994). Convergence rates for single hidden layer feedforward networks. *Neural Netw.* **7** 147–158.

[41] NADARAYA, E. A. (1964). On estimating regression. *Theory Probab. Appl.* **9** 141–142.

[42] 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

[43] SHEN, G., JIAO, Y., LIN, Y., HOROWITZ, J. L. and HUANG, J. (2021). Deep quantile regression: Mitigating the curse of dimensionality through composition. ArXiv preprint. Available at arXiv:2107.04907.

[44] SHEN, G., JIAO, Y., LIN, Y. and HUANG, J. (2021). Robust nonparametric regression with deep neural networks. ArXiv preprint. Available at arXiv:2107.10343.

[45] 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

[46] 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

[47] 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

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

[49] TELGARSKY, M. (2016). Benefits of depth in neural networks. *Proc. Mach. Learn. Res.* **49** 1517–1539.

[50] 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

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

[52] 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

[53] WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26** 359–372. MR0185765

[54] YANG, Z., BAI, Y. and MEI, S. (2021). Exact gap between generalization error and uniform convergence in random feature models. In *International Conference on Machine Learning* 11704–11715. PMLR.

[55] 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

[56] YAROTSKY, D. (2018). Optimal approximation of continuous functions by very deep ReLU networks. *Proc. Mach. Learn. Res.* **75** 639–649.

[57] ZHONG, Q., MUELLER, J. and WANG, J.-L. (2022). Deep learning for the partially linear Cox model. *Ann. Statist.* **50** 1348–1375. MR4441123 https://doi.org/10.1214/21-aos2153