

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

Continuous-time targeted minimum loss-based estimation of intervention-specific mean outcomes	HELENE C. RYTGAARD, THOMAS A. GERDS AND MARK J. VAN DER LAAN	2469
Central limit theorem and bootstrap approximation in high dimensions: Near $1/\sqrt{n}$ rates via implicit smoothing	MILES E. LOPES	2492
Testing goodness-of-fit and conditional independence with approximate co-sufficient sampling	RINA FOYGEL BARBER AND LUCAS JANSON	2514
New Edgeworth-type expansions with finite sample guarantees	MAYYA ZHILOVA	2545
Improved central limit theorem and bootstrap approximations in high dimensions VICTOR CHERNOZHUKOV, DENIS CHETVERIKOV, KENGO KATO AND YUTA KOIKE		2562
Bounds on the conditional and average treatment effect with unobserved confounding factors	STEVE YADLOWSKY, HONGSEOK NAMKOONG, SANJAY BASU, JOHN DUCHI AND LU TIAN	2587
Affine-equivariant inference for multivariate location under L_p loss functions ALEXANDER DÜRRE AND DAVY PAINDAVEINE		2616
Is infinity that far? A Bayesian nonparametric perspective of finite mixture models RAFFAELE ARGIENTO AND MARIA DE IORIO		2641
Global and individualized community detection in inhomogeneous multilayer networks SHUXIAO CHEN, SIFAN LIU AND ZONGMING MA		2664
Asymptotic analysis of synchrosqueezing transform—toward statistical inference with nonlinear-type time-frequency analysis	MATT SOURISSEAU, HAU-TIENG WU AND ZHOU ZHOU	2694
Approximate kernel PCA: Computational versus statistical trade-off BHARATH K. SRIPERUMBUDUR AND NICHOLAS STERGE		2713
Relaxing the Gaussian assumption in shrinkage and SURE in high dimension MAX FATHI, LARRY GOLDSTEIN, GESINE REINERT AND ADRIEN SAUMARD		2737
Nonparametric bivariate density estimation for censored lifetimes . . .	SAM EFROMOVICH	2767
Linear biomarker combination for constrained classification YIJIAN HUANG AND MARTIN G. SANDA		2793
The interpolation phase transition in neural networks: Memorization and generalization under lazy training	ANDREA MONTANARI AND YIQIAO ZHONG	2816
A sieve stochastic gradient descent estimator for online nonparametric regression in Sobolev ellipsoids	TIANYU ZHANG AND NOAH SIMON	2848
Nonparametric Bayesian inference for reversible multidimensional diffusions MATTEO GIORDANO AND KOLYAN RAY		2872
Scalable estimation and inference for censored quantile regression process XUMING HE, XIAOOU PAN, KEAN MING TAN AND WEN-XIN ZHOU		2899
A study of orthogonal array-based designs under a broad class of space-filling criteria GUANZHOU CHEN AND BOXIN TANG		2925
Two-sample testing of high-dimensional linear regression coefficients via complementary sketching	FENGAN GAO AND TENGYAO WANG	2950

continued

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles—*Continued from front cover*

- Nonparametric regression on Lie groups with measurement errors
JEONG MIN JEON, BYEONG U. PARK AND INGRID VAN KEILEGOM 2973
- Locally associated graphical models and mixed convex exponential families
STEFFEN LAURITZEN AND PIOTR ZWIERNIK 3009
- Estimation of time series models using residuals dependence measures
CARLOS VELASCO 3039
- Rate-optimal cluster-randomized designs for spatial interference . . . MICHAEL P. LEUNG 3064

Correction note

- “Asymptotic spectral theory for nonlinear time series”
YI ZHANG, XIAOFENG SHAO AND WEIBIAO WU 3088

THE ANNALS OF STATISTICS

Vol. 50, No. 5, pp. 2469–3089 October 2022

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

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

IMS OFFICERS

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

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

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

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

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

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

IMS EDITORS

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

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

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

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

Statistical Science. *Editor:* Sonia Petrone, Department of Decision Sciences, Università Bocconi, 20100 Milano MI, Italy

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

The Annals of Statistics [ISSN 0090-5364 (print); ISSN 2168-8966 (online)], Volume 50, Number 5, October 2022. 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.

CONTINUOUS-TIME TARGETED MINIMUM LOSS-BASED ESTIMATION OF INTERVENTION-SPECIFIC MEAN OUTCOMES

BY HELENE C. RYTGAARD^{1,a}, THOMAS A. GERDS^{1,b} AND MARK J. VAN DER LAAN^{2,c}

¹Section of Biostatistics, University of Copenhagen, ^ahely@sund.ku.dk, ^btagteam@sund.ku.dk

²Division of Biostatistics and Center for Targeted Machine Learning and Causal Inference, University of California, Berkeley, ^claan@berkeley.edu

This paper generalizes the targeted minimum loss-based estimation (TMLE) framework to allow for estimating the effects of time-varying interventions in settings where both interventions, covariates, and outcome can happen at subject-specific time-points on an arbitrarily fine time-scale. TMLE is a general template for constructing asymptotically linear substitution estimators for smooth low-dimensional parameters in infinite-dimensional models. Existing longitudinal TMLE methods are developed for data where observations are made on a discrete time-grid.

We consider a continuous-time counting process model where intensity measures track the monitoring of subjects, and focus on a low-dimensional target parameter defined as the intervention-specific mean outcome at the end of follow-up. To construct our TMLE algorithm for the given statistical estimation problem, we derive an expression for the efficient influence curve and represent the target parameter as a functional of intensities and conditional expectations. The high-dimensional nuisance parameters of our model are estimated and updated in an iterative manner according to separate targeting steps for the involved intensities and conditional expectations.

The resulting estimator solves the efficient influence curve equation. We state a general efficiency theorem and describe a highly adaptive lasso estimator for nuisance parameters that allows us to establish asymptotic linearity and efficiency of our estimator under minimal conditions on the underlying statistical model.

REFERENCES

- ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer Series in Statistics. Springer, New York. MR1198884 <https://doi.org/10.1007/978-1-4612-4348-9>
- ANDERSEN, J. T., PETERSEN, M., JIMENEZ-SOLE, E., BROEDBAEK, K., ANDERSEN, E. W., ANDERSEN, N. L., AFZAL, S., TORP-PEDERSEN, C., KEIDING, N. et al. (2013). Trimethoprim use in early pregnancy and the risk of miscarriage: A register-based nationwide cohort study. *Epidemiol. Infect.* **141** 1749–1755.
- BANG, H. and ROBINS, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61** 962–972. MR2216189 <https://doi.org/10.1111/j.1541-0420.2005.00377.x>
- BIBAUT, A. F. and VAN DER LAAN, M. J. (2019). July. Fast rates for empirical risk minimization over càdlàg functions with bounded sectional variation norm. Available at [arXiv:1907.09244](https://arxiv.org/abs/1907.09244).
- 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
- CHAKRABORTY, B. and MOODIE, E. E. M. (2013). *Statistical Methods for Dynamic Treatment Regimes: Reinforcement Learning, Causal Inference, and Personalized Medicine*. Statistics for Biology and Health. Springer, New York. MR3112454 <https://doi.org/10.1007/978-1-4614-7428-9>
- DAWID, A. P. and DIDELEZ, V. (2010). Identifying the consequences of dynamic treatment strategies: A decision-theoretic overview. *Stat. Surv.* **4** 184–231. MR2740837 <https://doi.org/10.1214/10-SS081>

MSC2020 subject classifications. 62G05.

Key words and phrases. Targeted minimum loss-based estimation (TMLE), time-varying confounding, continuous-time interventions, semiparametric model, efficient estimation, causal inference.

- GILL, R. D. (1994). Lectures on survival analysis. In *Lectures on Probability Theory (Saint-Flour, 1992). Lecture Notes in Math.* **1581** 115–241. Springer, Berlin. [MR1307414](#) <https://doi.org/10.1007/BFb0073873>
- GILL, R. D. and JOHANSEN, S. (1990). A survey of product-integration with a view toward application in survival analysis. *Ann. Statist.* **18** 1501–1555. [MR1074422](#) <https://doi.org/10.1214/aos/1176347865>
- GILL, R. D. and ROBINS, J. M. (2001). Causal inference for complex longitudinal data: The continuous case. *Ann. Statist.* **29** 1785–1811. [MR1891746](#) <https://doi.org/10.1214/aos/1015345962>
- GILL, R. D., VAN DER LAAN, M. J. and WELLNER, J. A. (1995). Inefficient estimators of the bivariate survival function for three models. *Ann. Inst. Henri Poincaré Probab. Stat.* **31** 545–597. [MR1338452](#)
- HERNÁN, M. A. (2010). The hazards of hazard ratios. *Epidemiology* **21** 13.
- HERNÁN, M. A. and ROBINS, J. M. (2020). *Causal Inference*. CRC Press/CRC, Boca Raton, FL.
- HERNÁN, M. A., LANOY, E., COSTAGLIOLA, D. and ROBINS, J. M. (2006). Comparison of dynamic treatment regimes via inverse probability weighting. *Basic Clin. Pharmacol. Toxicol.* **98** 237–242.
- KARIM, M. E., GUSTAFSON, P., PETKAU, J., TREMLETT, H. and LONG-TERM BENEFITS AND ADVERSE EFFECTS OF BETA-INTERFERON FOR MULTIPLE SCLEROSIS (BEAMS) STUDY GROUP (2016). Comparison of statistical approaches for dealing with immortal time bias in drug effectiveness studies. *Am. J. Epidemiol.* **184** 325–335.
- KESSING, L. V., RYTGAARD, H. C., GERDS, T. A., BERK, M., EKSTRØM, C. T. and ANDERSEN, P. K. (2019). New drug candidates for depression – a nationwide population-based study. *Acta Psychiatr. Scand.* **139** 68–77.
- LENDLE, S. D., SCHWAB, J., PETERSEN, M. L. and VAN DER LAAN, M. J. (2017). Ltmle: An R package implementing targeted minimum loss-based estimation for longitudinal data. *J. Stat. Softw.* **81** 1–21.
- LOK, J. J. (2008). Statistical modeling of causal effects in continuous time. *Ann. Statist.* **36** 1464–1507. [MR2418664](#) <https://doi.org/10.1214/009053607000000820>
- MARTINUSSEN, T., VANSTEELENDT, S. and ANDERSEN, P. K. (2020). Subtleties in the interpretation of hazard contrasts. *Lifetime Data Anal.* **26** 833–855. [MR4148449](#) <https://doi.org/10.1007/s10985-020-09501-5>
- MURPHY, S. A. (2003). Optimal dynamic treatment regimes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 331–366. [MR1983752](#) <https://doi.org/10.1111/1467-9868.00389>
- MURPHY, S. A. (2005). An experimental design for the development of adaptive treatment strategies. *Stat. Med.* **24** 1455–1481. [MR2137651](#) <https://doi.org/10.1002/sim.2022>
- MURPHY, S. A., VAN DER LAAN, M. J. and ROBINS, J. M. (2001). Marginal mean models for dynamic regimes. *J. Amer. Statist. Assoc.* **96** 1410–1423. [MR1946586](#) <https://doi.org/10.1198/016214501753382327>
- PETERSEN, M., SCHWAB, J., GRUBER, S., BLASER, N., SCHOMAKER, M. and VAN DER LAAN, M. (2014). Targeted maximum likelihood estimation for dynamic and static longitudinal marginal structural working models. *J. Causal Inference* **2** 147–185. [MR4289419](#) <https://doi.org/10.1515/jci-2013-0007>
- 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** 1393–1512. [MR0877758](#) [https://doi.org/10.1016/0270-0255\(86\)90088-6](https://doi.org/10.1016/0270-0255(86)90088-6)
- ROBINS, J. M. (1987). Addendum to: “A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect” [Math. Modelling **7** (1986), 1393–1512; [MR0877758](#) (87m:92078)]. *Comput. Math. Appl.* **14** 923–945. [MR0922792](#) [https://doi.org/10.1016/0898-1221\(87\)90238-0](https://doi.org/10.1016/0898-1221(87)90238-0)
- ROBINS, J. (1989a). The control of confounding by intermediate variables. *Stat. Med.* **8** 679–701.
- ROBINS, J. M. (1989b). The analysis of randomized and non-randomized aids treatment trials using a new approach to causal inference in longitudinal studies. *Health Serv. Outcomes Res. Methodol.* 113–159.
- ROBINS, J. (1992). Estimation of the time-dependent accelerated failure time model in the presence of confounding factors. *Biometrika* **79** 321–334. [MR1185134](#) <https://doi.org/10.1093/biomet/79.2.321>
- ROBINS, J. M. (1998). Marginal structural models. in: 1997 proceedings of the american statistical association, section on bayesian statistical science (pp. 1–10). Retrieved from.
- ROBINS, J. M. (2000a). 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
- ROBINS, J. M. (2000b). Robust estimation in sequentially ignorable missing data and causal inference models. In *Proceedings of the American Statistical Association* **1999** 6–10.
- ROBINS, J. M. (2002). Analytic methods for estimating hiv-treatment and cofactor effects. In *Methodological Issues in AIDS Behavioral Research* 213–288. Springer, Berlin.
- ROBINS, J. M., HERNÁN, M. A. and SIEBERT, U. (2004). Effects of multiple interventions. *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors* **1** 2191–2230.
- ROBINS, J., ORELLANA, L. and ROTNITZKY, A. (2008). Estimation and extrapolation of optimal treatment and testing strategies. *Stat. Med.* **27** 4678–4721. [MR2528576](#) <https://doi.org/10.1002/sim.3301>

- RØYSLAND, K. (2011). A martingale approach to continuous-time marginal structural models. *Bernoulli* **17** 895–915. [MR2817610 https://doi.org/10.3150/10-BEJ303](https://doi.org/10.3150/10-BEJ303)
- RØYSLAND, K. (2012). Counterfactual analyses with graphical models based on local independence. *Ann. Statist.* **40** 2162–2194. [MR3059080 https://doi.org/10.1214/12-AOS1031](https://doi.org/10.1214/12-AOS1031)
- RYTGAARD, H. C., GERDS, T. A and VAN DER LAAN, M. J. (2022). Supplement to “Continuous-time targeted minimum loss-based estimation of intervention-specific mean outcomes.” <https://doi.org/10.1214/21-AOS2114SUPP>
- STITELMAN, O. M., DE GRUTTOLA, V. and VAN DER LAAN, M. J. (2012). A general implementation of TMLE for longitudinal data applied to causal inference in survival analysis. *Int. J. Biostat.* **8** Art. 26, front matter+37. [MR2997676 https://doi.org/10.1515/1557-4679.1334](https://doi.org/10.1515/1557-4679.1334)
- TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. [MR1379242](https://doi.org/10.1111/j.1467-9868.1996.tb00540.x)
- 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. (2010a). Targeted maximum likelihood based causal inference. I. *Int. J. Biostat.* **6** Art. 2, 44. [MR2595112 https://doi.org/10.2202/1557-4679.1241](https://doi.org/10.2202/1557-4679.1241)
- VAN DER LAAN, M. J. (2010b). Targeted maximum likelihood based causal inference. I. *Int. J. Biostat.* **6** Art. 2, 44. [MR2595112 https://doi.org/10.2202/1557-4679.1241](https://doi.org/10.2202/1557-4679.1241)
- VAN DER LAAN, M. J. (2017). A generally efficient targeted minimum loss based estimator based on the highly adaptive Lasso. *Int. J. Biostat.* **13** 20150097, 35. [MR3724476 https://doi.org/10.1515/ijb-2015-0097](https://doi.org/10.1515/ijb-2015-0097)
- VAN DER LAAN, M. J. and GRUBER, S. (2012). Targeted minimum loss based estimation of causal effects of multiple time point interventions. *Int. J. Biostat.* **8** Art. 9, 41. [MR2923282 https://doi.org/10.1515/1557-4679.1370](https://doi.org/10.1515/1557-4679.1370)
- VAN DER LAAN, M. J. and PETERSEN, M. L. (2007). Causal effect models for realistic individualized treatment and intention to treat rules. *Int. J. Biostat.* **3** Art. 3, 54. [MR2306841 https://doi.org/10.2202/1557-4679.1022](https://doi.org/10.2202/1557-4679.1022)
- 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](https://doi.org/10.2202/1544-6115.1309)
- 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](https://doi.org/10.1007/978-0-387-21700-0)
- VAN DER LAAN, M. J. and ROSE, S. (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](https://doi.org/10.1007/978-1-4419-9782-1)
- VAN DER LAAN, M. J. and ROSE, S. (2018). *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies*. Springer Series in Statistics. Springer, Cham. [MR3791826 https://doi.org/10.1007/978-3-319-65304-4](https://doi.org/10.1007/978-3-319-65304-4)
- 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](https://doi.org/10.2202/1557-4679.1043)
- 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](https://doi.org/10.1017/CBO9780511802256)
- 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](https://doi.org/10.1524/stnd.2006.24.3.351)
- ZHANG, B., TSIATIS, A. A., LABER, E. B. and DAVIDIAN, M. (2012). A robust method for estimating optimal treatment regimes. *Biometrics* **68** 1010–1018. [MR3040007 https://doi.org/10.1111/j.1541-0420.2012.01763.x](https://doi.org/10.1111/j.1541-0420.2012.01763.x)
- ZHANG, B., TSIATIS, A. A., LABER, E. B. and DAVIDIAN, M. (2013). Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika* **100** 681–694. [MR3094445 https://doi.org/10.1093/biomet/ast014](https://doi.org/10.1093/biomet/ast014)

CENTRAL LIMIT THEOREM AND BOOTSTRAP APPROXIMATION IN HIGH DIMENSIONS: NEAR $1/\sqrt{n}$ RATES VIA IMPLICIT SMOOTHING

BY MILES E. LOPES^a

Department of Statistics, University of California, Davis, ^amelopes@ucdavis.edu

Nonasymptotic bounds for Gaussian and bootstrap approximation have recently attracted significant interest in high-dimensional statistics. This paper studies Berry–Esseen bounds for such approximations with respect to the multivariate Kolmogorov distance, in the context of a sum of n random vectors that are p -dimensional and i.i.d. Up to now, a growing line of work has established bounds with mild logarithmic dependence on p . However, the problem of developing corresponding bounds with near $n^{-1/2}$ dependence on n has remained largely unresolved. Within the setting of random vectors that have sub-Gaussian or subexponential entries, this paper establishes bounds with near $n^{-1/2}$ dependence, for both Gaussian and bootstrap approximation. In addition, the proofs are considerably distinct from other recent approaches, and make use of an “implicit smoothing” operation in the Lindeberg interpolation.

REFERENCES

- ANDERSON, N. H., HALL, P. and TITTERINGTON, D. M. (1998). Edgeworth expansions in very-high-dimensional problems. *J. Statist. Plann. Inference* **70** 1–18. [MR1649439](#) [https://doi.org/10.1016/S0378-3758\(97\)00171-7](https://doi.org/10.1016/S0378-3758(97)00171-7)
- BELLONI, A., CHERNOZHUKOV, V., CHETVERIKOV, D., HANSEN, C. and KATO, K. (2018). High-dimensional econometrics and regularized GMM. Available at [arXiv:1806.01888](#).
- BENTKUS, V. (1990). Smooth approximations of the norm and differentiable functions with bounded support in the Banach space l_{∞}^k . *Liet. Mat. Rink.* **30** 489–499. [MR1082475](#) <https://doi.org/10.1007/BF00970805>
- BENTKUS, V. (2003). On the dependence of the Berry–Esseen bound on dimension. *J. Statist. Plann. Inference* **113** 385–402. [MR1965117](#) [https://doi.org/10.1016/S0378-3758\(02\)00094-0](https://doi.org/10.1016/S0378-3758(02)00094-0)
- BENTKUS, V. (2004). A Lyapunov type bound in R^d . *Teor. Veroyatn. Primen.* **49** 400–410. [MR2144310](#) <https://doi.org/10.1137/S0040585X97981123>
- CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. [MR3161448](#) <https://doi.org/10.1214/13-AOS1161>
- CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017a). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. [MR3693963](#) <https://doi.org/10.1214/16-AOP1113>
- CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017b). Detailed proof of Nazarov’s inequality. Available at [arXiv:1711.10696](#).
- CHERNOZHUKOV, V., CHETVERIKOV, D. and KOIKE, Y. (2020). Nearly optimal central limit theorem and bootstrap approximations in high dimensions. Available at [arXiv:2012.09513](#).
- CHERNOZHUKOV, V., CHETVERIKOV, D., KATO, K. and KOIKE, Y. (2019). Improved central limit theorem and bootstrap approximations in high dimensions. Available at [arXiv:1912.10529](#).
- DAS, D. and LAHIRI, S. (2021). Central Limit Theorem in high dimensions: The optimal bound on dimension growth rate. *Trans. Amer. Math. Soc.* **374** 6991–7009. [MR4315595](#) <https://doi.org/10.1090/tran/8459>
- DENG, H. (2020). Slightly conservative bootstrap for maxima of sums. Available at [arXiv:2007.15877](#).
- DENG, H. and ZHANG, C.-H. (2020). Beyond Gaussian approximation: Bootstrap for maxima of sums of independent random vectors. *Ann. Statist.* **48** 3643–3671. [MR4185823](#) <https://doi.org/10.1214/20-AOS1946>
- EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26. [MR0515681](#)
- ESSEEN, C. G. (1956). A moment inequality with an application to the central limit theorem. *Skand. Aktuarietidskr.* **39** 160–170. [MR0090166](#) <https://doi.org/10.1080/03461238.1956.10414946>

- FANG, X. and KOIKE, Y. (2020). Large-dimensional central limit theorem with fourth-moment error bounds on convex sets and balls. Available at [arXiv:2009.00339](https://arxiv.org/abs/2009.00339).
- FANG, X. and KOIKE, Y. (2021). High-dimensional central limit theorems by Stein's method. *Ann. Appl. Probab.* **31** 1660–1686. [MR4312842 https://doi.org/10.1214/20-aap1629](https://doi.org/10.1214/20-aap1629)
- KOIKE, Y. (2021). Notes on the dimension dependence in high-dimensional central limit theorems for hyperrectangles. *Jpn. J. Stat. Data Sci.* **4** 257–297. [MR4273258 https://doi.org/10.1007/s42081-020-00096-7](https://doi.org/10.1007/s42081-020-00096-7)
- KUCHIBHOTLA, A. K. and CHAKRABORTTY, A. (2020). Moving beyond sub-Gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression. Available at [arXiv:1804.02605](https://arxiv.org/abs/1804.02605).
- KUCHIBHOTLA, A. K., MUKHERJEE, S. and BANERJEE, D. (2021a). High-dimensional CLT: Improvements, non-uniform extensions and large deviations. *Bernoulli* **27** 192–217. [MR4177366 https://doi.org/10.3150/20-BEJ1233](https://doi.org/10.3150/20-BEJ1233)
- KUCHIBHOTLA, A. K. and RINALDO, A. (2020). High-dimensional CLT for sums of non-degenerate random vectors: $n^{-1/2}$ -rate. Available at [arXiv:2009.13673](https://arxiv.org/abs/2009.13673).
- LOPES, M. E. (2020). Central limit theorem and bootstrap approximation in high dimensions with near $1/\sqrt{n}$ rates. Available at [arXiv:2009.06004](https://arxiv.org/abs/2009.06004).
- LOPES, M. E., LIN, Z. and MÜLLER, H.-G. (2020). Bootstrapping max statistics in high dimensions: Near-parametric rates under weak variance decay and application to functional and multinomial data. *Ann. Statist.* **48** 1214–1229. [MR4102694 https://doi.org/10.1214/19-AOS1844](https://doi.org/10.1214/19-AOS1844)
- LOPES, M. E. and YAO, J. (2022). A sharp lower-tail bound for Gaussian maxima with application to bootstrap methods in high dimensions. *Electron. J. Stat.* **16** 58–83. [MR4359356 https://doi.org/10.1214/21-ejs1961](https://doi.org/10.1214/21-ejs1961)
- NAZAROV, F. (2003). On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a Gaussian measure. In *Geometric Aspects of Functional Analysis. Lecture Notes in Math.* **1807** 169–187. Springer, Berlin. [MR2083397 https://doi.org/10.1007/978-3-540-36428-3_15](https://doi.org/10.1007/978-3-540-36428-3_15)
- O'DONNELL, R., SERVEDIO, R. A. and TAN, L.-Y. (2019). Fooling polytopes. In *STOC'19—Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing* 614–625. ACM, New York. [MR4003369 https://doi.org/10.1145/3313276.3316321](https://doi.org/10.1145/3313276.3316321)
- RAIČ, M. (2019). A multivariate Berry–Esseen theorem with explicit constants. *Bernoulli* **25** 2824–2853. [MR4003566 https://doi.org/10.3150/18-BEJ1072](https://doi.org/10.3150/18-BEJ1072)
- SENATOV, V. V. (1980). Some uniform estimates of the convergence rate in the multidimensional central limit theorem. *Teor. Veroyatn. Primen.* **25** 757–770. [MR0595137](https://doi.org/10.1007/978-1-4757-2545-2)
- VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer, New York. [MR1385671 https://doi.org/10.1007/978-1-4757-2545-2](https://doi.org/10.1007/978-1-4757-2545-2)
- VERSHYNIN, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics **47**. Cambridge Univ. Press, Cambridge. [MR3837109 https://doi.org/10.1017/9781108231596](https://doi.org/10.1017/9781108231596)

TESTING GOODNESS-OF-FIT AND CONDITIONAL INDEPENDENCE WITH APPROXIMATE CO-SUFFICIENT SAMPLING

BY RINA FOYGEL BARBER^{1,a} AND LUCAS JANSON^{2,b}

¹*Department of Statistics, University of Chicago, rina@uchicago.edu*

²*Department of Statistics, Harvard University, ljanson@fas.harvard.edu*

Goodness-of-fit (GoF) testing is ubiquitous in statistics, with direct ties to model selection, confidence interval construction, conditional independence testing, and multiple testing, just to name a few applications. While testing the GoF of a simple (point) null hypothesis provides an analyst great flexibility in the choice of test statistic while still ensuring validity, most GoF tests for composite null hypotheses are far more constrained, as the test statistic must have a tractable distribution over the entire null model space. A notable exception is *co-sufficient sampling* (CSS): resampling the data conditional on a sufficient statistic for the null model guarantees valid GoF testing using any test statistic the analyst chooses. But CSS testing requires the null model to have a compact (in an information-theoretic sense) sufficient statistic, which only holds for a very limited class of models; even for a null model as simple as logistic regression, CSS testing is powerless. In this paper, we leverage the concept of *approximate sufficiency* to generalize CSS testing to essentially any parametric model with an asymptotically efficient estimator; we call our extension “approximate CSS” (aCSS) testing. We quantify the finite-sample Type I error inflation of aCSS testing and show that it is vanishing under standard maximum likelihood asymptotics, for any choice of test statistic. We apply our proposed procedure both theoretically and in simulation to a number of models of interest to demonstrate its finite-sample Type I error and power.

REFERENCES

- [1] AGRESTI, A. (1992). A survey of exact inference for contingency tables. *Statist. Sci.* **7** 131–177. [MR1173420](#)
- [2] AGRESTI, A. (2001). Exact inference for categorical data: Recent advances and continuing controversies. *Stat. Med.* **20** 2709–2722.
- [3] BACHOC, F. (2014). Asymptotic analysis of the role of spatial sampling for covariance parameter estimation of Gaussian processes. *J. Multivariate Anal.* **125** 1–35. [MR3163828](#) <https://doi.org/10.1016/j.jmva.2013.11.015>
- [4] BARBER, R. F. and CANDÈS, E. J. (2015). Controlling the false discovery rate via knockoffs. *Ann. Statist.* **43** 2055–2085. [MR3375876](#) <https://doi.org/10.1214/15-AOS1337>
- [5] BARBER, R. F. and JANSON, L. (2022). Supplement to “Testing goodness-of-fit and conditional independence with approximate co-sufficient sampling.” <https://doi.org/10.1214/22-AOS2187SUPP>
- [6] BARTLETT, M. S. (1937). Properties of sufficiency and statistical tests. *Proc. R. Soc. Lond. Ser. A, Math. Phys. Sci.* **160** 268–282.
- [7] BAYARRI, M. J. and BERGER, J. O. (1999). Quantifying surprise in the data and model verification. In *Bayesian Statistics, 6 (Alcoceber, 1998)* 53–82. Oxford Univ. Press, New York. [MR1723493](#)
- [8] BELL, C. B. (1984). Inference for goodness-of-fit problems with nuisance parameters (applications to signal detection). *J. Statist. Plann. Inference* **9** 273–284. [MR0748257](#) [https://doi.org/10.1016/0378-3758\(84\)90003-X](https://doi.org/10.1016/0378-3758(84)90003-X)
- [9] BELTRÁN-BELTRÁN, J. I. and O’REILLY, F. J. (2019). On goodness of fit tests for the Poisson, negative binomial and binomial distributions. *Statist. Papers* **60** 1–18. [MR3905436](#) <https://doi.org/10.1007/s00362-016-0820-5>

MSC2020 subject classifications. Primary 62F03; secondary 62B05.

Key words and phrases. Goodness-of-fit test, approximate sufficiency, co-sufficiency, conditional randomization test, model-X, conditional independence testing, high-dimensional inference.

- [10] BERRETT, T. B., WANG, Y., BARBER, R. F. and SAMWORTH, R. J. (2020). The conditional permutation test for independence while controlling for confounders. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 175–197. [MR4060981](#)
- [11] BESAG, J. and CLIFFORD, P. (1989). Generalized Monte Carlo significance tests. *Biometrika* **76** 633–642. [MR1041408](#) <https://doi.org/10.1093/biomet/76.4.633>
- [12] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A nonasymptotic theory of independence*. Oxford Univ. Press, Oxford. [MR3185193](#) <https://doi.org/10.1093/acprof:oso/9780199535255.001.0001>
- [13] BRONIATOWSKI, M. and CARON, V. (2019). Conditional inference in parametric models. *J. SFdS* **160** 48–66. [MR3987789](#)
- [14] CANDÈS, E., FAN, Y., JANSON, L. and LV, J. (2018). Panning for gold: ‘model-X’ knockoffs for high dimensional controlled variable selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 551–577. [MR3798878](#) <https://doi.org/10.1111/rssb.12265>
- [15] CONTRERAS-CRISTÁN, A., LOCKHART, R. A., STEPHENS, M. A. and SUN, S. Z. (2019). On the use of priors in goodness-of-fit tests. *Canad. J. Statist.* **47** 560–579. [MR4035789](#) <https://doi.org/10.1002/cjs.11512>
- [16] COX, D. R. and REID, N. (1987). Parameter orthogonality and approximate conditional inference. *J. Roy. Statist. Soc. Ser. B* **49** 1–39. [MR0893334](#)
- [17] DIACONIS, P., HOLMES, S. and SHAHSHAHANI, M. (2013). Sampling from a manifold. In *Advances in Modern Statistical Theory and Applications: A Festschrift in Honor of Morris L. Eaton*. Inst. Math. Stat. (IMS) Collect. **10** 102–125. IMS, Beachwood, OH. [MR3586941](#)
- [18] DURBIN, J. (1961). Some methods of constructing exact tests. *Biometrika* **48** 41–55. [MR0126313](#) <https://doi.org/10.1093/biomet/48.1-2.41>
- [19] 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>
- [20] ENGEN, S. and LILLEGÅRD, M. (1997). Stochastic simulations conditioned on sufficient statistics. *Biometrika* **84** 235–240. [MR1450203](#) <https://doi.org/10.1093/biomet/84.1.235>
- [21] HAZRA, A. (2013). An exact Kolmogorov–Smirnov test for the negative Binomial distribution with unknown probability of success. *Res. Rev.: J. Stat.* **2** 1–13.
- [22] HUANG, D. and JANSON, L. (2020). Relaxing the assumptions of knockoffs by conditioning. *Ann. Statist.* **48** 3021–3042. [MR4152633](#) <https://doi.org/10.1214/19-AOS1920>
- [23] JANKOVÁ, J., SHAH, R. D., BÜHLMANN, P. and SAMWORTH, R. J. (2020). Goodness-of-fit testing in high dimensional generalized linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 773–795. [MR4112784](#) <https://doi.org/10.1111/rssb.12371>
- [24] KALBFLEISCH, J. D. and SPROTT, D. A. (1970). Application of likelihood methods to models involving large numbers of parameters. *J. Roy. Statist. Soc. Ser. B* **32** 175–208. [MR0270474](#)
- [25] KOLASSA, J. E. (2003). Algorithms for approximate conditional inference. *Stat. Comput.* **13** 121–126. [MR1963328](#) <https://doi.org/10.1023/A:1023252308207>
- [26] KOUSATHANAS, A., LEUENBERGER, C., HELFER, J., QUINODOZ, M., FOLL, M. and WEGMANN, D. (2016). Likelihood-free inference in high-dimensional models. *Genetics* **203** 893–904. <https://doi.org/10.1534/genetics.116.187567>
- [27] KUMAR, A. and PATHAK, P. K. (1977). Sufficiency and tests of goodness of fit. *Scand. J. Stat.* **4** 39–43. [MR0431506](#)
- [28] LE CAM, L. (1960). Locally asymptotically normal families of distributions. Certain approximations to families of distributions and their use in the theory of estimation and testing hypotheses. *Univ. California Publ. Statist.* **3** 37–98. [MR0126903](#)
- [29] LE CAM, L. (2012). *Asymptotic Methods in Statistical Decision Theory*. Springer Series in Statistics. Springer, New York.
- [30] LI, K.-C. (1991). Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.* **86** 316–342. [MR1137117](#)
- [31] LILLEGÅRD, M. (2001). Tests based on Monte Carlo simulations conditioned on maximum likelihood estimates of nuisance parameters. *J. Stat. Comput. Simul.* **71** 1–10. [MR1888292](#) <https://doi.org/10.1080/00949650108812130>
- [32] LILLEGÅRD, M. and ENGEN, S. (1999). Exact confidence intervals generated by conditional parametric bootstrapping. *J. Appl. Stat.* **26** 447–459. [MR1703739](#) <https://doi.org/10.1080/02664769922331>
- [33] LINDQVIST, B. H. and RANNESTAD, B. (2011). Monte Carlo exact goodness-of-fit tests for nonhomogeneous Poisson processes. *Appl. Stoch. Models Bus. Ind.* **27** 329–341. [MR2849567](#) <https://doi.org/10.1002/asmb.841>

- [34] LINDQVIST, B. H. and TARALDSEN, G. (2013). Exact statistical inference for some parametric nonhomogeneous Poisson processes. *J. Iran. Stat. Soc. (JIRSS)* **12** 113–126. MR3048729 <https://doi.org/10.1214/13-aos1083>
- [35] LOCKHART, R. A. (2012). Conditional limit laws for goodness-of-fit tests. *Bernoulli* **18** 857–882. MR2948905 <https://doi.org/10.3150/11-BEJ366>
- [36] LOCKHART, R. A., O'REILLY, F. and STEPHENS, M. (2009). Exact conditional tests and approximate bootstrap tests for the von Mises distribution. *J. Stat. Theory Pract.* **3** 543–554. MR2750456 <https://doi.org/10.1080/15598608.2009.10411945>
- [37] LOCKHART, R. A., O'REILLY, F. J. and STEPHENS, M. A. (2007). Use of the Gibbs sampler to obtain conditional tests, with applications. *Biometrika* **94** 992–998. MR2416805 <https://doi.org/10.1093/biomet/asm065>
- [38] MARCUS, R., PERITZ, E. and GABRIEL, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* **63** 655–660. MR0468056 <https://doi.org/10.1093/biomet/63.3.655>
- [39] O'REILLY, F. and GRACIA-MEDRANO, L. (2006). On the conditional distribution of goodness-of-fit tests. *Comm. Statist. Theory Methods* **35** 541–549. MR2274070 <https://doi.org/10.1080/03610920500476622>
- [40] ROSENBAUM, P. R. (1984). Conditional permutation tests and the propensity score in observational studies. *J. Amer. Statist. Assoc.* **79** 565–574. MR0763575
- [41] SANTOS, J. D. and FILHO, N. L. S. (2019). A Metropolis algorithm to obtain co-sufficient samples with applications in conditional tests. *Comm. Statist. Simulation Comput.* **48** 2655–2659. MR4001227 <https://doi.org/10.1080/03610918.2018.1458140>
- [42] STEPHENS, M. A. (2012). Goodness-of-fit and sufficiency: Exact and approximate tests. *Methodol. Comput. Appl. Probab.* **14** 785–791. MR2966320 <https://doi.org/10.1007/s11009-011-9267-2>
- [43] TIAN, X. and TAYLOR, J. (2018). Selective inference with a randomized response. *Ann. Statist.* **46** 679–710. MR3782381 <https://doi.org/10.1214/17-AOS1564>
- [44] 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>

NEW EDGEWORTH-TYPE EXPANSIONS WITH FINITE SAMPLE GUARANTEES

BY MAYYA ZHILOVA^a

School of Mathematics, Georgia Institute of Technology, ^amzhilova@math.gatech.edu

We establish higher-order nonasymptotic expansions for a difference between probability distributions of sums of i.i.d. random vectors in a Euclidean space. The derived bounds are uniform over two classes of sets: the set of all Euclidean balls and the set of all half-spaces. These results allow to account for an impact of higher-order moments or cumulants of the considered distributions; the obtained error terms depend on a sample size and a dimension explicitly. The new inequalities outperform accuracy of the normal approximation in existing Berry–Esseen inequalities under very general conditions. Under some symmetry assumptions on the probability distribution of random summands, the obtained results are optimal in terms of the ratio between the dimension and the sample size. The new technique which we developed for establishing nonasymptotic higher-order expansions can be interesting by itself. Using the new higher-order inequalities, we study accuracy of the non-parametric bootstrap approximation and propose a bootstrap score test under possible model misspecification. The results of the paper also include explicit error bounds for general elliptic confidence regions for an expected value of the random summands, and optimality of the Gaussian anticoncentration inequality over the set of all Euclidean balls.

REFERENCES

- [1] ARLOT, S., BLANCHARD, G. and ROQUAIN, E. (2010). Some nonasymptotic results on resampling in high dimension. I. Confidence regions. *Ann. Statist.* **38** 51–82. [MR2589316](#) <https://doi.org/10.1214/08-AOS667>
- [2] BALL, K. (1993). The reverse isoperimetric problem for Gaussian measure. *Discrete Comput. Geom.* **10** 411–420. [MR1243336](#) <https://doi.org/10.1007/BF02573986>
- [3] BARBE, P. and BERTAIL, P. (1995). *The Weighted Bootstrap. Lecture Notes in Statistics* **98**. Springer, New York. [MR2195545](#) <https://doi.org/10.1007/978-1-4612-2532-4>
- [4] BELLONI, A., BUGNI, F. and CHERNOZHUKOV, V. (2018). Subvector inference in partially identified models with many moment inequalities. Preprint. Available at [arXiv:1806.11466](#).
- [5] BENTKUS, V. (2003). On the dependence of the Berry–Esseen bound on dimension. *J. Statist. Plann. Inference* **113** 385–402. [MR1965117](#) [https://doi.org/10.1016/S0378-3758\(02\)00094-0](https://doi.org/10.1016/S0378-3758(02)00094-0)
- [6] BERRY, A. C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.* **49** 122–136. [MR0003498](#) <https://doi.org/10.2307/1990053>
- [7] BHATTACHARYA, R. N. and RAO, R. R. (2010). *Normal Approximation and Asymptotic Expansions. Classics in Applied Mathematics* **64**. SIAM, Philadelphia, PA. [MR3396213](#) <https://doi.org/10.1137/1.9780898719895.ch1>
- [8] BICKEL, P. J. and DOKSUM, K. A. (2015). *Mathematical Statistics—Basic Ideas and Selected Topics. Vol. 1*, 2nd ed. *Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. [MR3445928](#)
- [9] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford. [MR3185193](#) <https://doi.org/10.1093/acprof:oso/9780199535255.001.0001>
- [10] BOX, G. E. P. (1976). Science and statistics. *J. Amer. Statist. Assoc.* **71** 791–799. [MR0431440](#)
- [11] CHATTERJEE, S. (2006). A generalization of the Lindeberg principle. *Ann. Probab.* **34** 2061–2076. [MR2294976](#) <https://doi.org/10.1214/009117906000000575>

MSC2020 subject classifications. Primary 62E17, 62F40; secondary 62F25.

Key words and phrases. Edgeworth series, dependence on dimension, higher-order accuracy, multivariate Berry–Esseen inequality, chi-square approximation, finite sample inference, anticoncentration inequality, bootstrap, elliptic confidence sets, linear contrasts, bootstrap score test, model misspecification.

- [12] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. [MR3161448](#) <https://doi.org/10.1214/13-AOS1161>
- [13] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* **162** 47–70. [MR3350040](#) <https://doi.org/10.1007/s00440-014-0565-9>
- [14] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. [MR3693963](#) <https://doi.org/10.1214/16-AOP1113>
- [15] CRAMÉR, H. (1928). On the composition of elementary errors. *Scand. Actuar. J.* **1928** 13–74.
- [16] EDGEWORTH, F. Y. (1896). The asymmetrical probability-curve. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **41** 90–99.
- [17] EDGEWORTH, F. Y. (1905). The law of error. In *Pros. Camb. Philos. Soc* **20** 16–65.
- [18] EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26. [MR0515681](#)
- [19] 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>
- [20] ESSEEN, C.-G. (1942). On the Liapounoff limit of error in the theory of probability. *Ark. Mat. Astron. Fys.* **28A** 19. [MR0011909](#)
- [21] FREEDMAN, D. A. (2007). How can the score test be inconsistent? *Amer. Statist.* **61** 291–295. [MR2411784](#) <https://doi.org/10.1198/000313007X243061>
- [22] FRIENDLY, M., MONETTE, G. and FOX, J. (2013). Elliptic insights: Understanding statistical methods through elliptical geometry. *Statist. Sci.* **28** 1–39. [MR3075337](#) <https://doi.org/10.1214/12-STS402>
- [23] GUSTAFSON, P. (2001). On measuring sensitivity to parametric model misspecification. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 81–94. [MR1811992](#) <https://doi.org/10.1111/1467-9868.00277>
- [24] 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>
- [25] HOROWITZ, J. L. (2001). The bootstrap. *Handb. Econom.* **5** 3159–3228.
- [26] KLIVANS, A. R., O'DONNELL, R. and SERVEDIO, R. A. (2008). Learning geometric concepts via Gaussian surface area. In *Foundations of Computer Science, 2008. FOCS'08. IEEE 49th Annual IEEE Symposium on* 541–550. IEEE Comput. Soc., Los Alamitos.
- [27] KOLASSA, J. E. (1994). *Series Approximation Methods in Statistics*. Lecture Notes in Statistics **88**. Springer, New York. [MR1295242](#) <https://doi.org/10.1007/978-1-4757-4275-6>
- [28] LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data*. Springer Series in Statistics. Springer, New York. [MR2001447](#) <https://doi.org/10.1007/978-1-4757-3803-2>
- [29] LINDBERGH, J. W. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Z.* **15** 211–225. [MR1544569](#) <https://doi.org/10.1007/BF01494395>
- [30] LIU, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models. *Ann. Statist.* **16** 1696–1708. [MR0964947](#) <https://doi.org/10.1214/aos/1176351062>
- [31] LOPES, M. E. (2022). Central limit theorem and bootstrap approximation in high dimensions: $1/\sqrt{n}$ rates via implicit smoothing. *Ann. Statist.*, to appear. [arXiv:2009.06004](#).
- [32] MAMMEN, E. (1993). Bootstrap and wild bootstrap for high-dimensional linear models. *Ann. Statist.* **21** 255–285. [MR1212176](#) <https://doi.org/10.1214/aos/1176349025>
- [33] MCCULLAGH, P. (1987). *Tensor Methods in Statistics*. Monographs on Statistics and Applied Probability. CRC Press, London. [MR0907286](#)
- [34] MCCULLAGH, P. and NELDER, J. A. (1983). *Generalized Linear Models*. Monographs on Statistics and Applied Probability. CRC Press, London. [MR0727836](#) <https://doi.org/10.1007/978-1-4899-3244-0>
- [35] PORTNOY, S. (1986). On the central limit theorem in R^p when $p \rightarrow \infty$. *Probab. Theory Related Fields* **73** 571–583. [MR0863546](#) <https://doi.org/10.1007/BF00324853>
- [36] PRÆSTGAARD, J. and WELLNER, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.* **21** 2053–2086. [MR1245301](#)
- [37] RADHAKRISHNA RAO, C. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proc. Camb. Philos. Soc.* **44** 50–57. [MR0024111](#)
- [38] RAO, C. R. (2005). Score test: Historical review and recent developments. In *Advances in Ranking and Selection, Multiple Comparisons, and Reliability*. Stat. Ind. Technol. 3–20. Birkhäuser, Boston, MA. [MR2111500](#) https://doi.org/10.1007/0-8176-4422-9_1
- [39] SHEVTSOVA, I. G. (2014). On absolute constants in inequalities of Berry–Esseen type. *Dokl. Akad. Nauk* **456** 650–654. [MR3287912](#) <https://doi.org/10.1134/s1064562414030338>
- [40] SKOVGAARD, I. M. (1986). On multivariate Edgeworth expansions. *Int. Stat. Rev.* **54** 169–186. [MR0962933](#) <https://doi.org/10.2307/1403142>

- [41] TCHEBYCHEFF, P. (1890). Sur deux théorèmes relatifs aux probabilités. *Acta Math.* **14** 305–315. [MR1554802](#) <https://doi.org/10.1007/BF02413327>
- [42] TROTTER, H. F. (1959). An elementary proof of the central limit theorem. *Arch. Math.* **10** 226–234. [MR0108847](#) <https://doi.org/10.1007/BF01240790>
- [43] WANG, M., DAO DUC, K., FISCHER, J. and SONG, Y. S. (2017). Operator norm inequalities between tensor unfoldings on the partition lattice. *Linear Algebra Appl.* **520** 44–66. [MR3611456](#) <https://doi.org/10.1016/j.laa.2017.01.017>
- [44] WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50** 1–25. [MR0640163](#) <https://doi.org/10.2307/1912526>
- [45] WIT, E., VAN DEN HEUVEL, E. and ROMEIJN, J.-W. (2012). ‘All models are wrong...’: An introduction to model uncertainty. *Stat. Neerl.* **66** 217–236. [MR2955417](#) <https://doi.org/10.1111/j.1467-9574.2012.00530.x>
- [46] ZHILOVA, M. (2020). Nonclassical Berry–Esseen inequalities and accuracy of the bootstrap. *Ann. Statist.* **48** 1922–1939. [MR4134780](#) <https://doi.org/10.1214/18-AOS1802>
- [47] ZHILOVA, M. (2022). Supplement to “New Edgeworth-type expansions with finite sample guarantees.” <https://doi.org/10.1214/22-AOS2192SUPP>

IMPROVED CENTRAL LIMIT THEOREM AND BOOTSTRAP APPROXIMATIONS IN HIGH DIMENSIONS

BY VICTOR CHERNOZHUKOV^{1,a}, DENIS CHETVERIKOV^{2,b}, KENGO KATO^{3,c} AND YUTA KOIKE^{4,d}

¹Department of Economics and Operations Research Center, MIT, avchern@mit.edu

²Department of Economics, UCLA, chetverikov@econ.ucla.edu

³Department of Statistics and Data Science, Cornell University, kk976@cornell.edu

⁴Department, Mathematics and Informatics Center and Graduate School of Mathematical Sciences, The University of Tokyo, kyuta@ms.u-tokyo.ac.jp

This paper deals with the Gaussian and bootstrap approximations to the distribution of the max statistic in high dimensions. This statistic takes the form of the maximum over components of the sum of independent random vectors and its distribution plays a key role in many high-dimensional estimation and testing problems. Using a novel iterative randomized Lindeberg method, the paper derives new bounds for the distributional approximation errors. These new bounds substantially improve upon existing ones and simultaneously allow for a larger class of bootstrap methods.

REFERENCES

- [1] ANDREWS, D. W. K. and SHI, X. (2013). Inference based on conditional moment inequalities. *Econometrica* **81** 609–666. [MR3043344](#) <https://doi.org/10.3982/ECTA9370>
- [2] BAI, Y., SANTOS, A. and SHAIKH, A. (2019). A practical method for testing many moment inequalities. Becker Friedman Institute for Economics Working Paper.
- [3] BELLONI, A., CHERNOZHUKOV, V., CHETVERIKOV, D., HANSEN, C. and KATO, K. (2018). High-dimensional econometrics and regularized GMM. [arXiv:1806.01888](#).
- [4] BELLONI, A., CHERNOZHUKOV, V., CHETVERIKOV, D. and WEI, Y. (2018). Uniformly valid post-regularization confidence regions for many functional parameters in z-estimation framework. *Ann. Statist.* **46** 3643–3675. [MR3852664](#) <https://doi.org/10.1214/17-AOS1671>
- [5] BELLONI, A., CHERNOZHUKOV, V. and KATO, K. (2015). Uniform post-selection inference for least absolute deviation regression and other Z-estimation problems. *Biometrika* **102** 77–94. [MR3335097](#) <https://doi.org/10.1093/biomet/asu056>
- [6] CHATTERJEE, S. (2006). A generalization of the Lindeberg principle. *Ann. Probab.* **34** 2061–2076. [MR2294976](#) <https://doi.org/10.1214/009117906000000575>
- [7] CHEN, X. and KATO, K. (2019). Randomized incomplete U -statistics in high dimensions. *Ann. Statist.* **47** 3127–3156. [MR4025737](#) <https://doi.org/10.1214/18-AOS1773>
- [8] CHEN, X. and KATO, K. (2020). Jackknife multiplier bootstrap: Finite sample approximations to the U -process supremum with applications. *Probab. Theory Related Fields* **176** 1097–1163. [MR4087490](#) <https://doi.org/10.1007/s00440-019-00936-y>
- [9] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. [MR3161448](#) <https://doi.org/10.1214/13-AOS1161>
- [10] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Anti-concentration and honest, adaptive confidence bands. *Ann. Statist.* **42** 1787–1818. [MR3262468](#) <https://doi.org/10.1214/14-AOS1235>
- [11] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* **162** 47–70. [MR3350040](#) <https://doi.org/10.1007/s00440-014-0565-9>
- [12] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. [MR3693963](#) <https://doi.org/10.1214/16-AOP1113>

- [13] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2019). Inference on causal and structural parameters using many moment inequalities. *Rev. Econ. Stud.* **86** 1867–1900. [MR4009488](#) <https://doi.org/10.1093/restud/rdy065>
- [14] CHERNOZHUKOV, V., CHETVERIKOV, D. and KOIKE, Y. (2020). Nearly optimal central limit theorem and bootstrap approximations in high dimensions. [arXiv:2012.09513](#).
- [15] CHERNOZHUKOV, V., CHETVERIKOV, D., KATO, K. and KOIKE, Y. (2022). Supplement to “Improved central limit theorem and bootstrap approximations in high dimensions.” <https://doi.org/10.1214/22-AOS2193SUPP>
- [16] CHESHER, A. and ROSEN, A. M. (2020). Generalized instrumental variable models, methods, and applications. In *Handbook of Econometrics, Vol. 7A. Handbooks in Econom.* 1–110. Elsevier/North-Holland, Amsterdam. [MR4254179](#) <https://doi.org/10.1016/bs.hoe.2019.11.001>
- [17] CHETVERIKOV, D. (2018). Adaptive tests of conditional moment inequalities. *Econometric Theory* **34** 186–227. [MR3739240](#) <https://doi.org/10.1017/S0266466617000184>
- [18] CHETVERIKOV, D. (2019). Testing regression monotonicity in econometric models. *Econometric Theory* **35** 729–776. [MR3974977](#) <https://doi.org/10.1017/s0266466618000282>
- [19] CHETVERIKOV, D., WILHELM, D. and KIM, D. (2018). An adaptive test of stochastic monotonicity. Cemmap working paper CWP24/18.
- [20] DENG, H. and ZHANG, C.-H. (2020). Beyond Gaussian approximation: Bootstrap for maxima of sums of independent random vectors. *Ann. Statist.* **48** 3643–3671. [MR4185823](#) <https://doi.org/10.1214/20-AOS1946>
- [21] FANG, X. and KOIKE, Y. (2021). High-dimensional central limit theorems by Stein’s method. *Ann. Appl. Probab.* **31** 1660–1686. [MR4312842](#) <https://doi.org/10.1214/20-aap1629>
- [22] FATHI, M. (2019). Stein kernels and moment maps. *Ann. Probab.* **47** 2172–2185. [MR3980918](#) <https://doi.org/10.1214/18-AOP1305>
- [23] HANSEN, P. R. (2005). A test for superior predictive ability. *J. Bus. Econom. Statist.* **23** 365–380. [MR2206007](#) <https://doi.org/10.1198/073500105000000063>
- [24] HANSEN, P. R., LUNDE, A. and NASON, J. M. (2011). The model confidence set. *Econometrica* **79** 453–497. [MR2809377](#) <https://doi.org/10.3982/ECTA5771>
- [25] KOIKE, Y. (2019). High-dimensional central limit theorems for homogeneous sums. [arXiv:1902.03809](#).
- [26] KOIKE, Y. (2021). Notes on the dimension dependence in high-dimensional central limit theorems for hyperrectangles. *Jpn. J. Stat. Data Sci.* **4** 257–297. [MR4273258](#) <https://doi.org/10.1007/s42081-020-00096-7>
- [27] KONING, N. and BEKKER, P. (2019). Exact testing of many moment inequalities against multiple violations. [arXiv:1904.12775](#).
- [28] KUCHIBHOTLA, A. and RINALDO, A. (2020). High-dimensional CLT for sums of non-degenerate random vectors: $n^{-1/2}$ rate. [arXiv:2009.13673](#).
- [29] LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. *Springer Texts in Statistics*. Springer, New York. [MR2135927](#)
- [30] LOPES, M. (2020). Central limit theorem and bootstrap approximation in high dimensions with near $1/\sqrt{n}$ rates. [arXiv:2009.06004](#).
- [31] LOPES, M. E., LIN, Z. and MÜLLER, H.-G. (2020). Bootstrapping max statistics in high dimensions: Near-parametric rates under weak variance decay and application to functional and multinomial data. *Ann. Statist.* **48** 1214–1229. [MR4102694](#) <https://doi.org/10.1214/19-AOS1844>
- [32] MAMMEN, E. (1993). Bootstrap and wild bootstrap for high-dimensional linear models. *Ann. Statist.* **21** 255–285. [MR1212176](#) <https://doi.org/10.1214/aos/1176349025>
- [33] O’DONNELL, R., SERVEDIO, R. A. and TAN, L.-Y. (2019). Fooling polytopes. In *STOC’19—Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing* 614–625. ACM, New York. [MR4003369](#) <https://doi.org/10.1145/3313276.3316321>
- [34] ROMANO, J. P., SHAIKH, A. M. and WOLF, M. (2014). A practical two-step method for testing moment inequalities. *Econometrica* **82** 1979–2002. [MR3268401](#) <https://doi.org/10.3982/ECTA11011>
- [35] 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>
- [36] 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>
- [37] WHITE, H. (2000). A reality check for data snooping. *Econometrica* **68** 1097–1126. [MR1779145](#) <https://doi.org/10.1111/1468-0262.00152>
- [38] ZHANG, D. and WU, W. B. (2017). Gaussian approximation for high dimensional time series. *Ann. Statist.* **45** 1895–1919. [MR3718156](#) <https://doi.org/10.1214/16-AOS1512>

- [39] ZHANG, X. and CHENG, G. (2018). Gaussian approximation for high dimensional vector under physical dependence. *Bernoulli* **24** 2640–2675. [MR3779697](#) <https://doi.org/10.3150/17-BEJ939>

BOUNDS ON THE CONDITIONAL AND AVERAGE TREATMENT EFFECT WITH UNOBSERVED CONFOUNDING FACTORS

BY STEVE YADLOWSKY^{1,a}, HONGSEOK NAMKOONG^{2,b}, SANJAY BASU^{3,c},
JOHN DUCHI^{4,d} AND LU TIAN^{5,e}

¹Google Research, Brain Team, ayadlowsky@google.com

²Decision, Risk, and Operations Division, Columbia Business School, bnamkoong@gsb.columbia.edu

³Research and Development, Waymark, csanjay.basu@waymarkcare.com

⁴Statistics and Electrical Engineering, Stanford University, djduchi@stanford.edu

⁵Biomedical Data Science, Stanford University, elutian@stanford.edu

For observational studies, we study the sensitivity of causal inference when treatment assignments may depend on unobserved confounders. We develop a loss minimization approach for estimating bounds on the conditional average treatment effect (CATE) when unobserved confounders have a bounded effect on the odds ratio of treatment selection. Our approach is scalable and allows flexible use of model classes in estimation, including non-parametric and black-box machine learning methods. Based on these bounds for the CATE, we propose a sensitivity analysis for the average treatment effect (ATE). Our semiparametric estimator extends/bounds the augmented inverse propensity weighted (AIPW) estimator for the ATE under bounded unobserved confounding. By constructing a Neyman orthogonal score, our estimator of the bound for the ATE is a regular root- n estimator so long as the nuisance parameters are estimated at the $o_p(n^{-1/4})$ rate. We complement our methodology with optimality results showing that our proposed bounds are tight in certain cases. We demonstrate our method on simulated and real data examples, and show accurate coverage of our confidence intervals in practical finite sample regimes with rich covariate information.

REFERENCES

- [1] ABADIE, A. and IMBENS, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica* **74** 235–267. [MR2194325 https://doi.org/10.1111/j.1468-0262.2006.00655.x](https://doi.org/10.1111/j.1468-0262.2006.00655.x)
- [2] ARONOW, P. M. and LEE, D. K. K. (2013). Interval estimation of population means under unknown but bounded probabilities of sample selection. *Biometrika* **100** 235–240. [MR3034337 https://doi.org/10.1093/biomet/ass064](https://doi.org/10.1093/biomet/ass064)
- [3] ATHEY, S. and IMBENS, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proc. Natl. Acad. Sci. USA* **113** 7353–7360. [MR3531135 https://doi.org/10.1073/pnas.1510489113](https://doi.org/10.1073/pnas.1510489113)
- [4] ATHEY, S., TIBSHIRANI, J. and WAGER, S. (2019). Generalized random forests. *Ann. Statist.* **47** 1148–1178. [MR3909963 https://doi.org/10.1214/18-AOS1709](https://doi.org/10.1214/18-AOS1709)
- [5] BANG, H. and ROBINS, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61** 962–972. [MR2216189 https://doi.org/10.1111/j.1541-0420.2005.00377.x](https://doi.org/10.1111/j.1541-0420.2005.00377.x)
- [6] BOSCO, J. L., SILLIMAN, R. A., THWIN, S. S., GEIGER, A. M., BUIST, D. S., PROUT, M. N., YOOD, M. U., HAQUE, R., WEI, F. et al. (2010). A most stubborn bias: No adjustment method fully resolves confounding by indication in observational studies. *J. Clin. Epidemiol.* **63** 64–74.
- [7] BOYD, S. and VANDENBERGHE, L. (2004). *Convex Optimization*. Cambridge Univ. Press, Cambridge. [MR2061575 https://doi.org/10.1017/CBO9780511804441](https://doi.org/10.1017/CBO9780511804441)
- [8] BRUMBACK, B. A., HERNÁN, M. A., HANEUSE, S. J. P. A. and ROBINS, J. M. (2004). Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. *Stat. Med.* **23** 749–767.

- [9] CHEN, T. and GUESTRIN, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16 785–794. ACM, New York, NY, USA.
- [10] CHEN, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Handb. Econom.* **6** 5549–5632.
- [11] CHEN, X. and CHRISTENSEN, T. M. (2015). Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. *J. Econometrics* **188** 447–465. [MR3383220 https://doi.org/10.1016/j.jeconom.2015.03.010](https://doi.org/10.1016/j.jeconom.2015.03.010)
- [12] CHEN, X. and WHITE, H. (1999). Improved rates and asymptotic normality for nonparametric neural network estimators. *IEEE Trans. Inf. Theory* **45** 682–691. [MR1677026 https://doi.org/10.1109/18.749011](https://doi.org/10.1109/18.749011)
- [13] 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](https://doi.org/10.1111/ectj.12097)
- [14] COLOMA, P. M., TRIFIRÒ, G., SCHUEMIE, M. J., GINI, R., HERINGS, R., HIPPISEY-COX, J., MAZ-ZAGLIA, G., PICELLI, G., CORRAO, G. et al. (2012). Electronic healthcare databases for active drug safety surveillance: Is there enough leverage? *Pharmacoepidemiol. Drug Saf.* **21** 611–621.
- [15] CORNFIELD, J., HAENSZEL, W., HAMMOND, E. C., LILIENFELD, A. M., SHIMKIN, M. B. and WYN-DER, E. L. (1959). Smoking and lung cancer: Recent evidence and a discussion of some questions. *J. Natl. Cancer Inst.* **22** 173–203.
- [16] DAUBECHIES, I. (1992). *Ten Lectures on Wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics **61**. SIAM, Philadelphia, PA. [MR1162107 https://doi.org/10.1137/1.9781611970104](https://doi.org/10.1137/1.9781611970104)
- [17] FOGARTY, C. B. and SMALL, D. S. (2016). Sensitivity analysis for multiple comparisons in matched observational studies through quadratically constrained linear programming. *J. Amer. Statist. Assoc.* **111** 1820–1830. [MR3601738 https://doi.org/10.1080/01621459.2015.1120675](https://doi.org/10.1080/01621459.2015.1120675)
- [18] FRANKS, A. M., D'AMOUR, A. and FELLER, A. (2020). Flexible sensitivity analysis for observational studies without observable implications. *J. Amer. Statist. Assoc.* **115** 1730–1746. [MR4189753 https://doi.org/10.1080/01621459.2019.1604369](https://doi.org/10.1080/01621459.2019.1604369)
- [19] GEMAN, S. and HWANG, C.-R. (1982). Nonparametric maximum likelihood estimation by the method of sieves. *Ann. Statist.* **10** 401–414. [MR0653512](https://doi.org/10.1214/aos/1176340632)
- [20] GYÖRFI, L., KOHLER, M., KRZYŻAK, A. and WALK, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. Springer, Berlin.
- [21] HAHN, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica* **66** 315–331. [MR1612242 https://doi.org/10.2307/2998560](https://doi.org/10.2307/2998560)
- [22] 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](https://doi.org/10.1198/jcgs.2010.08162)
- [23] HIRANO, K., IMBENS, G. W. and RIDDER, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71** 1161–1189. [MR1995826 https://doi.org/10.1111/1468-0262.00442](https://doi.org/10.1111/1468-0262.00442)
- [24] IMBENS, G. W. (2003). Sensitivity to exogeneity assumptions in program evaluation. *Am. Econ. Rev.* **93** 126–132.
- [25] IMBENS, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Rev. Econ. Stat.* **86** 4–29.
- [26] IMBENS, G. W. and RUBIN, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge Univ. Press, New York. [MR3309951 https://doi.org/10.1017/CBO9781139025751](https://doi.org/10.1017/CBO9781139025751)
- [27] KALLUS, N., MAO, X. and ZHOU, A. (2019). Interval estimation of individual-level causal effects under unobserved confounding. In *The 22nd International Conference on Artificial Intelligence and Statistics* 2281–2290.
- [28] KALLUS, N. and ZHOU, A. (2018). Confounding-robust policy improvement. Available at <https://papers.nips.cc/paper/2018/hash/3a09a524440d44d7f19870070a5ad42f-Abstract.html>.
- [29] KENNEDY, E. H. (2020). Optimal doubly robust estimation of heterogeneous causal effects. arXiv preprint. Available at [arXiv:2004.14497 \[math.ST\]](https://arxiv.org/abs/2004.14497).
- [30] KÜNZEL, S. R., SEKHON, J. S., BICKEL, P. J. and YU, B. (2017). Meta-learners for estimating heterogeneous treatment effects using machine learning. Available at <https://www.pnas.org/doi/10.1073/pnas.1804597116>.
- [31] LEE, B. K., LESSLER, J. and STUART, E. A. (2011). Weight trimming and propensity score weighting. *PLoS ONE* **6** e18174. <https://doi.org/10.1371/journal.pone.0018174>
- [32] LUNBERGER, D. G. (1969). *Optimization by Vector Space Methods*. Wiley, New York. [MR0238472](https://doi.org/10.1002/9781118160697)
- [33] MIRATRIG, L. W., WAGER, S. and ZUBIZARRETA, J. R. (2018). Shape-constrained partial identification of a population mean under unknown probabilities of sample selection. *Biometrika* **105** 103–114. [MR3768868 https://doi.org/10.1093/biomet/asx077](https://doi.org/10.1093/biomet/asx077)

- [34] NEWHEY, W. K. (1994). Kernel estimation of partial means and a general variance estimator. *Econometric Theory* **10** 233–253. [MR1293201](#) <https://doi.org/10.1017/S0266466600008409>
- [35] NEWHEY, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica* **62** 1349–1382. [MR1303237](#) <https://doi.org/10.2307/2951752>
- [36] NEWHEY, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *J. Econometrics* **79** 147–168. [MR1457700](#) [https://doi.org/10.1016/S0304-4076\(97\)00011-0](https://doi.org/10.1016/S0304-4076(97)00011-0)
- [37] NEYMAN, J. (1959). Optimal asymptotic tests of composite statistical hypotheses. *Probab. Stat.* **416**.
- [38] 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>
- [39] NORTON, E. C., DOWD, B. E. and MACIEJEWSKI, M. L. (2018). Odds ratios-current best practice and use. *JAMA* **320** 84–85. <https://doi.org/10.1001/jama.2018.6971>
- [40] RICHARDSON, A., HUDGENS, M. G., GILBERT, P. B. and FINE, J. P. (2014). Nonparametric bounds and sensitivity analysis of treatment effects. *Statist. Sci.* **29** 596–618. [MR3300361](#) <https://doi.org/10.1214/14-STS499>
- [41] ROBINS, J. M., ROTNITZKY, A. and SCHARFSTEIN, D. O. (2000). Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials* (Minneapolis, MN, 1997). *IMA Vol. Math. Appl.* **116** 1–94. Springer, New York. [MR1731681](#) https://doi.org/10.1007/978-1-4612-1284-3_1
- [42] ROCKAFELLAR, R. T. and WETS, R. J.-B. (1998). *Variational Analysis. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **317**. Springer, Berlin. [MR1491362](#) <https://doi.org/10.1007/978-3-642-02431-3>
- [43] ROSENBAUM, P. R. (2002). *Observational Studies*, 2nd ed. *Springer Series in Statistics*. Springer, New York. [MR1899138](#) <https://doi.org/10.1007/978-1-4757-3692-2>
- [44] ROSENBAUM, P. R. (2002). Covariance adjustment in randomized experiments and observational studies. *Statist. Sci.* **17** 286–327. [MR1962487](#) <https://doi.org/10.1214/ss/1042727942>
- [45] ROSENBAUM, P. R. (2010). *Design of Observational Studies. Springer Series in Statistics*. Springer, New York. [MR2561612](#) <https://doi.org/10.1007/978-1-4419-1213-8>
- [46] ROSENBAUM, P. R. (2011). A new u-statistic with superior design sensitivity in matched observational studies. *Biometrics* **67** 1017–1027. [MR2829236](#) <https://doi.org/10.1111/j.1541-0420.2010.01535.x>
- [47] ROSENBAUM, P. R. (2014). Weighted M -statistics with superior design sensitivity in matched observational studies with multiple controls. *J. Amer. Statist. Assoc.* **109** 1145–1158. [MR3265687](#) <https://doi.org/10.1080/01621459.2013.879261>
- [48] SCHARFSTEIN, D. O., ROTNITZKY, A. and ROBINS, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J. Amer. Statist. Assoc.* **94** 1096–1146. [MR1731478](#) <https://doi.org/10.2307/2669923>
- [49] SCHUMAKER, L. L. (2007). *Spline Functions: Basic Theory*, 3rd ed. *Cambridge Mathematical Library*. Cambridge Univ. Press, Cambridge. [MR2348176](#) <https://doi.org/10.1017/CBO9780511618994>
- [50] SHEN, C., LI, X., LI, L. and WERE, M. C. (2011). Sensitivity analysis for causal inference using inverse probability weighting. *Biom. J.* **53** 822–837. [MR2861489](#) <https://doi.org/10.1002/bimj.201100042>
- [51] STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360. [MR0594650](#)
- [52] TIMAN, A. F. (1963). *Theory of Approximation of Functions of a Real Variable. A Pergamon Press Book*. The Macmillan Company, New York. [MR0192238](#)
- [53] TSIATIS, A. A. and DAVIDIAN, M. (2007). Comment: Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data [MR2420458]. *Statist. Sci.* **22** 569–573. [MR2420466](#) <https://doi.org/10.1214/07-STS227B>
- [54] VANDERWEELE, T. J. and DING, P. (2017). Sensitivity analysis in observational research: Introducing the E-value. *Ann. Intern. Med.* **167** 268–274. <https://doi.org/10.7326/M16-2607>
- [55] WAGER, S. and ATHEY, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *J. Amer. Statist. Assoc.* **113** 1228–1242. [MR3862353](#) <https://doi.org/10.1080/01621459.2017.1319839>
- [56] WAGER, S. and WALTHER, G. (2015). Adaptive concentration of regression trees, with application to random forests. Available at [arXiv:1503.06388](https://arxiv.org/abs/1503.06388) [math.ST].
- [57] YADLOWSKY, S., NAMKOONG, H., BASU, S., DUCHI, J. and TIAN, L. (2022). Supplement to “Bounds on the Conditional and Average Treatment Effect with Unobserved Confounding Factors.” <https://doi.org/10.1214/22-AOS2195SUPP>
- [58] ZHAO, Q., SMALL, D. S. and BHATTACHARYA, B. B. (2019). Sensitivity analysis for inverse probability weighting estimators via the percentile bootstrap. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 735–761. [MR3997099](#)

AFFINE-EQUIVARIANT INFERENCE FOR MULTIVARIATE LOCATION UNDER L_p LOSS FUNCTIONS

BY ALEXANDER DÜRRE^a AND DAVY PAINDAVEINE^b

ECARES and Department of Mathematics, Université libre de Bruxelles, ^aAlexander.Durre@ulb.be,
^bDavy.Paindaveine@ulb.be

We consider the fundamental problem of estimating the location of a d -variate probability measure under an L_p loss function. The naive estimator, that minimizes the usual empirical L_p risk, has a known asymptotic behavior but suffers from several deficiencies for $p \neq 2$, the most important one being the lack of equivariance under general affine transformations. In this work, we introduce a collection of L_p location estimators $\hat{\mu}_n^{p,\ell}$ that minimize the size of suitable ℓ -dimensional data-based simplices. For $\ell = 1$, these estimators reduce to the naive ones, whereas, for $\ell = d$, they are equivariant under affine transformations. Irrespective of ℓ , these estimators reduce to the sample mean for $p = 2$, whereas for $p = 1$, the estimators provide the well-known spatial median and Oja median for $\ell = 1$ and $\ell = d$, respectively. Under very mild assumptions, we derive an explicit Bahadur representation result for $\hat{\mu}_n^{p,\ell}$ and establish asymptotic normality. We prove that, quite remarkably, the asymptotic behavior of the estimators does not depend on ℓ under spherical symmetry, so that the affine equivariance for $\ell = d$ is achieved at no cost in terms of efficiency. To allow for large sample size n and/or large dimension d , we introduce a version of our estimators relying on incomplete U-statistics. Under a centro-symmetry assumption, we also define companion tests $\phi_n^{p,\ell}$ for the problem of testing the null hypothesis that the location μ of the underlying probability measure coincides with a given location μ_0 . For any p , affine invariance is achieved for $\ell = d$. For any ℓ and p , we derive explicit expressions for the asymptotic power of these tests under contiguous local alternatives, which reveals that asymptotic relative efficiencies with respect to traditional parametric Gaussian procedures for hypothesis testing coincide with those obtained for point estimation. We illustrate finite-sample relevance of our asymptotic results through Monte Carlo exercises and also treat a real data example.

REFERENCES

- ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley, New York.
- BLOM, G. (1976). Some properties of incomplete U-statistics. *Biometrika* **63** 573–580. [MR0474582](#)
- BROWN, B. M. (1983). Statistical uses of the spatial median. *J. Roy. Statist. Soc. Ser. B* **45** 25–30. [MR0701072](#)
- CARDOT, H., CÉNAC, P. and GODICHON-BAGGIONI, A. (2017). Online estimation of the geometric median in Hilbert spaces: Nonasymptotic confidence balls. *Ann. Statist.* **45** 591–614. [MR3650394](#) <https://doi.org/10.1214/16-AOS1460>
- CARDOT, H., CÉNAC, P. and ZITT, P.-A. (2013). Efficient and fast estimation of the geometric median in Hilbert spaces with an averaged stochastic gradient algorithm. *Bernoulli* **19** 18–43. [MR3019484](#) <https://doi.org/10.3150/11-BEJ390>
- DAVIES, P. L. (1987). Asymptotic behaviour of S -estimates of multivariate location parameters and dispersion matrices. *Ann. Statist.* **15** 1269–1292. [MR0902258](#) <https://doi.org/10.1214/aos/1176350505>
- DONOHU, D. L. and GASKO, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.* **20** 1803–1827. [MR1193313](#) <https://doi.org/10.1214/aos/1176348890>

MSC2020 subject classifications. Primary 62G05, 62G20; secondary 62H15, 60D05.

Key words and phrases. Affine equivariance/affine invariance, contiguous alternatives, L_p loss functions, multivariate location functionals, Oja median, random simplices, spatial median.

- DÜRRE, A. and PAINDAVEINE, D. (2021). On the consistency of incomplete U-statistics under infinite second-order moments. [arXiv:2112.14666](https://arxiv.org/abs/2112.14666).
- DÜRRE, A. and PAINDAVEINE, D. (2022). Supplement to “Affine-equivariant inference for multivariate location under L_p loss functions.” <https://doi.org/10.1214/22-AOS2199SUPP>
- ENQVIST, E. (1978). On sampling from sets of random variables with application to incomplete U-statistics. Dissertation, Lund Univ.
- FISCHER, D., MOSLER, K., MÖTTÖNEN, J., NORDHAUSEN, K., POKOTYLO, O. and VOGEL, D. (2020). Computing the Oja median in R: The package OjaNP. *J. Stat. Softw.* **92**.
- HETTMANSPERGER, T. P., MÖTTÖNEN, J. and OJA, H. (1997). Affine-invariant multivariate one-sample signed-rank tests. *J. Amer. Statist. Assoc.* **92** 1591–1600. [MR1615268 https://doi.org/10.2307/2965430](https://doi.org/10.2307/2965430)
- HETTMANSPERGER, T. P., NYBLOM, J. and OJA, H. (1994). Affine invariant multivariate one-sample sign tests. *J. Roy. Statist. Soc. Ser. B* **56** 221–234. [MR1257809](https://doi.org/10.2307/2287809)
- JANSON, S. (1984). The asymptotic distributions of incomplete U-statistics. *Probab. Theory Related Fields* **66** 495–505. [MR0753810](https://doi.org/10.1007/BF01233810)
- KONEN, D. and PAINDAVEINE, D. (2022). Multivariate ρ -quantiles: A spatial approach. *Bernoulli* **28** 1912–1934. [MR4411516 https://doi.org/10.3150/21-bej1404](https://doi.org/10.3150/21-bej1404)
- LOPUHAÄ, H. P. (1999). Asymptotics of reweighted estimators of multivariate location and scatter. *Ann. Statist.* **27** 1638–1665. [MR1742503 https://doi.org/10.1214/aos/1017939145](https://doi.org/10.1214/aos/1017939145)
- MAGYAR, A. and TYLER, D. E. (2011). The asymptotic efficiency of the spatial median for elliptically symmetric distributions. *Sankhya B* **73** 165–192. [MR2924190 https://doi.org/10.1007/s13571-011-0032-x](https://doi.org/10.1007/s13571-011-0032-x)
- MINSKER, S. (2015). Geometric median and robust estimation in Banach spaces. *Bernoulli* **21** 2308–2335. [MR3378468 https://doi.org/10.3150/14-BEJ645](https://doi.org/10.3150/14-BEJ645)
- MÖTTÖNEN, J., NORDHAUSEN, K. and OJA, H. (2010). Asymptotic theory of the spatial median. In *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in Honor of Professor Jana Jurečková*. *Inst. Math. Stat. (IMS) Collect.* **7** 182–193. IMS, Beachwood, OH. [MR2808378](https://doi.org/10.2307/2808378)
- MÖTTÖNEN, J. and OJA, H. (1995). Multivariate spatial sign and rank methods. *J. Nonparametr. Stat.* **5** 201–213. [MR1346895 https://doi.org/10.1080/10485259508832643](https://doi.org/10.1080/10485259508832643)
- MÖTTÖNEN, J., OJA, H. and TIENARI, J. (1997). On the efficiency of multivariate spatial sign and rank tests. *Ann. Statist.* **25** 542–552. [MR1439313 https://doi.org/10.1214/aos/1031833663](https://doi.org/10.1214/aos/1031833663)
- NIINIMAA, A. and OJA, H. (1995). On the influence functions of certain bivariate medians. *J. Roy. Statist. Soc. Ser. B* **57** 565–574. [MR1341324](https://doi.org/10.2307/2341324)
- OJA, H. (1983). Descriptive statistics for multivariate distributions. *Statist. Probab. Lett.* **1** 327–332. [MR0721446 https://doi.org/10.1016/0167-7152\(83\)90054-8](https://doi.org/10.1016/0167-7152(83)90054-8)
- OJA, H. (1999). Affine invariant multivariate sign and rank tests and corresponding estimates: A review. *Scand. J. Stat.* **26** 319–343. [MR1712063 https://doi.org/10.1111/1467-9469.00152](https://doi.org/10.1111/1467-9469.00152)
- OJA, H. (2010). *Multivariate Nonparametric Methods with R. An Approach Based on Spatial Signs and Ranks. Lecture Notes in Statistics* **199**. Springer, New York. [MR2598854 https://doi.org/10.1007/978-1-4419-0468-3](https://doi.org/10.1007/978-1-4419-0468-3)
- OLLILA, E., OJA, H. and HETTMANSPERGER, T. P. (2002). Estimates of regression coefficients based on the sign covariance matrix. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 447–466. [MR1924300 https://doi.org/10.1111/1467-9868.00344](https://doi.org/10.1111/1467-9868.00344)
- PAINDAVEINE, D. (2022). On the measure of anchored Gaussian simplices, with applications to multivariate medians. *Bernoulli* **28** 965–996. [MR4388926 https://doi.org/10.3150/21-bej1373](https://doi.org/10.3150/21-bej1373)
- ZUO, Y. (2003). Projection-based depth functions and associated medians. *Ann. Statist.* **31** 1460–1490. [MR2012822 https://doi.org/10.1214/aos/1065705115](https://doi.org/10.1214/aos/1065705115)

IS INFINITY THAT FAR? A BAYESIAN NONPARAMETRIC PERSPECTIVE OF FINITE MIXTURE MODELS

BY RAFFAELE ARGIENTO^{1,a} AND MARIA DE IORIO^{2,b}

¹Department of Economics, Università degli Studi di Bergamo, raffaele.argiento@unibg.it

²Yong Loo Lin School of Medicine, National University of Singapore, mdi@nus.edu.sg

Mixture models are one of the most widely used statistical tools when dealing with data from heterogeneous populations. Following a Bayesian nonparametric perspective, we introduce a new class of priors: the Normalized Independent Point Process. We investigate the probabilistic properties of this new class and present many special cases. In particular, we provide an explicit formula for the distribution of the implied partition, as well as the posterior characterization of the new process in terms of the superposition of two discrete measures. We also provide consistency results. Moreover, we design both a marginal and a conditional algorithm for finite mixture models with a random number of components. These schemes are based on an auxiliary variable MCMC, which allows handling the otherwise intractable posterior distribution and overcomes the challenges associated with the Reversible Jump algorithm. We illustrate the performance and the potential of our model in a simulation study and on real data applications.

REFERENCES

- [1] ALDOUS, D. J. (1985). Exchangeability and related topics. In *École D'été de Probabilités de Saint-Flour, XIII—1983. Lecture Notes in Math.* **1117** 1–198. Springer, Berlin. [MR0883646](#) <https://doi.org/10.1007/BFb0099421>
- [2] ARGIENTO, R., CREMASCHI, A. and VANNUCCI, M. (2020). Hierarchical normalized completely random measures to cluster grouped data. *J. Amer. Statist. Assoc.* **115** 318–333. [MR4078466](#) <https://doi.org/10.1080/01621459.2019.1594833>
- [3] ARGIENTO, R. and DE IORIO, M. (2022). Supplement to “Is infinity that far? A Bayesian nonparametric perspective of finite mixture models.” <https://doi.org/10.1214/22-AOS2201SUPP>
- [4] BODIN, B., IORIO, M. D. and ARGIENTO, R. (2020). AntMAN: Anthology of Mixture ANalysis tools.
- [5] CALLENS, T., GALBUSERA, P., MATTHYSEN, E., DURAND, E. Y., GITHIRU, M., HUYGHE, J. R. and LENS, L. (2011). Genetic signature of population fragmentation varies with mobility in seven bird species of a fragmented Kenyan cloud forest. *Mol. Ecol.* **20** 1829–1844.
- [6] CHARALAMBIDES, C. A. (2005). *Combinatorial Methods in Discrete Distributions*. Wiley Series in Probability and Statistics. Wiley Interscience, Hoboken, NJ. [MR2131068](#) <https://doi.org/10.1002/0471733180>
- [7] DALEY, D. J. and VERE-JONES, D. (2008). *An Introduction to the Theory of Point Processes. Vol. II: General Theory and Structure*, 2nd ed. *Probability and Its Applications (New York)*. Springer, New York. [MR2371524](#) <https://doi.org/10.1007/978-0-387-49835-5>
- [8] DELLAPORTAS, P. and PAPAGEORGIOU, I. (2006). Multivariate mixtures of normals with unknown number of components. *Stat. Comput.* **16** 57–68. [MR2224189](#) <https://doi.org/10.1007/s11222-006-5338-6>
- [9] DEVROYE, L. (2009). Random variate generation for exponentially and polynomially tilted stable distributions. *ACM Trans. Model. Comput. Simul.* **19** 18.
- [10] ESCOBAR, M. D. and WEST, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90** 577–588. [MR1340510](#)
- [11] FALUSH, D., STEPHENS, M. and PRITCHARD, J. K. (2003). Inference of population structure using multi-locus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164** 1567–1587.

MSC2020 subject classifications. Primary 62F15, 60G57; secondary 62G07, 92C20.

Key words and phrases. Bayesian mixture models, Bayesian clustering, Dirichlet process, mixture of finite mixtures, Markov chain Monte Carlo methods.

- [12] FAVARO, S., NIPOTI, B. and TEH, Y. W. (2015). Random variate generation for Laguerre-type exponentially tilted α -stable distributions. *Electron. J. Stat.* **9** 1230–1242. [MR3355756](#) <https://doi.org/10.1214/15-EJS1033>
- [13] FAVARO, S. and TEH, Y. W. (2013). MCMC for normalized random measure mixture models. *Statist. Sci.* **28** 335–359. [MR3135536](#) <https://doi.org/10.1214/13-STS422>
- [14] FRÜHWIRTH-SCHNATTER, S. (2006). *Finite Mixture and Markov Switching Models. Springer Series in Statistics*. Springer, New York. [MR2265601](#)
- [15] FRÜHWIRTH-SCHNATTER, S., CELEUX, G. and ROBERT, C. P., eds. (2019). *Handbook of Mixture Analysis. Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. CRC Press, Boca Raton, FL. [MR3889980](#)
- [16] FRÜHWIRTH-SCHNATTER, S. and MALSINER-WALLI, G. (2019). From here to infinity: Sparse finite versus Dirichlet process mixtures in model-based clustering. *Adv. Data Anal. Classif.* **13** 33–64. [MR3935190](#) <https://doi.org/10.1007/s11634-018-0329-y>
- [17] FRÜHWIRTH-SCHNATTER, S., MALSINER-WALLI, G. and GRÜN, B. (2021). Generalized mixtures of finite mixtures and telescoping sampling. *Bayesian Anal.* **16** 1279–1307. [MR4381135](#) <https://doi.org/10.1214/21-BA1294>
- [18] GALBUSERA, P., LENS, L., SCHENCK, T., WAIYAKI, E. and MATTHYSEN, E. (2000). Genetic variability and gene flow in the globally, critically-endangered Taita thrush. *Conserv. Genet.* **1** 45–55.
- [19] GELFAND, A. E. and KOTTAS, A. (2002). A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models. *J. Comput. Graph. Statist.* **11** 289–305. [MR1938136](#) <https://doi.org/10.1198/106186002760180518>
- [20] GOUTIS, C. and ROBERT, C. P. (1998). Model choice in generalised linear models: A Bayesian approach via Kullback–Leibler projections. *Biometrika* **85** 29–37. [MR1627250](#) <https://doi.org/10.1093/biomet/85.1.29>
- [21] GREEN, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82** 711–732. [MR1380810](#) <https://doi.org/10.1093/biomet/82.4.711>
- [22] GRÜN, B., MALSINER-WALLI, G. and FRÜHWIRTH-SCHNATTER, S. (2022). How many data clusters are in the Galaxy data set? *Adv. Data Anal. Classif.* **16** 325–349. [MR4440853](#) <https://doi.org/10.1007/s11634-021-00461-8>
- [23] GUHA, A., HO, N. and NGUYEN, X. (2021). On posterior contraction of parameters and interpretability in Bayesian mixture modeling. *Bernoulli* **27** 2159–2188. [MR4303879](#) <https://doi.org/10.3150/20-BEJ1275>
- [24] HOFERT, M. (2011). Sampling exponentially tilted stable distributions. *ACM Trans. Model. Comput. Simul.* **22** Art. 3, 11. [MR2955859](#) <https://doi.org/10.1145/2043635.2043638>
- [25] HUELSENBECK, J. P. and ANDOLFATTO, P. (2007). Inference of population structure under a Dirichlet process model. *Genetics* **175** 1787–1802.
- [26] HUELSENBECK, J. P., ANDOLFATTO, P. and HUELSENBECK, E. T. (2011). Structurama: Bayesian inference of population structure. *Evol. Bioinform.* **7** 55–59. <https://doi.org/10.4137/EBO.S6761>
- [27] ISHWARAN, H. and JAMES, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.* **96** 161–173. [MR1952729](#) <https://doi.org/10.1198/016214501750332758>
- [28] ISHWARAN, H. and ZAREPOUR, M. (2002). Exact and approximate sum representations for the Dirichlet process. *Canad. J. Statist.* **30** 269–283. [MR1926065](#) <https://doi.org/10.2307/3315951>
- [29] JACOD, J. and SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **288**. Springer, Berlin. [MR0959133](#) <https://doi.org/10.1007/978-3-662-02514-7>
- [30] JAMES, L. F., LIJOI, A. and PRÜNSTER, I. (2009). Posterior analysis for normalized random measures with independent increments. *Scand. J. Stat.* **36** 76–97. [MR2508332](#) <https://doi.org/10.1111/j.1467-9469.2008.00609.x>
- [31] KALLI, M., GRIFFIN, J. E. and WALKER, S. G. (2011). Slice sampling mixture models. *Stat. Comput.* **21** 93–105. [MR2746606](#) <https://doi.org/10.1007/s11222-009-9150-y>
- [32] KOMÁREK, A. (2009). A new R package for Bayesian estimation of multivariate normal mixtures allowing for selection of the number of components and interval-censored data. *Comput. Statist. Data Anal.* **53** 3932–3947. [MR2744295](#) <https://doi.org/10.1016/j.csda.2009.05.006>
- [33] LIJOI, A. and PRÜNSTER, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics. Camb. Ser. Stat. Probab. Math.* **28** 80–136. Cambridge Univ. Press, Cambridge. [MR2730661](#)
- [34] MALSINER-WALLI, G., FRÜHWIRTH-SCHNATTER, S. and GRÜN, B. (2016). Model-based clustering based on sparse finite Gaussian mixtures. *Stat. Comput.* **26** 303–324. [MR3439375](#) <https://doi.org/10.1007/s11222-014-9500-2>

- [35] MALSINER-WALLI, G., FRÜHWIRTH-SCHNATTER, S. and GRÜN, B. (2017). Identifying mixtures of mixtures using Bayesian estimation. *J. Comput. Graph. Statist.* **26** 285–295. MR3640186 <https://doi.org/10.1080/10618600.2016.1200472>
- [36] MCLACHLAN, G. and PEEL, D. (2000). *Finite Mixture Models*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. Wiley Interscience, New York. MR1789474 <https://doi.org/10.1002/0471721182>
- [37] MILLER, J. W. and HARRISON, M. T. (2018). Mixture models with a prior on the number of components. *J. Amer. Statist. Assoc.* **113** 340–356. MR3803469 <https://doi.org/10.1080/01621459.2016.1255636>
- [38] NEAL, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *J. Comput. Graph. Statist.* **9** 249–265. MR1823804 <https://doi.org/10.2307/1390653>
- [39] 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>
- [40] NOBILE, A. (1994). *Bayesian Analysis of Finite Mixture Distributions*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—Carnegie Mellon Univ. MR2692049
- [41] NOBILE, A. (2004). On the posterior distribution of the number of components in a finite mixture. *Ann. Statist.* **32** 2044–2073. MR2102502 <https://doi.org/10.1214/009053604000000788>
- [42] PITMAN, J. (1996). Blackwell–Macqueen urn scheme. In *Statistics, Probability, and Game Theory: Papers in Honor of David Blackwell* **30** 245.
- [43] PITMAN, J. (2006). *Combinatorial Stochastic Processes*. Lecture Notes in Math. **1875**. Springer, Berlin. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002, With a foreword by Jean Picard. MR2245368
- [44] POLLARD, H. (1946). The representation of e^{-x^λ} as a Laplace integral. *Bull. Amer. Math. Soc.* **52** 908–910. MR0018286 <https://doi.org/10.1090/S0002-9904-1946-08672-3>
- [45] PRITCHARD, J. K., STEPHENS, M. and DONNELLY, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* **155** 945–959.
- [46] PRITCHARD, J. K. and WEN, W. (2003). Documentation for STRUCTURE software: Version 2.3.X.
- [47] RICHARDSON, S. and GREEN, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *J. Roy. Statist. Soc. Ser. B* **59** 731–792. MR1483213 <https://doi.org/10.1111/1467-9868.00095>
- [48] ROEDER, K. (1990). Density estimation with confidence sets exemplified by superclusters and voids in the galaxies. *J. Amer. Statist. Assoc.* **85** 617–624.
- [49] ROUSSEAU, J. and MENGENSEN, K. (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 689–710. MR2867454 <https://doi.org/10.1111/j.1467-9868.2011.00781.x>
- [50] STEPHENS, M. (2000). Bayesian analysis of mixture models with an unknown number of components—an alternative to reversible jump methods. *Ann. Statist.* **28** 40–74. MR1762903 <https://doi.org/10.1214/aos/1016120364>

GLOBAL AND INDIVIDUALIZED COMMUNITY DETECTION IN INHOMOGENEOUS MULTILAYER NETWORKS

BY SHUXIAO CHEN^{1,a}, SIFAN LIU^{2,c} AND ZONGMING MA^{1,b}

¹Department of Statistics, University of Pennsylvania, ^ashuxiaoc@wharton.upenn.edu, ^bzongming@wharton.upenn.edu

²Department of Statistics, Stanford University, ^csfliu@stanford.edu

In network applications, it has become increasingly common to obtain datasets in the form of multiple networks observed on the same set of subjects, where each network is obtained in a related but different experiment condition or application scenario. Such datasets can be modeled by multilayer networks where each layer is a separate network itself while different layers are associated and share some common information. The present paper studies community detection in a stylized yet informative inhomogeneous multilayer network model. In our model, layers are generated by different stochastic block models, the community structures of which are (random) perturbations of a common global structure while the connecting probabilities in different layers are not related. Focusing on the symmetric two block case, we establish minimax rates for both *global estimation* of the common structure and *individualized estimation* of layerwise community structures. Both minimax rates have sharp exponents. In addition, we provide an efficient algorithm that is simultaneously asymptotic minimax optimal for both estimation tasks under mild conditions. The optimal rates depend on the *parity* of the number of most informative layers, a phenomenon that is caused by inhomogeneity across layers. The method is extended to handle multiple and potentially asymmetric community cases. We demonstrate its effectiveness on both simulated examples and a real multimodal single-cell dataset.

REFERENCES

- [1] ABBE, E. (2017). Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** Paper No. 177, 86. [MR3827065](#)
- [2] ABBE, E., BANDEIRA, A. S. and HALL, G. (2016). Exact recovery in the stochastic block model. *IEEE Trans. Inf. Theory* **62** 471–487. [MR3447993](#) <https://doi.org/10.1109/TIT.2015.2490670>
- [3] ABBE, E. and SANDON, C. (2015). Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In 2015 *IEEE 56th Annual Symposium on Foundations of Computer Science—FOCS 2015* 670–688. IEEE Computer Soc., Los Alamitos, CA. [MR3473334](#) <https://doi.org/10.1109/FOCS.2015.47>
- [4] ABBE, E. and SANDON, C. (2018). Proof of the achievability conjectures for the general stochastic block model. *Comm. Pure Appl. Math.* **71** 1334–1406. [MR3812075](#) <https://doi.org/10.1002/cpa.21719>
- [5] ARROYO, J., ATHREYA, A., CAPE, J., CHEN, G., PRIEBE, C. E. and VOGELSTEIN, J. T. (2021). Inference for multiple heterogeneous networks with a common invariant subspace. *J. Mach. Learn. Res.* **22** Paper No. 142, 49. [MR4318498](#)
- [6] ASSOUD, P. (1983). Deux remarques sur l’estimation. *C. R. Acad. Sci. Paris Sér. I Math.* **296** 1021–1024. [MR0777600](#)
- [7] BARKAS, N., PETUKHOV, V., NIKOLAEVA, D., LOZINSKY, Y., DEMHARTER, S., KHODOSEVICH, K. and KHARCHENKO, P. V. (2019). Joint analysis of heterogeneous single-cell RNA-seq dataset collections. *Nat. Methods* **16** 695–698. <https://doi.org/10.1038/s41592-019-0466-z>
- [8] BHATTACHARYYA, S. and CHATTERJEE, S. (2018). Spectral clustering for multiple sparse networks: I. arXiv preprint [arXiv:1805.10594](https://arxiv.org/abs/1805.10594).

MSC2020 subject classifications. Primary 62H30, 91D30; secondary 62C20, 90B15.

Key words and phrases. Integrative data analysis, minimax rate, Rényi divergence, spectral clustering, stochastic block model.

- [9] BOCCALETTI, S., BIANCONI, G., CRIADO, R., DEL GENIO, C. I., GÓMEZ-GARDEÑES, J., ROMANCE, M., SENDIÑA-NADAL, I., WANG, Z. and ZANIN, M. (2014). The structure and dynamics of multilayer networks. *Phys. Rep.* **544** 1–122. [MR3270140 https://doi.org/10.1016/j.physrep.2014.07.001](https://doi.org/10.1016/j.physrep.2014.07.001)
- [10] BORDENAVE, C., LELARGE, M. and MASSOULIÉ, L. (2015). Non-backtracking spectrum of random graphs: Community detection and non-regular Ramanujan graphs. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science—FOCS 2015* 1347–1357. IEEE Computer Soc., Los Alamitos, CA. [MR3473374 https://doi.org/10.1109/FOCS.2015.86](https://doi.org/10.1109/FOCS.2015.86)
- [11] CARUANA, R. (1997). Multitask learning. *Mach. Learn.* **28** 41–75.
- [12] CHEN, S., LIU, S. and MA, Z. (2022). Supplement to “Global and individualized community detection in inhomogeneous multilayer networks.” <https://doi.org/10.1214/22-AOS2202SUPP>
- [13] CHEN, Y. and XU, J. (2016). Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *J. Mach. Learn. Res.* **17** Paper No. 27, 57. [MR3491121](https://doi.org/10.1137/16M1121)
- [14] CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Stat.* **23** 493–507. [MR0057518 https://doi.org/10.1214/aoms/1177729330](https://doi.org/10.1214/aoms/1177729330)
- [15] CURRAN, P. J. and HUSSONG, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychol. Methods* **14** 81–100. <https://doi.org/10.1037/a0015914>
- [16] DECELLE, A., KRZAKALA, F., MOORE, C. and ZDEBOROVÁ, L. (2011). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* **84** 066106.
- [17] DE DOMENICO, M. (2017). Multilayer modeling and analysis of human brain networks. *GigaScience* **6** 1–8. <https://doi.org/10.1093/gigascience/gix004>
- [18] DE DOMENICO, M., LANCICHINETTI, A., ARENAS, A. and ROSVALL, M. (2015). Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Phys. Rev. X* **5** 011027.
- [19] DE DOMENICO, M., SOLÉ-RIBALTA, A., COZZO, E., KIVELÄ, M., MORENO, Y., PORTER, M. A., GÓMEZ, S. and ARENAS, A. (2013). Mathematical formulation of multilayer networks. *Phys. Rev. X* **3** 041022.
- [20] DICKISON, M. E., MAGNANI, M. and ROSSI, L. (2016). *Multilayer Social Networks*. Cambridge Univ. Press, New York. [MR3618060 https://doi.org/10.1017/CBO9781139941907](https://doi.org/10.1017/CBO9781139941907)
- [21] GAO, C. and MA, Z. (2019). Testing equivalence of clustering. arXiv preprint [arXiv:1910.12797](https://arxiv.org/abs/1910.12797).
- [22] GAO, C. and MA, Z. (2021). Minimax rates in network analysis: Graphon estimation, community detection and hypothesis testing. *Statist. Sci.* **36** 16–33. [MR4194201 https://doi.org/10.1214/19-STS736](https://doi.org/10.1214/19-STS736)
- [23] GAO, C., MA, Z., ZHANG, A. Y. and ZHOU, H. H. (2017). Achieving optimal misclassification proportion in stochastic block models. *J. Mach. Learn. Res.* **18** Paper No. 60, 45. [MR3687603](https://doi.org/10.1137/17M1164052)
- [24] GAO, C., MA, Z., ZHANG, A. Y. and ZHOU, H. H. (2018). Community detection in degree-corrected block models. *Ann. Statist.* **46** 2153–2185. [MR3845014 https://doi.org/10.1214/17-AOS1615](https://doi.org/10.1214/17-AOS1615)
- [25] GAO, L. L., BIEN, J. and WITTEN, D. (2020). Are clusterings of multiple data views independent? *Biostatistics* **21** 692–708. [MR4164052 https://doi.org/10.1093/biostatistics/kxz001](https://doi.org/10.1093/biostatistics/kxz001)
- [26] GAO, L. L., WITTEN, D. and BIEN, J. (2019). Testing for association in multi-view network data. arXiv preprint [arXiv:1909.11640](https://arxiv.org/abs/1909.11640).
- [27] GARCIA, J. O., ASHOURVAN, A., MULDOON, S. F., VETTEL, J. M. and BASSETT, D. S. (2018). Applications of community detection techniques to brain graphs: Algorithmic considerations and implications for neural function. *Proc. IEEE Inst. Electr. Electron. Eng.* **106** 846–867. <https://doi.org/10.1109/JPROC.2017.2786710>
- [28] GOLDENBERG, A., ZHENG, A. X., FIENBERG, S. E. and AIROLDI, E. M. (2010). *A Survey of Statistical Network Models*. Now Publishers Inc..
- [29] HAGHVERDI, L., LUN, A. T. L., MORGAN, M. D. and MARIONI, J. C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36** 421–427. <https://doi.org/10.1038/nbt.4091>
- [30] HAO, Y., HAO, S., ANDERSEN-NISSEN, E., MAUCK III, W. M., ZHENG, S., BUTLER, A., LEE, M. J., WILK, A. J., DARBY, C. et al. (2021). Integrated analysis of multimodal single-cell data. *Cell*.
- [31] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. [MR0144363](https://doi.org/10.1137/16M1121)
- [32] HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. [MR0718088 https://doi.org/10.1016/0378-8733\(83\)90021-7](https://doi.org/10.1016/0378-8733(83)90021-7)
- [33] JACKSON, M. O. (2008). *Social and Economic Networks*. Princeton Univ. Press, Princeton, NJ. [MR2435744](https://doi.org/10.1137/16M1121)
- [34] JACKSON, M. O. (2011). An overview of social networks and economic applications. In *Handbook of Social Economics* **1** 511–585. Elsevier.

- [35] JING, B.-Y., LI, T., LYU, Z. and XIA, D. (2021). Community detection on mixture multilayer networks via regularized tensor decomposition. *Ann. Statist.* **49** 3181–3205. [MR4352527](#) <https://doi.org/10.1214/21-aos2079>
- [36] KAIROUZ, P., MCMAHAN, H. B., AVENT, B., BELLET, A., BENNIS, M., BHAGOJI, A. N., BONAWITZ, K., CHARLES, Z., CORMODE, G. et al. (2019). Advances and open problems in federated learning. arXiv preprint [arXiv:1912.04977](#).
- [37] KIVELÄ, M., ARENAS, A., BARTHELEMY, M., GLEESON, J. P., MORENO, Y. and PORTER, M. A. (2014). Multilayer networks. *J. Complex Netw.* **2** 203–271.
- [38] KRIVELEVICH, M. and SUDAKOV, B. (2003). The largest eigenvalue of sparse random graphs. *Combin. Probab. Comput.* **12** 61–72. [MR1967486](#) <https://doi.org/10.1017/S0963548302005424>
- [39] KUMAR, A., RAI, P. and DAUME, H. (2011). Co-regularized multi-view spectral clustering. In *Advances in Neural Information Processing Systems* 1413–1421.
- [40] KUMAR, A., SABHARWAL, Y. and SEN, S. (2004). A simple linear time $(1 + \varepsilon)$ -approximation algorithm for k-means clustering in any dimensions. In *45th Annual IEEE Symposium on Foundations of Computer Science* 454–462. IEEE.
- [41] LE, C. M., LEVIN, K. and LEVINA, E. (2018). Estimating a network from multiple noisy realizations. *Electron. J. Stat.* **12** 4697–4740. [MR3894068](#) <https://doi.org/10.1214/18-ejs1521>
- [42] LE, C. M., LEVINA, E. and VERSHYNIN, R. (2017). Concentration and regularization of random graphs. *Random Structures Algorithms* **51** 538–561. [MR3689343](#) <https://doi.org/10.1002/rsa.20713>
- [43] LEI, J. (2020). Tail bounds for matrix quadratic forms and bias adjusted spectral clustering in multi-layer stochastic block models. arXiv preprint [arXiv:2003.08222](#).
- [44] LEI, J., CHEN, K. and LYNCH, B. (2020). Consistent community detection in multi-layer network data. *Biometrika* **107** 61–73. [MR4064140](#) <https://doi.org/10.1093/biomet/asz068>
- [45] LI, X., CHEN, Y. and XU, J. (2021). Convex relaxation methods for community detection. *Statist. Sci.* **36** 2–15. [MR4194200](#) <https://doi.org/10.1214/19-STS715>
- [46] LÜ, L. and ZHOU, T. (2011). Link prediction in complex networks: A survey. *Phys. A, Stat. Mech. Appl.* **390** 1150–1170.
- [47] MA, C., WANG, K., CHI, Y. and CHEN, Y. (2018). Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval and matrix completion. In *International Conference on Machine Learning* 3345–3354. PMLR.
- [48] MONTANARI, A. (2015). Finding one community in a sparse graph. *J. Stat. Phys.* **161** 273–299. [MR3401018](#) <https://doi.org/10.1007/s10955-015-1338-2>
- [49] MOSSEL, E., NEEMAN, J. and SLY, A. (2012). Stochastic block models and reconstruction. arXiv preprint [arXiv:1202.1499](#).
- [50] MOSSEL, E., NEEMAN, J. and SLY, A. (2015). Reconstruction and estimation in the planted partition model. *Probab. Theory Related Fields* **162** 431–461. [MR3383334](#) <https://doi.org/10.1007/s00440-014-0576-6>
- [51] MOSSEL, E., NEEMAN, J. and SLY, A. (2015). Consistency thresholds for the planted bisection model. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing* 69–75.
- [52] MOSSEL, E., NEEMAN, J. and SLY, A. (2018). A proof of the block model threshold conjecture. *Combinatorica* **38** 665–708. [MR3876880](#) <https://doi.org/10.1007/s00493-016-3238-8>
- [53] MUCHA, P. J., RICHARDSON, T., MACON, K., PORTER, M. A. and ONNELA, J.-P. (2010). Community structure in time-dependent, multiscale, and multiplex networks. *Science* **328** 876–878. With supplementary material available online. [MR2662590](#) <https://doi.org/10.1126/science.1184819>
- [54] NEEMAN, J. and NETRAPALLI, P. (2014). Non-reconstructability in the stochastic block model. arXiv preprint [arXiv:1404.6304](#).
- [55] NEUMANN, J. v. (1928). Zur Theorie der Gesellschaftsspiele. *Math. Ann.* **100** 295–320. [MR1512486](#) <https://doi.org/10.1007/BF01448847>
- [56] PAUL, S. and CHEN, Y. (2016). Consistent community detection in multi-relational data through restricted multi-layer stochastic blockmodel. *Electron. J. Stat.* **10** 3807–3870. [MR3579677](#) <https://doi.org/10.1214/16-EJS1211>
- [57] PAUL, S. and CHEN, Y. (2020). A random effects stochastic block model for joint community detection in multiple networks with applications to neuroimaging. *Ann. Appl. Stat.* **14** 993–1029. [MR4117838](#) <https://doi.org/10.1214/20-AOAS1339>
- [58] PAUL, S. and CHEN, Y. (2020). Spectral and matrix factorization methods for consistent community detection in multi-layer networks. *Ann. Statist.* **48** 230–250. [MR4065160](#) <https://doi.org/10.1214/18-AOS1800>
- [59] PEIXOTO, T. P. (2015). Inferring the mesoscale structure of layered, edge-valued, and time-varying networks. *Phys. Rev. E* **92** 042807.

- [60] RÉNYI, A. (1961). On measures of entropy and information. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I* 547–561. Univ. California Press, Berkeley, CA. [MR0132570](#)
- [61] SION, M. (1958). On general minimax theorems. *Pacific J. Math.* **8** 171–176. [MR0097026](#)
- [62] STANLEY, N., SHAI, S., TAYLOR, D. and MUCHA, P. J. (2016). Clustering network layers with the strata multilayer stochastic block model. *IEEE Trans. Netw. Sci. Eng.* **3** 95–105. [MR3515211](#) <https://doi.org/10.1109/TNSE.2016.2537545>
- [63] STOECKIUS, M., HAFEMEISTER, C., STEPHENSON, W., HOUCK-LOOMIS, B., CHATTOPADHYAY, P. K., SWERDLOW, H., SATIJA, R. and SMIBERT, P. (2017). Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14** 865–868. <https://doi.org/10.1038/nmeth.4380>
- [64] STUART, T., BUTLER, A., HOFFMAN, P., HAFEMEISTER, C., PAPALEXI, E., MAUCK III, W. M., HAO, Y., STOECKIUS, M., SMIBERT, P. et al. (2019). Comprehensive integration of single-cell data. *Cell* **177** 1888–1902.
- [65] VARADHAN, S. R. S. (1984). *Large Deviations and Applications*. CBMS-NSF Regional Conference Series in Applied Mathematics **46**. SIAM, Philadelphia, PA. [MR0758258](#) <https://doi.org/10.1137/1.9781611970241.bm>
- [66] WANG, M. and ZENG, Y. (2019). Multiway clustering via tensor block models. In *Advances in Neural Information Processing Systems* 715–725.
- [67] WASSERMAN, S. and FAUST, K. (1994). *Social Network Analysis: Methods and Applications* **8**. Cambridge Univ. Press.
- [68] XU, M., JOG, V. and LOH, P.-L. (2020). Optimal rates for community estimation in the weighted stochastic block model. *Ann. Statist.* **48** 183–204. [MR4065158](#) <https://doi.org/10.1214/18-AOS1797>
- [69] YUN, S.-Y. and PROUTIERE, A. (2014). Community detection via random and adaptive sampling. In *Conference on Learning Theory* 138–175.
- [70] YUN, S.-Y. and PROUTIERE, A. (2016). Optimal cluster recovery in the labeled stochastic block model. In *Advances in Neural Information Processing Systems* 965–973.
- [71] ZHANG, A. Y. and ZHOU, H. H. (2016). Minimax rates of community detection in stochastic block models. *Ann. Statist.* **44** 2252–2280. [MR3546450](#) <https://doi.org/10.1214/15-AOS1428>
- [72] ZHANG, X., XUE, S. and ZHU, J. (2020). A flexible latent space model for multilayer networks. In *International Conference on Machine Learning PMLR*.
- [73] ZHU, B., CHEN, S., BAI, Y., CHEN, H., MUKHERJEE, N., VAZQUEZ, G., MCILWAIN, D. R., TZANKOV, A., LEE, I. T. et al. (2021). Robust single-cell matching and multi-modal analysis using shared and distinct features reveals orchestrated immune responses. *bioRxiv*.

ASYMPTOTIC ANALYSIS OF SYNCHROSQUEEZING TRANSFORM—TOWARD STATISTICAL INFERENCE WITH NONLINEAR-TYPE TIME-FREQUENCY ANALYSIS

BY MATT SOURISSEAU^{1,a}, HAU-TIENG WU^{2,b} AND ZHOU ZHOU^{3,c}

¹Department of Mathematics, University of Toronto, ^asourisse@math.utoronto.ca

²Department of Mathematics and Department of Statistical Science, Duke University, ^bhauwu@math.duke.edu

³Department of Statistics, University of Toronto, ^czhou@utstat.toronto.edu

We provide a statistical analysis of a tool in nonlinear-type time-frequency analysis, the synchrosqueezing transform (SST), for both the null and nonnull cases. The intricate nonlinear interaction of different quantities in SST is quantified by carefully analyzing relevant multivariate complex Gaussian random variables. Specifically, we provide the quotient distribution of dependent and improper complex Gaussian random variables. Then a central limit theorem result for SST is established. As an example, we provide a block bootstrap scheme based on the established SST theory to test if a given time series contains oscillatory components.

REFERENCES

- [1] ALIAN, A. A. and PHOTOPLETHYSMOGRAPHY, K. H. S. (2014). *Bailliere's Best Pract. Res., Clin. Anaesthesiol.* **28** 395–406.
- [2] AUGER, F. and FLANDRIN, P. (1995). Improving the readability of time-frequency and time-scale representations by the reassignment method. *IEEE Trans. Signal Process.* **43** 1068–1089.
- [3] BAXLEY, R. J., WALKENHORST, B. T. and ACOSTA-MARUM, G. (2010). Complex Gaussian ratio distribution with applications for error rate calculation in fading channels with imperfect CSI. In *IEEE Global Communications Conference*.
- [4] BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. [MR1325392](#)
- [5] BERNSTEIN, D. S. (2009). *Matrix Mathematics: Theory, Facts, and Formulas*, 2nd ed. Princeton Univ. Press, Princeton, NJ. [MR2513751](#) <https://doi.org/10.1515/9781400833344>
- [6] BROCKWELL, P. J. (2001). Continuous-time ARMA processes. In *Stochastic Processes: Theory and Methods. Handbook of Statist.* **19** 249–276. North-Holland, Amsterdam. [MR1861726](#) [https://doi.org/10.1016/S0169-7161\(01\)19011-5](https://doi.org/10.1016/S0169-7161(01)19011-5)
- [7] BROCKWELL, P. J. and DAVIS, R. A. (2016). *Introduction to Time Series and Forecasting*, 3rd ed. *Springer Texts in Statistics*. Springer, Cham. [MR3526245](#) <https://doi.org/10.1007/978-3-319-29854-2>
- [8] CANDÈS, E. J., CHARLTON, P. R. and HELGASON, H. (2008). Detecting highly oscillatory signals by chirplet path pursuit. *Appl. Comput. Harmon. Anal.* **24** 14–40. [MR2379113](#) <https://doi.org/10.1016/j.acha.2007.04.003>
- [9] CHASSANDE-MOTTIN, E., FLANDRIN, P. and AUGER, F. (1998). On the statistics of spectrogram reassignment vectors. *Multidimens. Syst. Signal Process.* **9** 355–362.
- [10] CHEN, Y.-C., CHENG, M.-Y. and WU, H.-T. (2014). Non-parametric and adaptive modelling of dynamic periodicity and trend with heteroscedastic and dependent errors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 651–682. [MR3210732](#) <https://doi.org/10.1111/rssb.12039>
- [11] CHEN, Z. and WU, H.-T. (2021). Disentangling modes with crossover instantaneous frequencies by synchrosqueezed chirplet transforms, from theory to application. *Appl. Comput. Harmon. Anal.* **62** 84–122. <https://doi.org/10.1016/j.acha.2022.08.004>
- [12] DAUBECHIES, I. (1992). *Ten Lectures on Wavelets. CBMS-NSF Regional Conference Series in Applied Mathematics* **61**. SIAM, Philadelphia, PA. [MR1162107](#) <https://doi.org/10.1137/1.9781611970104>

MSC2020 subject classifications. 60K35.

Key words and phrases. Synchrosqueezing transform, time series analysis, kernel regression, complex Gaussian random vector, CR calculation.

- [13] DAUBECHIES, I., LU, J. and WU, H.-T. (2011). Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool. *Appl. Comput. Harmon. Anal.* **30** 243–261. [MR2754779](#) <https://doi.org/10.1016/j.acha.2010.08.002>
- [14] DAUBECHIES, I., WANG, Y. and WU, H. (2016). ConceFT: Concentration of frequency and time via a multitapered synchrosqueezed transform. *Philos. Trans. R. Soc. Lond. Ser. A, Math. Phys. Sci.* **374** 20150193, 19. [MR3479895](#) <https://doi.org/10.1098/rsta.2015.0193>
- [15] FAN, J. and YAO, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer Series in Statistics. Springer, New York. [MR1964455](#) <https://doi.org/10.1007/b97702>
- [16] FLANDRIN, P. (1999). *Time-Frequency/Time-Scale Analysis. Wavelet Analysis and Its Applications* **10**. Academic Press, San Diego, CA. [MR1681043](#)
- [17] GABOR, D. (1946). Theory of communication. Part 1: The analysis of information. *J. Inst. Electr. Eng.*, **3** **93** 429–441.
- [18] GEL'FAND, I. M. and VILENKIN, N. YA. (1964). *Generalized Functions. Vol. 4: Applications of Harmonic Analysis*. Academic Press, New York. [MR0173945](#)
- [19] HALLIN, M. (1978). Mixed autoregressive-moving average multivariate processes with time-dependent coefficients. *J. Multivariate Anal.* **8** 567–572. [MR0520964](#) [https://doi.org/10.1016/0047-259X\(78\)90034-9](https://doi.org/10.1016/0047-259X(78)90034-9)
- [20] HAMILTON, J. D. (1994). *Time Series Analysis*. Princeton Univ. Press, Princeton, NJ. [MR1278033](#)
- [21] HUANG, N. E., SHEN, Z., LONG, S. R., WU, M. C., SHIH, H. H., ZHENG, Q., YEN, N.-C., TUNG, C. C. and LIU, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **454** 903–995. [MR1631591](#) <https://doi.org/10.1098/rspa.1998.0193>
- [22] HUANG, Z., ZHANG, J., ZHAO, T. and SUN, Y. (2015). Synchrosqueezing s-transform and its application in seismic spectral decomposition. *IEEE Trans. Geosci. Remote Sens.* **PP(99)** 1–9.
- [23] KODERA, K., VILLEDARY, C. D. and GENDRIN, R. (1976). A new method for the numerical analysis of non-stationary signals. *Phys. Earth Planet. Inter.* **12** 142–150.
- [24] KORALOV, L. B. and SINAI, Y. G. (2007). *Theory of Probability and Random Processes*, 2nd ed. *Universitext*. Springer, Berlin. [MR2343262](#) <https://doi.org/10.1007/978-3-540-68829-7>
- [25] KREUTZ-DELGADO, K. (2009). The complex gradient operator and the $\mathbb{C}\mathbb{R}$ -calculus. ArXiv e-prints.
- [26] LEBEDEV, N. N. and SILVERMAN, R. A. (1972). *Special Functions & Their Applications*. Dover Books on Mathematics, Dover.
- [27] LIN, C.-Y., SU, L. and WU, H.-T. (2018). Wave-shape function analysis: When cepstrum meets time-frequency analysis. *J. Fourier Anal. Appl.* **24** 451–505. [MR3776331](#) <https://doi.org/10.1007/s00041-017-9523-0>
- [28] OBERLIN, T., MEIGNEN, S. and PERRIER, V. (2015). Second-order synchrosqueezing transform or invertible reassignment? Towards ideal time-frequency representations. *IEEE Trans. Signal Process.* **63** 1335–1344. [MR3312176](#) <https://doi.org/10.1109/TSP.2015.2391077>
- [29] PHAM-GIA, T., TURKKAN, N. and MARCHAND, E. (2006). Density of the ratio of two normal random variables and applications. *Comm. Statist. Theory Methods* **35** 1569–1591. [MR2328495](#) <https://doi.org/10.1080/03610920600683689>
- [30] PICINBONO, B. (1997). On instantaneous amplitude and phase of signals. *IEEE Trans. Signal Process.* **45** 552–560.
- [31] PRIESTLEY, M. (1967). Power spectral analysis of non-stationary random processes. *J. Sound Vib.* **6** 86–97.
- [32] PRIESTLEY, M. B. (1981). *Spectral Analysis and Time Series. Vol. 1. Probability and Mathematical Statistics*. Academic Press, London-New York. [MR0628735](#)
- [33] ROSENBLATT, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA* **42** 43–47. [MR0074711](#) <https://doi.org/10.1073/pnas.42.1.43>
- [34] SCHREIER, P. J. and SCHARF, L. L. (2010). *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge Univ. Press, Cambridge. [MR2769061](#) <https://doi.org/10.1017/CBO9780511815911>
- [35] SOURISSEAU, M., WU, H.-T. and ZHOU, Z. (2022). Supplement to “Asymptotic analysis of synchrosqueezing transform—toward statistical inference with nonlinear-type time-frequency analysis.” <https://doi.org/10.1214/22-AOS2203SUPP>
- [36] THAKUR, G., BREVD, E., FUCKAR, N. S. and WU, H.-T. (2013). The synchrosqueezing algorithm for time-varying spectral analysis: Robustness properties and new paleoclimate applications. *Signal Process.* **93** 1079–1094.
- [37] THAKUR, G. and WU, H.-T. (2011). Synchrosqueezing-based recovery of instantaneous frequency from nonuniform samples. *SIAM J. Math. Anal.* **43** 2078–2095. [MR2837495](#) <https://doi.org/10.1137/100798818>

- [38] WU, H. (2013). Instantaneous frequency and wave shape functions (I). *Appl. Comput. Harmon. Anal.* **35** 181–199. [MR3062471](#) <https://doi.org/10.1016/j.acha.2012.08.008>
- [39] WU, H.-T. (2020). Current state of nonlinear-type time–frequency analysis and applications to high-frequency biomedical signals. *Curr. Opin. Syst. Biol.* **23** 8–21.
- [40] WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. [MR2172215](#) <https://doi.org/10.1073/pnas.0506715102>
- [41] XIAO, H. and WU, W. B. (2012). Covariance matrix estimation for stationary time series. *Ann. Statist.* **40** 466–493. [MR3014314](#) <https://doi.org/10.1214/11-AOS967>
- [42] XIAO, J. and FLANDRIN, P. (2007). Multitaper time-frequency reassignment for nonstationary spectrum estimation and chirp enhancement. *IEEE Trans. Signal Process.* **55** 2851–2860. [MR2473558](#) <https://doi.org/10.1109/TSP.2007.893961>
- [43] YANG, H. (2015). Synchrosqueezed wave packet transforms and diffeomorphism based spectral analysis for 1D general mode decompositions. *Appl. Comput. Harmon. Anal.* **39** 33–66. [MR3343801](#) <https://doi.org/10.1016/j.acha.2014.08.004>
- [44] YANG, H. (2018). Statistical analysis of synchrosqueezed transforms. *Appl. Comput. Harmon. Anal.* **45** 526–550. [MR3842645](#) <https://doi.org/10.1016/j.acha.2017.01.001>
- [45] YANG, J. and ZHOU, Z. (2022). Spectral inference under complex temporal dynamics. *J. Amer. Statist. Assoc.* **117** 133–155. [MR4399075](#) <https://doi.org/10.1080/01621459.2020.1764365>
- [46] ZHOU, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *J. Amer. Statist. Assoc.* **108** 726–740. [MR3174655](#) <https://doi.org/10.1080/01621459.2013.787184>

APPROXIMATE KERNEL PCA: COMPUTATIONAL VERSUS STATISTICAL TRADE-OFF

BY BHARATH K. SRIPERUMBUDUR^a AND NICHOLAS STERGE^b

Department of Statistics, Pennsylvania State University, ^abks18@psu.edu, ^bnsterge@us.flowtraders.com

Kernel methods are powerful learning methodologies that allow to perform nonlinear data analysis. Despite their popularity, they suffer from poor scalability in big data scenarios. Various approximation methods, including random feature approximation, have been proposed to alleviate the problem. However, the statistical consistency of most of these approximate kernel methods is not well understood except for kernel ridge regression wherein it has been shown that the random feature approximation is not only computationally efficient but also statistically consistent with a minimax optimal rate of convergence. In this paper, we investigate the efficacy of random feature approximation in the context of kernel principal component analysis (KPCA) by studying the trade-off between computational and statistical behaviors of approximate KPCA. We show that the approximate KPCA is both computationally and statistically efficient compared to KPCA in terms of the error associated with reconstructing a kernel function based on its projection onto the corresponding eigenspaces. The analysis hinges on Bernstein-type inequalities for the operator and Hilbert–Schmidt norms of a self-adjoint Hilbert–Schmidt operator-valued U-statistics, which are of independent interest.

REFERENCES

- ALAOU, A. and MAHONEY, M. (2015). Fast randomized kernel ridge regression with statistical guarantees. In *Advances in Neural Information Processing Systems* 28 (C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama and R. Garnett, eds.) 775–783. Curran Associates, Red Hook.
- ARONSAJN, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68** 337–404. [MR0051437 https://doi.org/10.2307/1990404](https://doi.org/10.2307/1990404)
- BACH, F. (2013). Sharp analysis of low-rank kernel matrix approximations. In *Proc. of the 26th Annual Conference on Learning Theory* (S. Shalev-Shwartz and I. Steinwart, eds.). *Proceedings of Machine Learning Research* **30** 185–209. PMLR.
- BACH, F. and JORDAN, M. I. (2005). Predictive low-rank decomposition for kernel methods. In *Proc. of the 22nd International Conference on Machine Learning* (L. D. Raedt and S. Wrobel, eds.) 33–40.
- BLANCHARD, G., BOUSQUET, O. and ZWALD, L. (2007). Statistical properties of kernel principal component analysis. *Mach. Learn.* **66** 259–294.
- CAPONNETTO, A. and DE VITO, E. (2007). Optimal rates for the regularized least-squares algorithm. *Found. Comput. Math.* **7** 331–368. [MR2335249 https://doi.org/10.1007/s10208-006-0196-8](https://doi.org/10.1007/s10208-006-0196-8)
- CHANG, P. C. and WU, J. L. (2015). A critical feature extraction by kernel PCA in stock trading model. *Soft Comput.* **19** 1393–1408.
- COHN, D. L. (2013). *Measure Theory*, 2nd ed. *Birkhäuser Advanced Texts: Basler Lehrbücher*. [Birkhäuser Advanced Texts: Basel Textbooks]. Birkhäuser/Springer, New York. [MR3098996 https://doi.org/10.1007/978-1-4614-6956-8](https://doi.org/10.1007/978-1-4614-6956-8)
- DE MOURA, N. N. and DE SEIXAS, J. M. (2017). Novelty detection in passive sonar systems using a kernel approach. In *13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery* 116–122.
- DIESTEL, J. and UHL, J. J. JR. (1977). *Vector Measures*. *Mathematical Surveys*, No. 15. Amer. Math. Soc., Providence, R.I. [MR0453964](https://doi.org/10.1007/978-1-4614-6956-8)
- DINCULEANU, N. (2000). *Vector Integration and Stochastic Integration in Banach Spaces*. *Pure and Applied Mathematics (New York)*. Wiley Interscience, New York. [MR1782432 https://doi.org/10.1002/9781118033012](https://doi.org/10.1002/9781118033012)

MSC2020 subject classifications. Primary 62H25; secondary 62G05.

Key words and phrases. Principal component analysis, kernel PCA, random feature approximation, reproducing kernel Hilbert space, covariance operator, Bernstein’s inequality.

- DRINEAS, P. and MAHONEY, M. W. (2005). On the Nyström method for approximating a Gram matrix for improved kernel-based learning. *J. Mach. Learn. Res.* **6** 2153–2175. [MR2249884](#)
- FINE, S. and SCHEINBERG, K. (2001). Efficient SVM training using low-rank kernel representations. *J. Mach. Learn. Res.* **2** 243–264.
- HALKO, N., MARTINSSON, P. G. and TROPP, J. A. (2011). Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Rev.* **53** 217–288. [MR2806637](#) <https://doi.org/10.1137/090771806>
- HOFFMANN, H. (2007). Kernel PCA for novelty detection. *Pattern Recognit.* **40** 863–874.
- JADE, A. M., SRIKANTH, B., JAYARAMAN, V. K., KULKARNI, B. D., JOG, J. P. and PRIYA, L. (2003). Feature extraction and denoising using kernel PCA. *Chem. Eng. Sci.* **58** 4441–4448.
- JOLLIFFE, I. T. (1986). *Principal Component Analysis*. *Springer Series in Statistics*. Springer, New York. [MR0841268](#) <https://doi.org/10.1007/978-1-4757-1904-8>
- KIM, K. I., FRANZ, M. O. and SCHÖLKOPF, B. (2005). Iterative kernel principal component analysis for image modeling. *IEEE Trans. Pattern Anal. Mach. Intell.* **27** 1351–1366.
- KIMELDORF, G. and WAHBA, G. (1971). Some results on Tchebycheffian spline functions. *J. Math. Anal. Appl.* **33** 82–95. [MR0290013](#) [https://doi.org/10.1016/0022-247X\(71\)90184-3](https://doi.org/10.1016/0022-247X(71)90184-3)
- KUMAR, S., MOHRI, M. and TALWALKAR, A. (2009). Ensemble Nyström method. In *Advances in Neural Information Processing Systems 22* (Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams and A. Culotta, eds.) 1060–1068. Curran Associates, Red Hook.
- LAMPERT, C. (2009). Kernel methods in computer vision. *Foundations and Trends in Computer Graphics and Vision* **4** 193–285.
- LI, Z., KRUGER, U., XIE, L., ALMANSOORI, A. and SU, H. (2015). Adaptive KPCA modeling on nonlinear systems. *IEEE Trans. Signal Process.* **63** 2364–2376. [MR3332006](#) <https://doi.org/10.1109/TSP.2015.2412913>
- LOPEZ-PAZ, D., SRA, S., SMOLA, A., GHAHRAMANI, Z. and SCHÖLKOPF, B. (2014). Randomized nonlinear component analysis. In *Proceedings of the 31st International Conference on Machine Learning* (E. P. Xing and T. Jebara, eds.). *Proceedings of Machine Learning Research* **32** 1359–1367. PMLR.
- MIKA, S., SCHÖLKOPF, B., SMOLA, A. J., MÜLLER, K.-R., SCHOLZ, M. and RÄTSCH, G. (1999). Kernel PCA and de-noising in feature spaces. In *Advances in Neural Information Processing Systems 11* (M. J. Kearns, S. A. Solla and D. A. Cohn, eds.) 536–542. MIT Press, Cambridge.
- PETER, M., MINOI, J.-L. and HIPINY, I. H. M. (2019). 3D face recognition using kernel-based PCA approach. In *Computational Science and Technology* (R. Alfred, Y. Lim, A. A. Ibrahim and P. Anthony, eds.) 77–86. Springer Singapore, Singapore.
- PHOPHALIA, A. and MITRA, S. K. (2017). 3D MR image denoising using rough set and kernel PCA method. *Magn. Reson. Imaging* **36** 135–145.
- RAHIMI, A. and RECHT, B. (2008). Random features for large-scale kernel machines. In *Advances in Neural Information Processing Systems 20* (J. C. Platt, D. Koller, Y. Singer and S. T. Roweis, eds.) 1177–1184. Curran Associates, Red Hook.
- REED, M. and SIMON, B. (1980). *Methods of Modern Mathematical Physics. Functional analysis I*, 2nd ed. Academic Press [Harcourt Brace Jovanovich, Publishers], New York. [MR0751959](#)
- RUDI, A., CAMORIANO, R. and ROSASCO, L. (2015). Less is more: Nyström computational regularization. In *Advances in Neural Information Processing Systems 28* (C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama and R. Garnett, eds.) 1657–1665. Curran Associates, Red Hook.
- RUDI, A., CANAS, G. and ROSASCO, L. (2013). On the sample complexity of subspace learning. In *Advances in Neural Information Processing Systems 26* (C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani and K. Q. Weinberger, eds.) 2067–2075. Curran Associates, Red Hook.
- RUDI, A. and ROSASCO, L. (2017). Generalization properties of learning with random features. In *Advances in Neural Information Processing Systems 30* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, eds.) 3215–3225. Curran Associates, Red Hook.
- SAMUEL, R. T. and CAO, Y. (2016). Nonlinear process fault detection and identification using kernel PCA and kernel density estimation. *Syst. Sci. Control Eng.* **4** 165–174.
- SCHÖLKOPF, B. and SMOLA, A. J. (2002). *Learning with Kernels*. MIT Press, Cambridge, MA.
- SCHÖLKOPF, B., SMOLA, A. and MÜLLER, K. R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* **10** 1299–1319.
- SHAWE-TAYLOR, J., WILLIAMS, C. K. I., CRISTIANINI, N. and KANDOLA, J. (2005). On the eigenspectrum of the Gram matrix and the generalization error of kernel-PCA. *IEEE Trans. Inf. Theory* **51** 2510–2522. [MR2246374](#) <https://doi.org/10.1109/TIT.2005.850052>
- SMOLA, A. J. and SCHÖLKOPF, B. (2000). Sparse greedy matrix approximation for machine learning. In *Proc. 17th International Conference on Machine Learning* 911–918. Morgan Kaufmann, San Francisco, CA.
- SRIPERUMBUDUR, B. K. and STERGE, N. (2017). Statistical consistency of kernel PCA with random features. Available at <https://arxiv.org/pdf/1706.06296v1.pdf>.

- SRIPERUMBUDUR, B. K. and STERGE, N. (2018). Approximate kernel PCA using random features: Computational vs. statistical trade-off. Available at <https://arxiv.org/pdf/1706.06296v2.pdf>.
- SRIPERUMBUDUR, B. K. and STERGE, N. (2022). Supplement to “Approximate Kernel PCA: Computational versus Statistical Trade-off.” <https://doi.org/10.1214/22-AOS2204SUPP>
- STEINWART, I. and CHRISTMANN, A. (2008). *Support Vector Machines*. Springer, New York.
- TEIXEIRA, A. R., TOMÉ, A. M., STADLTHANNER, K. and LANG, E. W. (2008). KPCA denoising and the pre-image problem revisited. *Digit. Signal Process.* **18** 568–580.
- ULLAH, E., MIANY, P., MARINOV, T. V. and ARORA, R. (2018). Streaming kernel PCA with $\tilde{O}(\sqrt{n})$ random features. In *Advances in Neural Information Processing Systems* (S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi and R. Garnett, eds.) **31** 7311–7321. Curran Associates, Red Hook.
- WENDLAND, H. (2005). *Scattered Data Approximation. Cambridge Monographs on Applied and Computational Mathematics* **17**. Cambridge Univ. Press, Cambridge. [MR2131724](#)
- WILLIAMS, C. K. I. and SEEGER, M. (2001). Using the Nyström method to speed up kernel machines. In *Advances in Neural Information Processing Systems* 13 (V. T. T. K. Leen T. G. Diettrich, ed.) 682–688. MIT Press, Cambridge, MA.
- YANG, Y., PILANCI, M. and WAINWRIGHT, M. J. (2017). Randomized sketches for kernels: Fast and optimal nonparametric regression. *Ann. Statist.* **45** 991–1023. [MR3662446](#) <https://doi.org/10.1214/16-AOS1472>
- YANG, T., LI, Y., MAHDAVI, M., JIN, R. and ZHOU, Z.-H. (2012). Nyström method vs random Fourier features: A theoretical and empirical comparison. In *Advances in Neural Information Processing Systems* 25 (F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger, eds.) 476–484. Curran Associates, Red Hook.
- ZWALD, L. and BLANCHARD, G. (2006). On the convergence of eigenspaces in kernel principal component analysis. In *Advances in Neural Information Processing Systems* 18 (Y. Weiss, B. Schölkopf and J. C. Platt, eds.) 1649–1656. MIT Press.

RELAXING THE GAUSSIAN ASSUMPTION IN SHRINKAGE AND SURE IN HIGH DIMENSION

BY MAX FATHI^{1,a}, LARRY GOLDSTEIN^{2,b}, GESINE REINERT^{3,c} AND ADRIEN SAUMARD^{4,d}

¹Université Paris Cité and Sorbonne Université, CNRS, Laboratoire Jacques-Louis Lions & Laboratoire de Probabilités, Statistique et Modélisation, ^amfathi@lpsm.paris

²Department of Mathematics, University of Southern California, ^blarry@math.usc.edu

³Department of Statistics, University of Oxford, ^creinert@stats.ox.ac.uk

⁴Université de Rennes, Ensai, CREST-UMR 9194, ^dadrien.saumard@ensai.fr

Shrinkage estimation is a fundamental tool of modern statistics, pioneered by Charles Stein upon his discovery of the famous paradox involving the multivariate Gaussian. A large portion of the subsequent literature only considers the efficiency of shrinkage, and that of an associated procedure known as Stein’s Unbiased Risk Estimate, or SURE, in the Gaussian setting of that original work. We investigate what extensions to the domain of validity of shrinkage and SURE can be made away from the Gaussian through the use of tools developed in the probabilistic area now known as Stein’s method. We show that shrinkage is efficient away from the Gaussian under very mild conditions on the distribution of the noise. SURE is also proved to be adaptive under similar assumptions, and in particular in a way that retains the classical asymptotics of Pinsker’s theorem. Notably, shrinkage and SURE are shown to be efficient under mild distributional assumptions, and particularly for general isotropic log-concave measures.

REFERENCES

- [1] ADAMCZAK, R., GUÉDON, O., LATAŁA, R., LITVAK, A. E., OLESZKIEWICZ, K., PAJOR, A. and TOMCZAK-JAEGERMANN, N. (2012). Moment estimates for convex measures. *Electron. J. Probab.* **17** no. 101, 19. [MR3005719](https://doi.org/10.1214/EJP.v17-2150) <https://doi.org/10.1214/EJP.v17-2150>
- [2] AVERKAMP, R. and HOUDRÉ, C. (2003). Wavelet thresholding for non-necessarily Gaussian noise: Idealism. *Ann. Statist.* **31** 110–151. [MR1962501](https://doi.org/10.1214/aos/1046294459) <https://doi.org/10.1214/aos/1046294459>
- [3] AVERKAMP, R. and HOUDRÉ, C. (2005). Wavelet thresholding for nonnecessarily Gaussian noise: Functionality. *Ann. Statist.* **33** 2164–2193. [MR2211083](https://doi.org/10.1214/009053605000000471) <https://doi.org/10.1214/009053605000000471>
- [4] BAYATI, M., ERDOĞDU, M. and MONTANARI, A. (2013). Estimating lasso risk and noise level. In *Advances in Neural Information Processing Systems*.
- [5] BOBKOV, S. G. and NAZAROV, F. L. (2003). On convex bodies and log-concave probability measures with unconditional basis. In *Geometric Aspects of Functional Analysis. Lecture Notes in Math.* **1807** 53–69. Springer, Berlin. [MR2083388](https://doi.org/10.1007/978-3-540-36428-3_6) https://doi.org/10.1007/978-3-540-36428-3_6
- [6] CACOULOS, T. and PAPATHANASIOU, V. (1992). Lower variance bounds and a new proof of the central limit theorem. *J. Multivariate Anal.* **43** 173–184. [MR1193610](https://doi.org/10.1016/0047-259X(92)90032-B) [https://doi.org/10.1016/0047-259X\(92\)90032-B](https://doi.org/10.1016/0047-259X(92)90032-B)
- [7] CANDÈS, E. J., SING-LONG, C. A. and TRZASKO, J. D. (2013). Unbiased risk estimates for singular value thresholding and spectral estimators. *IEEE Trans. Signal Process.* **61** 4643–4657. [MR3105401](https://doi.org/10.1109/TSP.2013.2270464) <https://doi.org/10.1109/TSP.2013.2270464>
- [8] CATTIAUX, P. and GUILLIN, A. (2020). On the Poincaré constant of log-concave measures. In *Geometric Aspects of Functional Analysis. Vol. I. Lecture Notes in Math.* **2256** 171–217. Springer, Cham. [MR4175748](https://doi.org/10.1007/978-3-030-36020-7_9) https://doi.org/10.1007/978-3-030-36020-7_9
- [9] CELLIER, D., FOURDRINIER, D. and ROBERT, C. (1989). Robust shrinkage estimators of the location parameter for elliptically symmetric distributions. *J. Multivariate Anal.* **29** 39–52. [MR0991055](https://doi.org/10.1016/0047-259X(89)90075-4) [https://doi.org/10.1016/0047-259X\(89\)90075-4](https://doi.org/10.1016/0047-259X(89)90075-4)

- [10] CHATTERJEE, S. (2009). Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields* **143** 1–40. [MR2449121](#) <https://doi.org/10.1007/s00440-007-0118-6>
- [11] CHEN, Y. (2021). An almost constant lower bound of the isoperimetric coefficient in the KLS conjecture. *Geom. Funct. Anal.* **31** 34–61. [MR4244847](#) <https://doi.org/10.1007/s00039-021-00558-4>
- [12] CHEN, Y., WIESEL, A. and HERO, A. O. III (2011). Robust shrinkage estimation of high-dimensional covariance matrices. *IEEE Trans. Signal Process.* **59** 4097–4107. [MR2865971](#) <https://doi.org/10.1109/TSP.2011.2138698>
- [13] COURTADE, T. A., FATHI, M. and PANANJADY, A. (2019). Existence of Stein kernels under a spectral gap, and discrepancy bounds. *Ann. Inst. Henri Poincaré Probab. Stat.* **55** 777–790. [MR3949953](#) <https://doi.org/10.1214/18-aihp898>
- [14] DELYON, B. and JUDITSKY, A. (1996). On minimax wavelet estimators. *Appl. Comput. Harmon. Anal.* **3** 215–228. [MR1400080](#) <https://doi.org/10.1006/acha.1996.0017>
- [15] DÖBLER, C. (2015). New Berry–Esseen and Wasserstein bounds in the CLT for non-randomly centered random sums by probabilistic methods. *ALEA Lat. Am. J. Probab. Math. Stat.* **12** 863–902. [MR3453299](#)
- [16] DONOHO, D. L. and JOHNSTONE, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90** 1200–1224. [MR1379464](#)
- [17] ELDAR, Y. C. (2009). Generalized SURE for exponential families: Applications to regularization. *IEEE Trans. Signal Process.* **57** 471–481. [MR2603376](#) <https://doi.org/10.1109/TSP.2008.2008212>
- [18] EVANS, L. C. (1998). *Partial Differential Equations. Graduate Studies in Mathematics* **19**. Amer. Math. Soc., Providence, RI. [MR1625845](#)
- [19] EVANS, S. N. and STARK, P. B. (1996). Shrinkage estimators, Skorokhod’s problem and stochastic integration by parts. *Ann. Statist.* **24** 809–815. [MR1394989](#) <https://doi.org/10.1214/aos/1032894466>
- [20] FATHI, M. (2019). Stein kernels and moment maps. *Ann. Probab.* **47** 2172–2185. [MR3980918](#) <https://doi.org/10.1214/18-AOP1305>
- [21] FATHI, M., GOLDSTEIN, L., REINERT, G. and SAUMARD, A. (2022). Supplement to “Relaxing the Gaussian assumption in shrinkage and SURE in high dimension.” <https://doi.org/10.1214/22-AOS2208SUPP>
- [22] FOURDRINIER, D., STRAWDERMAN, W. E. and WELLS, M. T. (2018). *Shrinkage Estimation. Springer Series in Statistics*. Springer, Cham. [MR3887633](#) <https://doi.org/10.1007/978-3-030-02185-6>
- [23] GOLDSTEIN, L. (2007). L^1 bounds in normal approximation. *Ann. Probab.* **35** 1888–1930. [MR2349578](#) <https://doi.org/10.1214/009117906000001123>
- [24] GOLDSTEIN, L. and REINERT, G. (1997). Stein’s method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.* **7** 935–952. [MR1484792](#) <https://doi.org/10.1214/aoap/1043862419>
- [25] GOLDSTEIN, L. and REINERT, G. (2005). Zero biasing in one and higher dimensions, and applications. In *Stein’s Method and Applications. Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **5** 1–18. Singapore Univ. Press, Singapore. [MR2201883](#) https://doi.org/10.1142/9789812567673_0001
- [26] JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I* 361–379. Univ. California Press, Berkeley, CA. [MR0133191](#)
- [27] JOAG-DEV, K. and PROSCHAN, F. (1983). Negative association of random variables, with applications. *Ann. Statist.* **11** 286–295. [MR0684886](#) <https://doi.org/10.1214/aos/1176346079>
- [28] KLARTAG, B. (2013). Poincaré inequalities and moment maps. *Ann. Fac. Sci. Toulouse Math.* (6) **22** 1–41. [MR3247770](#) <https://doi.org/10.5802/afst.1366>
- [29] LANDSMAN, Z., VANDUFFEL, S. and YAO, J. (2015). Some Stein-type inequalities for multivariate elliptical distributions and applications. *Statist. Probab. Lett.* **97** 54–62. [MR3299751](#) <https://doi.org/10.1016/j.spl.2014.11.005>
- [30] LEDOUX, M., NOURDIN, I. and PECCATI, G. (2015). Stein’s method, logarithmic Sobolev and transport inequalities. *Geom. Funct. Anal.* **25** 256–306. [MR3320893](#) <https://doi.org/10.1007/s00039-015-0312-0>
- [31] LEE, Y. T. and VEMPALA, S. (2017). The KLS conjecture. In *Current Developments in Mathematics*.
- [32] LI, K.-C. (1985). From Stein’s unbiased risk estimates to the method of generalized cross validation. *Ann. Statist.* **13** 1352–1377. [MR0811497](#) <https://doi.org/10.1214/aos/1176349742>
- [33] MIJOULE, G., REINERT, G. and SWAN, Y. (2018). Stein operators, kernels and discrepancies for multivariate continuous distributions. [arXiv:1806.03478](#).
- [34] NAHAPETIAN, B. S. and PETROSIAN, A. N. (1992). Martingale-difference Gibbs random fields and central limit theorem. *Ann. Acad. Sci. Fenn. Ser. A I Math.* **17** 105–110. [MR1162153](#) <https://doi.org/10.5186/aasfm.1992.1713>
- [35] NOURDIN, I. and PECCATI, G. (2012). *Normal Approximations with Malliavin Calculus: From Stein’s Method to Universality. Cambridge Tracts in Mathematics* **192**. Cambridge Univ. Press, Cambridge. [MR2962301](#) <https://doi.org/10.1017/CBO9781139084659>

- [36] NUSSBAUM, M. (1996). The Pinsker bound: A review. In *Encyclopedia of Statistical Sciences* (S. Kotz, ed.) Wiley, New York, NY.
- [37] PAOURIS, G. (2012). Small ball probability estimates for log-concave measures. *Trans. Amer. Math. Soc.* **364** 287–308. [MR2833584](#) <https://doi.org/10.1090/S0002-9947-2011-05411-5>
- [38] PINSKER, M. S. Optimal filtration of square-integrable signals in Gaussian noise. [MR0624591](#)
- [39] SAUMARD, A. and WELLNER, J. A. (2014). Log-concavity and strong log-concavity: A review. *Stat. Surv.* **8** 45–114. [MR3290441](#) <https://doi.org/10.1214/14-SS107>
- [40] SRIVASTAVA, M. S. and BILODEAU, M. (1989). Stein estimation under elliptical distributions. *J. Multivariate Anal.* **28** 247–259. [MR0991949](#) [https://doi.org/10.1016/0047-259X\(89\)90108-5](https://doi.org/10.1016/0047-259X(89)90108-5)
- [41] STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, Vol. I* 197–206. Univ. California Press, Berkeley–Los Angeles, CA. [MR0084922](#)
- [42] STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. [MR0630098](#)
- [43] UTEV, S. A. (1989). Probabilistic problems connected with an integro-differential inequality. *Sibirsk. Mat. Zh.* **30** 182–186, 220. [MR1010851](#) <https://doi.org/10.1007/BF00971508>
- [44] VEMPALA, S. (2010). Recent progress and open problems in algorithmic convex geometry. In *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2010)*.
- [45] WASSERMAN, L. (2006). *All of Nonparametric Statistics. Springer Texts in Statistics*. Springer, New York. [MR2172729](#)
- [46] ZHANG, X.-P. and MITA, D. (1998). Adaptive denoising based on SURE risk. *IEEE Signal Process. Lett.* **5** 265–267.

NONPARAMETRIC BIVARIATE DENSITY ESTIMATION FOR CENSORED LIFETIMES

BY SAM EFROMOVICH^a

Department of Mathematical Sciences, University of Texas at Dallas, ^aefrom@utdallas.edu

It is well known that estimation of a bivariate cumulative distribution function of a pair of right censored lifetimes presents challenges unparalleled to the univariate case where a product-limit Kaplan–Meyer’s methodology typically yields optimal estimation, and the literature on optimal estimation of the joint probability density is next to none. The paper, for the first time in the survival analysis literature, develops the theory and methodology of sharp minimax and adaptive nonparametric estimation of the joint density under the mean integrated squared error (MISE) criterion. The theory shows how an underlying joint density, together with the bivariate distribution of censoring variables, affect the estimation, and what and how may or may not be estimated in the presence of censoring. Practical example illustrates the problem.

REFERENCES

- AKRITAS, M. G. and VAN KEILEGOM, I. (2003). Estimation of bivariate and marginal distributions with censored data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 457–471. [MR1983758](#) <https://doi.org/10.1111/1467-9868.00396>
- ALBU, N., BARANI, N. and CONSTANTIN, M. (2021). Choosing an economical solution for water aeration. *Hydraul.* **145** 32–37.
- CAMPBELL, G. (1981). Nonparametric bivariate estimation with randomly censored data. *Biometrika* **68** 417–422. [MR0626401](#) <https://doi.org/10.1093/biomet/68.2.417>
- COLLETT, D. (2003). *Modelling Survival Data in Medical Research*. CRC Press, Boca Raton, FL.
- CROWDER, M. (2012). *Multivariate Survival Analysis and Competing Risks. Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. [MR2964342](#) <https://doi.org/10.1201/b11893>
- DABROWSKA, D. M. (1988). Kaplan–Meier estimate on the plane. *Ann. Statist.* **16** 1475–1489. [MR0964934](#) <https://doi.org/10.1214/aos/1176351049>
- DABROWSKA, D., DUFFY, D. and ZHANG, Z. (1998). Hazard and density estimation from bivariate censored data. *J. Nonparametr. Stat.* **10** 67–93.
- EFROMOVICH, S. Y. and PINSKER, M. S. (1982). Estimation of square-integrable probability density of a random variable. *Problemy Peredachi Informatsii* **18** 19–38. [MR0711898](#)
- EFROMOVICH, S. (1985). Adaptive estimation of a density with unknown smoothness. *Theory Probab. Appl.* **30** 557–568.
- EFROMOVICH, S. (1989). On sequential nonparametric estimation of a density. *Theory Probab. Appl.* **34** 228–239.
- EFROMOVICH, S. (1999). *Nonparametric Curve Estimation: Methods, Theory, and Applications. Springer Series in Statistics*. Springer, New York. [MR1705298](#)
- EFROMOVICH, S. (2001). Density estimation under random censorship and order restrictions: From asymptotic to small samples. *J. Amer. Statist. Assoc.* **96** 667–684. [MR1946433](#) <https://doi.org/10.1198/016214501753168334>
- EFROMOVICH, S. (2018). *Missing and Modified Data in Nonparametric Estimation. Monographs on Statistics and Applied Probability* **156**. CRC Press, Boca Raton, FL. [MR3752670](#)
- FREES, E., CARRIERE, J. and VALDEZ, E. (1995). Annuity valuation with dependent mortality. *Actuar. Res. Clearing House* **2** 31–80.
- 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>

- GOLUBEV, G. K. (1991). LAN in problems of non-parametric estimation of functions and lower bounds for quadratic risks. *Probl. Inf. Transm.* **36** 152–157.
- HOFFMANN, M. and LEPSKI, O. (2002). Random rates in anisotropic regression. *Ann. Statist.* **30** 325–396. [MR1902892 https://doi.org/10.1214/aos/1021379858](https://doi.org/10.1214/aos/1021379858)
- HOUGAARD, P. (2000). *Analysis of Multivariate Survival Data. Statistics for Biology and Health*. Springer, New York. [MR1777022 https://doi.org/10.1007/978-1-4612-1304-8](https://doi.org/10.1007/978-1-4612-1304-8)
- KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481. [MR0093867](https://doi.org/10.2307/2343181)
- KAHANE, J. (1968). *Some Random Series of Functions. Cambridge Series in Advance Mathematics*. Some Random Series of Functions, Lexington, MA. [MR0254888](https://doi.org/10.1017/CBO9780511524488)
- KOOPERBERG, C. (1998). Bivariate density estimation with an application to survival analysis. *J. Comput. Graph. Statist.* **7** 322–341.
- LI, J. and MA, S. (2013). *Survival Analysis in Medicine and Genetics. Chapman & Hall/CRC Biostatistics Series*. CRC Press, Boca Raton, FL. [MR3616722](https://doi.org/10.1201/b12722)
- LOPEZ, O. (2012). A generalization of the Kaplan–Meier estimator for analyzing bivariate mortality under right-censoring and left-truncation with applications in model-checking for survival copula models. *Insurance Math. Econom.* **51** 505–516. [MR2975191 https://doi.org/10.1016/j.insmatheco.2012.07.009](https://doi.org/10.1016/j.insmatheco.2012.07.009)
- MOORE, D. (2016). *Applied Survival Analysis Using R*. Springer, New York.
- NIKOLSKII, S. M. (1975). *Approximation of Functions of Several Variables and Imbedding Theorems. Die Grundlehren der Mathematischen Wissenschaften, Band 205*. Springer, New York. [MR0374877](https://doi.org/10.1007/978-1-4612-1304-8)
- OAKES, D. (1989). Bivariate survival models induced by frailties. *J. Amer. Statist. Assoc.* **84** 487–493. [MR1010337](https://doi.org/10.2307/2343181)
- PINSKER, M. (1980). Optimal filtration of functions from L_2 in Gaussian noise. *Probl. Inf. Transm.* **16** 970–983.
- PRENTICE, R. L. (2016). Higher dimensional Clayton–Oakes models for multivariate failure time data. *Biometrika* **103** 231–236. [MR3465834 https://doi.org/10.1093/biomet/asv057](https://doi.org/10.1093/biomet/asv057)
- PRENTICE, R. L. and CAI, J. (1992). Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika* **79** 495–512. [MR1187604 https://doi.org/10.1093/biomet/79.3.495](https://doi.org/10.1093/biomet/79.3.495)
- PRENTICE, R. L. and ZHAO, S. (2018). Nonparametric estimation of the multivariate survivor function: The multivariate Kaplan–Meier estimator. *Lifetime Data Anal.* **24** 3–27. [MR3742904 https://doi.org/10.1007/s10985-016-9383-y](https://doi.org/10.1007/s10985-016-9383-y)
- PRUITT, R. C. (1993). Identifiability of bivariate survival curves from censored data. *J. Amer. Statist. Assoc.* **88** 573–579. [MR1224383](https://doi.org/10.2307/2343181)
- ROSSO, D., LARSON, L. and STENSTROM, M. (2008). Aeration of large-scale municipal wastewater treatment plants: State of the art. *Water Sci. Technol.* **57** 973–978.
- SARAIVA, E. F., SUZUKI, A. K. and MILAN, L. A. (2018). Bayesian computational methods for sampling from the posterior distribution of a bivariate survival model, based on AMH Copula in the presence of right-censored data. *Entropy* **20** 642. [MR3882496 https://doi.org/10.3390/e20090642](https://doi.org/10.3390/e20090642)
- SEOK, J., TIAN, L. and WONG, W. (2014). Density estimation on multivariate censored data with optional Polya tree. *Biostatistics* **15** 182–195.
- TSAI, W.-Y., LEURGANS, S. and CROWLEY, J. (1986). Nonparametric estimation of a bivariate survival function in the presence of censoring. *Ann. Statist.* **14** 1351–1365. [MR0868304 https://doi.org/10.1214/aos/1176350162](https://doi.org/10.1214/aos/1176350162)
- WANG, W. and WELLS, M. T. (1997). Nonparametric estimators of the bivariate survival function under simplified censoring conditions. *Biometrika* **84** 863–880. [MR1625051 https://doi.org/10.1093/biomet/84.4.863](https://doi.org/10.1093/biomet/84.4.863)
- WASSERMAN, L. (2006). *All of Nonparametric Statistics. Springer Texts in Statistics*. Springer, New York. [MR2172729](https://doi.org/10.1007/978-1-4612-1304-8)
- WELLS, M. T. and YEO, K. P. (1996). Density estimation with bivariate censored data. *J. Amer. Statist. Assoc.* **91** 1566–1574. [MR1439096 https://doi.org/10.2307/2291582](https://doi.org/10.2307/2291582)

LINEAR BIOMARKER COMBINATION FOR CONSTRAINED CLASSIFICATION

BY YIJIAN HUANG^{1,a} AND MARTIN G. SANDA^{2,b}

¹*Department of Biostatistics and Bioinformatics, Emory University, ayhuang5@emory.edu*

²*Department of Urology, Emory University, martinsanda@emory.edu*

Multiple biomarkers are often combined to improve disease diagnosis. The uniformly optimal combination, that is, with respect to all reasonable performance metrics, unfortunately requires excessive distributional modeling, to which the estimation can be sensitive. An alternative strategy is rather to pursue local optimality with respect to a specific performance metric. Nevertheless, existing methods may not target clinical utility of the intended medical test, which usually needs to operate above a certain sensitivity or specificity level, or do not have their statistical properties well studied and understood. In this article, we develop and investigate a linear combination method to maximize the clinical utility empirically for such a constrained classification. The combination coefficient is shown to have cube root asymptotics. The convergence rate and limiting distribution of the predictive performance are subsequently established, exhibiting robustness of the method in comparison with others. An algorithm with sound statistical justification is devised for efficient and high-quality computation. Simulations corroborate the theoretical results, and demonstrate good statistical and computational performance. Illustration with a clinical study on aggressive prostate cancer detection is provided.

REFERENCES

- BAHADUR, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Stat.* **37** 577–580. [MR0189095](https://doi.org/10.1214/aoms/1177699450)
<https://doi.org/10.1214/aoms/1177699450>
- BERTSIMAS, D., KING, A. and MAZUMDER, R. (2016). Best subset selection via a modern optimization lens. *Ann. Statist.* **44** 813–852. [MR3476618](https://doi.org/10.1214/15-AOS1388) <https://doi.org/10.1214/15-AOS1388>
- CATALONA, W. J., PARTIN, A. W., SANDA, M. G., WEI, J. T., KLEE, G. G., BANGMA, C. H., SLAWIN, K. M., MARKS, L. S., LOEB, S. et al. (2011). A multicenter study of [−2]pro-prostate specific antigen combined with prostate specific antigen and free prostate specific antigen for prostate cancer detection in the 2.0 to 10.0 Ng/ml prostate specific antigen range. *J. Urol.* **185** 1650–1655.
- CATALONA, W. J., PARTIN, A. W., SLAWIN, K. M., BRAWER, M. K., FLANIGAN, R. C., PATEL, A., RICHIE, J. P., DEKERNION, J. B., WALSH, P. C. et al. (1998). Use of the percentage of free prostate-specific antigen to enhance differentiation of prostate cancer from benign prostatic disease: A prospective multicenter clinical trial. *JAMA* **279** 1542–1547.
- CHEN, B., LI, P., QIN, J. and YU, T. (2016). Using a monotonic density ratio model to find the asymptotically optimal combination of multiple diagnostic tests. *J. Amer. Statist. Assoc.* **111** 861–874. [MR3538711](https://doi.org/10.1080/01621459.2015.1066681)
<https://doi.org/10.1080/01621459.2015.1066681>
- CHEN, X., VEXLER, A. and MARKATOU, M. (2015). Empirical likelihood ratio confidence interval estimation of best linear combinations of biomarkers. *Comput. Statist. Data Anal.* **82** 186–198. [MR3282175](https://doi.org/10.1016/j.csda.2014.09.010)
<https://doi.org/10.1016/j.csda.2014.09.010>
- EGUCHI, S. and COPAS, J. (2002). A class of logistic-type discriminant functions. *Biometrika* **89** 1–22. [MR1888342](https://doi.org/10.1093/biomet/89.1.1) <https://doi.org/10.1093/biomet/89.1.1>
- ELLIOTT, G. and LIELI, R. P. (2013). Predicting binary outcomes. *J. Econometrics* **174** 15–26. [MR3036958](https://doi.org/10.1016/j.jeconom.2013.01.003)
<https://doi.org/10.1016/j.jeconom.2013.01.003>
- FLORIOS, K. and SKOURAS, S. (2008). Exact computation of max weighted score estimators. *J. Econometrics* **146** 86–91. [MR2459645](https://doi.org/10.1016/j.jeconom.2008.05.018) <https://doi.org/10.1016/j.jeconom.2008.05.018>

- FONG, Y., YIN, S. and HUANG, Y. (2016). Combining biomarkers linearly and nonlinearly for classification using the area under the ROC curve. *Stat. Med.* **35** 3792–3809. [MR3538048](#) <https://doi.org/10.1002/sim.6956>
- HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. *Springer Series in Statistics*. Springer, New York. [MR2722294](#) <https://doi.org/10.1007/978-0-387-84858-7>
- KIEFER, J. (1967). On Bahadur's representation of sample quantiles. *Ann. Math. Stat.* **38** 1323–1342. [MR0217844](#) <https://doi.org/10.1214/aoms/1177698690>
- 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). *Introduction to Empirical Processes and Semiparametric Inference*. *Springer Series in Statistics*. Springer, New York. [MR2724368](#) <https://doi.org/10.1007/978-0-387-74978-5>
- LIN, H., ZHOU, L., PENG, H. and ZHOU, X.-H. (2011). Selection and combination of biomarkers using ROC method for disease classification and prediction. *Canad. J. Statist.* **39** 324–343. [MR2839483](#) <https://doi.org/10.1002/cjs.10107>
- LIPP, T. and BOYD, S. (2016). Variations and extension of the convex-concave procedure. *Optim. Eng.* **17** 263–287. [MR3511445](#) <https://doi.org/10.1007/s11081-015-9294-x>
- MA, S. and HUANG, J. (2007). Combining multiple markers for classification using ROC. *Biometrics* **63** 751–757. [MR2395712](#) <https://doi.org/10.1111/j.1541-0420.2006.00731.x>
- MANSKI, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *J. Econometrics* **3** 205–228. [MR0436905](#) [https://doi.org/10.1016/0304-4076\(75\)90032-9](https://doi.org/10.1016/0304-4076(75)90032-9)
- MANSKI, C. F. (1985). Semiparametric analysis of discrete response. Asymptotic properties of the maximum score estimator. *J. Econometrics* **27** 313–333. [MR0788628](#) [https://doi.org/10.1016/0304-4076\(85\)90009-0](https://doi.org/10.1016/0304-4076(85)90009-0)
- MCINTOSH, M. W. and PEPE, M. S. (2002). Combining several screening tests: Optimality of the risk score. *Biometrics* **58** 657–664. [MR1926119](#) <https://doi.org/10.1111/j.0006-341X.2002.00657.x>
- MEISNER, A., CARONE, M., PEPE, M. S. and KERR, K. F. (2021). Combining biomarkers by maximizing the true positive rate for a fixed false positive rate. *Biom. J.* **63** 1223–1240. [MR4307731](#) <https://doi.org/10.1002/bimj.202000210>
- OU, F.-S., ZENG, D. and CAI, J. (2016). Quantile regression models for current status data. *J. Statist. Plann. Inference* **178** 112–127. [MR3531531](#) <https://doi.org/10.1016/j.jspi.2016.06.001>
- PEPE, M. S., CAI, T. and LONGTON, G. (2006). Combining predictors for classification using the area under the receiver operating characteristic curve. *Biometrics* **62** 221–229, 319. [MR2226576](#) <https://doi.org/10.1111/j.1541-0420.2005.00420.x>
- PEPE, M. S. and THOMPSON, M. L. (2000). Combining diagnostic test results to increase accuracy. *Biostatistics* **1** 123–140.
- RUDIN, W. (1976). *Principles of Mathematical Analysis*, 3rd ed. *International Series in Pure and Applied Mathematics*. McGraw-Hill, New York-Auckland-Düsseldorf. [MR0385023](#)
- SANDA, M. G., FENG, Z., HOWARD, D. H., TOMLINS, S. A., SOKOLL, L. J., CHAN, D. W., REGAN, M. M., GROSKOPF, J., CHIPMAN, J. et al. (2017). Association between combined TMPRSS2:ERG and PCA3 RNA urinary testing and detection of aggressive prostate cancer. *JAMA Oncol.* **3** 1085–1093.
- SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. [MR0595165](#)
- VEXLER, A., LIU, A., SCHISTERMAN, E. F. and WU, C. (2006). Note on distribution-free estimation of maximum linear separation of two multivariate distributions. *J. Nonparametr. Stat.* **18** 145–158. [MR2229886](#) <https://doi.org/10.1080/10485250600662260>
- YAN, Q., BANTIS, L. E., STANFORD, J. L. and FENG, Z. (2018). Combining multiple biomarkers linearly to maximize the partial area under the ROC curve. *Stat. Med.* **37** 627–642. [MR3748766](#) <https://doi.org/10.1002/sim.7535>
- YUILLE, A. L. and RANGARAJAN, A. (2003). The concave-convex procedure. *Neural Comput.* **15** 915–936.

THE INTERPOLATION PHASE TRANSITION IN NEURAL NETWORKS: MEMORIZATION AND GENERALIZATION UNDER LAZY TRAINING

BY ANDREA MONTANARI^a AND YIQIAO ZHONG^b

Department of Electrical Engineering and Department of Statistics, Stanford University, ^amontanar@stanford.edu,
^byiqiao@stanford.edu

Modern neural networks are often operated in a strongly overparametrized regime: they comprise so many parameters that they can interpolate the training set, even if actual labels are replaced by purely random ones. Despite this, they achieve good prediction error on unseen data: interpolating the training set does not lead to a large generalization error. Further, overparametrization appears to be beneficial in that it simplifies the optimization landscape. Here, we study these phenomena in the context of two-layers neural networks in the neural tangent (NT) regime. We consider a simple data model, with isotropic covariates vectors in d dimensions, and N hidden neurons. We assume that both the sample size n and the dimension d are large, and they are polynomially related. Our first main result is a characterization of the eigenstructure of the empirical NT kernel in the overparametrized regime $Nd \gg n$. This characterization implies as a corollary that the minimum eigenvalue of the empirical NT kernel is bounded away from zero as soon as $Nd \gg n$ and, therefore, the network can exactly interpolate arbitrary labels in the same regime.

Our second main result is a characterization of the generalization error of NT ridge regression including, as a special case, min- ℓ_2 norm interpolation. We prove that, as soon as $Nd \gg n$, the test error is well approximated by the one of kernel ridge regression with respect to the infinite-width kernel. The latter is in turn well approximated by the error of polynomial ridge regression, whereby the regularization parameter is increased by a “self-induced” term related to the high-degree components of the activation function. The polynomial degree depends on the sample size and the dimension (in particular on $\log n / \log d$).

REFERENCES

- [1] ALLEN-ZHU, Z., LI, Y. and LIANG, Y. (2019). Learning and generalization in overparameterized neural networks, going beyond two layers. *Adv. Neural Inf. Process. Syst.* **32** 6158–6169.
- [2] ALLEN-ZHU, Z., LI, Y. and SONG, Z. (2019). A convergence theory for deep learning via overparameterization. In *International Conference on Machine Learning* 242–252.
- [3] ARORA, S., DU, S., HU, W., LI, Z. and WANG, R. (2019). Fine-grained analysis of optimization and generalization for overparameterized two-layer neural networks. In *International Conference on Machine Learning* 322–332. PMLR.
- [4] ATKINSON, K. and HAN, W. (2012). *Spherical Harmonics and Approximations on the Unit Sphere: An Introduction. Lecture Notes in Math.* **2044**. Springer, Heidelberg. MR2934227 <https://doi.org/10.1007/978-3-642-25983-8>
- [5] BACH, F. (2017). Breaking the curse of dimensionality with convex neural networks. *J. Mach. Learn. Res.* **18** Paper No. 19, 53 pp. MR3634886
- [6] 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 pp. MR3960917

MSC2020 subject classifications. Primary 62J07; secondary 62J05.

Key words and phrases. Neural tangent kernel, memorization, overfitting, overparametrization, kernel ridge regression.

- [7] BARTLETT, P. L., LONG, P. M., LUGOSI, G. and TSIGLER, A. (2020). Benign overfitting in linear regression. *Proc. Natl. Acad. Sci. USA* **117** 30063–30070. MR4263288 <https://doi.org/10.1073/pnas.1907378117>
- [8] BARTLETT, P. L., MONTANARI, A. and RAKHLIN, A. (2021). Deep learning: A statistical viewpoint. *Acta Numer.* **30** 87–201. MR4295218 <https://doi.org/10.1017/S0962492921000027>
- [9] BAUM, E. B. (1988). On the capabilities of multilayer perceptrons. *J. Complexity* **4** 193–215. MR0960066 [https://doi.org/10.1016/0885-064X\(88\)90020-9](https://doi.org/10.1016/0885-064X(88)90020-9)
- [10] BELKIN, M., HSU, D., MA, S. and MANDAL, S. (2019). Reconciling modern machine-learning practice and the classical bias-variance trade-off. *Proc. Natl. Acad. Sci. USA* **116** 15849–15854. MR3997901 <https://doi.org/10.1073/pnas.1903070116>
- [11] BELKIN, M., HSU, D. and XU, J. (2020). Two models of double descent for weak features. *SIAM J. Math. Data Sci.* **2** 1167–1180. MR4186534 <https://doi.org/10.1137/20M1336072>
- [12] BIETTI, A. and MAIRAL, J. (2019). On the inductive bias of neural tangent kernels. Preprint. Available at [arXiv:1905.12173](https://arxiv.org/abs/1905.12173).
- [13] BUBECK, S., ELDAN, R., LEE, Y. T. and MIKULINCER, D. (2020). Network size and weights size for memorization with two-layers neural networks. Preprint. Available at [arXiv:2006.02855](https://arxiv.org/abs/2006.02855).
- [14] CAO, Y. and GU, Q. (2019). Generalization bounds of stochastic gradient descent for wide and deep neural networks. *Adv. Neural Inf. Process. Syst.* **32** 10836–10846.
- [15] CHEN, Z., CAO, Y., ZOU, D. and GU, Q. (2019). How much over-parameterization is sufficient to learn deep relu networks? Preprint. Available at [arXiv:1911.12360](https://arxiv.org/abs/1911.12360).
- [16] CHIZAT, L., OYALLON, E. and BACH, F. (2019). On lazy training in differentiable programming. In *Advances in Neural Information Processing Systems* 2937–2947.
- [17] COSTAS, E. and CHRISTOPHER, F. (2014). *Spherical Harmonics in p Dimensions*. World Scientific, Singapore.
- [18] COVER, T. M. (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Trans. Electron. Comput.* **3** 326–334.
- [19] DANIELY, A. (2019). Neural networks learning and memorization with (almost) no over-parameterization. Preprint. Available at [arXiv:1911.09873](https://arxiv.org/abs/1911.09873).
- [20] DANIELY, A. (2020). Memorizing Gaussians with no over-parameterization via gradient descent on neural networks. Preprint. Available at [arXiv:2003.12895](https://arxiv.org/abs/2003.12895).
- [21] DU, S. S., ZHAI, X., POZOS, B. and SINGH, A. (2018). Gradient descent provably optimizes over-parameterized neural networks. In *International Conference on Learning Representations*.
- [22] E, W., MA, C. and WU, L. (2020). A comparative analysis of optimization and generalization properties of two-layer neural network and random feature models under gradient descent dynamics. *Sci. China Math.* **63** 1235–1258. MR4119555 <https://doi.org/10.1007/s11425-019-1628-5>
- [23] GHORBANI, B., MEI, S., MISIAKIEWICZ, T. and MONTANARI, A. (2021). When do neural networks outperform kernel methods? *J. Stat. Mech. Theory Exp.* **12** Paper No. 124009, 110 pp. MR4412837 <https://doi.org/10.1088/1742-5468/ac3a81>
- [24] GHORBANI, B., MEI, S., MISIAKIEWICZ, T. and MONTANARI, A. (2021). Linearized two-layers neural networks in high dimension. *Ann. Statist.* **49** 1029–1054. MR4255118 <https://doi.org/10.1214/20-aos1990>
- [25] HASTIE, T., MONTANARI, A., ROSSET, S. and TIBSHIRANI, R. J. (2022). Surprises in high-dimensional ridgeless least squares interpolation. *Ann. Statist.* **50** 949–986. MR4404925 <https://doi.org/10.1214/21-aos2133>
- [26] JACOT, A., GABRIEL, F. and HONGLER, C. (2018). Neural tangent kernel: Convergence and generalization in neural networks. In *Advances in Neural Information Processing Systems* 8571–8580.
- [27] JI, Z. and TELGARSKY, M. (2019). Polylogarithmic width suffices for gradient descent to achieve arbitrarily small test error with shallow ReLU networks. In *International Conference on Learning Representations*.
- [28] KOWALCZYK, A. (1994). Counting function theorem for multi-layer networks. In *Advances in Neural Information Processing Systems* 375–382.
- [29] LEE, J., XIAO, L., SCHOENHOLZ, S., BAHRI, Y., NOVAK, R., SOHL-DICKSTEIN, J. and PENNINGTON, J. (2019). Wide neural networks of any depth evolve as linear models under gradient descent. *Adv. Neural Inf. Process. Syst.* **32** 8572–8583.
- [30] LIANG, T. and RAKHLIN, A. (2020). Just interpolate: Kernel “ridgeless” regression can generalize. *Ann. Statist.* **48** 1329–1347. MR4124325 <https://doi.org/10.1214/19-AOS1849>
- [31] LIANG, T., RAKHLIN, A. and ZHAI, X. (2020). On the multiple descent of minimum-norm interpolants and restricted lower isometry of kernels. In *Proceedings of Thirty Third Conference on Learning Theory* (J. Abernethy and S. Agarwal, eds.). *Proceedings of Machine Learning Research* **125** 2683–2711. PMLR. Available at [arXiv:1908.10292](https://arxiv.org/abs/1908.10292).

- [32] LIU, C., ZHU, L. and BELKIN, M. (2020). On the linearity of large non-linear models: When and why the tangent kernel is constant. In *Advances in Neural Information Processing Systems* (H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan and H. Lin, eds.) **33** 15954–15964. Curran Associates, Red Hook.
- [33] MEI, S., MISIAKIEWICZ, T. and MONTANARI, A. (2022). Generalization error of random feature and kernel methods: Hypercontractivity and kernel matrix concentration. *Appl. Comput. Harmon. Anal.* **59** 3–84. [MR4412180](#) <https://doi.org/10.1016/j.acha.2021.12.003>
- [34] MEI, S. and MONTANARI, A. (2022). The generalization error of random features regression: Precise asymptotics and the double descent curve. *Comm. Pure Appl. Math.* **75** 667–766. [MR4400901](#) <https://doi.org/10.1002/cpa.22008>
- [35] MONTANARI, A., RUAN, F., SOHN, Y. and YAN, J. (2019). The generalization error of max-margin linear classifiers: High-dimensional asymptotics in the overparametrized regime. Preprint. Available at [arXiv:1911.01544](#).
- [36] MONTANARI, A. and ZHONG, Y. (2022). Supplement to “The interpolation phase transition in neural networks: Memorization and generalization under lazy training.” <https://doi.org/10.1214/22-AOS2211SUPP>
- [37] NEYSHABUR, B., TOMIOKA, R. and SREBRO, N. (2015). In search of the real inductive bias: On the role of implicit regularization in deep learning. In *ICLR (Workshop)*.
- [38] NITANDA, A., CHINOT, G. and SUZUKI, T. (2019). Gradient descent can learn less over-parameterized two-layer neural networks on classification problems. Preprint. Available at [arXiv:1905.09870](#).
- [39] OYMAK, S. and SOLTANOLKOTABI, M. (2020). Towards moderate overparameterization: Global convergence guarantees for training shallow neural networks. *IEEE J. Sel. Areas Inf. Theory* **1** 84–105.
- [40] PARK, S., LEE, J., YUN, C. and SHIN, J. (2021). Provable memorization via deep neural networks using sub-linear parameters. In *Conference on Learning Theory* 3627–3661. PMLR.
- [41] RAHIMI, A. and RECHT, B. (2008). Random features for large-scale kernel machines. In *Advances in Neural Information Processing Systems* 1177–1184.
- [42] SAKURAI, A. (1992). n - h -1 networks store no less n^*h+1 examples, but sometimes no more. In [*Proceedings 1992 IJCNN International Joint Conference on Neural Networks* **3** 936–941.
- [43] SCHOENBERG, I. J. (1942). Positive definite functions on spheres. *Duke Math. J.* **9** 96–108. [MR0005922](#)
- [44] SOLTANOLKOTABI, M., JAVANMARD, A. and LEE, J. D. (2019). Theoretical insights into the optimization landscape of over-parameterized shallow neural networks. *IEEE Trans. Inf. Theory* **65** 742–769. [MR3904911](#) <https://doi.org/10.1109/TIT.2018.2854560>
- [45] VARDI, G., YEHUDAI, G. and SHAMIR, O. (2021). On the optimal memorization power of ReLU neural networks. Preprint. Available at [arXiv:2110.03187](#).
- [46] VERSHYNIN, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge Univ. Press, Cambridge. [MR2963170](#)
- [47] VERSHYNIN, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics **47**. Cambridge Univ. Press, Cambridge. [MR3837109](#) <https://doi.org/10.1017/9781108231596>
- [48] VERSHYNIN, R. (2020). Memory capacity of neural networks with threshold and rectified linear unit activations. *SIAM J. Math. Data Sci.* **2** 1004–1033. [MR4164497](#) <https://doi.org/10.1137/20M1314884>
- [49] YUN, C., SRA, S. and JADBABAIE, A. (2019). Small ReLU networks are powerful memorizers: A tight analysis of memorization capacity. In *Advances in Neural Information Processing Systems* 15558–15569.
- [50] ZHANG, C., BENGIO, S., HARDT, M., RECHT, B. and VINYALS, O. (2016). Understanding deep learning requires rethinking generalization. Preprint. Available at [arXiv:1611.03530](#).
- [51] ZOU, D., CAO, Y., ZHOU, D. and GU, Q. (2020). Gradient descent optimizes over-parameterized deep ReLU networks. *Mach. Learn.* **109** 467–492. [MR4075425](#) <https://doi.org/10.1007/s10994-019-05839-6>

A SIEVE STOCHASTIC GRADIENT DESCENT ESTIMATOR FOR ONLINE NONPARAMETRIC REGRESSION IN SOBOLEV ELLIPSOIDS

BY TIANYU ZHANG^a AND NOAH SIMON^b

Department of Biostatistics, University of Washington, ^azty@uw.edu, ^bnrsimon@uw.edu

The goal of regression is to recover an unknown underlying function that best links a set of predictors to an outcome from noisy observations. In nonparametric regression, one assumes that the regression function belongs to a prespecified infinite-dimensional function space (the hypothesis space). In the online setting, when the observations come in a stream, it is computationally-preferable to iteratively update an estimate rather than refitting an entire model repeatedly. Inspired by nonparametric sieve estimation and stochastic approximation methods, we propose a sieve stochastic gradient descent estimator (Sieve-SGD) when the hypothesis space is a Sobolev ellipsoid. We show that Sieve-SGD has rate-optimal mean squared error (MSE) under a set of simple and direct conditions. The proposed estimator can be constructed with a low computational (time and space) expense: We also formally show that Sieve-SGD requires almost minimal memory usage among all statistically rate-optimal estimators.

REFERENCES

- [1] ARORA, S. and BARAK, B. (2009). *Computational Complexity: A Modern Approach*. Cambridge Univ. Press, Cambridge. [MR2500087](#) <https://doi.org/10.1017/CBO9780511804090>
- [2] BABICHEV, D. and BACH, F. (2018). Constant step size stochastic gradient descent for probabilistic modeling. *Stat* **1050** 21.
- [3] BACH, F. and MOULINES, E. (2013). Non-strongly-convex smooth stochastic approximation with convergence rate $o(1/n)$. In *Advances in Neural Information Processing Systems* 773–781.
- [4] BELKIN, M., HSU, D., MA, S. and MANDAL, S. (2019). Reconciling modern machine-learning practice and the classical bias-variance trade-off. *Proc. Natl. Acad. Sci. USA* **116** 15849–15854. [MR3997901](#) <https://doi.org/10.1073/pnas.1903070116>
- [5] BERLINET, A. and THOMAS-AGNAN, C. (2011). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer, Berlin.
- [6] BORKAR, V. S. (2009). *Stochastic Approximation: A Dynamical Systems Viewpoint* **48**. Springer, Berlin.
- [7] BOTTOU, L. (2010). Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010* 177–186. Physica-Verlag/Springer, Heidelberg. [MR3362066](#)
- [8] 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>
- [9] CALANDRIELLO, D., LAZARIC, A. and VALKO, M. (2017). Efficient second-order online kernel learning with adaptive embedding. In *Advances in Neural Information Processing Systems* 6140–6150.
- [10] CAPONNETTO, A. and DE VITO, E. (2007). Optimal rates for the regularized least-squares algorithm. *Found. Comput. Math.* **7** 331–368. [MR2335249](#) <https://doi.org/10.1007/s10208-006-0196-8>
- [11] CHRISTMANN, A. and STEINWART, I. (2008). Support vector machines.
- [12] CUCKER, F. and SMALE, S. (2002). On the mathematical foundations of learning. *Bull. Amer. Math. Soc. (N.S.)* **39** 1–49. [MR1864085](#) <https://doi.org/10.1090/S0273-0979-01-00923-5>
- [13] DIEULEVEUT, A. and BACH, F. (2016). Nonparametric stochastic approximation with large step-sizes. *Ann. Statist.* **44** 1363–1399. [MR3519927](#) <https://doi.org/10.1214/15-AOS1391>
- [14] DUCHI, J. C. (2014). Multiple Optimality Guarantees in Statistical Learning Ph.D. thesis UC Berkeley.
- [15] EUBANK, R. L. and SPECKMAN, P. (1990). Curve fitting by polynomial-trigonometric regression. *Biometrika* **77** 1–9. [MR1049403](#) <https://doi.org/10.1093/biomet/77.1.1>

- [16] FASSHAUER, G. E. and MCCOURT, M. J. (2015). *Kernel-Based Approximation Methods Using Matlab* **19**. World Scientific, Singapore.
- [17] FROSTIG, R., GE, R., KAKADE, S. M. and SIDFORD, A. (2015). Competing with the empirical risk minimizer in a single pass. In *Conference on Learning Theory* 728–763.
- [18] GAILLARD, P. and GERCHINOVITZ, S. (2015). A chaining algorithm for online nonparametric regression. In *Conference on Learning Theory* 764–796.
- [19] GAO, C., MA, Z., REN, Z. and ZHOU, H. H. (2015). Minimax estimation in sparse canonical correlation analysis. *Ann. Statist.* **43** 2168–2197. [MR3396982](#) <https://doi.org/10.1214/15-AOS1332>
- [20] GAO, C., MA, Z. and ZHOU, H. H. (2017). Sparse CCA: Adaptive estimation and computational barriers. *Ann. Statist.* **45** 2074–2101. [MR3718162](#) <https://doi.org/10.1214/16-AOS1519>
- [21] GEER, S. A. and VAN DE GEER, S. (2000). *Empirical Processes in M-Estimation* **6**. Cambridge Univ. Press, Cambridge.
- [22] GYÖRFI, L., KOHLER, M., KRZYZAK, A. and WALK, H. (2006). *A Distribution-Free Theory of Nonparametric Regression*. Springer, Berlin.
- [23] HALL, P. and OPSOMER, J. D. (2005). Theory for penalised spline regression. *Biometrika* **92** 105–118. [MR2158613](#) <https://doi.org/10.1093/biomet/92.1.105>
- [24] HÄRDLE, W., KERKYACHARIAN, G., PICARD, D. and TSYBAKOV, A. (2012). *Wavelets, Approximation, and Statistical Applications* **129**. Springer, Berlin.
- [25] HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. *Springer Series in Statistics*. Springer, New York. [MR2722294](#) <https://doi.org/10.1007/978-0-387-84858-7>
- [26] HERNÁNDEZ, E. and WEISS, G. (1996). *A First Course on Wavelets. Studies in Advanced Mathematics*. CRC Press, Boca Raton, FL. [MR1408902](#) <https://doi.org/10.1201/9781420049985>
- [27] KENNEDY, R. A., SADEGHI, P., KHALID, Z. and MCEWEN, J. D. (2013). Classification and construction of closed-form kernels for signal representation on the 2-sphere. In *Wavelets and Sparsity XV* **8858** 88580M. International Society for Optics and Photonics.
- [28] KOLMOGOROV, A. N. and TIHOMIROV, V. M. (1959). ε -entropy and ε -capacity of sets in function spaces. *Uspekhi Mat. Nauk* **14** 3–86. [MR0112032](#)
- [29] KOPPEL, A., WARNELL, G., STUMP, E. and RIBEIRO, A. (2019). Parsimonious online learning with kernels via sparse projections in function space. *J. Mach. Learn. Res.* **20** 83–126.
- [30] KUSHNER, H. J. and YIN, G. G. (2003). *Stochastic Approximation and Recursive Algorithms and Applications*, 2nd ed. *Applications of Mathematics (New York)* **35**. Springer, New York. [MR1993642](#)
- [31] LIANG, T. and RAKHLIN, A. (2020). Just interpolate: Kernel “ridgeless” regression can generalize. *Ann. Statist.* **48** 1329–1347. [MR4124325](#) <https://doi.org/10.1214/19-AOS1849>
- [32] LU, J., HOI, S. C., WANG, J., ZHAO, P. and LIU, Z.-Y. (2016). Large scale online kernel learning. *J. Mach. Learn. Res.* **17** 1613–1655.
- [33] MA, Z. and WU, Y. (2015). Computational barriers in minimax submatrix detection. *Ann. Statist.* **43** 1089–1116. [MR3346698](#) <https://doi.org/10.1214/14-AOS1300>
- [34] MARTEAU-FEREY, U., BACH, F. and RUDI, A. (2019). Globally convergent Newton methods for Ill-conditioned generalized self-concordant losses. In *Advances in Neural Information Processing Systems* 7634–7644.
- [35] MARTEAU-FEREY, U., OSTROVSKII, D., BACH, F. and RUDI, A. (2019). Beyond least-squares: Fast rates for regularized empirical risk minimization through self-concordance. In *Conference on Learning Theory* 2294–2340. PMLR.
- [36] MICHEL, V. (2012). *Lectures on Constructive Approximation: Fourier, Spline, and Wavelet Methods on the Real Line, the Sphere, and the Ball*. Springer, Berlin.
- [37] NEMIROVSKI, A. (2000). Topics in non-parametric statistics. In *Lectures on Probability Theory and Statistics (Saint-Flour, 1998). Lecture Notes in Math.* **1738** 85–277. Springer, Berlin. [MR1775640](#)
- [38] NEMIROVSKIY, A. S. and YUDIN, D. B. (1983). *Problem Complexity and Method Efficiency in Optimization. Wiley-Interscience Series in Discrete Mathematics*. Wiley, New York. [MR0702836](#)
- [39] NOVAK, E. and WOZNIAKOWSKI, H. (2008). *Tractability of Multivariate Problems. Vol. 1: Linear Information*.
- [40] RAKHLIN, A. and SRIDHARAN, K. (2015). Online nonparametric regression with general loss functions. arXiv preprint. Available at [arXiv:1501.06598](#).
- [41] RASKUTTI, G., YU, B. and WAINWRIGHT, M. J. (2009). Lower bounds on minimax rates for nonparametric regression with additive sparsity and smoothness. *Adv. Neural Inf. Process. Syst.* **22** 1563–1570.
- [42] SHEN, X. (1997). On methods of sieves and penalization. *Ann. Statist.* **25** 2555–2591. [MR1604416](#) <https://doi.org/10.1214/aos/1030741085>
- [43] SI, S., KUMAR, S. and LI, Y. (2018). Nonlinear online learning with adaptive Nyström approximation. arXiv preprint. Available at [arXiv:1802.07887](#).

- [44] STEINWART, I. and SCOVEL, C. (2012). Mercer’s theorem on general domains: On the interaction between measures, kernels, and RKHSs. *Constr. Approx.* **35** 363–417. [MR2914365](#) <https://doi.org/10.1007/s00365-012-9153-3>
- [45] STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360. [MR0594650](#)
- [46] STONE, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13** 689–705. [MR0790566](#) <https://doi.org/10.1214/aos/1176349548>
- [47] SUN, H. (2005). Mercer theorem for RKHS on noncompact sets. *J. Complexity* **21** 337–349. [MR2138444](#) <https://doi.org/10.1016/j.jco.2004.09.002>
- [48] TARRÈS, P. and YAO, Y. (2014). Online learning as stochastic approximation of regularization paths: Optimality and almost-sure convergence. *IEEE Trans. Inf. Theory* **60** 5716–5735. [MR3252416](#) <https://doi.org/10.1109/TIT.2014.2332531>
- [49] TSYBAKOV, A. (2008). *Introduction to Nonparametric Estimation*. Springer, Berlin.
- [50] VEMPALA, S. S. (2005). *The Random Projection Method* **65**. American Mathematical Soc., Providence.
- [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] 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>
- [53] WANG, T., BERTHET, Q. and SAMWORTH, R. J. (2016). Statistical and computational trade-offs in estimation of sparse principal components. *Ann. Statist.* **44** 1896–1930. [MR3546438](#) <https://doi.org/10.1214/15-AOS1369>
- [54] WOOD, S. N. (2017). *Generalized Additive Models: An Introduction with R. Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. [MR3726911](#)
- [55] YING, Y. and PONTIL, M. (2008). Online gradient descent learning algorithms. *Found. Comput. Math.* **8** 561–596. [MR2443089](#) <https://doi.org/10.1007/s10208-006-0237-y>
- [56] YUAN, M. and CAI, T. T. (2010). A reproducing kernel Hilbert space approach to functional linear regression. *Ann. Statist.* **38** 3412–3444. [MR2766857](#) <https://doi.org/10.1214/09-AOS772>
- [57] YUAN, M. and ZHOU, D.-X. (2016). Minimax optimal rates of estimation in high dimensional additive models. *Ann. Statist.* **44** 2564–2593. [MR3576554](#) <https://doi.org/10.1214/15-AOS1422>
- [58] ZHANG, T. and SIMON, N. (2021). An online projection estimator for nonparametric regression in reproducing kernel Hilbert spaces. arXiv preprint. Available at [arXiv:2104.00780](#).
- [59] ZHANG, T. and SIMON, N. (2022). Regression in tensor product spaces by the method of sieves. arXiv preprint. Available at [arXiv:2206.02994](#).
- [60] ZHANG, T. and SIMON, N. (2022). Supplement to “A Sieve Stochastic Gradient Descent estimator for online nonparametric regression in Sobolev ellipsoids.” <https://doi.org/10.1214/22-AOS2212SUPP>
- [61] ZHANG, Y., WAINWRIGHT, M. J. and JORDAN, M. I. (2014). Lower bounds on the performance of polynomial-time algorithms for sparse linear regression. In *Conference on Learning Theory* 921–948.

NONPARAMETRIC BAYESIAN INFERENCE FOR REVERSIBLE MULTIDIMENSIONAL DIFFUSIONS

BY MATTEO GIORDANO^{1,a} AND KOLYAN RAY^{2,b}

¹Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, ^amg846@cam.ac.uk

²Department of Mathematics, Imperial College London, ^bkolyan.ray@imperial.ac.uk

We study nonparametric Bayesian models for reversible multidimensional diffusions with periodic drift. For continuous observation paths, reversibility is exploited to prove a general posterior contraction rate theorem for the drift gradient vector field under approximation-theoretic conditions on the induced prior for the invariant measure. The general theorem is applied to Gaussian priors and p -exponential priors, which are shown to converge to the truth at the optimal nonparametric rate over Sobolev smoothness classes in any dimension.

REFERENCES

- [1] ABRAHAM, K. (2019). Nonparametric Bayesian posterior contraction rates for scalar diffusions with high-frequency data. *Bernoulli* **25** 2696–2728. MR4003562 <https://doi.org/10.3150/18-BEJ1067>
- [2] ABRAHAM, K. and NICKL, R. (2019). On statistical Calderón problems. *Math. Stat. Learn.* **2** 165–216. MR4130599
- [3] 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 pp. MR3774703 <https://doi.org/10.1088/1361-6420/aaacac>
- [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 WANG, S. (2021). Laplace priors and spatial inhomogeneity in Bayesian inverse problems. Preprint. Available at [arXiv:2112.05679](https://arxiv.org/abs/2112.05679).
- [6] AURZADA, F. (2007). On the lower tail probabilities of some random sequences in l_p . *J. Theoret. Probab.* **20** 843–858. MR2359058 <https://doi.org/10.1007/s10959-007-0095-9>
- [7] BAKRY, D., GENTIL, I. and LEDOUX, M. (2014). *Analysis and Geometry of Markov Diffusion Operators. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **348**. Springer, Cham. MR3155209 <https://doi.org/10.1007/978-3-319-00227-9>
- [8] BASS, R. F. (2011). *Stochastic Processes. Cambridge Series in Statistical and Probabilistic Mathematics* **33**. Cambridge Univ. Press, Cambridge. MR2856623 <https://doi.org/10.1017/CBO9780511997044>
- [9] BATZ, P., RUTTOR, A. and OPPER, M. (2018). Approximate Bayes learning of stochastic differential equations. *Phys. Rev. E* **98** 022109, 19 pp. MR3862381 <https://doi.org/10.1103/physreve.98.022109>
- [10] BERS, L., JOHN, F. and SCHECHTER, M. (1964). *Partial Differential Equations. Lectures in Applied Mathematics III*. Wiley & Sons, Inc., New York–London–Sydney. MR0163043
- [11] BESKOS, A., PAPASPILIOPOULOS, O., ROBERTS, G. O. and FEARNHED, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 333–382. MR2278331 <https://doi.org/10.1111/j.1467-9868.2006.00552.x>
- [12] BOLHUIS, P. G., CHANDLER, D., DELLAGO, C. and GEISLER, P. L. (2002). Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* **53** 291–318. <https://doi.org/10.1146/annurev.physchem.53.082301.113146>
- [13] BOURDAUD, G. and SICKEL, W. (2011). Composition operators on function spaces with fractional order of smoothness. In *Harmonic Analysis and Nonlinear Partial Differential Equations. RIMS Kôkyûroku Bessatsu*, B26 93–132. Res. Inst. Math. Sci. (RIMS), Kyoto. MR2883848

MSC2020 subject classifications. Primary 62G20; secondary 62F15, 60J60.

Key words and phrases. Bayesian nonparametrics, multidimensional diffusions, reversibility, Gaussian processes, Laplace prior.

- [14] 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>
- [15] DALALYAN, A. and REISS, M. (2007). Asymptotic statistical equivalence for ergodic diffusions: The multidimensional case. *Probab. Theory Related Fields* **137** 25–47. MR2278451 <https://doi.org/10.1007/s00440-006-0502-7>
- [16] DALALYAN, A. S. (2017). Theoretical guarantees for approximate sampling from smooth and log-concave densities. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 651–676. MR3641401 <https://doi.org/10.1111/rssb.12183>
- [17] 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>
- [18] DASHTI, M., LAW, K. J. H., STUART, A. M. and VOSS, J. (2013). MAP estimators and their consistency in Bayesian nonparametric inverse problems. *Inverse Probl.* **29** 095017, 27 pp. MR3104933 <https://doi.org/10.1088/0266-5611/29/9/095017>
- [19] DASHTI, M. and STUART, A. M. (2017). The Bayesian approach to inverse problems. In *Handbook of Uncertainty Quantification*. Vols. 1, 2, 3 311–428. Springer, Cham. MR3839555
- [20] DELYON, B. and JUDITSKY, A. (1996). On minimax wavelet estimators. *Appl. Comput. Harmon. Anal.* **3** 215–228. MR1400080 <https://doi.org/10.1006/acha.1996.0017>
- [21] DIRKSEN, S. (2015). Tail bounds via generic chaining. *Electron. J. Probab.* **20** Paper No. 53, 29 pp. MR3354613 <https://doi.org/10.1214/EJP.v20-3760>
- [22] DURMUS, A. and MOULINES, É. (2017). Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *Ann. Appl. Probab.* **27** 1551–1587. MR3678479 <https://doi.org/10.1214/16-AAP1238>
- [23] DURMUS, A. and MOULINES, É. (2019). High-dimensional Bayesian inference via the unadjusted Langevin algorithm. *Bernoulli* **25** 2854–2882. MR4003567 <https://doi.org/10.3150/18-BEJ1073>
- [24] DURMUS, A., MOULINES, É. and PEREYRA, M. (2018). Efficient Bayesian computation by proximal Markov chain Monte Carlo: When Langevin meets Moreau. *SIAM J. Imaging Sci.* **11** 473–506. MR3763089 <https://doi.org/10.1137/16M1108340>
- [25] 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>
- [26] GARCÍA-PORTUGUÉS, E., SØRENSEN, M., MARDIA, K. V. and HAMELRYCK, T. (2019). Langevin diffusions on the torus: Estimation and applications. *Stat. Comput.* **29** 1–22. MR3905536 <https://doi.org/10.1007/s11222-017-9790-2>
- [27] GARRALDA, N., LLATSER, I., CABELLOS-APARICIO, A., ALARCÓN, E. and PIEROBON, M. (2011). Diffusion-based physical channel identification in molecular nanonetworks. *Nano Commun. Netw.* **2** 196–204. <https://doi.org/10.1016/j.nancom.2011.07.001>
- [28] 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>
- [29] 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>
- [30] GINÉ, E. and NICKL, R. (2011). Rates of contraction for posterior distributions in L^r -metrics, $1 \leq r \leq \infty$. *Ann. Statist.* **39** 2883–2911. MR3012395 <https://doi.org/10.1214/11-AOS924>
- [31] 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>
- [32] GIORDANO, M. and NICKL, R. (2020). Consistency of Bayesian inference with Gaussian process priors in an elliptic inverse problem. *Inverse Probl.* **36** 085001, 35 pp. MR4151406 <https://doi.org/10.1088/1361-6420/ab7d2a>
- [33] GIORDANO, M. and RAY, K. (2022). Supplement to “Nonparametric Bayesian inference for reversible multidimensional diffusions.” <https://doi.org/10.1214/22-AOS2213SUPP>
- [34] GIORDANO, M., RAY, K. and SCHMIDT-HIEBER, J. (2022). On the inability of Gaussian process regression to optimally learn compositional functions. Preprint. Available at [arXiv:2205.07764](https://arxiv.org/abs/2205.07764).
- [35] GOBET, E., HOFFMANN, M. and REISS, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.* **32** 2223–2253. MR2102509 <https://doi.org/10.1214/009053604000000797>
- [36] GUGUSHVILI, S. and SPREIJ, P. (2014). Nonparametric Bayesian drift estimation for multidimensional stochastic differential equations. *Lith. Math. J.* **54** 127–141. MR3212631 <https://doi.org/10.1007/s10986-014-9232-1>
- [37] HELIN, T. and BURGER, M. (2015). Maximum *a posteriori* probability estimates in infinite-dimensional Bayesian inverse problems. *Inverse Probl.* **31** 085009, 22 pp. MR3377106 <https://doi.org/10.1088/0266-5611/31/8/085009>

- [38] JIA, J., PENG, J. and GAO, J. (2016). Bayesian approach to inverse problems for functions with a variable-index Besov prior. *Inverse Probl.* **32** 085006, 32 pp. MR3535665 <https://doi.org/10.1088/0266-5611/32/8/085006>
- [39] KOLEHMAINEN, V., LASSAS, M., NIINIMÄKI, K. and SILTANEN, S. (2012). Sparsity-promoting Bayesian inversion. *Inverse Probl.* **28** 025005, 28 pp. MR2876856 <https://doi.org/10.1088/0266-5611/28/2/025005>
- [40] KRAMERS, H. A. (1940). Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **7** 284–304. MR0002962
- [41] LASANEN, S. (2012). Non-Gaussian statistical inverse problems. Part I: Posterior distributions. *Inverse Probl. Imaging* **6** 215–266. MR2942739 <https://doi.org/10.3934/ipi.2012.6.215>
- [42] 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>
- [43] LEPORINI, D. and PESQUET, J. C. (2001). Bayesian wavelet denoising: Besov priors and non-Gaussian noises. *Signal Process.* **81** 55–67. Special section on Markov Chain Monte Carlo (MCMC) Methods for Signal Processing. [https://doi.org/10.1016/S0165-1684\(00\)00190-0](https://doi.org/10.1016/S0165-1684(00)00190-0)
- [44] LEPSKI, O. V., MAMMEN, E. and SPOKOINY, V. G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: An approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.* **25** 929–947. MR1447734 <https://doi.org/10.1214/aos/1069362731>
- [45] MARIUCCI, E., RAY, K. and SZABÓ, B. (2020). A Bayesian nonparametric approach to log-concave density estimation. *Bernoulli* **26** 1070–1097. MR4058361 <https://doi.org/10.3150/19-BEJ1139>
- [46] 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>
- [47] 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>
- [48] NICKL, R. and RAY, K. (2020). Nonparametric statistical inference for drift vector fields of multidimensional diffusions. *Ann. Statist.* **48** 1383–1408. MR4124327 <https://doi.org/10.1214/19-AOS1851>
- [49] 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>
- [50] NICKL, R. and WANG, S. (2020). On polynomial-time computation of high-dimensional posterior measures by Langevin-type algorithms. Preprint. Available at [arXiv:2009.05298](https://arxiv.org/abs/2009.05298).
- [51] PAPASPILIOPOULOS, O., POKERN, Y., ROBERTS, G. O. and STUART, A. M. (2012). Nonparametric estimation of diffusions: A differential equations approach. *Biometrika* **99** 511–531. MR2966767 <https://doi.org/10.1093/biomet/ass034>
- [52] PARDOUX, E. and VERETENNIKOV, A. Y. (2001). On the Poisson equation and diffusion approximation. I. *Ann. Probab.* **29** 1061–1085. MR1872736 <https://doi.org/10.1214/aop/1015345596>
- [53] PEREYRA, M. (2016). Proximal Markov chain Monte Carlo algorithms. *Stat. Comput.* **26** 745–760. MR3515019 <https://doi.org/10.1007/s11222-015-9567-4>
- [54] PINSKI, F. J. and STUART, A. M. (2010). Transition paths in molecules: Gradient descent in pathspace. *J. Chem. Phys.* **132** 184104.
- [55] PINSKI, F. J., STUART, A. M. and THEIL, F. (2012). Γ -limit for transition paths of maximal probability. *J. Stat. Phys.* **146** 955–974. MR2902449 <https://doi.org/10.1007/s10955-012-0443-8>
- [56] POKERN, Y., STUART, A. M. and VAN ZANTEN, J. H. (2013). Posterior consistency via precision operators for Bayesian nonparametric drift estimation in SDEs. *Stochastic Process. Appl.* **123** 603–628. MR3003365 <https://doi.org/10.1016/j.spa.2012.08.010>
- [57] RANTALA, M., VÄNSKÄ, S., JÄRVENPÄÄ, S., KALKE, M., LASSAS, M., MÖBERG, J. and SILTANEN, S. (2006). Wavelet-based reconstruction for limited-angle X-ray tomography. *IEEE Trans. Med. Imag.* **25** 210–217. <https://doi.org/10.1109/TMI.2005.862206>
- [58] RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning. Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR2514435
- [59] RAY, K. (2013). Bayesian inverse problems with non-conjugate priors. *Electron. J. Stat.* **7** 2516–2549. MR3117105 <https://doi.org/10.1214/13-EJS851>
- [60] REVUZ, D. and YOR, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **293**. Springer, Berlin. MR1725357 <https://doi.org/10.1007/978-3-662-06400-9>
- [61] RUTTOR, A., BATZ, P. and OPPER, M. (2013). Approximate Gaussian process inference for the drift function in stochastic differential equations. In *Advances in Neural Information Processing Systems* **26** 2040–2048.

- [62] SAKHAEI, E. and ENTEZARI, A. (2015). Spline-based sparse tomographic reconstruction with Besov priors. In *Medical Imaging 2015: Image Processing* (S. Ourselin and M. A. Styner, eds.) **9413** 101–108. International Society for Optics and Photonics. SPIE. <https://doi.org/10.1117/12.2082797>
- [63] SCHAUER, M., VAN DER MEULEN, F. and VAN ZANTEN, H. (2017). Guided proposals for simulating multi-dimensional diffusion bridges. *Bernoulli* **23** 2917–2950. [MR3648050 https://doi.org/10.3150/16-BEJ833](https://doi.org/10.3150/16-BEJ833)
- [64] SCHMEISSER, H.-J. and TRIEBEL, H. (1987). *Topics in Fourier Analysis and Function Spaces*. A Wiley-Interscience Publication. Wiley, Chichester. [MR0891189](https://doi.org/10.1002/9780470210000)
- [65] SCHMISSER, E. (2013). Penalized nonparametric drift estimation for a multidimensional diffusion process. *Statistics* **47** 61–84. [MR3023015 https://doi.org/10.1080/02331888.2011.591931](https://doi.org/10.1080/02331888.2011.591931)
- [66] SCHUSS, Z. (1980). Singular perturbation methods in stochastic differential equations of mathematical physics. *SIAM Rev.* **22** 119–155. [MR0564560 https://doi.org/10.1137/1022024](https://doi.org/10.1137/1022024)
- [67] STRAUCH, C. (2015). Sharp adaptive drift estimation for ergodic diffusions: The multivariate case. *Stochastic Process. Appl.* **125** 2562–2602. [MR3332848 https://doi.org/10.1016/j.spa.2015.02.003](https://doi.org/10.1016/j.spa.2015.02.003)
- [68] STRAUCH, C. (2016). Exact adaptive pointwise drift estimation for multidimensional ergodic diffusions. *Probab. Theory Related Fields* **164** 361–400. [MR3449393 https://doi.org/10.1007/s00440-014-0614-4](https://doi.org/10.1007/s00440-014-0614-4)
- [69] STRAUCH, C. (2018). Adaptive invariant density estimation for ergodic diffusions over anisotropic classes. *Ann. Statist.* **46** 3451–3480. [MR3852658 https://doi.org/10.1214/17-AOS1664](https://doi.org/10.1214/17-AOS1664)
- [70] TALAGRAND, M. (2014). *Upper and Lower Bounds for Stochastic Processes: Modern Methods and Classical Problems*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]* **60**. Springer, Heidelberg. [MR3184689 https://doi.org/10.1007/978-3-642-54075-2](https://doi.org/10.1007/978-3-642-54075-2)
- [71] VAN WAAIJ, J. (2019). Adaptive posterior contraction rates for empirical Bayesian drift estimation of a diffusion. Preprint. Available at [arXiv:1909.12710](https://arxiv.org/abs/1909.12710).
- [72] VAN WAAIJ, J. and VAN ZANTEN, H. (2016). Gaussian process methods for one-dimensional diffusions: Optimal rates and adaptation. *Electron. J. Stat.* **10** 628–645. [MR3471991 https://doi.org/10.1214/16-EJS1117](https://doi.org/10.1214/16-EJS1117)
- [73] VÄNSKÄ, S., LASSAS, M. and SILTANEN, S. (2009). Statistical X-ray tomography using empirical Besov priors. *Int. J. Tomogr. Stat.* **11** 3–32. [MR2505762](https://doi.org/10.1007/s11222-009-9161-1)
- [74] VAN DER MEULEN, F. and SCHAUER, M. (2017). Bayesian estimation of discretely observed multi-dimensional diffusion processes using guided proposals. *Electron. J. Stat.* **11** 2358–2396. [MR3656495 https://doi.org/10.1214/17-EJS1290](https://doi.org/10.1214/17-EJS1290)
- [75] VAN DER MEULEN, F. H., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2006). Convergence rates of posterior distributions for Brownian semimartingale models. *Bernoulli* **12** 863–888. [MR2265666 https://doi.org/10.3150/bj/1161614950](https://doi.org/10.3150/bj/1161614950)
- [76] VAN DER VAART, A. and VAN ZANTEN, H. (2007). Bayesian inference with rescaled Gaussian process priors. *Electron. J. Stat.* **1** 433–448. [MR2357712 https://doi.org/10.1214/07-EJS098](https://doi.org/10.1214/07-EJS098)
- [77] 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/10.1214/090536070000000613](https://doi.org/10.1214/090536070000000613)

SCALABLE ESTIMATION AND INFERENCE FOR CENSORED QUANTILE REGRESSION PROCESS

BY XUMING HE^{1,a}, XIAOOU PAN^{2,c}, KEAN MING TAN^{1,b} AND WEN-XIN ZHOU^{2,d}

¹Department of Statistics, University of Michigan, ^axmhe@umich.edu, ^bkeanming@umich.edu

²Department of Mathematical Sciences, University of California, San Diego, ^cxip024@ucsd.edu, ^dwez243@ucsd.edu

Censored quantile regression (CQR) has become a valuable tool to study the heterogeneous association between a possibly censored outcome and a set of covariates, yet computation and statistical inference for CQR have remained a challenge for large-scale data with many covariates. In this paper, we focus on a smoothed martingale-based sequential estimating equations approach, to which scalable gradient-based algorithms can be applied. Theoretically, we provide a unified analysis of the smoothed sequential estimator and its penalized counterpart in increasing dimensions. When the covariate dimension grows with the sample size at a sublinear rate, we establish the uniform convergence rate (over a range of quantile indexes) and provide a rigorous justification for the validity of a multiplier bootstrap procedure for inference. In high-dimensional sparse settings, our results considerably improve the existing work on CQR by relaxing an exponential term of sparsity. We also demonstrate the advantage of the smoothed CQR over existing methods with both simulated experiments and data applications.

REFERENCES

- [1] ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer Series in Statistics. Springer, New York. MR1198884 <https://doi.org/10.1007/978-1-4612-4348-9>
- [2] BARBE, P. and BERTAIL, P. (1995). *The Weighted Bootstrap*. Lecture Notes in Statistics **98**. Springer, New York. MR2195545 <https://doi.org/10.1007/978-1-4612-2532-4>
- [3] BARRODALE, I. and ROBERTS, F. (1974). Solution of an overdetermined system of equations in the ℓ_1 norm. *Commun. ACM* **17** 319–320.
- [4] BELLONI, A. and CHERNOZHUKOV, V. (2011). ℓ_1 -penalized quantile regression in high-dimensional sparse models. *Ann. Statist.* **39** 82–130. MR2797841 <https://doi.org/10.1214/10-AOS827>
- [5] BRADIC, J., FAN, J. and JIANG, J. (2011). Regularization for Cox’s proportional hazards model with NP-dimensionality. *Ann. Statist.* **39** 3092–3120. MR3012402 <https://doi.org/10.1214/11-AOS911>
- [6] BUCHINSKY, M. and HAHN, J. (1998). An alternative estimator for the censored quantile regression model. *Econometrica* **66** 653–671. MR1627038 <https://doi.org/10.2307/2998578>
- [7] CAI, T., HUANG, J. and TIAN, L. (2009). Regularized estimation for the accelerated failure time model. *Biometrics* **65** 394–404. MR2751463 <https://doi.org/10.1111/j.1541-0420.2008.01074.x>
- [8] CHATTERJEE, S. and BOSE, A. (2005). Generalized bootstrap for estimating equations. *Ann. Statist.* **33** 414–436. MR2157808 <https://doi.org/10.1214/009053604000000904>
- [9] CHERNOZHUKOV, V. and HONG, H. (2002). Three-step censored quantile regression and extramarital affairs. *J. Amer. Statist. Assoc.* **97** 872–882. MR1941416 <https://doi.org/10.1198/016214502388618663>
- [10] DE BACKER, M., EL GHOUGH, A. and VAN KEILEGOM, I. (2019). An adapted loss function for censored quantile regression. *J. Amer. Statist. Assoc.* **114** 1126–1137. MR4011767 <https://doi.org/10.1080/01621459.2018.1469996>
- [11] DE BACKER, M., EL GHOUGH, A. and VAN KEILEGOM, I. (2020). Linear censored quantile regression: A novel minimum-distance approach. *Scand. J. Stat.* **47** 1275–1306. MR4178194 <https://doi.org/10.1111/sjos.12475>
- [12] DE CASTRO, L., GALVAO, A. F., KAPLAN, D. M. and LIU, X. (2019). Smoothed GMM for quantile models. *J. Econometrics* **213** 121–144. MR4013218 <https://doi.org/10.1016/j.jeconom.2019.04.008>

MSC2020 subject classifications. Primary 62J05, 62J07; secondary 62F40.

Key words and phrases. Censored quantile regression, smoothing, high-dimensional survival data, nonasymptotic theory, weighted bootstrap.

- [13] DICKSON, E. R., GRAMBSCH, P. M., FLEMING, T. R., FISHER, L. D. and LANGWORTHY, A. (1989). Prognosis in primary biliary cirrhosis: Model for decision making. *Hepatology* **10** 1–7. <https://doi.org/10.1002/hep.1840100102>
- [14] EFRON, B. (1967). The two sample problem with censored data. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 4: Biology and Problems of Health* 831–853.
- [15] 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](https://doi.org/10.1198/016214501753382273)
- [16] FAN, J. and LI, R. (2002). Variable selection for Cox’s proportional hazards model and frailty model. *Ann. Statist.* **30** 74–99. [MR1892656 https://doi.org/10.1214/aos/1015362185](https://doi.org/10.1214/aos/1015362185)
- [17] FEI, Z., ZHENG, Q., HONG, H. G. and LI, Y. (2021). Inference for high dimensional censored quantile regression. *J. Amer. Statist. Assoc.* <https://doi.org/10.1080/01621459.2021.1957900>
- [18] FERNANDES, M., GUERRE, E. and HORTA, E. (2021). Smoothing quantile regressions. *J. Bus. Econom. Statist.* **39** 338–357. [MR4187194 https://doi.org/10.1080/07350015.2019.1660177](https://doi.org/10.1080/07350015.2019.1660177)
- [19] FLEMING, T. R. and HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. Wiley, New York. [MR1100924](https://doi.org/10.1002/9781118160607)
- [20] FYGENSON, M. and RITOV, Y. (1994). Monotone estimating equations for censored data. *Ann. Statist.* **22** 732–746. [MR1292538 https://doi.org/10.1214/aos/1176325493](https://doi.org/10.1214/aos/1176325493)
- [21] GILL, R. D. and JOHANSEN, S. (1990). A survey of product-integration with a view toward application in survival analysis. *Ann. Statist.* **18** 1501–1555. [MR1074422 https://doi.org/10.1214/aos/1176347865](https://doi.org/10.1214/aos/1176347865)
- [22] GU, Y., FAN, J., KONG, L., MA, S. and ZOU, H. (2018). ADMM for high-dimensional sparse penalized quantile regression. *Technometrics* **60** 319–331. [MR3847169 https://doi.org/10.1080/00401706.2017.1345703](https://doi.org/10.1080/00401706.2017.1345703)
- [23] HE, X., PAN, X., TAN, K. M. and ZHOU, W.-X. (2021). Smoothed quantile regression with large-scale inference. *J. Econometrics* <https://doi.org/10.1016/j.jeconom.2021.07.010>
- [24] HE, X., PAN, X., TAN, K. M. and ZHOU, W.-X. (2022). Supplement to “Scalable estimation and inference for censored quantile regression process.” <https://doi.org/10.1214/22-AOS2214SUPP>
- [25] HONORÉ, B., KHAN, S. and POWELL, J. L. (2002). Quantile regression under random censoring. *J. Econometrics* **109** 67–105. [MR1899693 https://doi.org/10.1016/S0304-4076\(01\)00142-7](https://doi.org/10.1016/S0304-4076(01)00142-7)
- [26] HU, F. and KALBFLEISCH, J. D. (2000). The estimating function bootstrap. *Canad. J. Statist.* **28** 449–499. [MR1793106 https://doi.org/10.2307/3315958](https://doi.org/10.2307/3315958)
- [27] HUANG, J., MA, S. and XIE, H. (2006). Regularized estimation in the accelerated failure time model with high-dimensional covariates. *Biometrics* **62** 813–820. [MR2247210 https://doi.org/10.1111/j.1541-0420.2006.00562.x](https://doi.org/10.1111/j.1541-0420.2006.00562.x)
- [28] HUANG, Y. (2010). Quantile calculus and censored regression. *Ann. Statist.* **38** 1607–1637. [MR2662354 https://doi.org/10.1214/09-AOS771](https://doi.org/10.1214/09-AOS771)
- [29] JIN, Z., YING, Z. and WEI, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88** 381–390. [MR1844838 https://doi.org/10.1093/biomet/88.2.381](https://doi.org/10.1093/biomet/88.2.381)
- [30] KAPLAN, D. M. and SUN, Y. (2017). Smoothed estimating equations for instrumental variables quantile regression. *Econometric Theory* **33** 105–157. [MR3574862 https://doi.org/10.1017/S0266466615000407](https://doi.org/10.1017/S0266466615000407)
- [31] KLEINBAUM, D. G. and KLEIN, M. (2012). *Survival Analysis: A Self-Learning Text*, 3rd ed. *Statistics for Biology and Health*. Springer, New York. [MR2882858 https://doi.org/10.1007/978-1-4419-6646-9](https://doi.org/10.1007/978-1-4419-6646-9)
- [32] KOENKER, R. (2005). *Quantile Regression*. *Econometric Society Monographs* **38**. Cambridge Univ. Press, Cambridge. [MR2268657 https://doi.org/10.1017/CBO9780511754098](https://doi.org/10.1017/CBO9780511754098)
- [33] KOENKER, R. (2008). Censored quantile regression redux. *J. Stat. Softw.* **38** 1–25.
- [34] KOENKER, R., CHERNOZHUKOV, V., HE, X. and PENG, L. (2017). *Handbook of Quantile Regression*. CRC Press, New York.
- [35] KOENKER, R. and GELING, O. (2001). Reappraising medfly longevity: A quantile regression survival analysis. *J. Amer. Statist. Assoc.* **96** 458–468. [MR1939348 https://doi.org/10.1198/016214501753168172](https://doi.org/10.1198/016214501753168172)
- [36] KOENKER, R. and MIZERA, I. (2014). Convex optimization in R. *J. Stat. Softw.* **60**.
- [37] KOENKER, R. and NG, P. (2005). A Frisch–Newton algorithm for sparse quantile regression. *Acta Math. Appl. Sin. Engl. Ser.* **21** 225–236. [MR2141542 https://doi.org/10.1007/s10255-005-0231-1](https://doi.org/10.1007/s10255-005-0231-1)
- [38] LENG, C. and TONG, X. (2013). A quantile regression estimator for censored data. *Bernoulli* **19** 344–361. [MR3019498 https://doi.org/10.3150/11-BEJ388](https://doi.org/10.3150/11-BEJ388)
- [39] MA, S. and KOSOROK, M. R. (2005). Robust semiparametric M-estimation and the weighted bootstrap. *J. Multivariate Anal.* **96** 190–217. [MR2202406 https://doi.org/10.1016/j.jmva.2004.09.008](https://doi.org/10.1016/j.jmva.2004.09.008)
- [40] NEOCLEOUS, T., VANDEN BRANDEN, K. and PORTNOY, S. (2006). Correction to: “Censored regression quantiles” [J. Amer. Statist. Assoc. **98**(464) (2003), 1001–1012; [MR2041488](https://doi.org/10.1080/016214503000000087)] by Portnoy. *J. Amer. Statist. Assoc.* **101** 860–861. [MR2281250 https://doi.org/10.1198/016214506000000087](https://doi.org/10.1198/016214506000000087)

- [41] PARZEN, M. I., WEI, L. J. and YING, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81** 341–350. MR1294895 <https://doi.org/10.1093/biomet/81.2.341>
- [42] PENG, L. (2012). Self-consistent estimation of censored quantile regression. *J. Multivariate Anal.* **105** 368–379. MR2877523 <https://doi.org/10.1016/j.jmva.2011.10.005>
- [43] PENG, L. (2021). Quantile regression for survival data. *Annu. Rev. Stat. Appl.* **8** 413–437. MR4243554 <https://doi.org/10.1146/annurev-statistics-042720-020233>
- [44] PENG, L. and HUANG, Y. (2008). Survival analysis with quantile regression models. *J. Amer. Statist. Assoc.* **103** 637–649. MR2435468 <https://doi.org/10.1198/016214508000000355>
- [45] PORTNOY, S. (2003). Censored regression quantiles. *J. Amer. Statist. Assoc.* **98** 1001–1012. MR2041488 <https://doi.org/10.1198/016214503000000954>
- [46] PORTNOY, S. and KOENKER, R. (1997). The Gaussian hare and the Laplacian tortoise: Computability of squared-error versus absolute-error estimators. *Statist. Sci.* **12** 279–300. MR1619189 <https://doi.org/10.1214/ss/1030037960>
- [47] PORTNOY, S. and LIN, G. (2010). Asymptotics for censored regression quantiles. *J. Nonparametr. Stat.* **22** 115–130. MR2598957 <https://doi.org/10.1080/10485250903105009>
- [48] POWELL, J. L. (1984). Least absolute deviations estimation for the censored regression model. *J. Econometrics* **25** 303–325. MR0752444 [https://doi.org/10.1016/0304-4076\(84\)90004-6](https://doi.org/10.1016/0304-4076(84)90004-6)
- [49] POWELL, J. L. (1986). Censored regression quantiles. *J. Econometrics* **32** 143–155. MR0853049 [https://doi.org/10.1016/0304-4076\(86\)90016-3](https://doi.org/10.1016/0304-4076(86)90016-3)
- [50] SHEDDEN, K., TAYLOR, J. M., ENKEMANN, S. A. (2008). Gene expression-based survival prediction in lung adenocarcinoma: A multi-site, blinded validation study. *Nat. Med.* **14** 822–827.
- [51] SHERWOOD, B. and MAIDMAN, A. (2020). Package ‘rqPen’, version 2.2.2. Reference manual. <https://cran.r-project.org/web/packages/rqPen/rqPen.pdf>.
- [52] SHOWS, J. H., LU, W. and ZHANG, H. H. (2010). Sparse estimation and inference for censored median regression. *J. Statist. Plann. Inference* **140** 1903–1917. MR2606727 <https://doi.org/10.1016/j.jspi.2010.01.043>
- [53] SUN, X., PENG, L., HUANG, Y. and LAI, H. J. (2016). Generalizing quantile regression for counting processes with applications to recurrent events. *J. Amer. Statist. Assoc.* **111** 145–156. MR3494649 <https://doi.org/10.1080/01621459.2014.995795>
- [54] TAN, K. M., WANG, L. and ZHOU, W.-X. (2022). High-dimensional quantile regression: Convolution smoothing and concave regularization. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 205–233. MR4400395 <https://doi.org/10.1111/rssb.12485>
- [55] THERNEAU, T. M., GRAMBSCH, P. M. and FLEMING, T. R. (1990). Martingale-based residuals for survival models. *Biometrika* **77** 147–160. MR1049416 <https://doi.org/10.1093/biomet/77.1.147>
- [56] TIAN, L., ZUCKER, D. and WEI, L. J. (2005). On the Cox model with time-varying regression coefficients. *J. Amer. Statist. Assoc.* **100** 172–183. MR2156827 <https://doi.org/10.1198/016214504000000845>
- [57] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242
- [58] VOLGUSHEV, S., WAGENER, J. and DETTE, H. (2014). Censored quantile regression processes under dependence and penalization. *Electron. J. Stat.* **8** 2405–2447. MR3278338 <https://doi.org/10.1214/14-EJS54>
- [59] WANG, H. J. and WANG, L. (2009). Locally weighted censored quantile regression. *J. Amer. Statist. Assoc.* **104** 1117–1128. MR2562007 <https://doi.org/10.1198/jasa.2009.tm08230>
- [60] WANG, H. J., ZHOU, J. and LI, Y. (2013). Variable selection for censored quantile regression. *Statist. Sinica* **23** 145–167. MR3076162
- [61] WHANG, Y.-J. (2006). Smoothed empirical likelihood methods for quantile regression models. *Econometric Theory* **22** 173–205. MR2230386 <https://doi.org/10.1017/S0266466606060087>
- [62] WU, Y., MA, Y. and YIN, G. (2015). Smoothed and corrected score approach to censored quantile regression with measurement errors. *J. Amer. Statist. Assoc.* **110** 1670–1683. MR3449063 <https://doi.org/10.1080/01621459.2014.989323>
- [63] YANG, X., NARISSETY, N. N. and HE, X. (2018). A new approach to censored quantile regression estimation. *J. Comput. Graph. Statist.* **27** 417–425. MR3816276 <https://doi.org/10.1080/10618600.2017.1385469>
- [64] YING, Z., JUNG, S. H. and WEI, L. J. (1995). Survival analysis with median regression models. *J. Amer. Statist. Assoc.* **90** 178–184. MR1325125
- [65] YU, L., LIN, N. and WANG, L. (2017). A parallel algorithm for large-scale nonconvex penalized quantile regression. *J. Comput. Graph. Statist.* **26** 935–939. MR3765357 <https://doi.org/10.1080/10618600.2017.1328366>
- [66] ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. MR2604701 <https://doi.org/10.1214/09-AOS729>

- [67] ZHENG, Q., PENG, L. and HE, X. (2018). High dimensional censored quantile regression. *Ann. Statist.* **46** 308–343. [MR3766954](#) <https://doi.org/10.1214/17-AOS1551>
- [68] ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. [MR2279469](#) <https://doi.org/10.1198/016214506000000735>
- [69] ZOU, H. and LI, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.* **36** 1509–1533. [MR2435443](#) <https://doi.org/10.1214/009053607000000802>

A STUDY OF ORTHOGONAL ARRAY-BASED DESIGNS UNDER A BROAD CLASS OF SPACE-FILLING CRITERIA

BY GUANZHOU CHEN^a AND BOXIN TANG^b

Department of Statistics and Actuarial Science, Simon Fraser University, ^agca58@sfu.ca, ^bboxint@sfu.ca

Space-filling designs based on orthogonal arrays are attractive for computer experiments for they can be easily generated with desirable low-dimensional stratification properties. Nonetheless, it is not very clear how they behave and how to construct good such designs under other space-filling criteria. In this paper, we justify orthogonal array-based designs under a broad class of space-filling criteria, which include commonly used distance-, orthogonality- and discrepancy-based measures. To identify designs with even better space-filling properties, we partition orthogonal array-based designs into classes by allowable level permutations and show that the average performance of each class of designs is determined by two types of stratifications, with one of them being achieved by strong orthogonal arrays of strength 2+. Based on these results, we investigate various new and existing constructions of space-filling orthogonal array-based designs, including some strong orthogonal arrays of strength 2+ and mappable nearly orthogonal arrays.

REFERENCES

- AI, M.-Y. and ZHANG, R.-C. (2004). Projection justification of generalized minimum aberration for asymmetrical fractional factorial designs. *Metrika* **60** 279–285. [MR2189756](#) <https://doi.org/10.1007/s001840300310>
- CHEN, G. and TANG, B. (2022). Using nonregular designs to generate space-filling designs. *J. Statist. Plann. Inference* **221** 201–211. [MR4426122](#) <https://doi.org/10.1016/j.jspi.2022.04.007>
- CHENG, C.-S., HE, Y. and TANG, B. (2021). Minimal second order saturated designs and their applications to space-filling designs. *Statist. Sinica* **31** 867–890. [MR4286198](#) <https://doi.org/10.5705/ss.20>
- FANG, K.-T., LI, R. and SUDJANTO, A. (2006). *Design and Modeling for Computer Experiments*. Chapman & Hall/CRC Computer Science and Data Analysis Series. CRC Press/CRC, Boca Raton, FL. [MR2223960](#)
- FANG, K.-T., LIU, M.-Q., QIN, H. and ZHOU, Y.-D. (2018). *Theory and Application of Uniform Experimental Designs*. *Lecture Notes in Statistics* **221**. Science Press Beijing, Beijing; Springer, Singapore. With a foreword by Yuan Wang. [MR3837569](#) <https://doi.org/10.1007/978-981-13-2041-5>
- HE, Y., CHENG, C.-S. and TANG, B. (2018). Strong orthogonal arrays of strength two plus. *Ann. Statist.* **46** 457–468. [MR3782373](#) <https://doi.org/10.1214/17-AOS1555>
- HE, Y. and TANG, B. (2013). Strong orthogonal arrays and associated Latin hypercubes for computer experiments. *Biometrika* **100** 254–260. [MR3034340](#) <https://doi.org/10.1093/biomet/ass065>
- HE, Y. and TANG, B. (2014). A characterization of strong orthogonal arrays of strength three. *Ann. Statist.* **42** 115–128. [MR3226159](#) <https://doi.org/10.1214/14-AOS1225>
- HICKERNELL, F. J. (1998). A generalized discrepancy and quadrature error bound. *Math. Comp.* **67** 299–322. [MR1433265](#) <https://doi.org/10.1090/S0025-5718-98-00894-1>
- JOHNSON, M. E., MOORE, L. M. and YLVISAKER, D. (1990). Minimax and maximin distance designs. *J. Statist. Plann. Inference* **26** 131–148. [MR1079258](#) [https://doi.org/10.1016/0378-3758\(90\)90122-B](https://doi.org/10.1016/0378-3758(90)90122-B)
- LI, W., LIU, M.-Q. and TANG, B. (2021). A method of constructing maximin distance designs. *Biometrika* **108** 845–855. [MR4341355](#) <https://doi.org/10.1093/biomet/asaa089>
- LIN, C. D., BINGHAM, D., SITTER, R. R. and TANG, B. (2010). A new and flexible method for constructing designs for computer experiments. *Ann. Statist.* **38** 1460–1477. [MR2662349](#) <https://doi.org/10.1214/09-AOS757>
- MCKAY, M. D., BECKMAN, R. J. and CONOVER, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21** 239–245. [MR0533252](#) <https://doi.org/10.2307/1268522>

MSC2020 subject classifications. 62K99, 62K15.

Key words and phrases. Computer experiment, digital net, mappable nearly orthogonal array, strong orthogonal array, supersaturated design.

- MUKERJEE, R., SUN, F. and TANG, B. (2014). Nearly orthogonal arrays mappable into fully orthogonal arrays. *Biometrika* **101** 957–963. [MR3286928](#) <https://doi.org/10.1093/biomet/asu042>
- NIEDERREITER, H. (1987). Point sets and sequences with small discrepancy. *Monatsh. Math.* **104** 273–337. [MR0918037](#) <https://doi.org/10.1007/BF01294651>
- OWEN, A. B. (1992). Orthogonal arrays for computer experiments, integration and visualization. *Statist. Sinica* **2** 439–452. [MR1187952](#)
- OWEN, A. B. (1994). Controlling correlations in Latin hypercube samples. *J. Amer. Statist. Assoc.* **89** 1517–1522.
- SANTNER, T. J., WILLIAMS, B. J. and NOTZ, W. I. (2018). *The Design and Analysis of Computer Experiments*. Springer Series in Statistics. Springer, New York. Second edition of [MR2160708]. [MR3887662](#)
- SHI, C. and TANG, B. (2019). Design selection for strong orthogonal arrays. *Canad. J. Statist.* **47** 302–314. [MR3952395](#) <https://doi.org/10.1002/cjs.11494>
- SOBOL', I. M. (1967). Distribution of points in a cube and approximate evaluation of integrals. *Zh. Vychisl. Mat. Mat. Fiz.* **7** 784–802. [MR0219238](#)
- STEINBERG, D. M. and LIN, D. K. J. (2006). A construction method for orthogonal Latin hypercube designs. *Biometrika* **93** 279–288. [MR2278083](#) <https://doi.org/10.1093/biomet/93.2.279>
- SUN, F. and TANG, B. (2017). A general rotation method for orthogonal Latin hypercubes. *Biometrika* **104** 465–472. [MR3698266](#) <https://doi.org/10.1093/biomet/asx022>
- SUN, C.-Y. and TANG, B. (2021). Uniform projection designs and strong orthogonal arrays. *J. Amer. Statist. Assoc.* <https://doi.org/10.1080/01621459.2021.1935268>
- SUN, F., WANG, Y. and XU, H. (2019). Uniform projection designs. *Ann. Statist.* **47** 641–661. [MR3909945](#) <https://doi.org/10.1214/18-AOS1705>
- TANG, B. (1993). Orthogonal array-based Latin hypercubes. *J. Amer. Statist. Assoc.* **88** 1392–1397. [MR1245375](#)
- TANG, B. (2001). Theory of J -characteristics for fractional factorial designs and projection justification of minimum G_2 -aberration. *Biometrika* **88** 401–407. [MR1844840](#) <https://doi.org/10.1093/biomet/88.2.401>
- TANG, B. and DENG, L.-Y. (1999). Minimum G_2 -aberration for nonregular fractional factorial designs. *Ann. Statist.* **27** 1914–1926. [MR1765622](#) <https://doi.org/10.1214/aos/1017939244>
- TIAN, Y. and XU, H. (2022). A minimum aberration-type criterion for selecting space-filling designs. *Biometrika* **109** 489–501. [MR4430970](#) <https://doi.org/10.1093/biomet/asab021>
- WANG, Y., SUN, F. and XU, H. (2022). On design orthogonality, maximin distance, and projection uniformity for computer experiments. *J. Amer. Statist. Assoc.* **117** 375–385. [MR4399092](#) <https://doi.org/10.1080/01621459.2020.1782221>
- WANG, L., XIAO, Q. and XU, H. (2018). Optimal maximin L_1 -distance Latin hypercube designs based on good lattice point designs. *Ann. Statist.* **46** 3741–3766. [MR3852667](#) <https://doi.org/10.1214/17-AOS1674>
- WANG, C., YANG, J. and LIU, M.-Q. (2022). Construction of strong group-orthogonal arrays. *Statist. Sinica* **32** 1225–1243. [MR4449892](#) <https://doi.org/10.5705/ss.202020.0110>
- WANG, Y., WANG, F., YUAN, Y. and XIAO, Q. (2021). Connecting U-type designs before and after level permutations and expansions. *J. Stat. Theory Pract.* **15** Paper No. 81, 19. [MR4324770](#) <https://doi.org/10.1007/s42519-021-00209-9>
- XIAO, Q. and XU, H. (2018). Construction of maximin distance designs via level permutation and expansion. *Statist. Sinica* **28** 1395–1414. [MR3821010](#)
- XU, H. (2003). Minimum moment aberration for nonregular designs and supersaturated designs. *Statist. Sinica* **13** 691–708. [MR1997169](#)
- XU, H. and WU, C. F. J. (2001). Generalized minimum aberration for asymmetrical fractional factorial designs [corrected republication of MR1863969]. *Ann. Statist.* **29** 1066–1077. [MR1869240](#) <https://doi.org/10.1214/aos/1013699993>
- XU, H. and WU, C. F. J. (2005). Construction of optimal multi-level supersaturated designs. *Ann. Statist.* **33** 2811–2836. [MR2253103](#) <https://doi.org/10.1214/009053605000000688>
- YE, K. Q. (1998). Orthogonal column Latin hypercubes and their application in computer experiments. *J. Amer. Statist. Assoc.* **93** 1430–1439. [MR1666638](#) <https://doi.org/10.2307/2670057>
- ZHOU, Y.-D., FANG, K.-T. and NING, J.-H. (2013). Mixture discrepancy for quasi-random point sets. *J. Complexity* **29** 283–301. [MR3062468](#) <https://doi.org/10.1016/j.jco.2012.11.006>
- ZHOU, Y.-D. and XU, H. (2014). Space-filling fractional factorial designs. *J. Amer. Statist. Assoc.* **109** 1134–1144. [MR3265686](#) <https://doi.org/10.1080/01621459.2013.873367>
- ZHOU, Y. and XU, H. (2015). Space-filling properties of good lattice point sets. *Biometrika* **102** 959–966. [MR3431565](#) <https://doi.org/10.1093/biomet/asv044>

TWO-SAMPLE TESTING OF HIGH-DIMENSIONAL LINEAR REGRESSION COEFFICIENTS VIA COMPLEMENTARY SKETCHING

BY FENGAN GAO^{1,a} AND TENG YAO WANG^{2,b}

¹*School of Data Science, Shanghai Center for Mathematical Sciences, Fudan University, fngao@fudan.edu.cn*

²*Department of Statistics, London School of Economics, t.wang59@lse.ac.uk*

We introduce a new method for two-sample testing of high-dimensional linear regression coefficients without assuming that those coefficients are individually estimable. The procedure works by first projecting the matrices of covariates and response vectors along directions that are complementary in sign in a subset of the coordinates, a process which we call “complementary sketching.” The resulting projected covariates and responses are aggregated to form two test statistics, which are shown to have essentially optimal asymptotic power under a Gaussian design when the difference between the two regression coefficients is sparse and dense respectively. Simulations confirm that our methods perform well in a broad class of settings and an application to a large single-cell RNA sequencing dataset demonstrates its utility in the real world.

REFERENCES

- 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>
- BAI, Z., HU, J., PAN, G. and ZHOU, W. (2015). Convergence of the empirical spectral distribution function of Beta matrices. *Bernoulli* **21** 1538–1574. [MR3352053](#) <https://doi.org/10.3150/14-BEJ613>
- BAI, F., ZHANG, P., FU, Y., CHEN, H., ZHANG, M., HUANG, Q., LI, D., LI, B. and WU, K. (2020). Targeting ANXA1 abrogates Treg-mediated immune suppression in triple-negative breast cancer. *J. Immunother. Cancer* **8**.
- BAYER, A. L., YU, A. and MALEK, T. R. (2007). Function of the IL-2R for thymic and peripheral CD4+ CD25+ Foxp3+ T regulatory cells. *J. Immunol.* **178** 4062–4071.
- BIRGÉ, L. (2001). An alternative point of view on Lepski’s method. In *State of the Art in Probability and Statistics (Leiden, 1999)*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **36** 113–133. IMS, Beachwood, OH. [MR1836557](#) <https://doi.org/10.1214/lnms/1215090065>
- 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>
- CARPENTIER, A. and VERZELEN, N. (2021). Optimal sparsity testing in linear regression model. *Bernoulli* **27** 727–750. [MR4255213](#) <https://doi.org/10.3150/20-bej1224>
- CARPENTIER, A., COLLIER, O., COMMINGES, L., TSYBAKOV, A. B. and WANG, Y. (2019). Minimax rate of testing in sparse linear regression. *Autom. Remote Control* **80** 1817–1834.
- CHARBONNIER, C., VERZELEN, N. and VILLERS, F. (2015). A global homogeneity test for high-dimensional linear regression. *Electron. J. Stat.* **9** 318–382. [MR3323203](#) <https://doi.org/10.1214/15-EJS999>
- CHEN, S. X., LI, J. and ZHONG, P.-S. (2019). Two-sample and ANOVA tests for high dimensional means. *Ann. Statist.* **47** 1443–1474. [MR3911118](#) <https://doi.org/10.1214/18-AOS1720>
- CHOW, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica* **28** 591–605. [MR0141193](#) <https://doi.org/10.2307/1910133>
- DICKER, L. H. (2014). Variance estimation in high-dimensional linear models. *Biometrika* **101** 269–284. [MR3215347](#) <https://doi.org/10.1093/biomet/ast065>
- DOEBBELER, M., KOENIG, C., KRZYŻAK, L., SEITZ, C., WILD, A., ULAS, T., BASSLER, K., KOPELYANSKIY, D., BUTTERHOF, A. et al. (2018). CD83 expression is essential for Treg cell differentiation and stability. *JCI Insight* **3**.

MSC2020 subject classifications. 62H15, 62J05.

Key words and phrases. Two-sample hypotheses testing, high-dimensional data, linear model, sparsity, mini-max detection.

- 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>
- EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499. [MR2060166](#) <https://doi.org/10.1214/009053604000000067>
- FAN, J., GUO, S. and HAO, N. (2012). Variance estimation using refitted cross-validation in ultrahigh dimensional regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 37–65. [MR2885839](#) <https://doi.org/10.1111/j.1467-9868.2011.01005.x>
- GAO, F. and WANG, T. (2022). Supplement to “Two-sample testing of high-dimensional linear regression coefficients via complementary sketching.” <https://doi.org/10.1214/22-AOS2216SUPP>
- HOMRIGHAUSEN, D. and McDONALD, D. (2013). The lasso, persistence, and cross-validation. In *International Conference on Machine Learning* 1031–1039, PMLR.
- INGSTER, Y. I. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods Statist.* **6** 47–69. [MR1456646](#)
- 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>
- KANNEL, W. B. and MCGEE, D. L. (1979). Diabetes and cardiovascular disease: The framingham study. *J. Am. Med. Assoc.* **241** 2035–2038.
- KIM, H.-J., BARNITZ, R. A., KRESLAVSKY, T., BROWN, F. D., MOFFETT, H., LEMIEUX, M. E., KAYGUSUZ, Y., MEISSNER, T., HOLDERRIED, T. A. et al. (2015). Stable inhibitory activity of regulatory T cells requires the transcription factor Helios. *Science* **350** 334–339.
- KRAFT, P. and HUNTER, D. J. (2009). Genetic risk prediction—are we there yet? *N. Engl. J. Med.* **360** 1701–1703.
- MAHONEY, M. W. (2011). Randomized algorithms for matrices and data. *Found. Trends Mach. Learn.* **3** 123–224.
- REID, S., TIBSHIRANI, R. and FRIEDMAN, J. (2016). A study of error variance estimation in Lasso regression. *Statist. Sinica* **26** 35–67. [MR3468344](#)
- STÄDLER, N. and MUKHERJEE, S. (2012). Two-sample testing in high-dimensional models. Preprint. Available at [arXiv:1210.4584](https://arxiv.org/abs/1210.4584).
- SUN, T. and ZHANG, C.-H. (2012). Scaled sparse linear regression. *Biometrika* **99** 879–898. [MR2999166](#) <https://doi.org/10.1093/biomet/ass043>
- SUO, C., DANN, E., GOH, I., JARDINE, L., KLESHCHEVNIKOV, V., PARK, J.-E., BOTTING, R. A., STEPHENSON, E., ENGELBERT, J. et al. (2022). Mapping the developing human immune system across organs. *Science* **376** eabo0510. <https://doi.org/10.1126/science.abo0510>
- TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. [MR1379242](#)
- TOOMER, K. H., LUI, J. B., ALTMAN, N. H., BAN, Y., CHEN, X. and MALEK, T. R. (2019). Essential and non-overlapping IL-2R α -dependent processes for thymic development and peripheral homeostasis of regulatory T cells. *Nat. Commun.* **10** 1–16.
- WALKER, L. S. K. (2013). Treg and CTLA-4: Two intertwining pathways to immune tolerance. *J. Autoimmun.* **45** 49–57. <https://doi.org/10.1016/j.jaut.2013.06.006>
- XIA, Y., CAI, T. and CAI, T. T. (2015). Testing differential networks with applications to the detection of gene-gene interactions. *Biometrika* **102** 247–266. [MR3371002](#) <https://doi.org/10.1093/biomet/asu074>
- XIA, Y., CAI, T. and CAI, T. T. (2018). Two-sample tests for high-dimensional linear regression with an application to detecting interactions. *Statist. Sinica* **28** 63–92. [MR3752252](#)
- XIA, Y., CAI, T. T. and SUN, W. (2020). GAP: A general framework for information pooling in two-sample sparse inference. *J. Amer. Statist. Assoc.* **115** 1236–1250. [MR4143462](#) <https://doi.org/10.1080/01621459.2019.1611585>
- ZHU, Y. and BRADIC, J. (2016). Two-sample testing in non-sparse high-dimensional linear models. Preprint. Available at [arXiv:1610.04580](https://arxiv.org/abs/1610.04580).

NONPARAMETRIC REGRESSION ON LIE GROUPS WITH MEASUREMENT ERRORS

BY JEONG MIN JEON^{1,a}, BYEONG U. PARK^{2,c} AND INGRID VAN KEILEGOM^{1,b}

¹*ORSTAT, KU Leuven, jeongmin.jeon@kuleuven.be, ingrid.vankeilegom@kuleuven.be*

²*Department of Statistics, Seoul National University, bupark@stats.snu.ac.kr*

This paper develops a foundation of methodology and theory for non-parametric regression with Lie group-valued predictors contaminated by measurement errors. Our methodology and theory are based on harmonic analysis on Lie groups, which is largely unknown in statistics. We establish a novel deconvolution regression estimator, and study its rate of convergence and asymptotic distribution. We also provide asymptotic confidence intervals based on the asymptotic distribution of the estimator and on the empirical likelihood technique. Several theoretical properties are also studied for a deconvolution density estimator, which is necessary to construct our regression estimator. The case of unknown measurement error distribution is also covered. We present practical details on implementation as well as the results of simulation studies for several Lie groups. A real data example is also provided.

REFERENCES

- [1] APPLEBAUM, D. (2014). *Probability on Compact Lie Groups. Probability Theory and Stochastic Modelling* **70**. Springer, Cham. MR3243650 <https://doi.org/10.1007/978-3-319-07842-7>
- [2] ASWANI, A., BICKEL, P. and TOMLIN, C. (2011). Regression on manifolds: Estimation of the exterior derivative. *Ann. Statist.* **39** 48–81. MR2797840 <https://doi.org/10.1214/10-AOS823>
- [3] BERESTOVSKIĬ, V. N. and SVIRKIN, V. M. (2010). The Laplace operator on normal homogeneous Riemannian manifolds. *Siberian Adv. Math.* **20** 231–255.
- [4] BHATTACHARYA, A. and DUNSON, D. B. (2010). Nonparametric Bayesian density estimation on manifolds with applications to planar shapes. *Biometrika* **97** 851–865. MR2746156 <https://doi.org/10.1093/biomet/asq044>
- [5] CHAKRABORTY, A. and PANARETOS, V. M. (2017). Regression with genuinely functional errors-in-covariates. arXiv preprint, [arXiv:1712.04290](https://arxiv.org/abs/1712.04290).
- [6] CHANG, T. (1989). Spherical regression with errors in variables. *Ann. Statist.* **17** 293–306. MR0981451 <https://doi.org/10.1214/aos/1176347017>
- [7] CHEN, C., GUO, S. and QIAO, X. (2022). Functional linear regression: Dependence and error contamination. *J. Bus. Econom. Statist.* **40** 444–457. MR4356585 <https://doi.org/10.1080/07350015.2020.1832503>
- [8] CHEN, M., JIANG, H., LIAO, W. and ZHAO, T. (2022). Nonparametric regression on low-dimensional manifolds using deep ReLU networks: Function approximation and statistical recovery. arXiv preprint. Available at [arXiv:1908.01842v5](https://arxiv.org/abs/1908.01842v5).
- [9] CHEN, S. X. and VAN KEILEGOM, I. (2009). A review on empirical likelihood methods for regression. *TEST* **18** 415–447. MR2566404 <https://doi.org/10.1007/s11749-009-0159-5>
- [10] CHENG, M.-Y. and WU, H.-T. (2013). Local linear regression on manifolds and its geometric interpretation. *J. Amer. Statist. Assoc.* **108** 1421–1434. MR3174718 <https://doi.org/10.1080/01621459.2013.827984>
- [11] COMTE, F., SAMSON, A. and STIRNEMANN, J. J. (2014). Deconvolution estimation of onset of pregnancy with replicate observations. *Scand. J. Stat.* **41** 325–345. MR3207174 <https://doi.org/10.1111/sjos.12029>

MSC2020 subject classifications. Primary 62G08; secondary 62G20.

Key words and phrases. Deconvolution, errors-in-variables, Lie groups, manifold-valued data, measurement errors.

- [12] CORNEA, E., ZHU, H., KIM, P. and IBRAHIM, J. G. (2017). Regression models on Riemannian symmetric spaces. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 463–482. MR3611755 <https://doi.org/10.1111/rssb.12169>
- [13] DATTFNER, 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>
- [14] DELAIGLE, A. (2014). Nonparametric kernel methods with errors-in-variables: Constructing estimators, computing them, and avoiding common mistakes. *Aust. N. Z. J. Stat.* **56** 105–124. MR3226432 <https://doi.org/10.1111/anzs.12066>
- [15] DELAIGLE, A., FAN, J. and CARROLL, R. J. (2009). A design-adaptive local polynomial estimator for the errors-in-variables problem. *J. Amer. Statist. Assoc.* **104** 348–359. MR2504382 <https://doi.org/10.1198/jasa.2009.0114>
- [16] DELAIGLE, A., HALL, P. and JAMSHIDI, F. (2015). Confidence bands in non-parametric errors-in-variables regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 149–169. MR3299403 <https://doi.org/10.1111/rssb.12067>
- [17] DELAIGLE, A., HALL, P. and MEISTER, A. (2008). On deconvolution with repeated measurements. *Ann. Statist.* **36** 665–685. MR2396811 <https://doi.org/10.1214/009053607000000884>
- [18] DELAIGLE, A. and VAN KEILEGOM, I. (2021). Deconvolution with unknown error distribution. In *Handbook on Measurement Error Models* (G. Yi, A. Delaigle and P. Gustafson, eds.) CRC Press/CRC, Boca Raton.
- [19] DI MARZIO, M., PANZERA, A. and TAYLOR, C. C. (2009). Local polynomial regression for circular predictors. *Statist. Probab. Lett.* **79** 2066–2075. MR2571770 <https://doi.org/10.1016/j.spl.2009.06.014>
- [20] 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
- [21] DUNCAN, B. N., LAMSAL, L. N., THOMPSON, A. M., YOSHIDA, Y., LU, Z., STREETS, D. G., HURWITZ, M. M. and PICKERING, K. E. (2016). A space-based, high-resolution view of notable changes in urban NO_x pollution around the world (2005–2014). *J. Geophys. Res., Atmos.* **121** 976–996.
- [22] FAN, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272. MR1126324 <https://doi.org/10.1214/aos/1176348248>
- [23] FAN, J. (1991). Asymptotic normality for deconvolution kernel density estimators. *Sankhyā Ser. A* **53** 97–110. MR1177770
- [24] FAN, J. and TRUONG, Y. K. (1993). Nonparametric regression with errors in variables. *Ann. Statist.* **21** 1900–1925. MR1245773 <https://doi.org/10.1214/aos/1176349402>
- [25] FARAUT, J. (2008). *Analysis on Lie Groups. Cambridge Studies in Advanced Mathematics* **110**. Cambridge Univ. Press, Cambridge. MR2426516 <https://doi.org/10.1017/CBO9780511755170>
- [26] FOLLAND, G. B. (2016). *A Course in Abstract Harmonic Analysis*. CRC Press/CRC, Boca Raton, FL.
- [27] GAO, F., HUANG, X.-Y., JACOBS, N. A. and WANG, H. (2015). Assimilation of wind speed and direction observations: Results from real observation experiments. *Tellus, Ser. A Dyn. Meteorol. Oceanogr.* **67** 27132.
- [28] GARCÍA-PORTUGUÉS, E., CRUJEIRAS, R. M. and GONZÁLEZ-MANTEIGA, W. (2013). Exploring wind direction and SO₂ concentration by circular-linear density estimation. *Stoch. Environ. Res. Risk Assess.* **27** 1055–1067.
- [29] GARCÍA-PORTUGUÉS, E., CRUJEIRAS, R. M. and GONZÁLEZ-MANTEIGA, W. (2013). Kernel density estimation for directional-linear data. *J. Multivariate Anal.* **121** 152–175. MR3090475 <https://doi.org/10.1016/j.jmva.2013.06.009>
- [30] GONZALEZ-MANTEIGA, W., HENRY, G. and RODRIGUEZ, D. (2012). Partly linear models on Riemannian manifolds. *J. Appl. Stat.* **39** 1797–1809. MR2935559 <https://doi.org/10.1080/02664763.2012.683169>
- [31] GRAFAKOS, L. (2008). *Classical Fourier Analysis*, 2nd ed. *Graduate Texts in Mathematics* **249**. Springer, New York. MR2445437
- [32] HALL, P., WATSON, G. S. and CABRERA, J. (1987). Kernel density estimation with spherical data. *Biometrika* **74** 751–762. MR0919843 <https://doi.org/10.1093/biomet/74.4.751>
- [33] HEALY, D. M. JR., HENDRIKS, H. and KIM, P. T. (1998). Spherical deconvolution. *J. Multivariate Anal.* **67** 1–22. MR1659108 <https://doi.org/10.1006/jmva.1998.1757>
- [34] HENDRIKS, H. (1990). Nonparametric estimation of a probability density on a Riemannian manifold using Fourier expansions. *Ann. Statist.* **18** 832–849. MR1056339 <https://doi.org/10.1214/aos/1176347628>
- [35] HENRY, G. and RODRIGUEZ, D. (2009). Robust nonparametric regression on Riemannian manifolds. *J. Nonparametr. Stat.* **21** 611–628. MR2543576 <https://doi.org/10.1080/10485250902846439>
- [36] HEYER, H. (1977). *Probability Measures on Locally Compact Groups. Ergebnisse der Mathematik und Ihrer Grenzgebiete, Band 94*. Springer, Berlin. MR0501241
- [37] HJORT, N. L., MCKEAGUE, I. W. and VAN KEILEGOM, I. (2009). Extending the scope of empirical likelihood. *Ann. Statist.* **37** 1079–1111. MR2509068 <https://doi.org/10.1214/07-AOS555>

- [38] HUCKEMANN, S. F., KIM, P. T., KOO, J.-Y. and MUNK, A. (2010). Möbius deconvolution on the hyperbolic plane with application to impedance density estimation. *Ann. Statist.* **38** 2465–2498. MR2676895 <https://doi.org/10.1214/09-AOS783>
- [39] JADHAV, S. and MA, S. (2020). Functional measurement error in functional regression. *Canad. J. Statist.* **48** 238–258. MR4095262 <https://doi.org/10.1002/cjs>
- [40] 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>
- [41] JEON, J. M., PARK, B. U. and VAN KEILEGOM, I. (2022). Supplement to “Nonparametric regression on Lie groups with measurement errors.” <https://doi.org/10.1214/22-AOS2218SUPP>
- [42] JIAO, Y., SHEN, G., LIN, Y. and HUANG, J. (2022). Deep nonparametric regression on approximately low-dimensional manifolds. arXiv preprint. Available at [arXiv:2104.06708v4](https://arxiv.org/abs/2104.06708v4).
- [43] JOHANNES, J. (2009). Deconvolution with unknown error distribution. *Ann. Statist.* **37** 2301–2323. MR2543693 <https://doi.org/10.1214/08-AOS652>
- [44] KAPPUS, J. and MABON, G. (2014). Adaptive density estimation in deconvolution problems with unknown error distribution. *Electron. J. Stat.* **8** 2879–2904. MR3299125 <https://doi.org/10.1214/14-EJS976>
- [45] KATZNELSON, Y. (2004). *An Introduction to Harmonic Analysis*, 3rd ed. Cambridge Mathematical Library. Cambridge Univ. Press, Cambridge. MR2039503 <https://doi.org/10.1017/CBO9781139165372>
- [46] KERKYCHARIAN, G., PHAM NGOC, T. M. and PICARD, D. (2011). Localized spherical deconvolution. *Ann. Statist.* **39** 1042–1068. MR2816347 <https://doi.org/10.1214/10-AOS858>
- [47] KIM, P. T. (1998). Deconvolution density estimation on $SO(N)$. *Ann. Statist.* **26** 1083–1102. MR1635446 <https://doi.org/10.1214/aos/1024691089>
- [48] KIM, P. T. and KOO, J.-Y. (2002). Optimal spherical deconvolution. *J. Multivariate Anal.* **80** 21–42. MR1889831 <https://doi.org/10.1006/jmva.2000.1968>
- [49] KIM, P. T. and KOO, J.-Y. (2005). Statistical inverse problems on manifolds. *J. Fourier Anal. Appl.* **11** 639–653. MR2190676 <https://doi.org/10.1007/s00041-005-3041-1>
- [50] KIM, P. T. and RICHARDS, D. S. P. (2001). Deconvolution density estimation on compact Lie groups. In *Algebraic Methods in Statistics and Probability (Notre Dame, IN, 2000)*. *Contemp. Math.* **287** 155–171. Amer. Math. Soc., Providence, RI. MR1873674 <https://doi.org/10.1090/conm/287/04784>
- [51] KRAMER, L. (2011). The topology of a semisimple Lie group is essentially unique. *Adv. Math.* **228** 2623–2633. MR2838051 <https://doi.org/10.1016/j.aim.2011.07.019>
- [52] LANG, S. (1999). *Fundamentals of Differential Geometry*. *Graduate Texts in Mathematics* **191**. Springer, New York. MR1666820 <https://doi.org/10.1007/978-1-4612-0541-8>
- [53] LIN, Z. and YAO, F. (2019). Intrinsic Riemannian functional data analysis. *Ann. Statist.* **47** 3533–3577. MR4025751 <https://doi.org/10.1214/18-AOS1787>
- [54] LIN, Z. and YAO, F. (2021). Functional regression on the manifold with contamination. *Biometrika* **108** 167–181. MR4226196 <https://doi.org/10.1093/biomet/asaa041>
- [55] LUO, Z. M., KIM, P. T., KIM, T. Y. and KOO, J. Y. (2011). Deconvolution on the Euclidean motion group $\mathbb{SE}(3)$. *Inverse Probl.* **27** 035014. MR2772533 <https://doi.org/10.1088/0266-5611/27/3/035014>
- [56] MARRON, J. S. and ALONSO, A. M. (2014). Overview of object oriented data analysis. *Biom. J.* **56** 732–753. MR3258083 <https://doi.org/10.1002/bimj.201300072>
- [57] 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>
- [58] MINAKSHISUNDARAM, S. and PLEIJEL, Å. (1949). Some properties of the eigenfunctions of the Laplace-operator on Riemannian manifolds. *Canad. J. Math.* **1** 242–256. MR0031145 <https://doi.org/10.4153/cjm-1949-021-5>
- [59] MOORE, C. C. (1972). Groups with finite dimensional irreducible representations. *Trans. Amer. Math. Soc.* **166** 401–410. MR0302817 <https://doi.org/10.2307/1996058>
- [60] MYERS, D. F. (2016). *Pointwise and Uniform Convergence of Fourier Series on $SU(2)$* . ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—Missouri University of Science and Technology. MR3597695
- [61] NEUMANN, M. H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametr. Stat.* **7** 307–330. MR1460203 <https://doi.org/10.1080/10485259708832708>
- [62] NEUMANN, M. H. (2007). Deconvolution from panel data with unknown error distribution. *J. Multivariate Anal.* **98** 1955–1968. MR2396948 <https://doi.org/10.1016/j.jmva.2006.09.012>
- [63] OWEN, A. (2001). *Empirical Likelihood*. CRC Press/CRC, London.
- [64] PELLETIER, B. (2005). Kernel density estimation on Riemannian manifolds. *Statist. Probab. Lett.* **73** 297–304. MR2179289 <https://doi.org/10.1016/j.spl.2005.04.004>
- [65] PELLETIER, B. (2006). Non-parametric regression estimation on closed Riemannian manifolds. *J. Nonparametr. Stat.* **18** 57–67. MR2214065 <https://doi.org/10.1080/10485250500504828>
- [66] PROCESI, C. (2007). *Lie Groups: An Approach Through Invariants and Representations*. Universitext. Springer, New York. MR2265844

- [67] SAKURAI, J. J. and NAPOLITANO, J. (2017). *Modern Quantum Mechanics*. Cambridge Univ. Press, Cambridge.
- [68] SEI, T., SHIBATA, H., TAKEMURA, A., OHARA, K. and TAKAYAMA, N. (2013). Properties and applications of Fisher distribution on the rotation group. *J. Multivariate Anal.* **116** 440–455. [MR3049915](#) <https://doi.org/10.1016/j.jmva.2013.01.010>
- [69] SEPANSKI, M. R. (2007). *Compact Lie Groups. Graduate Texts in Mathematics* **235**. Springer, New York. [MR2279709](#) <https://doi.org/10.1007/978-0-387-49158-5>
- [70] SHAO, L., LIN, Z. and YAO, F. (2022). Intrinsic Riemannian functional data analysis for sparse longitudinal observations. *Ann. Statist.* **50** 1696–1721. [MR4441137](#) <https://doi.org/10.1214/22-aos2172>
- [71] SONG, W. (2011). Empirical likelihood confidence intervals for density function in errors-in-variables model. *J. Statist. Res.* **45** 95–110. [MR2934364](#)
- [72] STEFANSKI, L. and CARROLL, R. J. (1990). Deconvoluting kernel density estimators. *Statistics* **21** 169–184. [MR1054861](#) <https://doi.org/10.1080/02331889008802238>
- [73] TERRAS, A. (2016). *Harmonic Analysis on Symmetric Spaces—Higher Rank Spaces, Positive Definite Matrix Space and Generalizations*, 2nd ed. Springer, New York. [MR3496932](#) <https://doi.org/10.1007/978-1-4939-3408-9>
- [74] VAN DEN BAN, E., FLENSTED-JENSEN, M. and SCHLICHTKRULL, H. (1994). Basic harmonic analysis on pseudo-Riemannian symmetric spaces. In *Noncompact Lie Groups and Some of Their Applications* (E. A. Tanner and R. Wilson, eds.) Springer, Dordrecht.
- [75] VOLLRATH, A. (2010). The nonequispaced fast SO(3) Fourier transform, generalisations and applications Ph.D. thesis, Universität Lübeck.
- [76] WALKER, P. L. (1969). Lipschitz classes on finite dimensional groups. *Proc. Camb. Philos. Soc.* **66** 31–38. [MR0240565](#) <https://doi.org/10.1017/s0305004100044686>
- [77] WALTERS, P. (1982). *An Introduction to Ergodic Theory. Graduate Texts in Mathematics* **79**. Springer, New York-Berlin. [MR0648108](#)
- [78] WANG, J.-L., CHIOU, J.-M. and MÜLLER, H.-G. (2016). Functional data analysis. *Annu. Rev. Stat. Appl.* **3** 257–295.
- [79] YAN, L. and CHEN, X. (2014). Empirical likelihood for partly linear models with errors in all variables. *J. Multivariate Anal.* **130** 275–288. [MR3229538](#) <https://doi.org/10.1016/j.jmva.2014.06.007>
- [80] YUAN, Y., ZHU, H., LIN, W. and MARRON, J. S. (2012). Local polynomial regression for symmetric positive definite matrices. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 697–719. [MR2965956](#) <https://doi.org/10.1111/j.1467-9868.2011.01022.x>

LOCALLY ASSOCIATED GRAPHICAL MODELS AND MIXED CONVEX EXPONENTIAL FAMILIES

BY STEFFEN LAURITZEN^{1,a} AND PIOTR ZWIERNIK^{2,b}

¹*Department of Mathematical Sciences, University of Copenhagen, a.lauritzen@math.ku.dk*

²*Department of Statistical Sciences, University of Toronto, piotr.zwiernik@utoronto.ca*

The notion of multivariate total positivity has proved to be useful in finance and psychology but may be too restrictive in other applications. In this paper, we propose a concept of local association, where highly connected components in a graphical model are positively associated and study its properties. Our main motivation comes from gene expression data, where graphical models have become a popular exploratory tool. The models are instances of what we term *mixed convex exponential families* and we show that a *mixed dual likelihood estimator* has simple exact properties for such families as well as asymptotic properties similar to the maximum likelihood estimator. We further relax the positivity assumption by penalizing negative partial correlations in what we term the *positive graphical lasso*. Finally, we develop a GOLAZO algorithm based on block-coordinate descent that applies to a number of optimization procedures that arise in the context of graphical models, including the estimation problems described above. We derive results on existence of the optimum for such problems.

REFERENCES

- [1] AGRAWAL, R., ROY, U. and UHLER, C. (2019). Covariance matrix estimation under total positivity for portfolio selection. arXiv preprint. Available at [arXiv:1909.04222](https://arxiv.org/abs/1909.04222).
- [2] AGRETI, A. (1983). Testing marginal homogeneity for ordinal categorical variables. *Biometrics* **39** 505–510.
- [3] AGRETI, A. (2002). *Categorical Data Analysis*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley-Interscience, New York. [MR1914507 https://doi.org/10.1002/0471249688](https://doi.org/10.1002/0471249688)
- [4] BANERJEE, O., EL GHAOU, L. and D’ASPREMONT, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *J. Mach. Learn. Res.* **9** 485–516. [MR2417243](https://doi.org/10.1162/10997460853437243)
- [5] BARNDORFF-NIELSEN, O. E. (1978). *Information and Exponential Families in Statistical Theory*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, Chichester. [MR0489333](https://doi.org/10.1002/9780470316881)
- [6] BOEGE, T., KAHLE, T., KRETSCHMER, A. and RÖTTGER, F. (2021). The geometry of Gaussian double Markovian distributions. arXiv preprint. Available at [arXiv:2107.00134](https://arxiv.org/abs/2107.00134).
- [7] BREUER, K., FOROUSHANI, A. K., LAIRD, M. R., CHEN, C., SRIBNAIA, A., LO, R., WINSOR, G. L., HANCOCK, R. E. W., BRINKMAN, F. S. L. and LYNN, D. J. (2013). InnateDB: Systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Res.* **41** D1228–D1233.
- [8] BROWN, L. D. (1986). *Fundamentals of Statistical Exponential Families*. *IMS-Monographs IX*. IMS, Hayward, CA.
- [9] CHEN, J. and CHEN, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* **95** 759–771. [MR2443189 https://doi.org/10.1093/biomet/asn034](https://doi.org/10.1093/biomet/asn034)
- [10] CHRISTENSEN, E. S. (1989). Statistical properties of *I*-projections within exponential families. *Scand. J. Stat.* **16** 307–318. [MR1039285](https://doi.org/10.1080/00373718908839285)
- [11] COHEN, J., VAN MARGER, L. J., SUN, Y., ALLRED, E., LEVITON, A. and KOHANE, I. S. (2007). Perturbation of gene expression of the chromatin remodeling pathway in premature newborns at risk for bronchopulmonary dysplasia. *Genome Biol.* **8** R210. <https://doi.org/10.1186/gb-2007-8-10-r210>

MSC2020 subject classifications. Primary 62H05, 62H12; secondary 62H22.

Key words and phrases. Association, convex optimization, dual likelihood, exponential families, Gaussian distribution, graphical lasso, Kullback–Leibler divergence, mixed parametrization, positive correlations, structure learning.

- [12] COSTA, D. and CASTELO, R. (2016). Umbilical cord gene expression reveals the molecular architecture of the fetal inflammatory response in extremely preterm newborns. *Pediatr. Res.* **79** 473–481.
- [13] DJORDJLOVIĆ, V. and CHIOGNA, M. (2018). Searching for a source of difference in Gaussian graphical models. arXiv preprint. Available at [arXiv:1811.02503](https://arxiv.org/abs/1811.02503).
- [14] DRTON, M. and RICHARDSON, T. S. (2002). A new algorithm for maximum likelihood estimation in Gaussian graphical models for marginal independence. In *Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence* 184–191. Morgan Kaufmann, San Mateo.
- [15] EFRON, B. (1978). The geometry of exponential families. *Ann. Statist.* **6** 362–376. [MR0471152](#)
- [16] ESARY, J. D., PROSCHAN, F. and WALKUP, D. W. (1967). Association of random variables, with applications. *Ann. Math. Stat.* **38** 1466–1474. [MR0217826](#) <https://doi.org/10.1214/aoms/1177698701>
- [17] FALLAT, S., LAURITZEN, S., SADEGHI, K., UHLER, C., WERMUTH, N. and ZWIERNIK, P. (2017). Total positivity in Markov structures. *Ann. Statist.* **45** 1152–1184. [MR3662451](#) <https://doi.org/10.1214/16-AOS1478>
- [18] FAN, J., FENG, Y. and WU, Y. (2009). Network exploration via the adaptive lasso and SCAD penalties. *Ann. Appl. Stat.* **3** 521–541. [MR2750671](#) <https://doi.org/10.1214/08-AOS215>
- [19] FITZMAURICE, G. M. and LAIRD, N. M. (1993). A likelihood-based method for analysing longitudinal binary responses. *Biometrika* **80** 141–151.
- [20] FOYCEL, R. and DRTON, M. (2010). Extended Bayesian information criteria for Gaussian graphical models. In *Advances in Neural Information Processing Systems* 604–612.
- [21] GEYER, C. J. (1994). On the asymptotics of constrained M -estimation. *Ann. Statist.* **22** 1993–2010. [MR1329179](#) <https://doi.org/10.1214/aos/1176325768>
- [22] GLONEK, G. F. V. (1996). A class of regression models for multivariate categorical responses. *Biometrika* **83** 15–28. [MR1399152](#) <https://doi.org/10.1093/biomet/83.1.15>
- [23] GROSS, E. and SULLIVAN, S. (2018). The maximum likelihood threshold of a graph. *Bernoulli* **24** 386–407. [MR3706762](#) <https://doi.org/10.3150/16-BEJ881>
- [24] HORN, R. A. and JOHNSON, C. R. (2013). *Matrix Analysis*, 2nd ed. Cambridge Univ. Press, Cambridge. [MR2978290](#)
- [25] JOHNSON, R. W. (1996). Fitting percentage of body fat to simple body measurements. *J. Stat. Educ.* **4**.
- [26] KARLIN, S. and RINOTT, Y. (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *J. Multivariate Anal.* **10** 467–498. [MR0599685](#) [https://doi.org/10.1016/0047-259X\(80\)90065-2](https://doi.org/10.1016/0047-259X(80)90065-2)
- [27] KARLIN, S. and RINOTT, Y. (1983). M -matrices as covariance matrices of multinormal distributions. *Linear Algebra Appl.* **52/53** 419–438. [MR0709364](#) [https://doi.org/10.1016/0024-3795\(83\)80027-5](https://doi.org/10.1016/0024-3795(83)80027-5)
- [28] KAUFMANN, G. (1996). On a dualization of graphical Gaussian models. *Scand. J. Stat.* **23** 105–116. [MR1380485](#)
- [29] KHARCHENKO, P., CHURCH, G. M. and VITKUP, D. (2005). Expression dynamics of a cellular metabolic network. *Mol. Syst. Biol.* **1** 2005-0016.
- [30] KULLBACK, S. (1959). *Information Theory and Statistics*. Wiley, New York. [MR0103557](#)
- [31] KULLBACK, S. (1971). Marginal homogeneity of multidimensional contingency tables. *Ann. Math. Stat.* **42** 594–606. [MR0297084](#) <https://doi.org/10.1214/aoms/1177693409>
- [32] 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>
- [33] LAURITZEN, S., UHLER, C. and ZWIERNIK, P. (2019). Maximum likelihood estimation in Gaussian models under total positivity. *Ann. Statist.* **47** 1835–1863. [MR3953437](#) <https://doi.org/10.1214/17-AOS1668>
- [34] LAURITZEN, S., UHLER, C. and ZWIERNIK, P. (2021). Total positivity in exponential families with application to binary variables. *Ann. Statist.* **49** 1436–1459. [MR4298870](#) <https://doi.org/10.1214/20-aos2007>
- [35] LAURITZEN, S. L. (1996). *Graphical Models. Oxford Statistical Science Series* **17**. Clarendon, Oxford. [MR1419991](#)
- [36] LINDSKOG, F., MCNEIL, A. and SCHMOCK, U. (2003). Kendall’s tau for elliptical distributions. In *Credit Risk* 149–156. Springer, Berlin.
- [37] LIU, H., HAN, F., YUAN, M., LAFFERTY, J. and WASSERMAN, L. (2012). High-dimensional semiparametric Gaussian copula graphical models. *Ann. Statist.* **40** 2293–2326. [MR3059084](#) <https://doi.org/10.1214/12-AOS1037>
- [38] MALIOUTOV, D. M., JOHNSON, J. K. and WILLSKY, A. S. (2006). Walk-sums and belief propagation in Gaussian graphical models. *J. Mach. Learn. Res.* **7** 2031–2064. [MR2274432](#)
- [39] MANN, H. B. and WALD, A. (1943). On stochastic limit and order relationships. *Ann. Math. Stat.* **14** 217–226. [MR0009261](#) <https://doi.org/10.1214/aoms/1177731415>

- [40] MASON, M. J., FAN, G., PLATH, K., ZHOU, Q. and HORVATH, S. (2009). Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. *BMC Genomics* **10** 327. <https://doi.org/10.1186/1471-2164-10-327>
- [41] PEARL, J. and WERMUTH, N. (1994). When can association graphs admit a causal interpretation? In *Selecting Models from Data* 205–214 Springer, New York.
- [42] PITT, L. D. (1982). Positively correlated normal variables are associated. *Ann. Probab.* **10** 496–499. [MR0665603](https://doi.org/10.2307/2346033)
- [43] ROCKAFELLAR, R. T. (1970). *Convex Analysis. Princeton Mathematical Series* **28**. Princeton Univ. Press, Princeton. [MR0274683](https://doi.org/10.2307/2346033)
- [44] ROCKAFELLAR, R. T. and WETS, R. J.-B. (2009). *Variational Analysis* **317**. Springer, Berlin.
- [45] ROSSELL, D. and ZWIERNIK, P. (2021). Dependence in elliptical partial correlation graphs. *Electron. J. Stat.* **15** 4236–4263. [MR4312203](https://doi.org/10.1214/21-ejs1891) <https://doi.org/10.1214/21-ejs1891>
- [46] SLAWSKI, M. and HEIN, M. (2015). Estimation of positive definite M -matrices and structure learning for attractive Gaussian Markov random fields. *Linear Algebra Appl.* **473** 145–179. [MR3338330](https://doi.org/10.1016/j.laa.2014.04.020) <https://doi.org/10.1016/j.laa.2014.04.020>
- [47] SOLOFF, J. A., GUNTUBOYINA, A. and JORDAN, M. I. (2020). Covariance estimation with nonnegative partial correlations. arXiv preprint. Available at [arXiv:2007.15252](https://arxiv.org/abs/2007.15252).
- [48] SPEED, T. P. and KIIVERI, H. T. (1986). Gaussian Markov distributions over finite graphs. *Ann. Statist.* **14** 138–150. [MR0829559](https://doi.org/10.1214/aos/1176349846) <https://doi.org/10.1214/aos/1176349846>
- [49] SUNDBERG, R. (2019). *Statistical Modelling by Exponential Families. Institute of Mathematical Statistics Textbooks* **12**. Cambridge Univ. Press, Cambridge. [MR3969949](https://doi.org/10.1017/9781108604574) <https://doi.org/10.1017/9781108604574>
- [50] UHLER, C. (2012). Geometry of maximum likelihood estimation in Gaussian graphical models. *Ann. Statist.* **40** 238–261. [MR3014306](https://doi.org/10.1214/11-AOS957) <https://doi.org/10.1214/11-AOS957>
- [51] VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. [MR1652247](https://doi.org/10.1017/CBO9780511802256) <https://doi.org/10.1017/CBO9780511802256>
- [52] WEI, H., PERSSON, S., MEHTA, T., SRINIVASASAINAGENDRA, V., CHEN, L., PAGE, G. P., SOMERVILLE, C. and LORAINE, A. (2006). Transcriptional coordination of the metabolic network in Arabidopsis. *Plant Physiol.* **142** 762–774.

ESTIMATION OF TIME SERIES MODELS USING RESIDUALS DEPENDENCE MEASURES

BY CARLOS VELASCO^a

Department of Economics, Universidad Carlos III de Madrid, ^acarlos.velasco@uc3m.es

We propose new estimation methods for time series models, possibly noncausal and/or noninvertible, using serial dependence information from the characteristic function of model residuals. This allows to impose the *i.i.d.* or martingale difference assumptions on the model errors to identify the unknown location of the roots of the lag polynomials for ARMA models without resorting to higher order moments or distributional assumptions. We consider generalized spectral density and cumulative distribution functions to measure residuals dependence at an increasing number of lags under both assumptions and discuss robust inference to higher order dependence when only mean independence is imposed on model errors. We study the consistency and asymptotic distribution of parameter estimates and discuss efficiency when different restrictions on error dependence are used simultaneously, including serial uncorrelation. Optimal weighting of continuous moment conditions yields maximum likelihood efficiency under independence for unknown error distribution. We investigate numerical implementation and finite sample properties of the new classes of estimates.

REFERENCES

- ANATOLYEV, S. (2007). Optimal instruments in time series: A survey. *J. Econ. Surv.* **21** 143–173.
- ANDREWS, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59** 817–858. [MR1106513](#) <https://doi.org/10.2307/2938229>
- ANDREWS, B., DAVIS, R. A. and BREIDT, F. J. (2007). Rank-based estimation for all-pass time series models. *Ann. Statist.* **35** 844–869. [MR2336871](#) <https://doi.org/10.1214/009053606000001316>
- BLOOMFIELD, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika* **60** 217–226. [MR0323048](#) <https://doi.org/10.1093/biomet/60.2.217>
- BOX, G. E. P. and PIERCE, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Amer. Statist. Assoc.* **65** 1509–1526. [MR0273762](#)
- BREIDT, F. J., DAVIS, R. A. and TRINDADE, A. A. (2001). Least absolute deviation estimation for all-pass time series models. *Ann. Statist.* **29** 919–946. [MR1869234](#) <https://doi.org/10.1214/aos/1013699986>
- CARRASCO, M., CHERNOV, M., FLORENS, J.-P. and GHYSELS, E. (2007). Efficient estimation of general dynamic models with a continuum of moment conditions. *J. Econometrics* **140** 529–573. [MR2408918](#) <https://doi.org/10.1016/j.jeconom.2006.07.013>
- CHEN, B., CHOI, J. and ESCANCIANO, J. C. (2017). Testing for fundamental vector moving average representations. *Quant. Econ.* **8** 149–180. [MR3638602](#) <https://doi.org/10.3982/QE393>
- CHENG, Q. S. (1992). On the unique representation of non-Gaussian linear processes. *Ann. Statist.* **20** 1143–1145. [MR1165613](#) <https://doi.org/10.1214/aos/1176348677>
- DEDECKER, J., DOUKHAN, P., LANG, G., LEÓN R., J. R., LOUHICHI, S. and PRIEUR, C. (2007). *Weak Dependence: With Examples and Applications. Lecture Notes in Statistics* **190**. Springer, New York. [MR2338725](#)
- DOUKHAN, P. and LOUHICHI, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.* **84** 313–342. [MR1719345](#) [https://doi.org/10.1016/S0304-4149\(99\)00055-1](https://doi.org/10.1016/S0304-4149(99)00055-1)
- ESCANCIANO, J. C. and VELASCO, C. (2006a). Generalized spectral tests for the martingale difference hypothesis. *J. Econometrics* **134** 151–185. [MR2328319](#) <https://doi.org/10.1016/j.jeconom.2005.06.019>
- ESCANCIANO, J. C. and VELASCO, C. (2006b). Testing the martingale difference hypothesis using integrated regression functions. *Comput. Statist. Data Anal.* **51** 2278–2294. [MR2307501](#) <https://doi.org/10.1016/j.csda.2006.07.039>

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

Key words and phrases. Characteristic function, martingale difference, generalized method of moments, generalized spectral density, noncausal processes, noninvertible processes.

- FEUERVERGER, A. (1990). An efficiency result for the empirical characteristic function in stationary time-series models. *Canad. J. Statist.* **18** 155–161. [MR1067167](#) <https://doi.org/10.2307/3315564>
- FOKIANOS, K. and PITSILLOU, M. (2018). Testing independence for multivariate time series via the auto-distance correlation matrix. *Biometrika* **105** 337–352. [MR3804406](#) <https://doi.org/10.1093/biomet/asx082>
- GASSIAT, É. (1993). Adaptive estimation in noncausal stationary AR processes. *Ann. Statist.* **21** 2022–2042. [MR1245779](#) <https://doi.org/10.1214/aos/1176349408>
- GORODECKIĬ, V. V. (1977). The strong mixing property for linearly generated sequences. *Theory Probab. Appl.* **22** 411–413. [MR0438450](#)
- GOSPODINOV, N. and NG, S. (2015). Minimum distance estimation of possibly noninvertible moving average models. *J. Bus. Econom. Statist.* **33** 403–417. [MR3372667](#) <https://doi.org/10.1080/07350015.2014.955175>
- GOURIÉROUX, C. and ZAKOÏAN, J.-M. (2017). Local explosion modelling by non-causal process. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 737–756. [MR3641405](#) <https://doi.org/10.1111/rssb.12193>
- HANNAN, E. J. (1970). *Multiple Time Series*. Wiley, New York. [MR0279952](#)
- HONG, Y. (1999). Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach. *J. Amer. Statist. Assoc.* **94** 1201–1220. [MR1731483](#) <https://doi.org/10.2307/2669935>
- HONG, Y. and LEE, Y.-J. (2005). Generalized spectral tests for conditional mean models in time series with conditional heteroscedasticity of unknown form. *Rev. Econ. Stud.* **72** 499–541. [MR2129829](#) <https://doi.org/10.1111/j.1467-937X.2005.00341.x>
- HOSKING, J. R. M. (1981). Fractional differencing. *Biometrika* **68** 165–176. [MR0614953](#) <https://doi.org/10.1093/biomet/68.1.165>
- HUANG, J. and PAWITAN, Y. (2000). Quasi-likelihood estimation of non-invertible moving average processes. *Scand. J. Stat.* **27** 689–702. [MR1804170](#) <https://doi.org/10.1111/1467-9469.00216>
- KNIGHT, J. L. and YU, J. (2002). Empirical characteristic function in time series estimation. *Econometric Theory* **18** 691–721. [MR1906331](#) <https://doi.org/10.1017/S026646660218306X>
- KREISS, J.-P. (1987). On adaptive estimation in stationary ARMA processes. *Ann. Statist.* **15** 112–133. [MR0885727](#) <https://doi.org/10.1214/aos/1176350256>
- LANNE, M. and SAIKKONEN, P. (2011). Noncausal autoregressions for economic time series. *J. Time Ser. Econom.* **3** Art. 2. [MR2928654](#) <https://doi.org/10.2202/1941-1928.1080>
- LII, K. S. and ROSENBLATT, M. (1982). Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes. *Ann. Statist.* **10** 1195–1208. [MR0673654](#)
- LII, K.-S. and ROSENBLATT, M. (1992). An approximate maximum likelihood estimation for non-Gaussian non-minimum phase moving average processes. *J. Multivariate Anal.* **43** 272–299. [MR1193615](#) [https://doi.org/10.1016/0047-259X\(92\)90037-G](https://doi.org/10.1016/0047-259X(92)90037-G)
- LII, K.-S. and ROSENBLATT, M. (1996). Maximum likelihood estimation for non-Gaussian nonminimum phase ARMA sequences. *Statist. Sinica* **6** 1–22. [MR1379046](#)
- LIPPI, M. and REICHLIN, L. (1994). VAR analysis, nonfundamental representations, Blaschke matrices. *J. Econometrics* **63** 307–325. [MR1309986](#) [https://doi.org/10.1016/0304-4076\(93\)01570-C](https://doi.org/10.1016/0304-4076(93)01570-C)
- PINKSE, J. (1998). A consistent nonparametric test for serial independence. *J. Econometrics* **84** 205–231. [MR1630190](#) [https://doi.org/10.1016/S0304-4076\(97\)00084-5](https://doi.org/10.1016/S0304-4076(97)00084-5)
- RAMSEY, J. B. and MONTENEGRO, A. (1992). Identification and estimation of non-invertible non-Gaussian MA(q) processes. *J. Econometrics* **54** 301–320.
- ROSENBLATT, M. (2000). *Gaussian and Non-Gaussian Linear Time Series and Random Fields*. Springer Series in Statistics. Springer, New York. [MR1742357](#) <https://doi.org/10.1007/978-1-4612-1262-1>
- STINCHCOMBE, M. B. and WHITE, H. (1998). Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* **14** 295–325. [MR1628586](#) <https://doi.org/10.1017/S0266466698143013>
- SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. [MR2382665](#) <https://doi.org/10.1214/009053607000000505>
- TJØSTHEIM, D., OTNEIM, H. and STØVE, B. (2022). Statistical dependence: Beyond Pearson’s ρ . *Statist. Sci.* **37** 90–109. [MR4371097](#) <https://doi.org/10.1214/21-sts823>
- VELASCO, C. (2022). Supplement to “Estimation of time series models using residuals dependence measures.” <https://doi.org/10.1214/22-AOS2220SUPPA>, <https://doi.org/10.1214/22-AOS2220SUPPB>
- VELASCO, C. and LOBATO, I. N. (2018). Frequency domain minimum distance inference for possibly non-invertible and noncausal ARMA models. *Ann. Statist.* **46** 555–579. [MR3782377](#) <https://doi.org/10.1214/17-AOS1560>
- YAO, S., ZHANG, X. and SHAO, X. (2018). Testing mutual independence in high dimension via distance covariance. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 455–480. [MR3798874](#) <https://doi.org/10.1111/rssb.12259>
- YOSHIHARA, K. (1978). Moment inequalities for mixing sequences. *Kodai Math. J.* **1** 316–328. [MR0508749](#)

RATE-OPTIMAL CLUSTER-RANDOMIZED DESIGNS FOR SPATIAL INTERFERENCE

BY MICHAEL P. LEUNG^a

Department of Economics, University of California, Santa Cruz, ^aleungm@ucsc.edu

We consider a potential outcomes model in which interference may be present between any two units but the extent of interference diminishes with spatial distance. The causal estimand is the global average treatment effect, which compares outcomes under the counterfactuals that all or no units are treated. We study a class of designs in which space is partitioned into clusters that are randomized into treatment and control. For each design, we estimate the treatment effect using a Horvitz–Thompson estimator that compares the average outcomes of units with all or no neighbors treated, where the neighborhood radius is of the same order as the cluster size dictated by the design. We derive the estimator’s rate of convergence as a function of the design and degree of interference and use this to obtain estimator–design pairs that achieve near-optimal rates of convergence under relatively minimal assumptions on interference. We prove that the estimators are asymptotically normal and provide a variance estimator. For practical implementation of the designs, we suggest partitioning space using clustering algorithms.

REFERENCES

- [1] ANDREWS, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59** 817–858. [MR1106513](#) <https://doi.org/10.2307/2938229>
- [2] ANSELIN, L. (2001). Spatial econometrics. In *A Companion to Theoretical Econometrics*. Blackwell Companions Contemp. Econ. 310–330. Blackwell, Malden, MA. [MR2041363](#)
- [3] ARONOW, P., SAMII, C. and WANG, Y. (2020). Design-based inference for spatial experiments with interference. arXiv preprint [arXiv:2010.13599](#).
- [4] ARONOW, P. M. and SAMII, C. (2017). Estimating average causal effects under general interference, with application to a social network experiment. *Ann. Appl. Stat.* **11** 1912–1947. [MR3743283](#) <https://doi.org/10.1214/16-AOAS1005>
- [5] BAIRD, S., BOHREN, J., MCINTOSH, C. and ÖZLER, B. (2018). Optimal design of experiments in the presence of interference. *Rev. Econ. Stat.* **100** 844–860.
- [6] BASSE, G. W. and AIROLDI, E. M. (2018). Model-assisted design of experiments in the presence of network-correlated outcomes. *Biometrika* **105** 849–858. [MR3877869](#) <https://doi.org/10.1093/biomet/asy036>
- [7] BASSE, G. W., FELLER, A. and TOULIS, P. (2019). Randomization tests of causal effects under interference. *Biometrika* **106** 487–494. [MR3949317](#) <https://doi.org/10.1093/biomet/asy072>
- [8] BLATTMAN, C., GREEN, D., ORTEGA, D. and TOBÓN, S. (2021). Place-based interventions at scale: The direct and spillover effects of policing and city services on crime. *J. Eur. Econ. Assoc.* **19** 2022–2051.
- [9] CHIN, A. (2019). Regression adjustments for estimating the global treatment effect in experiments with interference. *J. Causal Inference* **7** Art. No. 20180026, 36. [MR4350071](#) <https://doi.org/10.1515/jci-2018-0026>
- [10] CHOI, D. (2017). Estimation of monotone treatment effects in network experiments. *J. Amer. Statist. Assoc.* **112** 1147–1155. [MR3735366](#) <https://doi.org/10.1080/01621459.2016.1194845>
- [11] CLIFF, A. and ORD, J. (1973). *Spatial Autocorrelation*. Pion, London.
- [12] CLIFF, A. D. and ORD, J. K. (1981). *Spatial Processes: Models & Applications*. Pion Ltd., London. [MR0632256](#)
- [13] CONLEY, T. G. (1999). GMM estimation with cross sectional dependence. *J. Econometrics* **92** 1–45. [MR1707000](#) [https://doi.org/10.1016/S0304-4076\(98\)00084-0](https://doi.org/10.1016/S0304-4076(98)00084-0)

- [14] DONNELLY, C., WOODROFFE, R., COX, D., BOURNE, J., GETTINBY, G., LE FEVRE, A., MCINERNEY, J. and MORRISON, I. (2003). Impact of localized badger culling on tuberculosis incidence in British cattle. *Nature* **426** 834–837.
- [15] ECKLES, D., KARRER, B. and UGANDER, J. (2017). Design and analysis of experiments in networks: Reducing bias from interference. *J. Causal Inference* **5** Art. No. 20150021. [MR4323809 https://doi.org/10.1515/jci-2015-0021](https://doi.org/10.1515/jci-2015-0021)
- [16] FORASTIERE, L., AIROLDI, E. M. and MEALLI, F. (2021). Identification and estimation of treatment and interference effects in observational studies on networks. *J. Amer. Statist. Assoc.* **116** 901–918. [MR4270033 https://doi.org/10.1080/01621459.2020.1768100](https://doi.org/10.1080/01621459.2020.1768100)
- [17] GETIS, A. (2008). A history of the concept of spatial autocorrelation: A geographer’s perspective. *Geogr. Anal.* **40** 297–309.
- [18] GIFFIN, A., REICH, B., YANG, S. and RAPPOLD, A. (2020). Generalized propensity score approach to causal inference with spatial interference. arXiv preprint [arXiv:2007.00106](https://arxiv.org/abs/2007.00106).
- [19] HARSHAW, C., SÄVJE, F., EISENSTAT, D., MIRROKNI, V. and POUGET-ABADIE, J. (2021). Design and analysis of bipartite experiments under a linear exposure-response model. arXiv preprint [arXiv:2103.06392](https://arxiv.org/abs/2103.06392).
- [20] HAYES, R. and MOULTON, L. (2017). *Cluster Randomised Trials*. CRC Press/CRC.
- [21] HU, Y., LI, S. and WAGER, S. (2022). Average direct and indirect causal effects under interference. *Biometrika*. To appear.
- [22] IMBENS, G. W. and RUBIN, D. B. (2015). *Causal Inference—for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge Univ. Press, New York. [MR3309951 https://doi.org/10.1017/CBO9781139025751](https://doi.org/10.1017/CBO9781139025751)
- [23] JAGADEESAN, R., PILLAI, N. S. and VOLFOVSKY, A. (2020). Designs for estimating the treatment effect in networks with interference. *Ann. Statist.* **48** 679–712. [MR4102672 https://doi.org/10.1214/18-AOS1807](https://doi.org/10.1214/18-AOS1807)
- [24] JENISH, N. (2016). Spatial semiparametric model with endogenous regressors. *Econometric Theory* **32** 714–739. [MR3506437 https://doi.org/10.1017/S0266466614000905](https://doi.org/10.1017/S0266466614000905)
- [25] JENISH, N. and PRUCHA, I. R. (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields. *J. Econometrics* **150** 86–98. [MR2525996 https://doi.org/10.1016/j.jeconom.2009.02.009](https://doi.org/10.1016/j.jeconom.2009.02.009)
- [26] JENISH, N. and PRUCHA, I. R. (2011). On spatial processes and asymptotic inference under near-epoch dependence. U. Maryland working paper.
- [27] JENISH, N. and PRUCHA, I. R. (2012). On spatial processes and asymptotic inference under near-epoch dependence. *J. Econometrics* **170** 178–190. [MR2955948 https://doi.org/10.1016/j.jeconom.2012.05.022](https://doi.org/10.1016/j.jeconom.2012.05.022)
- [28] LAHIRI, S. N. (1996). On inconsistency of estimators based on spatial data under infill asymptotics. *Sankhyā Ser. A* **58** 403–417. [MR1659130 https://doi.org/10.2307/1400051](https://doi.org/10.2307/1400051)
- [29] LAHIRI, S. N. (2003). Central limit theorems for weighted sums of a spatial process under a class of stochastic and fixed designs. *Sankhyā* **65** 356–388. [MR2028905 https://doi.org/10.2307/1400051](https://doi.org/10.2307/1400051)
- [30] LAHIRI, S. N. and ZHU, J. (2006). Resampling methods for spatial regression models under a class of stochastic designs. *Ann. Statist.* **34** 1774–1813. [MR2283717 https://doi.org/10.1214/009053606000000551](https://doi.org/10.1214/009053606000000551)
- [31] LEUNG, M. (2022). Network cluster-robust inference. arXiv preprint [arXiv:2103.01470](https://arxiv.org/abs/2103.01470).
- [32] LEUNG, M. P. (2022). Supplement to “Rate-optimal cluster-randomized designs for spatial interference.” <https://doi.org/10.1214/22-AOS2224SUPP>
- [33] LEUNG, M. P. (2022). Causal inference under approximate neighborhood interference. *Econometrica* **90** 267–293. [MR4376833 https://doi.org/10.3982/ECTA17841](https://doi.org/10.3982/ECTA17841)
- [34] MANSKI, C. F. (2013). Identification of treatment response with social interactions. *Econom. J.* **16** S1–S23. [MR3030060 https://doi.org/10.1111/j.1368-423X.2012.00368.x](https://doi.org/10.1111/j.1368-423X.2012.00368.x)
- [35] MIGUEL, E. and KREMER, M. (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica* **72** 159–217.
- [36] PALER, L., SAMII, C., LISIECKI, M. and MOREL, A. (2015). Social and environmental impact of the community rangers program in Aceh. Technical report, World Bank, Washington, DC.
- [37] PARK, C. and KANG, H. (2021). Assumption-lean analysis of cluster randomized trials in infectious diseases for intent-to-treat effects and spillover effects among a vulnerable subpopulation. *J. Amer. Statist. Assoc.* To appear.
- [38] PENG, R., SUN, H. and ZANETTI, L. (2017). Partitioning well-clustered graphs: Spectral clustering works! *SIAM J. Comput.* **46** 710–743. [MR3629439 https://doi.org/10.1137/15M1047209](https://doi.org/10.1137/15M1047209)
- [39] POLLMANN, M. (2020). Causal inference for spatial treatments. arXiv preprint [arXiv:2011.00373](https://arxiv.org/abs/2011.00373).

- [40] POUGET-ABADIE, J., MIRROKNI, V., PARKES, D. and AIROLDI, E. (2018). Optimizing cluster-based randomized experiments under monotonicity. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* 2090–2099.
- [41] ROSS, N. (2011). Fundamentals of Stein’s method. *Probab. Surv.* **8** 210–293. **MR2861132** <https://doi.org/10.1214/11-PS182>
- [42] SÄVJE, F. (2021). Causal inference with misspecified exposure mappings. arXiv preprint [arXiv:2103.06471](https://arxiv.org/abs/2103.06471).
- [43] SÄVJE, F., ARONOW, P. M. and HUDGENS, M. G. (2021). Average treatment effects in the presence of unknown interference. *Ann. Statist.* **49** 673–701. **MR4255103** <https://doi.org/10.1214/20-aos1973>
- [44] SUSSMAN, D. and AIROLDI, E. (2017). Elements of estimation theory for causal effects in the presence of network interference. arXiv preprint [arXiv:1702.03578](https://arxiv.org/abs/1702.03578).
- [45] TOULIS, P. and KAO, E. (2013). Estimation of causal peer influence effects. In *International Conference on Machine Learning* 1489–1497.
- [46] UGANDER, J., KARRER, B., BACKSTROM, L. and KLEINBERG, J. (2013). Graph cluster randomization: Network exposure to multiple universes. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 329–337.
- [47] UGANDER, J. and YIN, H. (2020). Randomized graph cluster randomization. arXiv preprint [arXiv:2009.02297](https://arxiv.org/abs/2009.02297).
- [48] VALCU, M. and KEMPENAERS, B. (2010). Spatial autocorrelation: An overlooked concept in behavioral ecology. *Behav. Ecol.* **21** 902–905. <https://doi.org/10.1093/beheco/arq107>
- [49] VERBITSKY-SAVITZ, N. and RAUDENBUSH, S. (2012). Causal inference under interference in spatial settings: A case study evaluating community policing program in Chicago. *Epidemiol. Methods* **1** 107–130.
- [50] VIVIANO, D. (2020). Experimental design under network interference. arXiv preprint [arXiv:2003.08421](https://arxiv.org/abs/2003.08421).
- [51] ZIGLER, C. M. and PAPADOGEORGOU, G. (2021). Bipartite causal inference with interference. *Statist. Sci.* **36** 109–123. **MR4194206** <https://doi.org/10.1214/19-STS749>

CORRECTION NOTE: “ASYMPTOTIC SPECTRAL THEORY FOR NONLINEAR TIME SERIES”

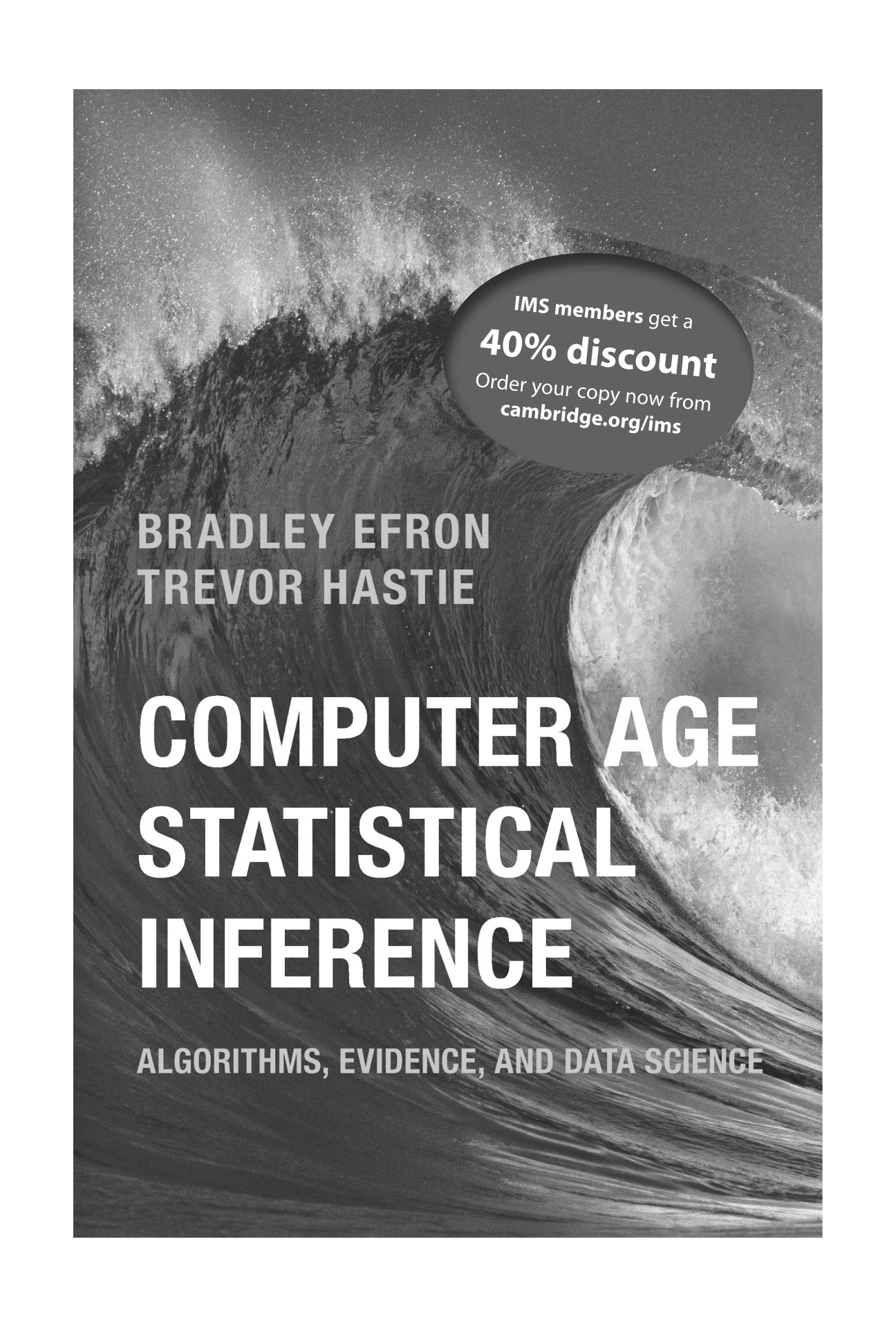
BY YI ZHANG^{1,a}, XIAOFENG SHAO^{1,b} AND WEIBIAO WU^{2,c}

¹*Department of Statistics, University of Illinois Urbana-Champaign, ^ayiz19@illinois.edu, ^bxshao@illinois.edu*

²*Department of Statistics, University of Chicago, ^cwbwu@galton.uchicago.edu*

REFERENCES

- [1] ROSENBLATT, M. (1985). *Stationary Sequences and Random Fields*. Birkhäuser, Boston, MA.
- [2] SHAO, X. and WU, W. B. (2007). Asymptotic spectral theory for nonlinear time series. *Ann. Statist.* **35** 1773–1801. [MR2351105 https://doi.org/10.1214/009053606000001479](https://doi.org/10.1214/009053606000001479)



IMS members get a
40% discount

Order your copy now from
cambridge.org/ims

BRADLEY EFRON
TREVOR HASTIE

COMPUTER AGE STATISTICAL INFERENCE

ALGORITHMS, EVIDENCE, AND DATA SCIENCE