

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

- Design admissibility and de la Garza phenomenon in multifactor experiments
HOLGER DETTE, XIN LIU AND RONG-XIAN YUE 1247
- Evidence factors from multiple, possibly invalid, instrumental variables
ANQI ZHAO, YOUJIN LEE, DYLAN S. SMALL AND BIKRAM KARMAKAR 1266
- Consistent order selection for ARFIMA processes
HSUEH-HAN HUANG, NGAI HANG CHAN, KUN CHEN AND CHING-KANG ING 1297
- Doubly debiased lasso: High-dimensional inference under hidden confounding
ZIJIAN GUO, DOMAGOJ ČEVID AND PETER BÜHLMANN 1320
- Deep learning for the partially linear Cox model
QIXIAN ZHONG, JONAS MUELLER AND JANE-LING WANG 1348
- Optimal difference-based variance estimators in time series: A general framework
KIN WAI CHAN 1376
- Ridge regression revisited: Debiasing, thresholding and bootstrap
YUNYI ZHANG AND DIMITRIS N. POLITIS 1401
- Cube root weak convergence of empirical estimators of a density level set
PHILIPPE BERTHET AND JOHN H. J. EINMAHL 1423
- Higher criticism to compare two large frequency tables, with sensitivity to possible rare
and weak differences DAVID L. DONOHO AND ALON KIPNIS 1447
- A new and flexible design construction for orthogonal arrays for modern applications
YUANZHEN HE, C. DEVON LIN AND FASHENG SUN 1473
- Consistency of Bayesian inference for multivariate max-stable distributions
SIMONE A. PADOAN AND STEFANO RIZZELLI 1490
- Minimax estimation of smooth densities in Wasserstein distance
JONATHAN NILES-WEED AND QUENTIN BERTHET 1519
- Large-scale inference with block structure. JIYAO KOU AND GUENTHER WALTHER 1541
- Uniform convergence of local Fréchet regression with applications to locating extrema
and time warping for metric space valued trajectories
YAQING CHEN AND HANS-GEORG MÜLLER 1573
- Cointegration in large VARs ANNA BYKHOVSKAYA AND VADIM GORIN 1593
- Partial recovery for top- k ranking: Optimality of MLE and SubOptimality of the spectral
method PINHAN CHEN, CHAO GAO AND ANDERSON Y. ZHANG 1618
- A data-adaptive method for estimating density level sets under shape conditions
ALBERTO RODRÍGUEZ-CASAL AND PAULA SAAVEDRA-NIEVES 1653
- A precise high-dimensional asymptotic theory for boosting and minimum- ℓ_1 -norm
interpolated classifiers TENGYUAN LIANG AND PRAGYA SUR 1669
- Intrinsic Riemannian functional data analysis for sparse longitudinal observations
LINGXUAN SHAO, ZHENHUA LIN AND FANG YAO 1696
- Spectral estimation of Hawkes processes from count data
FELIX CHEYSSON AND GABRIEL LANG 1722
- Model selection in the space of Gaussian models invariant by symmetry
PIOTR GRACZYK, HIDEYUKI ISHI,
BARTOSZ KOŁODZIEJEK AND HÉLÈNE MASSAM 1747

continued

THE ANNALS
of
STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles—Continued from front cover

Optimal full ranking from pairwise comparisons	
PINHAN CHEN, CHAO GAO AND ANDERSON Y. ZHANG	1775
Confidence regions near singular information and boundary points with applications to mixed models	
KARL OSKAR EKVALL AND MATTEO BOTTAI	1806
Computational barriers to estimation from low-degree polynomials	
TSELIL SCHRAMM AND ALEXANDER S. WEIN	1833

THE ANNALS OF STATISTICS

Vol. 50, No. 3, pp. 1247–1858 June 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: Krzysztof Burdzy, Department of Mathematics, University of Washington, Seattle, Washington 98195-4350, USA

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

Past President: Regina Y. Liu, Department of Statistics, Rutgers University, Piscataway, New Jersey 08854-8019, USA

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

Treasurer: Zhengjun Zhang, Department of Statistics, University of Wisconsin, Madison, Wisconsin 53706-1510, USA

Program Secretary: Annie Qu, Department of Statistics, University of California, Irvine, Irvine, California 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 3, June 2022. Published bimonthly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, Ohio 44094, USA. Periodicals postage paid at Cleveland, Ohio, and at additional mailing offices.

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

DESIGN ADMISSIBILITY AND DE LA GARZA PHENOMENON IN MULTIFACTOR EXPERIMENTS

BY HOLGER DETTE^{1,a}, XIN LIU^{2,b} AND RONG-XIAN YUE^{3,c}

¹Fakultät für Mathematik, Ruhr-Universität Bochum, ^aholger.dette@rub.de

²College of Science, Donghua University, ^bliuxin@dhu.edu.cn

³Department of Mathematics, Shanghai Normal University, ^cyue2@shnu.edu.cn

The determination of an optimal design for a given regression problem is an intricate optimization problem, especially for models with multivariate predictors. Design admissibility and invariance are main tools to reduce the complexity of the optimization problem and have been successfully applied for models with univariate predictors. In particular, several authors have developed sufficient conditions for the existence of minimally supported designs in univariate models, where the number of support points of the optimal design equals the number of parameters. These results generalize the celebrated de la Garza phenomenon (*Ann. Math. Statistics* **25** (1954) 123–130), which states that for a polynomial regression model of degree $k - 1$ any optimal design can be based on k points.

This paper provides—for the first time—extensions of these results for models with a multivariate predictor. In particular, we study a geometric characterization of the support points of an optimal design to provide sufficient conditions for the occurrence of the de la Garza phenomenon in models with multivariate predictors and characterize properties of admissible designs in terms of admissibility of designs in *conditional* univariate regression models.

REFERENCES

- ATKINSON, A. C., DONEV, A. N. and TOBIAS, R. D. (2007). *Optimum Experimental Designs, with SAS*. Oxford Statistical Science Series **34**. Oxford Univ. Press, Oxford. [MR2323647](#)
- BIEDERMANN, S., DETTE, H. and WOODS, D. C. (2011). Optimal design for additive partially nonlinear models. *Biometrika* **98** 449–458. [MR2806440](#) <https://doi.org/10.1093/biomet/asr001>
- DE CASTRO, Y., GAMBOA, F., HENRION, D., HESS, R. and LASSERRE, J.-B. (2019). Approximate optimal designs for multivariate polynomial regression. *Ann. Statist.* **47** 127–155. [MR3909929](#) <https://doi.org/10.1214/18-AOS1683>
- DE LA GARZA, A. (1954). Spacing of information in polynomial regression. *Ann. Math. Stat.* **25** 123–130. [MR0060777](#) <https://doi.org/10.1214/aoms/1177728851>
- DETTE, H. and GRIGORIEV, Y. (2014). E -optimal designs for second-order response surface models. *Ann. Statist.* **42** 1635–1656. [MR3262463](#) <https://doi.org/10.1214/14-AOS1241>
- DETTE, H. and MELAS, V. B. (2011). A note on the de la Garza phenomenon for locally optimal designs. *Ann. Statist.* **39** 1266–1281. [MR2816354](#) <https://doi.org/10.1214/11-AOS875>
- DETTE, H. and SCHORNING, K. (2013). Complete classes of designs for nonlinear regression models and principal representations of moment spaces. *Ann. Statist.* **41** 1260–1267. [MR3113810](#) <https://doi.org/10.1214/13-AOS1108>
- GRASSHOFF, U., GROSSMANN, H., HOLLING, H. and SCHWABE, R. (2007). Design optimality in multi-factor generalized linear models in the presence of an unrestricted quantitative factor. *J. Statist. Plann. Inference* **137** 3882–3893. [MR2368534](#) <https://doi.org/10.1016/j.jspi.2007.04.007>
- GRIGORIEV, Y. D., MELAS, V. B. and SHPILEV, P. V. (2018). Excess of locally D -optimal designs for Cobb-Douglas model. *Statist. Papers* **59** 1425–1439. [MR3878740](#) <https://doi.org/10.1007/s00362-018-1022-0>
- HAINES, L. M., KABERA, G. M. and NDLOVU, P. (2018). D -optimal designs for the two-variable binary logistic regression model with interaction. *J. Statist. Plann. Inference* **193** 136–150. [MR3713469](#) <https://doi.org/10.1016/j.jspi.2017.08.007>

MSC2020 subject classifications. Primary 62K05; secondary 62J05.

Key words and phrases. Admissibility, dual problem, conditional model, multifactor experiment, optimal design.

- HEILIGERS, B. (1992). Admissible experimental designs in multiple polynomial regression. *J. Statist. Plann. Inference* **31** 219–233. [MR1173547](#) [https://doi.org/10.1016/0378-3758\(92\)90031-M](https://doi.org/10.1016/0378-3758(92)90031-M)
- HU, L., YANG, M. and STUFKEN, J. (2015). Saturated locally optimal designs under differentiable optimality criteria. *Ann. Statist.* **43** 30–56. [MR3285599](#) <https://doi.org/10.1214/14-AOS1263>
- HUANG, S.-H., LO HUANG, M.-N. and LIN, C.-W. (2020). Optimal designs for binary response models with multiple nonnegative variables. *J. Statist. Plann. Inference* **206** 75–83. [MR4036696](#) <https://doi.org/10.1016/j.jspi.2019.09.006>
- KARLIN, S. and STUDDEN, W. J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics. Pure and Applied Mathematics, Vol. XV*. Interscience Publishers Wiley, New York. [MR0204922](#)
- KIEFER, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* **2** 849–879. [MR0356386](#)
- LANGSETH, H., NIELSEN, T. D., RUMÍ, R. and SALMERÓN, A. (2012). Mixtures of truncated basis functions. *Internat. J. Approx. Reason.* **53** 212–227. [MR2872909](#) <https://doi.org/10.1016/j.ijar.2011.10.004>
- MASOUDI, E., HOLLING, H. and WONG, W. K. (2017). Application of imperialist competitive algorithm to find minimax and standardized maximin optimal designs. *Comput. Statist. Data Anal.* **113** 330–345. [MR3662411](#) <https://doi.org/10.1016/j.csda.2016.06.014>
- PATAN, M. (2004). *Optimal Observation Strategies for Parameter Estimation of Distributed Systems. Lecture Notes in Control and Computer Science* **5**. Univ. Zielona Góra Press, Zielona Góra. [MR2119646](#)
- PUKELSHEIM, F. (2006). *Optimal Design of Experiments. Classics in Applied Mathematics* **50**. SIAM, Philadelphia, PA. [MR2224698](#) <https://doi.org/10.1137/1.9780898719109>
- RODRÍGUEZ, C., ORTIZ, I. and MARTÍNEZ, I. (2015). Locally and maximin optimal designs for multi-factor nonlinear models. *Statistics* **49** 1157–1168. [MR3378036](#) <https://doi.org/10.1080/02331888.2014.922562>
- SCHWABE, R. (1996). *Optimum Designs for Multi-Factor Models. Lecture Notes in Statistics* **113**. Springer, New York. [MR1479410](#) <https://doi.org/10.1007/978-1-4612-4038-9>
- YANG, M. (2010). On the de la Garza phenomenon. *Ann. Statist.* **38** 2499–2524. [MR2676896](#) <https://doi.org/10.1214/09-AOS787>
- YANG, M. and STUFKEN, J. (2009). Support points of locally optimal designs for nonlinear models with two parameters. *Ann. Statist.* **37** 518–541. [MR2488361](#) <https://doi.org/10.1214/07-AOS560>
- YANG, M. and STUFKEN, J. (2012). Identifying locally optimal designs for nonlinear models: A simple extension with profound consequences. *Ann. Statist.* **40** 1665–1681. [MR3015039](#) <https://doi.org/10.1214/12-AOS992>
- YANG, M., ZHANG, B. and HUANG, S. (2011). Optimal designs for generalized linear models with multiple design variables. *Statist. Sinica* **21** 1415–1430. [MR2827529](#) <https://doi.org/10.5705/ss.2009.115>
- ZHANG, Z. Z., WONG, W. K. and TAN, K. C. (2020). Competitive swarm optimizer with mutated agents for finding optimal designs for nonlinear regression models with multiple interacting factors. *Memet. Comput.* **12** 219–233.

EVIDENCE FACTORS FROM MULTIPLE, POSSIBLY INVALID, INSTRUMENTAL VARIABLES

BY ANQI ZHAO^{1,a}, YOUJIN LEE^{2,b}, DYLAN S. SMALL^{3,c} AND BIKRAM KARMAKAR^{4,d}

¹Department of Statistics and Data Science, National University of Singapore, ^astaza@nus.edu.sg

²Department of Biostatistics, Brown University, ^byoujin_lee@brown.edu

³Department of Statistics and Data Science, University of Pennsylvania, ^cdsmall@wharton.upenn.edu

⁴Department of Statistics, University of Florida, ^dbkarmakar@ufl.edu

Valid instrumental variables enable treatment effect inference even when selection into treatment is biased by unobserved confounders. When multiple candidate instruments are available, but some of them are possibly invalid, the previously proposed reinforced design enables one or more nearly independent valid analyses that depend on very different assumptions. That is, we can perform evidence factor analysis. However, the validity of the reinforced design depends crucially on the order in which multiple instrumental variable analyses are conducted. Motivated by the orthogonality of balanced factorial designs, we propose a balanced block design to offset the possible violation of the exclusion restriction by balancing the instruments against each other in the design, and demonstrate its utility for constructing approximate evidence factors under multiple analysis strategies free of the order imposition. We also propose a novel stratification method using multiple, nested candidate instruments, in which case the balanced block design is not applicable. We apply our proposed methods to evaluate (a) the effect of education on future earnings using instrumental variables arising from the disruption of education during World War II via the balanced block design, and (b) the causal effect of malaria on stunting among children in Western Kenya using three nested instruments.

REFERENCES

- SOCIO-ECONOMIC PANEL (SOEP) (2018). Data for years 1984–2018, v35i, SOEP.
- ANGRIST, J. D., IMBENS, G. W. and RUBIN, D. B. (1996). Identification of causal effects using instrumental variables. *J. Amer. Statist. Assoc.* **91** 444–455.
- ANGRIST, J. D. and KRUEGER, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *Q. J. Econ.* **106** 979–1014.
- ATEBA, F. F., DOUMBIA, S., TER KUILE, F. O., TERLOUW, D. J., LEFEBVRE, G., KARIUKI, S. and SMALL, D. S. (2021). The effect of malaria on stunting: An instrumental variables approach. *Trans. R. Soc. Trop. Med. Hyg.* In press.
- BECKER, B. J. (1994). Combining significance levels. In *The Handbook of Research Synthesis* (H. Cooper and L. V. Hedges, eds.) 215–230.
- BENJAMINI, Y. and HELLER, R. (2008). Screening for partial conjunction hypotheses. *Biometrics* **64** 1215–1222. [MR2522270 https://doi.org/10.1111/j.1541-0420.2007.00984.x](https://doi.org/10.1111/j.1541-0420.2007.00984.x)
- BOWDEN, J., DAVEY SMITH, G. and BURGESS, S. (2015). Mendelian randomization with invalid instruments: Effect estimation and bias detection through egger regression. *Int. J. Epidemiol.* **44** 512–525.
- BOWDEN, J., DAVEY SMITH, G., HAYCOCK, P. C. and BURGESS, S. (2016). Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40** 304–314.
- BRESLOW, N. (1970). A generalized Kruskal–Wallis test for comparing K samples subject to unequal patterns of censorship. *Biometrika* **57** 579–594.
- BURGESS, S., BUTTERWORTH, A. and THOMPSON, S. G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37** 658–665.

MSC2020 subject classifications. Primary 62G10; secondary 62K10.

Key words and phrases. Bias in observational studies, causal inference, exclusion restriction, nonparametric tests, replication, sensitivity analysis.

- BURGESS, S., DUDBRIDGE, F. and THOMPSON, S. G. (2016). Combining information on multiple instrumental variables in Mendelian randomization: Comparison of allele score and summarized data methods. *Stat. Med.* **35** 1880–1906. MR3513491 <https://doi.org/10.1002/sim.6835>
- BURGESS, S., BOWDEN, J., FALL, T., INGELSSON, E. and THOMPSON, S. G. (2017). Sensitivity analyses for robust causal inference from Mendelian randomization analyses with multiple genetic variants. *Epidemiology* **28** 30–42. <https://doi.org/10.1097/EDE.0000000000000559>
- CARD, D. (1993). Using geographic variation in college proximity to estimate the return to schooling. NBER Working Paper w4483.
- CARD, D. (1999). The causal effect of education on earnings (O. C. Ashenfelter and D. Card, eds.). *Handbook of Labor Economics* **3** 1801–1863. Elsevier, Amsterdam.
- CAUGHEY, D., DAFOE, A., LI, X. and MIRATRIX, L. (2021). Randomization inference beyond the sharp null: Bounded null hypotheses and quantiles of individual treatment effects. ArXiv preprint. Available at [arXiv:2101.09195](https://arxiv.org/abs/2101.09195).
- DAVIES, N. M., VON HINKE KESSLER SCHOLDER, S., FARBMACHER, H., BURGESS, S., WINDMEIJER, F. and DAVEY SMITH, G. (2015). The many weak instruments problem and Mendelian randomization. *Stat. Med.* **34** 454–468. MR3301588 <https://doi.org/10.1002/sim.6358>
- DEATON, A. (2010). Instruments, randomization, and learning about development. *J. Econ. Lit.* **48** 424–55.
- DEL GRECO M., F., MINELLI, C., SHEEHAN, N. A. and THOMPSON, J. R. (2015). Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Stat. Med.* **34** 2926–2940. MR3375990 <https://doi.org/10.1002/sim.6522>
- FISHER, R. A. (1926). The arrangement of field experiments. *J. Minist. Agric. G. B.* **33** 503–513.
- FOGARTY, C. B. and HASEGAWA, R. B. (2019). Extended sensitivity analysis for heterogeneous unmeasured confounding with an application to sibling studies of returns to education. *Ann. Appl. Stat.* **13** 767–796. MR3963552 <https://doi.org/10.1214/18-AOAS1215>
- GAIL, M. H., TAN, W. Y. and PIANTADOSI, S. (1988). Tests for no treatment effect in randomized clinical trials. *Biometrika* **75** 57–64. MR0932817 <https://doi.org/10.1093/biomet/75.1.57>
- GUO, Z., KANG, H., CAI, T. T. and SMALL, D. S. (2018). Confidence intervals for causal effects with invalid instruments by using two-stage hard thresholding with voting. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 793–815. MR3849344 <https://doi.org/10.1111/rssb.12275>
- HADLEY, J., POLSKY, D., MANDELBLATT, J. S., MITCHELL, J. M., WEEKS, J. C., WANG, Q., HWANG, Y.-T. and TEAM, O. R. (2003). An exploratory instrumental variable analysis of the outcomes of localized breast cancer treatments in a medicare population. *Health Econ.* **12** 171–186.
- HALSEY, A. H., HALSEY, A. H., ALBERT HENRY, H., HEATH, A. F., RIDGE, J. M. et al. (1980). *Origins and Destinations: Family, Class, and Education in Modern Britain*. Clarendon Press, Oxford; Oxford Univ. Press. New York.
- HAN, C. (2008). Detecting invalid instruments using L_1 -GMM. *Econom. Lett.* **101** 285–287. MR2477476 <https://doi.org/10.1016/j.econlet.2008.09.004>
- HARMON, C. and WALKER, I. (1995). Estimates of the economic return to schooling for the United Kingdom. *Am. Econ. Rev.* **85** 1278–1286.
- HASEGAWA, R. and SMALL, D. (2017). Sensitivity analysis for matched pair analysis of binary data: From worst case to average case analysis. *Biometrics* **73** 1424–1432. MR3744554 <https://doi.org/10.1111/biom.12688>
- HENG, S. and SMALL, D. S. (2021). Sharpening the Rosenbaum sensitivity bounds to address concerns about interactions between observed and unobserved covariates. *Statist. Sinica* **31** 2331–2353. MR4338087 <https://doi.org/10.1007/s12220-019-00351-8>
- HENG, S., SMALL, D. S. and ROSENBAUM, P. R. (2020). Finding the strength in a weak instrument in a study of cognitive outcomes produced by catholic high schools. *J. Roy. Statist. Soc. Ser. A* **183** 935–958.
- HSU, J. Y., SMALL, D. S. and ROSENBAUM, P. R. (2013). Effect modification and design sensitivity in observational studies. *J. Amer. Statist. Assoc.* **108** 135–148. MR3174608 <https://doi.org/10.1080/01621459.2012.742018>
- ICHINO, A. and WINTER-EBMER, R. (2004). The long-run educational cost of World War II. *J. Labor Econ.* **22** 57–87.
- IMBENS, G. W. (2003). Sensitivity to exogeneity assumptions in program evaluation. *Am. Econ. Rev.* **93** 126–132.
- IMBENS, G. W. (2010). Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). *J. Econ. Lit.* **48** 399–423.
- JACKSON, B. D. and BLACK, R. E. (2017). A literature review of the effect of malaria on stunting. *J. Nutr.* **147** 2163S–2168S. <https://doi.org/10.3945/jn.116.242289>
- KANG, H., ZHANG, A., CAI, T. T. and SMALL, D. S. (2016). Instrumental variables estimation with some invalid instruments and its application to Mendelian randomization. *J. Amer. Statist. Assoc.* **111** 132–144. MR3494648 <https://doi.org/10.1080/01621459.2014.994705>

- KANG, H., LEE, Y., CAI, T. T. and SMALL, D. S. (2021). Two robust tools for inference about causal effects with invalid instruments. *Biometrics*. (In press).
- KARMAKAR, B., DOUBENI, C. A. and SMALL, D. S. (2020). Evidence factors in a case-control study with application to the effect of flexible sigmoidoscopy screening on colorectal cancer. *Ann. Appl. Stat.* **14** 829–849. [MR4118331 https://doi.org/10.1214/20-AOAS1329](https://doi.org/10.1214/20-AOAS1329)
- KARMAKAR, B., FRENCH, B. and SMALL, D. S. (2019). Integrating the evidence from evidence factors in observational studies. *Biometrika* **106** 353–367. [MR3949308 https://doi.org/10.1093/biomet/asz003](https://doi.org/10.1093/biomet/asz003)
- KARMAKAR, B. and SMALL, D. S. (2020). Assessment of the extent of corroboration of an elaborate theory of a causal hypothesis using partial conjunctions of evidence factors. *Ann. Statist.* **48** 3283–3311. [MR4185809 https://doi.org/10.1214/19-AOS1929](https://doi.org/10.1214/19-AOS1929)
- KARMAKAR, B., SMALL, D. S. and ROSENBAUM, P. R. (2020). Using evidence factors to clarify exposure biomarkers. *Am. J. Epidemiol.* **189** 243–249.
- KARMAKAR, B., SMALL, D. S. and ROSENBAUM, P. R. (2021b). Reinforced designs: Multiple instruments plus control groups as evidence factors in an observational study of the effectiveness of Catholic schools. *J. Amer. Statist. Assoc.* **116** 82–92. [MR4227676 https://doi.org/10.1080/01621459.2020.1745811](https://doi.org/10.1080/01621459.2020.1745811)
- KOLESÁR, M., CHETTY, R., FRIEDMAN, J., GLAESER, E. and IMBENS, G. W. (2015). Identification and inference with many invalid instruments. *J. Bus. Econom. Statist.* **33** 474–484. [MR3416595 https://doi.org/10.1080/07350015.2014.978175](https://doi.org/10.1080/07350015.2014.978175)
- LORCH, S. A., BAIOCCHI, M., AHLBERG, C. E. and SMALL, D. S. (2012). The differential impact of delivery hospital on the outcomes of premature infants. *Pediatrics* **130** 270–278.
- NATTINO, G., LU, B., SHI, J., LEMESHOW, S. and XIANG, H. (2021). Triplet matching for estimating causal effects with three treatment arms: A comparative study of mortality by trauma center level. *J. Amer. Statist. Assoc.* **116** 44–53. [MR4227673 https://doi.org/10.1080/01621459.2020.1737078](https://doi.org/10.1080/01621459.2020.1737078)
- ROSENBAUM, P. R. (2001). Replicating effects and biases. *Amer. Statist.* **55** 223–227. [MR1963397 https://doi.org/10.1198/000313001317098220](https://doi.org/10.1198/000313001317098220)
- 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](https://doi.org/10.1214/ss/1042727942)
- ROSENBAUM, P. R. (2010a). *Design of Observational Studies*. Springer Series in Statistics. Springer, New York. [MR2561612 https://doi.org/10.1007/978-1-4419-1213-8](https://doi.org/10.1007/978-1-4419-1213-8)
- ROSENBAUM, P. R. (2010b). Evidence factors in observational studies. *Biometrika* **97** 333–345. [MR2650742 https://doi.org/10.1093/biomet/asq019](https://doi.org/10.1093/biomet/asq019)
- ROSENBAUM, P. R. (2011). Some approximate evidence factors in observational studies. *J. Amer. Statist. Assoc.* **106** 285–295. [MR2816721 https://doi.org/10.1198/jasa.2011.tm10422](https://doi.org/10.1198/jasa.2011.tm10422)
- ROSENBAUM, P. R. (2015). Two R packages for sensitivity analysis in observational studies. *Obs. Stud.* **1** 1–17.
- ROSENBAUM, P. R. (2017). The general structure of evidence factors in observational studies. *Statist. Sci.* **32** 514–530. [MR3730520 https://doi.org/10.1214/17-STS621](https://doi.org/10.1214/17-STS621)
- ROSENBAUM, P. R., ROSS, R. N. and SILBER, J. H. (2007). Minimum distance matched sampling with fine balance in an observational study of treatment for ovarian cancer. *J. Amer. Statist. Assoc.* **102** 75–83. [MR2345534 https://doi.org/10.1198/016214506000001059](https://doi.org/10.1198/016214506000001059)
- RUBIN, D. B. (1979). Using multivariate matched sampling and regression adjustment to control bias in observational studies. *J. Amer. Statist. Assoc.* **74** 318–328.
- SANDER, W. (1995). Schooling and quitting smoking. *Rev. Econ. Stat.* 191–199.
- SIMES, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73** 751–754. [MR0897872 https://doi.org/10.1093/biomet/73.3.751](https://doi.org/10.1093/biomet/73.3.751)
- SMALL, D. S. (2007). Sensitivity analysis for instrumental variables regression with overidentifying restrictions. *J. Amer. Statist. Assoc.* **102** 1049–1058. [MR2411664 https://doi.org/10.1198/016214507000000608](https://doi.org/10.1198/016214507000000608)
- SMALL, D. S. and ROSENBAUM, P. R. (2008). War and wages: The strength of instrumental variables and their sensitivity to unobserved biases. *J. Amer. Statist. Assoc.* **103** 924–933. [MR2528819 https://doi.org/10.1198/016214507000001247](https://doi.org/10.1198/016214507000001247)
- SPIEKER, A. J., GREEVY, R. A., NELSON, L. A. and MAYBERRY, L. S. (2020). Bounding the local average treatment effect in an instrumental variable analysis of engagement with a mobile intervention. ArXiv preprint. Available at [arXiv:2008.06473](https://arxiv.org/abs/2008.06473).
- TAN, Z. (2006). Regression and weighting methods for causal inference using instrumental variables. *J. Amer. Statist. Assoc.* **101** 1607–1618. [MR2279483 https://doi.org/10.1198/016214505000001366](https://doi.org/10.1198/016214505000001366)
- TIAN, J. and RAMDAS, A. (2019). ADDIS: An adaptive discarding algorithm for online FDR control with conservative nulls. In *Advances in Neural Information Processing Systems* 9388–9396.
- TUKEY, J. W. (1993). Tightening the clinical trial. *Control. Clin. Trials* **14** 266–285.
- VOORS, M. J., NILLESEN, E. E., VERWIMP, P., BULTE, E. H., LENSINK, R. and VAN SOEST, D. P. (2012). Violent conflict and behavior: A field experiment in Burundi. *Am. Econ. Rev.* **102** 941–64.

- WAGNER, G. G., FRICK, J. R. and SCHUPP, J. (2007). The German Socio-Economic Panel study (SOEP)- evolution, scope and enhancements. SOEP papers on Multidisciplinary Panel Data Research No. 1 DIW Berlin, The German Socio-Economic Panel (SOEP).
- WALKER, V. M., DAVIES, N. M., MARTIN, R. M. and KEHOE, P. G. (2020). Comparison of antihypertensive drug classes for dementia prevention. *Epidemiology* **31** 852.
- WANG, X., JIANG, Y., ZHANG, N. R. and SMALL, D. S. (2018). Sensitivity analysis and power for instrumental variable studies. *Biometrics* **74** 1150–1160. [MR3908133 https://doi.org/10.1111/biom.12873](https://doi.org/10.1111/biom.12873)
- WILSON, A. L., DHIMAN, R. C., KITRON, U., SCOTT, T. W., VAN DEN BERG, H. and LINDSAY, S. W. (2014). Benefit of insecticide-treated nets, curtains and screening on vector borne diseases, excluding malaria: A systematic review and meta-analysis. *PLoS Negl. Trop. Dis.* **8** e3228.
- WINDMEIJER, F., FARBMACHER, H., DAVIES, N. and SMITH, G. D. (2019). On the use of the Lasso for instrumental variables estimation with some invalid instruments. *J. Amer. Statist. Assoc.* **114** 1339–1350. [MR4011783 https://doi.org/10.1080/01621459.2018.1498346](https://doi.org/10.1080/01621459.2018.1498346)
- WU, J. and DING, P. (2021). Randomization Tests for Weak Null Hypotheses in Randomized Experiments. *J. Amer. Statist. Assoc.* **116** 1898–1913. [MR4353721 https://doi.org/10.1080/01621459.2020.1750415](https://doi.org/10.1080/01621459.2020.1750415)
- ZAYKIN, D. V., ZHIVOTOVSKY, L. A., WESTFALL, P. H. and WEIR, B. S. (2002). Truncated product method for combining P-values. *Genet. Epidemiol.* **22** 170–185.
- ZENG, S., LI, F. and DING, P. (2020). Is being the only child harmful to psychological health?: Evidence from an instrumental variable analysis of China's One-Child Policy. *J. Roy. Statist. Soc. Ser. A* **183** 1615–1635.
- ZENG, D., THOMSEN, M. R., NAYGA JR, R. M. and ROUSE, H. L. (2019). Neighbourhood convenience stores and childhood weight outcomes: An instrumental variable approach. *Appl. Econ.* **51** 288–302.
- ZHANG, K., SMALL, D. S., LORCH, S., SRINIVAS, S. and ROSENBAUM, P. R. (2011). Using split samples and evidence factors in an observational study of neonatal outcomes. *J. Amer. Statist. Assoc.* **106** 511–524. [MR2847966 https://doi.org/10.1198/jasa.2011.ap10604](https://doi.org/10.1198/jasa.2011.ap10604)
- ZHAO, Q., SMALL, D. S. and SU, W. (2019). Multiple testing when many p -values are uniformly conservative, with application to testing qualitative interaction in educational interventions. *J. Amer. Statist. Assoc.* **114** 1291–1304. [MR4011780 https://doi.org/10.1080/01621459.2018.1497499](https://doi.org/10.1080/01621459.2018.1497499)
- ZHAO, Q., LEE, Y., SMALL, D. S. and KARMAKAR, B. (2022). Supplement to “Evidence factors from multiple, possibly invalid, instrumental variables.” <https://doi.org/10.1214/21-AOS2148SUPP>
- ZUBIZARRETA, J. R., NEUMAN, M., SILBER, J. H. and ROSENBAUM, P. R. (2012). Contrasting evidence within and between institutions that provide treatment in an observational study of alternative forms of anesthesia. *J. Amer. Statist. Assoc.* **107** 901–915. [MR3010879 https://doi.org/10.1080/01621459.2012.682533](https://doi.org/10.1080/01621459.2012.682533)

CONSISTENT ORDER SELECTION FOR ARFIMA PROCESSES

BY HSUEH-HAN HUANG^{1,a}, NGAI HANG CHAN^{2,c}, KUN CHEN^{3,d} AND
 CHING-KANG ING^{1,b}

¹*Institute of Statistics, National Tsing Hua University, ^alink49177@gmail.com, ^bcking@stat.nthu.edu.tw*

²*Department of Statistics, The Chinese University of Hong Kong, ^cnhchan@sta.cuhk.edu.hk*

³*School of Statistics, Southwestern University of Finance and Economics, ^dchenkun@swufe.edu.cn*

Estimating the orders of the autoregressive fractionally integrated moving average (ARFIMA) model has been a long-standing problem in time series analysis. This paper tackles this challenge by establishing the consistency of the Bayesian information criterion (BIC) for ARFIMA models with independent errors. Since the memory parameter of the model can be any real number, this consistency result is valid for short memory, long memory and nonstationary time series. This paper further extends the consistency of the BIC to ARFIMA models with conditional heteroscedastic errors, thereby extending its applications to encompass many real-life situations. Finite-sample implications of the theoretical results are illustrated via numerical examples.

REFERENCES

- BAILLIE, R. T., KAPETANIOS, G. and PAPAILIAS, F. (2014). Modified information criteria and selection of long memory time series models. *Comput. Statist. Data Anal.* **76** 116–131. MR3209430 <https://doi.org/10.1016/j.csda.2013.04.012>
- BARDET, J.-M., KAMILA, K. and KENGNE, W. (2020). Consistent model selection criteria and goodness-of-fit test for common time series models. *Electron. J. Stat.* **14** 2009–2052. MR4091861 <https://doi.org/10.1214/20-EJS1709>
- BERAN, J. (1994). *Statistics for Long-Memory Processes. Monographs on Statistics and Applied Probability* **61**. CRC Press, New York. MR1304490
- BERAN, J., BHANSALI, R. J. and OCKER, D. (1998). On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes. *Biometrika* **85** 921–934. MR1666719 <https://doi.org/10.1093/biomet/85.4.921>
- BOLLERSLEV, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31** 307–327. MR0853051 [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)
- CHAN, N. H., HUANG, S.-F. and ING, C.-K. (2013). Moment bounds and mean squared prediction errors of long-memory time series. *Ann. Statist.* **41** 1268–1298. MR3113811 <https://doi.org/10.1214/13-AOS1110>
- CHAN, N. H. and ING, C.-K. (2011). Uniform moment bounds of Fisher’s information with applications to time series. *Ann. Statist.* **39** 1526–1550. MR2850211 <https://doi.org/10.1214/10-AOS861>
- GLOSTEN, L. R., JAGANNATHAN, R. and RUNKLE, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *J. Finance* **48** 1779–1801. <https://doi.org/10.1111/j.1540-6261.1993.tb05128.x>
- GUO, L., CHEN, H. F. and ZHANG, J. F. (1989). Consistent order estimation for linear stochastic feedback control systems (CARMA model). *Automatica J. IFAC* **25** 147–151. MR0986583 [https://doi.org/10.1016/0005-1098\(89\)90131-3](https://doi.org/10.1016/0005-1098(89)90131-3)
- HANNAN, E. J. (1980). The estimation of the order of an ARMA process. *Ann. Statist.* **8** 1071–1081. MR0585705
- HANNAN, E. J. and KAVALLIERIS, L. (1984). A method for autoregressive-moving average estimation. *Biometrika* **71** 273–280. MR0767155 <https://doi.org/10.1093/biomet/71.2.273>
- HANNAN, E. J. and RISSANEN, J. (1982). Recursive estimation of mixed autoregressive-moving average order. *Biometrika* **69** 81–94. MR0655673 <https://doi.org/10.1093/biomet/69.1.81>
- HOSOYA, Y. (2005). Fractional invariance principle. *J. Time Series Anal.* **26** 463–486. MR2163292 <https://doi.org/10.1111/j.1467-9892.2004.00411.x>

MSC2020 subject classifications. Primary 62M10; secondary 62M20.

Key words and phrases. ARFIMA models, Bayesian information criterion, conditional heteroscedastic errors, long memory and nonstationary time series, order selection.

- HUALDE, J. and ROBINSON, P. M. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models. *Ann. Statist.* **39** 3152–3181. [MR3012404](#) <https://doi.org/10.1214/11-AOS931>
- HUANG, D. W. and GUO, L. (1990). Estimation of nonstationary ARMAX models based on the Hannan-Rissanen method. *Ann. Statist.* **18** 1729–1756. [MR1074432](#) <https://doi.org/10.1214/aos/1176347875>
- HUANG, H.-H., CHAN, N. H., CHEN, K. and ING, C.-K. (2022). Supplement to “Consistent order selection for ARFIMA processes.” <https://doi.org/10.1214/21-AOS2149SUPP>
- ING, C.-K., SIN, C. and YU, S.-H. (2012). Model selection for integrated autoregressive processes of infinite order. *J. Multivariate Anal.* **106** 57–71. [MR2887680](#) <https://doi.org/10.1016/j.jmva.2011.10.008>
- ING, C.-K. and WEI, C.-Z. (2005). Order selection for same-realization predictions in autoregressive processes. *Ann. Statist.* **33** 2423–2474. [MR2211091](#) <https://doi.org/10.1214/009053605000000525>
- LING, S. (2007). Testing for change points in time series models and limiting theorems for NED sequences. *Ann. Statist.* **35** 1213–1237. [MR2341704](#) <https://doi.org/10.1214/009053606000001514>
- LING, S. and MCALEER, M. (2002a). Necessary and sufficient moment conditions for the GARCH(r, s) and asymmetric power GARCH(r, s) models. *Econometric Theory* **18** 722–729. [MR1906332](#) <https://doi.org/10.1017/S0266466602183071>
- LING, S. and MCALEER, M. (2002b). Stationarity and the existence of moments of a family of GARCH processes. *J. Econometrics* **106** 109–117. [MR1875529](#) [https://doi.org/10.1016/S0304-4076\(01\)00090-2](https://doi.org/10.1016/S0304-4076(01)00090-2)
- PAULSEN, J. (1984). Order determination of multivariate autoregressive time series with unit roots. *J. Time Series Anal.* **5** 115–127. [MR0758581](#) <https://doi.org/10.1111/j.1467-9892.1984.tb00381.x>
- SHIBATA, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statist.* **8** 147–164. [MR0557560](#)
- TSAY, R. S. (1984). Order selection in nonstationary autoregressive models. *Ann. Statist.* **12** 1425–1433. [MR0760697](#) <https://doi.org/10.1214/aos/1176346801>
- WEI, C. Z. (1992). On predictive least squares principles. *Ann. Statist.* **20** 1–42. [MR1150333](#) <https://doi.org/10.1214/aos/1176348511>

DOUBLY DEBIASED LASSO: HIGH-DIMENSIONAL INFERENCE UNDER HIDDEN CONFOUNDING

BY ZIJIAN GUO^{1,a}, DOMAGOJ ČEVID^{2,b} AND PETER BÜHLMANN^{2,c}

¹Department of Statistics, Rutgers University, ^azijguo@stat.rutgers.edu

²Seminar for Statistics, ETH Zürich, ^bcevid@stat.math.ethz.ch, ^cbuhlmann@stat.math.ethz.ch

Inferring causal relationships or related associations from observational data can be invalidated by the existence of hidden confounding. We focus on a high-dimensional linear regression setting, where the measured covariates are affected by hidden confounding and propose the *doubly debiased lasso* estimator for individual components of the regression coefficient vector. Our advocated method simultaneously corrects both the bias due to estimation of high-dimensional parameters as well as the bias caused by the hidden confounding. We establish its asymptotic normality and also prove that it is efficient in the Gauss–Markov sense. The validity of our methodology relies on a dense confounding assumption, that is, that every confounding variable affects many covariates. The finite sample performance is illustrated with an extensive simulation study and a genomic application. The method is implemented by the DDL package available from CRAN.

REFERENCES

- [1] BAI, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* **71** 135–171. [MR1956857](#) <https://doi.org/10.1111/1468-0262.00392>
- [2] BAI, J. and NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70** 191–221. [MR1926259](#) <https://doi.org/10.1111/1468-0262.00273>
- [3] BELLONI, A., CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. and HANSEN, C. (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica* **85** 233–298. [MR3611771](#) <https://doi.org/10.3982/ECTA12723>
- [4] BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Rev. Econ. Stud.* **81** 608–650. [MR3207983](#) <https://doi.org/10.1093/restud/rdt044>
- [5] BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. [MR2533469](#) <https://doi.org/10.1214/08-AOS620>
- [6] BOEF, A. G., DEKKERS, O. M., VANDENBROUCKE, J. P. and LE CESSIE, S. (2014). Sample size importantly limits the usefulness of instrumental variable methods, depending on instrument strength and level of confounding. *J. Clin. Epidemiol.* **67** 1258–1264.
- [7] BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer Series in Statistics. Springer, Heidelberg. [MR2807761](#) <https://doi.org/10.1007/978-3-642-20192-9>
- [8] BURGESS, S., SMALL, D. S. and THOMPSON, S. G. (2017). A review of instrumental variable estimators for Mendelian randomization. *Stat. Methods Med. Res.* **26** 2333–2355. [MR3712236](#) <https://doi.org/10.1177/0962280215597579>
- [9] CAI, T., LIU, W. and LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.* **106** 594–607. [MR2847973](#) <https://doi.org/10.1198/jasa.2011.tm10155>
- [10] CAI, T. T. and GUO, Z. (2017). Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *Ann. Statist.* **45** 615–646. [MR3650395](#) <https://doi.org/10.1214/16-AOS1461>
- [11] CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A. and CRAINICEANU, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd ed. *Monographs on Statistics and Applied Probability* **105**. CRC Press/CRC, Boca Raton, FL. [MR2243417](#) <https://doi.org/10.1201/9781420010138>

MSC2020 subject classifications. Primary 62E20, 62F12; secondary 62J07.

Key words and phrases. Causal inference, structural equation model, dense confounding, linear model, spectral deconfounding.

- [12] ČEVID, D., BÜHLMANN, P. and MEINSHAUSEN, N. (2020). Spectral deconfounding via perturbed sparse linear models. *J. Mach. Learn. Res.* **21** Paper No. 232, 41. MR4209518 <https://doi.org/10.22405/2226-8383-2020-21-1-221-232>
- [13] CHANDRASEKARAN, V., PARRILO, P. A. and WILLSKY, A. S. (2012). Latent variable graphical model selection via convex optimization. *Ann. Statist.* **40** 1935–1967. MR3059067 <https://doi.org/10.1214/11-AOS949>
- [14] CHERNOZHUKOV, V., CHETVERIKOV, D., DEMIRER, M., DUFLO, E., HANSEN, C., NEWEY, W. and ROBINS, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econom. J.* **21** C1–C68. MR3769544 <https://doi.org/10.1111/ectj.12097>
- [15] CHERNOZHUKOV, V., HANSEN, C. and LIAO, Y. (2017). A lava attack on the recovery of sums of dense and sparse signals. *Ann. Statist.* **45** 39–76. MR3611486 <https://doi.org/10.1214/16-AOS1434>
- [16] CHERNOZHUKOV, V., HANSEN, C. and SPINDLER, M. (2015). Valid post-selection and post-regularization inference: An elementary, general approach. *Ann. Rev. Econ.* **7** 649–688.
- [17] DEZEURE, R., BÜHLMANN, P. and ZHANG, C.-H. (2017). High-dimensional simultaneous inference with the bootstrap. *TEST* **26** 685–719. MR3713586 <https://doi.org/10.1007/s11749-017-0554-2>
- [18] FAN, J., FAN, Y. and LV, J. (2008). High dimensional covariance matrix estimation using a factor model. *J. Econometrics* **147** 186–197. MR2472991 <https://doi.org/10.1016/j.jeconom.2008.09.017>
- [19] FAN, J. and LIAO, Y. (2014). Endogeneity in high dimensions. *Ann. Statist.* **42** 872–917. MR3210990 <https://doi.org/10.1214/13-AOS1202>
- [20] FAN, J., LIAO, Y. and MINCHEVA, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 603–680. With 33 discussions by 57 authors and a reply by Fan, Liao and Mincheva. MR3091653 <https://doi.org/10.1111/rssb.12016>
- [21] FAN, J., LIAO, Y. and WANG, W. (2016). Projected principal component analysis in factor models. *Ann. Statist.* **44** 219–254. MR3449767 <https://doi.org/10.1214/15-AOS1364>
- [22] FARRELL, M. H. (2015). Robust inference on average treatment effects with possibly more covariates than observations. *J. Econometrics* **189** 1–23. MR3397349 <https://doi.org/10.1016/j.jeconom.2015.06.017>
- [23] GAGNON-BARTSCH, J. A. and SPEED, T. P. (2012). Using control genes to correct for unwanted variation in microarray data. *Biostatistics* **13** 539–552.
- [24] GAUTIER, E. and ROSE, C. (2011). High-dimensional instrumental variables regression and confidence sets. ArXiv preprint. Available at [arXiv:1105.2454](https://arxiv.org/abs/1105.2454).
- [25] GERARD, D. and STEPHENS, M. (2020). Empirical Bayes shrinkage and false discovery rate estimation, allowing for unwanted variation. *Biostatistics* **21** 15–32. MR4043843 <https://doi.org/10.1093/biostatistics/kxy029>
- [26] GOLD, D., LEDERER, J. and TAO, J. (2020). Inference for high-dimensional instrumental variables regression. *J. Econometrics* **217** 79–111. MR4093746 <https://doi.org/10.1016/j.jeconom.2019.09.009>
- [27] GÖTZE, F. and TIKHOMIROV, A. (2002). Asymptotic distribution of quadratic forms and applications. *J. Theoret. Probab.* **15** 423–475. MR1898815 <https://doi.org/10.1023/A:1014867011101>
- [28] GUERTIN, J. R., RAHME, E. and LELORIER, J. (2016). Performance of the high-dimensional propensity score in adjusting for unmeasured confounders. *Eur. J. Clin. Pharmacol.* **72** 1497–1505. <https://doi.org/10.1007/s00228-016-2118-x>
- [29] GUO, Z., ČEVID, D. and BÜHLMANN, P. (2022). Supplement to “Doubly debiased lasso: High-dimensional inference under hidden confounding.” <https://doi.org/10.1214/21-AOS2152SUPP>
- [30] GUO, Z., KANG, H., CAI, T. T. and SMALL, D. S. (2018). Confidence intervals for causal effects with invalid instruments by using two-stage hard thresholding with voting. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 793–815. MR3849344 <https://doi.org/10.1111/rssb.12275>
- [31] 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>
- [32] HAN, C. (2008). Detecting invalid instruments using L_1 -GMM. *Econom. Lett.* **101** 285–287. MR2477476 <https://doi.org/10.1016/j.econlet.2008.09.004>
- [33] JANKOVÁ, J. and VAN DE GEER, S. (2018). Semiparametric efficiency bounds for high-dimensional models. *Ann. Statist.* **46** 2336–2359. MR3845020 <https://doi.org/10.1214/17-AOS1622>
- [34] JAVANMARD, A. and MONTANARI, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909. MR3277152
- [35] JIA, J. and ROHE, K. (2015). Preconditioning the lasso for sign consistency. *Electron. J. Statist.* **9** 1150–1205. MR3354334 <https://doi.org/10.1214/15-EJS1029>
- [36] JOHNSON, W. E., LI, C. and RABINOVIC, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8** 118–127.

- [37] KANG, H., ZHANG, A., CAI, T. T. and SMALL, D. S. (2016). Instrumental variables estimation with some invalid instruments and its application to Mendelian randomization. *J. Amer. Statist. Assoc.* **111** 132–144. [MR3494648](#) <https://doi.org/10.1080/01621459.2014.994705>
- [38] LAM, C. and FAN, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Statist.* **37** 4254–4278. [MR2572459](#) <https://doi.org/10.1214/09-AOS720>
- [39] LAM, C. and YAO, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *Ann. Statist.* **40** 694–726. [MR2933663](#) <https://doi.org/10.1214/12-AOS970>
- [40] LAM, C., YAO, Q. and BATHIA, N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika* **98** 901–918. [MR2860332](#) <https://doi.org/10.1093/biomet/asr048>
- [41] LEEK, J. T., SCHARPF, R. B., BRAVO, H. C., SIMCHA, D., LANGMEAD, B., JOHNSON, W. E., GEMAN, D., BAGGERLY, K. and IRIZARRY, R. A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.* **11** 733–739.
- [42] LEEK, J. T., STOREY, J. D. (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3** 1–12.
- [43] LIN, W., FENG, R. and LI, H. (2015). Regularization methods for high-dimensional instrumental variables regression with an application to genetical genomics. *J. Amer. Statist. Assoc.* **110** 270–288. [MR3338502](#) <https://doi.org/10.1080/01621459.2014.908125>
- [44] LONSDALE, J., THOMAS, J., SALVATORE, M., PHILLIPS, R., LO, E., SHAD, S., HASZ, R., WALTERS, G., GARCIA, F. et al. (2013). The genotype-tissue expression (GTEx) project. *Nat. Genet.* **45** 580–585.
- [45] MANGHNANI, K., DRAKE, A., WAN, N. and HAQUE, I. (2018). METCC: METric learning for confounder control making distance matter in high dimensional biological analysis. ArXiv preprint. Available at [arXiv:1812.03188](https://arxiv.org/abs/1812.03188).
- [46] MCCARTHY, M. I., ABECASIS, G. R., CARDON, L. R., GOLDSTEIN, D. B., LITTLE, J., IOANNIDIS, J. P. and HIRSCHHORN, J. N. (2008). Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nat. Rev. Genet.* **9** 356–369.
- [47] MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. [MR2278363](#) <https://doi.org/10.1214/0090536060000000281>
- [48] NEYKOV, M., NING, Y., LIU, J. S. and LIU, H. (2018). A unified theory of confidence regions and testing for high-dimensional estimating equations. *Statist. Sci.* **33** 427–443. [MR3843384](#) <https://doi.org/10.1214/18-STS661>
- [49] NOVEMBRE, J., JOHNSON, T., BRYC, K., KUTALIK, Z., BOYKO, A. R., AUTON, A., INDAP, A., KING, K. S., BERGMANN, S. et al. (2008). Genes mirror geography within Europe. *Nature* **456** 98–101.
- [50] NOVEMBRE, J. and STEPHENS, M. (2008). Interpreting principal component analyses of spatial population genetic variation. *Nat. Genet.* **40** 646–649. <https://doi.org/10.1038/ng.139>
- [51] PEARL, J. (2009). *Causality: Models, Reasoning, and Inference*, 2nd ed. Cambridge Univ. Press, Cambridge. [MR2548166](#) <https://doi.org/10.1017/CBO9780511803161>
- [52] PRICE, A. L., PATTERSON, N. J., PLENGE, R. M., WEINBLATT, M. E., SHADICK, N. A. and REICH, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38** 904–909.
- [53] RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2010). Restricted eigenvalue properties for correlated Gaussian designs. *J. Mach. Learn. Res.* **11** 2241–2259. [MR2719855](#)
- [54] REID, S., TIBSHIRANI, R. and FRIEDMAN, J. (2016). A study of error variance estimation in Lasso regression. *Statist. Sinica* **26** 35–67. [MR3468344](#)
- [55] SHAH, R. D., FROT, B., THANEI, G.-A. and MEINSHAUSEN, N. (2020). Right singular vector projection graphs: Fast high dimensional covariance matrix estimation under latent confounding. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 361–389. [MR4084168](#) <https://doi.org/10.1111/rssb.12359>
- [56] SUBRAMANIAN, A., NARAYAN, R., CORSELLO, S. M., PECK, D. D., NATOLI, T. E., LU, X., GOULD, J., DAVIS, J. F., TUBELLI, A. A. et al. (2017). A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* **171** 1437–1452.
- [57] SUN, Y., ZHANG, N. R. and OWEN, A. B. (2012). Multiple hypothesis testing adjusted for latent variables, with an application to the AGEMAP gene expression data. *Ann. Appl. Stat.* **6** 1664–1688. [MR3058679](#) <https://doi.org/10.1214/12-AOAS561>
- [58] VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. and DEZEURE, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** 1166–1202. [MR3224285](#) <https://doi.org/10.1214/14-AOS1221>
- [59] VERSHYNIN, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge Univ. Press, Cambridge. [MR2963170](#)
- [60] WANG, J., ZHAO, Q., HASTIE, T. and OWEN, A. B. (2017). Confounder adjustment in multiple hypothesis testing. *Ann. Statist.* **45** 1863–1894. [MR3718155](#) <https://doi.org/10.1214/16-AOS1511>

- [61] WANG, W. and FAN, J. (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Ann. Statist.* **45** 1342–1374. [MR3662457](#) <https://doi.org/10.1214/16-AOS1487>
- [62] WANG, Y. and BLEI, D. M. (2019). The blessings of multiple causes. *J. Amer. Statist. Assoc.* **114** 1574–1596. [MR4047282](#) <https://doi.org/10.1080/01621459.2019.1686987>
- [63] WINDMEIJER, F., FARBMACHER, H., DAVIES, N. and SMITH, G. D. (2019). On the use of the Lasso for instrumental variables estimation with some invalid instruments. *J. Amer. Statist. Assoc.* **114** 1339–1350. [MR4011783](#) <https://doi.org/10.1080/01621459.2018.1498346>
- [64] WOOLDRIDGE, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. MIT Press, Cambridge, MA. [MR2768559](#)
- [65] YUAN, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *J. Mach. Learn. Res.* **11** 2261–2286. [MR2719856](#)
- [66] ZHANG, C.-H. and ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 217–242. [MR3153940](#) <https://doi.org/10.1111/rssb.12026>
- [67] ZHOU, S. (2009). Restricted eigenvalue conditions on subgaussian random matrices. ArXiv preprint. Available at [arXiv:0912.4045](#).
- [68] ZHU, Y. (2018). Sparse linear models and l_1 -regularized 2SLS with high-dimensional endogenous regressors and instruments. *J. Econometrics* **202** 196–213. [MR3778835](#) <https://doi.org/10.1016/j.jeconom.2017.10.002>

DEEP LEARNING FOR THE PARTIALLY LINEAR COX MODEL

BY QIXIAN ZHONG^{1,a}, JONAS MUELLER^{2,b} AND JANE-LING WANG^{3,c}

¹Department of Statistics and Data Science, School of Economics, Xiamen University, ^aqxzhong@xmu.edu.cn

²Amazon Web Services, ^bjonasmue@amazon.com

³Department of Statistics, University of California, Davis, ^cjanelwang@ucdavis.edu

While deep learning approaches to survival data have demonstrated empirical success in applications, most of these methods are difficult to interpret and mathematical understanding of them is lacking. This paper studies the partially linear Cox model, where the nonlinear component of the model is implemented using a deep neural network. The proposed approach is flexible and able to circumvent the curse of dimensionality, yet it facilitates interpretability of the effects of treatment covariates on survival. We establish asymptotic theories of maximum partial likelihood estimators and show that our nonparametric deep neural network estimator achieves the minimax optimal rate of convergence (up to a polylogarithmic factor). Moreover, we prove that the corresponding finite-dimensional estimator for treatment covariate effects is $\sqrt{n}$ -consistent, asymptotically normal and attains semiparametric efficiency. Extensive simulation studies and analyses of two real survival data sets show the proposed estimator produces confidence intervals with superior coverage as well as survival time predictions with superior concordance to actual survival times.

REFERENCES

- ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10** 1100–1120. [MR0673646](#)
- ANTHONY, M. and BARTLETT, P. L. (1999). *Neural Network Learning: Theoretical Foundations*. Cambridge Univ. Press, Cambridge. [MR1741038](#) <https://doi.org/10.1017/CBO9780511624216>
- BARRON, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans. Inf. Theory* **39** 930–945. [MR1237720](#) <https://doi.org/10.1109/18.256500>
- BARRON, A. R. (1994). Approximation and estimation bounds for artificial neural networks. *Mach. Learn.* **14** 115–133.
- BAUER, B. and KOHLER, M. (2019). On deep learning as a remedy for the curse of dimensionality in nonparametric regression. *Ann. Statist.* **47** 2261–2285. [MR3953451](#) <https://doi.org/10.1214/18-AOS1747>
- 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](#)
- CHAPFUWA, P., TAO, C., LI, C., PAGE, C., GOLDSTEIN, B., CARIN, L. and HENAO, R. (2018). Adversarial time-to-event modeling. In *Proceedings of the 35th International Conference on Machine Learning*.
- CHEN, S. and ZHOU, L. (2007). Local partial likelihood estimation in proportional hazards regression. *Ann. Statist.* **35** 888–916. [MR2336873](#) <https://doi.org/10.1214/009053606000001299>
- CHEN, K., GUO, S., SUN, L. and WANG, J.-L. (2010). Global partial likelihood for nonparametric proportional hazards models. *J. Amer. Statist. Assoc.* **105** 750–760. [MR2724858](#) <https://doi.org/10.1198/jasa.2010.tm08636>
- CHING, T., ZHU, X. and GARMIRE, L. X. (2018). Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data. *PLoS Comput. Biol.* **14** e1006076.
- COLLOBERT, R., WESTON, J., BOTTOU, L., KARLEN, M., KAVUKCUOGLU, K. and KUKSA, P. (2011). Natural language processing (almost) from scratch. *J. Mach. Learn. Res.* **12** 2493–2537.
- COX, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34** 187–220. [MR0341758](#)
- COX, D. R. (1975). Partial likelihood. *Biometrika* **62** 269–276. [MR0400509](#) <https://doi.org/10.1093/biomet/62.2.269>

MSC2020 subject classifications. Primary 62N02, 62G20; secondary 62C20, 62G08.

Key words and phrases. Censored data, survival analysis, neural network, minimax estimation, partial likelihood, semiparametric efficiency.

- COX, D. R. and OAKES, D. (1984). *Analysis of Survival Data. Monographs on Statistics and Applied Probability*. CRC Press, London. MR0751780 <https://doi.org/10.1201/9781315137438>
- CYBENKO, G. (1989). Approximation by superpositions of a sigmoidal function. *Math. Control Signals Systems* **2** 303–314. MR1015670 <https://doi.org/10.1007/BF02551274>
- DABROWSKA, D. M. (1997). Smoothed Cox regression. *Ann. Statist.* **25** 1510–1540. MR1463563 <https://doi.org/10.1214/aos/1031594730>
- DAVIDSON-PILON, C. (2019). lifelines: Survival analysis in Python. *J. Open Sour. Softw.* **4** 1317. <https://doi.org/10.21105/joss.01317>
- DOU, X. and LIANG, T. (2021). Training neural networks as learning data-adaptive kernels: Provable representation and approximation benefits. *J. Amer. Statist. Assoc.* **116** 1507–1520. MR4309289 <https://doi.org/10.1080/01621459.2020.1745812>
- DU, P., MA, S. and LIANG, H. (2010). Penalized variable selection procedure for Cox models with semiparametric relative risk. *Ann. Statist.* **38** 2092–2117. MR2676884 <https://doi.org/10.1214/09-AOS780>
- FARABET, C., COUPRIE, C., NAJMAN, L. and LECUN, Y. (2012). Learning hierarchical features for scene labeling. *IEEE Trans. Pattern Anal. Mach. Intell.* **35** 1915–1929.
- FARAGGI, D. and SIMON, R. (1995). A neural network model for survival data. *Stat. Med.* **14** 73–82. <https://doi.org/10.1002/sim.4780140108>
- FARRELL, M. H., LIANG, T. and MISRA, S. (2021). Deep neural networks for estimation and inference. *Econometrica* **89** 181–213. MR4220387 <https://doi.org/10.3982/ecta16901>
- 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
- FOEKENS, J. A., PETERS, H. A., LOOK, M. P., PORTENGEN, H., SCHMITT, M., KRAMER, M. D., BRÜNNER, N., JANICKE, F., MEIJER-VAN GELDER, M. E. et al. (2000). The urokinase system of plasminogen activation and prognosis in 2780 breast cancer patients. *Cancer Res.* **60** 636–643.
- GIUNCHIGLIA, E., NEMCHENKO, A. and VAN DER SCHAAR, M. (2018). RNN-SURV: A deep recurrent model for survival analysis. In *International Conference on Artificial Neural Networks* 23–32.
- GLOROT, X. and BENGIO, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the 30th International Conference on Artificial Intelligence and Statistics* 249–256.
- GOODFELLOW, I., BENGIO, Y. and COURVILLE, A. (2016). *Deep Learning. Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3617773
- GRAVES, A., MOHAMED, A.-R. and HINTON, G. (2013). Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* 6645–6649.
- GÜHRING, I., KUTYNIOK, G. and PETERSEN, P. (2020). Error bounds for approximations with deep ReLU neural networks in $W^{s,p}$ norms. *Anal. Appl. (Singap.)* **18** 803–859. MR4131039 <https://doi.org/10.1142/S0219530519410021>
- HAARBURGER, C., WEITZ, P., RIPPEL, O. and MERHOF, D. (2019). Image-based survival prediction for lung cancer patients using CNNs. In *2019 IEEE 16th International Symposium on Biomedical Imaging* 1197–1201.
- HAN, S., POOL, J., TRAN, J. and DALLY, W. (2015). Learning both weights and connections for efficient neural network. In *Advances in Neural Information Processing Systems* 1135–1143.
- HAO, J., KIM, Y., MALLAVARAPU, T., OH, J. H. and KANG, M. (2019). Interpretable deep neural network for cancer survival analysis by integrating genomic and clinical data. *BMC Med. Genom.* **12** 1–13.
- HÄRDLE, W., HALL, P. and ICHIMURA, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21** 157–178. MR1212171 <https://doi.org/10.1214/aos/1176349020>
- HARRELL, F. E., CALIFF, R. M., PRYOR, D. B., LEE, K. L. and ROSATI, R. A. (1982). Evaluating the yield of medical tests. *JAMA* **247** 2543–2546.
- HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models. Monographs on Statistics and Applied Probability* **43**. CRC Press, London. MR1082147
- HINTON, G., DENG, L., YU, D., DAHL, G. E., MOHAMED, A.-R., JAITLEY, N., SENIOR, A., VANHOUCHE, V., NGUYEN, P. et al. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Process. Mag.* **29** 82–97.
- HORNIK, K., STINCHCOMBE, M. and WHITE, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Netw.* **2** 359–366.
- HOROWITZ, J. L. (2009). *Semiparametric and Nonparametric Methods in Econometrics. Springer Series in Statistics*. Springer, New York. MR2535631 <https://doi.org/10.1007/978-0-387-92870-8>
- HOROWITZ, J. L. and MAMMEN, E. (2007). Rate-optimal estimation for a general class of nonparametric regression models with unknown link functions. *Ann. Statist.* **35** 2589–2619. MR2382659 <https://doi.org/10.1214/0090536070000000415>
- HOSMER, D. W., LEMESHOW, S. and MAY, S. (2008). *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR2383788 <https://doi.org/10.1002/9780470258019>

- HUANG, J. (1999). Efficient estimation of the partly linear additive Cox model. *Ann. Statist.* **27** 1536–1563. [MR1742499 https://doi.org/10.1214/aos/1017939141](https://doi.org/10.1214/aos/1017939141)
- ICHIMURA, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J. Econometrics* **58** 71–120. [MR1230981 https://doi.org/10.1016/0304-4076\(93\)90114-K](https://doi.org/10.1016/0304-4076(93)90114-K)
- IMAIZUMI, M. and FUKUMIZU, K. (2019). Deep neural networks learn non-smooth functions effectively. In *The 22nd International Conference on Artificial Intelligence and Statistics* 869–878.
- JIANG, J. and JIANG, X. (2011). Inference for partly linear additive Cox models. *Statist. Sinica* **21** 901–921. [MR2829860 https://doi.org/10.5705/ss.2011.039a](https://doi.org/10.5705/ss.2011.039a)
- KATZMAN, J. L., SHAHAM, U., CLONINGER, A., BATES, J., JIANG, T. and KLUGER, Y. (2018). DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med. Res. Methodol.* **18** 24.
- KINGMA, D. P. and BA, J. (2014). Adam: A method for stochastic optimization. Preprint. Available at [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- KOOPERBERG, C., STONE, C. J. and TRUONG, Y. K. (1995). The L_2 rate of convergence for hazard regression. *Scand. J. Stat.* **22** 143–157. [MR1339748](https://doi.org/10.2307/2347083)
- 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](https://doi.org/10.1007/978-0-387-74978-5)
- KRIZHEVSKY, A., SUTSKEVER, I. and HINTON, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* 1097–1105.
- LECUN, Y., BENGIO, Y. and HINTON, G. (2015). Deep learning. *Nature* **521** 436–444.
- LEE, J. Y. and DERONCOURT, F. (2016). Sequential short-text classification with recurrent and convolutional neural networks. Preprint. Available at [arXiv:1603.03827](https://arxiv.org/abs/1603.03827).
- LEE, C., ZAME, W. R., YOON, J. and VAN DER SCHAAR, M. (2018). Deephit: A deep learning approach to survival analysis with competing risks. In *Thirty-Second AAAI Conference on Artificial Intelligence*.
- LENGLART, E. (1977). Relation de domination entre deux processus. *Ann. Inst. Henri Poincaré B, Calc. Probab. Stat.* **13** 171–179. [MR0471069](https://doi.org/10.1016/0014-0139(77)90011-4)
- LESHNO, M., LIN, V. Y., PINKUS, A. and SCHOCKEN, S. (1993). Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Netw.* **6** 861–867.
- LI, H., BOIMEL, P., JANOPAUL-NAYLOR, J., ZHONG, H., XIAO, Y., BEN-JOSEF, E. and FAN, Y. (2019). Deep convolutional neural networks for imaging data based survival analysis of rectal cancer. In *2019 IEEE 16th International Symposium on Biomedical Imaging* 846–849. IEEE, New York.
- LIANG, S. and SRIKANT, R. (2016). Why deep neural networks for function approximation? Preprint. Available at [arXiv:1610.04161](https://arxiv.org/abs/1610.04161).
- LIU, J., ZHANG, R., ZHAO, W. and LV, Y. (2016). Variable selection in partially linear hazard regression for multivariate failure time data. *J. Nonparametr. Stat.* **28** 375–394. [MR3488604 https://doi.org/10.1080/10485252.2016.1163355](https://doi.org/10.1080/10485252.2016.1163355)
- MARTENS, J. (2010). Deep learning via Hessian-free optimization. In *International Conference on Machine Learning* **27** 735–742.
- MATSUO, K., PURUSHOTHAM, S., JIANG, B., MANDELBAUM, R. S., TAKIUCHI, T., LIU, Y. and ROMAN, L. D. (2019). Survival outcome prediction in cervical cancer: Cox models vs deep-learning model. *Am. J. Obstet. Gynecol.* **220** 381.e1–381.e14.
- MHASKAR, H. N. (1996). Neural networks for optimal approximation of smooth and analytic functions. *Neural Comput.* **8** 164–177.
- MHASKAR, H., LIAO, Q. and POGGIO, T. (2017). When and why are deep networks better than shallow ones? In *Proceedings of the AAAI Conference on Artificial Intelligence* **31**.
- NAIR, V. and HINTON, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. In *International Conference on Machine Learning* 807–814.
- O’SULLIVAN, F. (1993). Nonparametric estimation in the Cox model. *Ann. Statist.* **21** 124–145. [MR1212169 https://doi.org/10.1214/aos/1176349018](https://doi.org/10.1214/aos/1176349018)
- PASZKE, A., GROSS, S., MASSA, F., LERER, A., BRADBURY, J., CHANAN, G., KILLEEN, T., LIN, Z., GIMELSHEIN, N. et al. (2019). PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems* 8024–8035.
- PINKUS, A. (1999). Approximation theory of the MLP model in neural networks. In *Acta Numerica*, 1999. *Acta Numer.* **8** 143–195. Cambridge Univ. Press, Cambridge. [MR1819645 https://doi.org/10.1017/S0962492900002919](https://doi.org/10.1017/S0962492900002919)
- RAMACHANDRAN, P., ZOPH, B. and LE, Q. V. (2017). Searching for activation functions. Preprint. Available at [arXiv:1710.05941](https://arxiv.org/abs/1710.05941).
- RANGANATH, R., PEROTTE, A., ELHADAD, N. and BLEI, D. (2016). Deep survival analysis. In *Proceedings of Machine Learning Research* **56** 101–114.

- REN, K., QIN, J., ZHENG, L., YANG, Z., ZHANG, W., QIU, L. and YU, Y. (2019). Deep recurrent survival analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence* **33** 4798–4805.
- SARIKAYA, R., HINTON, G. E. and DEORAS, A. (2014). Application of deep belief networks for natural language understanding. *IEEE/ACM Trans. Audio Speech Lang. Process.* **22** 778–784.
- SASIENI, P. (1992a). Information bounds for the conditional hazard ratio in a nested family of regression models. *J. Roy. Statist. Soc. Ser. B* **54** 617–635. [MR1160487](#)
- SASIENI, P. (1992b). Nonorthogonal projections and their application to calculating the information in a partly linear Cox model. *Scand. J. Stat.* **19** 215–233. [MR1183198](#)
- SAXE, A. M., MCCLELLAND, J. L. and GANGULI, S. (2013). Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. Preprint. Available at [arXiv:1312.6120](#).
- SCHMIDT-HIEBER, J. (2017). Nonparametric regression using deep neural networks with ReLU activation function. Preprint. Available at [arXiv:1708.06633](#).
- SCHMIDT-HIEBER, J. (2020). Nonparametric regression using deep neural networks with ReLU activation function. *Ann. Statist.* **48** 1875–1897. [MR4134774](#) <https://doi.org/10.1214/19-AOS1875>
- SLEEPER, L. A. and HARRINGTON, D. P. (1990). Regression splines in the Cox model with application to covariate effects in liver disease. *J. Amer. Statist. Assoc.* **85** 941–949. <https://doi.org/10.1080/01621459.1990.10474965>
- SRINIVAS, S., SUBRAMANYA, A. and VENKATESH BABU, R. (2017). Training sparse neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* 138–145.
- SRIVASTAVA, N., HINTON, G., KRIZHEVSKY, A., SUTSKEVER, I. and SALAKHUTDINOV, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15** 1929–1958. [MR3231592](#)
- STONE, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13** 689–705. [MR0790566](#) <https://doi.org/10.1214/aos/1176349548>
- SZEGEDY, C., LIU, W., JIA, Y., SERMANET, P., REED, S., ANGUELOV, D., ERHAN, D., VANHOUCKE, V. and RABINOVICH, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 1–9.
- TELGARSKY, M. (2015). Representation benefits of deep feedforward networks. Preprint. Available at [arXiv:1509.08101](#).
- 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>
- TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. [MR2724359](#) <https://doi.org/10.1007/b13794>
- UNSER, M. (2019). A representer theorem for deep neural networks. *J. Mach. Learn. Res.* **20** Paper No. 110, 30 pp. [MR3990464](#) <https://doi.org/10.1093/biostatistics/kxx066>
- VAN DER VAART, A. W. (2000). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. [MR1652247](#) <https://doi.org/10.1017/CBO9780511802256>
- 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>
- WU, Q., ZHAO, H., ZHU, L. and SUN, J. (2020). Variable selection for high-dimensional partly linear additive Cox model with application to Alzheimer’s disease. *Stat. Med.* **39** 3120–3134. [MR4151923](#) <https://doi.org/10.1002/sim.8594>
- XIANG, A., LAPUERTA, P., RYUTOV, A., BUCKLEY, J. and AZEN, S. (2000). Comparison of the performance of neural network methods and Cox regression for censored survival data. *Comput. Statist. Data Anal.* **34** 243–257.
- YAROTSKY, D. (2017). Error bounds for approximations with deep ReLU networks. *Neural Netw.* **94** 103–114. <https://doi.org/10.1016/j.neunet.2017.07.002>
- YOUSEFI, S., AMROLLAHI, F., AMGAD, M., DONG, C., LEWIS, J. E., SONG, C., GUTMAN, D. A., HALANI, S. H., VEGA, J. E. V. et al. (2017). Predicting clinical outcomes from large scale cancer genomic profiles with deep survival models. *Sci. Rep.* **7** 11707. <https://doi.org/10.1038/s41598-017-11817-6>
- ZHONG, Q., MUELLER, J. and WANG, J.-L. (2022). Supplement to “Deep learning for the partially linear Cox model.” <https://doi.org/10.1214/21-AOS2153SUPP>

OPTIMAL DIFFERENCE-BASED VARIANCE ESTIMATORS IN TIME SERIES: A GENERAL FRAMEWORK

BY KIN WAI CHAN^a

Department of Statistics, The Chinese University of Hong Kong, ^akinwaichan@cuhk.edu.hk

Variance estimation is important for statistical inference. It becomes nontrivial when observations are masked by serial dependence structures and time-varying mean structures. Existing methods either ignore or sub-optimally handle these nuisance structures. This paper develops a general framework for the estimation of the long-run variance for time series with nonconstant means. The building blocks are difference statistics. The proposed class of estimators is general enough to cover many existing estimators. Necessary and sufficient conditions for consistency are investigated. The first asymptotically optimal estimator is derived. Our proposed estimator is theoretically proven to be invariant to arbitrary mean structures, which may include trends and a possibly divergent number of discontinuities.

REFERENCES

- ALEXOPOULOS, C., GOLDSMAN, D. and WILSON, J. R. (2011). Overlapping batch means: Something more for nothing? *Winter Simul. Conf.* 401–411.
- ALTISSIMO, F. and CORRADI, V. (2003). Strong rules for detecting the number of breaks in a time series. *J. Econometrics* **117** 207–244. [MR2008769](#) [https://doi.org/10.1016/S0304-4076\(03\)00147-7](https://doi.org/10.1016/S0304-4076(03)00147-7)
- ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York. [MR0283939](#)
- ANDREWS, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59** 817–858. [MR1106513](#) <https://doi.org/10.2307/2938229>
- BRADLEY, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** 107–144. [MR2178042](#) <https://doi.org/10.1214/154957805100000104>
- BROCKWELL, P. J. and DAVIS, R. A. (2006). *Time Series: Theory and Methods*. Springer Series in Statistics. Springer, New York. [MR2839251](#)
- CARLSTEIN, E. (1986). The use of subsamples for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist.* **14** 1171–1179. [MR0856813](#) <https://doi.org/10.1214/aos/1176350057>
- CHAN, K. W. (2022a). Mean-Structure and Autocorrelation Consistent Covariance Matrix Estimation. *J. Bus. Econom. Statist.* **40** 201–215. [MR4356567](#) <https://doi.org/10.1080/07350015.2020.1796397>
- CHAN, K. W. (2022b). Supplement to “Optimal difference-based variance estimators in time series: A general framework.” <https://doi.org/10.1214/21-AOS2154SUPP>
- CHAN, K. W. and YAU, C. Y. (2017a). Automatic optimal batch size selection for recursive estimators of time-average covariance matrix. *J. Amer. Statist. Assoc.* **112** 1076–1089. [MR3735361](#) <https://doi.org/10.1080/01621459.2016.1189337>
- CHAN, K. W. and YAU, C. Y. (2017b). High-order corrected estimator of asymptotic variance with optimal bandwidth. *Scand. J. Stat.* **44** 866–898. [MR3730019](#) <https://doi.org/10.1111/sjso.12279>
- CHEN, H. and SCHMEISER, B. (2013). I-SMOOTH: Iteratively smoothing mean-constrained and nonnegative piecewise-constant functions. *INFORMS J. Comput.* **25** 432–445. [MR3085324](#) <https://doi.org/10.1287/ijoc.1120.0512>
- CHEN, L., WANG, W. and WU, W. B. (2021). Inference of breakpoints in high-dimensional time series. *J. Amer. Statist. Assoc.* To appear.
- CHENG, C. H. and CHAN, K. W. (2022). *A general framework for constructing locally self-normalized multiple-change-point tests*. Manuscript.
- CRAINICEANU, C. M. and VOGELANG, T. J. (2007). Nonmonotonic power for tests of a mean shift in a time series. *J. Stat. Comput. Simul.* **77** 457–476. [MR2405424](#) <https://doi.org/10.1080/10629360600569394>

MSC2020 subject classifications. Primary 62G05; secondary 62G20.

Key words and phrases. Change point detection, nonlinear time series, optimal bandwidth selection, trend inference, variate difference method.

- CSÖRGŐ, M. and HORVÁTH, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley Series in Probability and Statistics. Wiley, Chichester. [MR2743035](#)
- DALLA, V., GIRAITIS, L. and PHILLIPS, P. C. B. (2015). *Testing Mean Stability of Heteroskedastic Time Series*. Manuscript.
- DEHLING, H., FRIED, R. and WENDLER, M. (2020). A robust method for shift detection in time series. *Biometrika* **107** 647–660. [MR4138981](#) <https://doi.org/10.1093/biomet/asaa004>
- DETTE, H., ECKLE, T. and VETTER, M. (2020). Multiscale change point detection for dependent data. *Scand. J. Stat.* **47** 1243–1274. [MR4178193](#) <https://doi.org/10.1111/sjos.12465>
- DETTE, H. and WU, W. (2019). Detecting relevant changes in the mean of nonstationary processes—a mass excess approach. *Ann. Statist.* **47** 3578–3608. [MR4025752](#) <https://doi.org/10.1214/19-AOS1811>
- GALLANT, A. R. (1987). *Nonlinear Statistical Models*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. Wiley, New York. [MR0921029](#) <https://doi.org/10.1002/9780470316719>
- GAO, Z., SHANG, Z., DU, P. and ROBERTSON, J. L. (2019). Variance change point detection under a smoothly-changing mean trend with application to liver procurement. *J. Amer. Statist. Assoc.* **114** 773–781. [MR3963179](#) <https://doi.org/10.1080/01621459.2018.1442341>
- GONÇALVES, S. and WHITE, H. (2002). The bootstrap of the mean for dependent heterogeneous arrays. *Econometric Theory* **18** 1367–1384. [MR1945417](#) <https://doi.org/10.1017/S0266466602186051>
- GÓRECKI, T., HORVÁTH, L. and KOKOSZKA, P. (2018). Change point detection in heteroscedastic time series. *Econom. Stat.* **7** 63–88. [MR3824127](#) <https://doi.org/10.1016/j.ecosta.2017.07.005>
- HALL, P., KAY, J. W. and TITTERINGTON, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* **77** 521–528. [MR1087842](#) <https://doi.org/10.1093/biomet/77.3.521>
- HORVÁTH, L., KOKOSZKA, P. and STEINEBACH, J. (1999). Testing for changes in multivariate dependent observations with an application to temperature changes. *J. Multivariate Anal.* **68** 96–119. [MR1668911](#) <https://doi.org/10.1006/jmva.1998.1780>
- IBRAGIMOV, I. A. (1962). Some limit theorems for stationary processes. *Teor. Veroyatn. Primen.* **7** 361–392. [MR0148125](#)
- JUHL, T. and XIAO, Z. (2009). Tests for changing mean with monotonic power. *J. Econometrics* **148** 14–24. [MR2494814](#) <https://doi.org/10.1016/j.jeconom.2008.08.020>
- KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217–1241. [MR1015147](#) <https://doi.org/10.1214/aos/1176347265>
- LEUNG, C. W. D. and CHAN, K. W. (2022). *Testing for variance changes under varying mean and serial correlation*. Manuscript.
- LEVINE, M. and TECUAPETLA-GÓMEZ, I. (2019). ACF estimation via difference schemes for a semiparametric model with m -dependent errors. ArXiv Preprint. Available at [arXiv:1905.04578](https://arxiv.org/abs/1905.04578).
- LOBATO, I. N. (2001). Testing that a dependent process is uncorrelated. *J. Amer. Statist. Assoc.* **96** 1066–1076. [MR1947254](#) <https://doi.org/10.1198/016214501753208726>
- NEWAY, W. K. and WEST, K. D. (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55** 703–708. [MR0890864](#) <https://doi.org/10.2307/1913610>
- PARZEN, E. (1957). On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Stat.* **28** 329–348. [MR0088833](#) <https://doi.org/10.1214/aoms/1177706962>
- PEŠTA, M. and WENDLER, M. (2020). Nuisance-parameter-free changepoint detection in non-stationary series. *TEST* **29** 379–408. [MR4095034](#) <https://doi.org/10.1007/s11749-019-00659-1>
- POLITIS, D. N. (2011). Higher-order accurate, positive semidefinite estimation of large-sample covariance and spectral density matrices. *Econometric Theory* **27** 703–744. [MR2822363](#) <https://doi.org/10.1017/S0266466610000484>
- POLITIS, D. N. and ROMANO, J. P. (1995). Bias-corrected nonparametric spectral estimation. *J. Time Series Anal.* **16** 67–103. [MR1323618](#) <https://doi.org/10.1111/j.1467-9892.1995.tb00223.x>
- POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling*. Springer Series in Statistics. Springer, New York. [MR1707286](#) <https://doi.org/10.1007/978-1-4612-1554-7>
- RICE, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12** 1215–1230. [MR0760684](#) <https://doi.org/10.1214/aos/1176346788>
- 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>
- RUPPERT, D., SHEATHER, S. J. and WAND, M. P. (1995). An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.* **90** 1257–1270. [MR1379468](#)
- SHAO, X. (2010). A self-normalized approach to confidence interval construction in time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 343–366. [MR2758116](#) <https://doi.org/10.1111/j.1467-9868.2009.00737.x>
- SHAO, X. (2015). Self-normalization for time series: A review of recent developments. *J. Amer. Statist. Assoc.* **110** 1797–1817. [MR3449074](#) <https://doi.org/10.1080/01621459.2015.1050493>

- SONG, W. T. and SCHMEISER, B. W. (1995). Optimal mean-squared-error batch sizes. *Manage. Sci.* **41** 110–123.
- TAQQU, M. and EBERLEIN (1986). *Dependence in Probability and Statistics*. Birkhäuser, Basel.
- TECUAPETLA-GÓMEZ, I. and MUNK, A. (2017). Autocovariance estimation in regression with a discontinuous signal and m -dependent errors: A difference-based approach. *Scand. J. Stat.* **44** 346–368. [MR3658518](#) <https://doi.org/10.1111/sjos.12256>
- TO, H. K. and CHAN, K. W. (2022). *Mean stationarity test in time series: A signal variance-based approach*. Manuscript.
- VATS, D. and FLEGAL, J. M. (2021). Lugsail lag windows for estimating time-average covariance matrices. *Biometrika*. To appear.
- VOLKONSKIĬ, V. A. and ROZANOV, Y. A. (1959). Some limit theorems for random functions. I. *Theory Probab. Appl.* **4** 178–197. [MR0121856](#) <https://doi.org/10.1137/1104015>
- WELCH, P. D. (1987). On the relationship between batch means, overlapping batch means and spectral estimation. *Winter Simul. Conf.* 320–323.
- WHITE, H. (1984). *Asymptotic Theory for Econometricians*. Academic Press, New York.
- WU, W. B. (2004). A test for detecting changes in mean. In *Time Series Analysis and Applications to Geophysical Systems* (D. R. Brillinger, E. A. Robinson and F. Schoenberg, eds.) **139** 105–122. Springer, New York.
- 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>
- WU, W. B. (2007). Strong invariance principles for dependent random variables. *Ann. Probab.* **35** 2294–2320. [MR2353389](#) <https://doi.org/10.1214/0091179070000000060>
- WU, W. B. (2011). Asymptotic theory for stationary processes. *Stat. Interface* **4** 207–226. [MR2812816](#) <https://doi.org/10.4310/SII.2011.v4.n2.a15>
- WU, W. B., WOODROOFE, M. and MENTZ, G. (2001). Isotonic regression: Another look at the changepoint problem. *Biometrika* **88** 793–804. [MR1859410](#) <https://doi.org/10.1093/biomet/88.3.793>
- WU, W. B. and ZHAO, Z. (2007). Inference of trends in time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 391–410. [MR2323759](#) <https://doi.org/10.1111/j.1467-9868.2007.00594.x>
- YAU, C. Y. and CHAN, K. W. (2016). New recursive estimators of the time-average variance constant. *Stat. Comput.* **26** 609–627. [MR3489860](#) <https://doi.org/10.1007/s11222-015-9548-7>
- ZHANG, T. and LAVITAS, L. (2018). Unsupervised self-normalized change-point testing for time series. *J. Amer. Statist. Assoc.* **113** 637–648. [MR3832215](#) <https://doi.org/10.1080/01621459.2016.1270214>
- ZHAO, Z. (2011). A self-normalized confidence interval for the mean of a class of nonstationary processes. *Biometrika* **98** 81–90. [MR2804211](#) <https://doi.org/10.1093/biomet/asq076>

RIDGE REGRESSION REVISITED: DEBIASING, THRESHOLDING AND BOOTSTRAP

BY YUNYI ZHANG^{1,a} AND DIMITRIS N. POLITIS^{2,b}

¹*Department of Mathematics, University of California, San Diego, ^ayuz334@ucsd.edu*

²*Department of Mathematics and Halicioglu Data Science Institute, University of California, San Diego, ^bdpolitis@ucsd.edu*

The success of the Lasso in the era of high-dimensional data can be attributed to its conducting an implicit model selection, that is, zeroing out regression coefficients that are not significant. By contrast, classical ridge regression cannot reveal a potential sparsity of parameters, and may also introduce a large bias under the high-dimensional setting. Nevertheless, recent work on the Lasso involves debiasing and thresholding, the latter in order to further enhance the model selection. As a consequence, ridge regression may be worth another look since—after debiasing and thresholding—it may offer some advantages over the Lasso, for example, it can be easily computed using a closed-form expression. In this paper, we define a debiased and thresholded ridge regression method, and prove a consistency result and a Gaussian approximation theorem. We further introduce a wild bootstrap algorithm to construct confidence regions and perform hypothesis testing for a linear combination of parameters. In addition to estimation, we consider the problem of prediction, and present a novel, hybrid bootstrap algorithm tailored for prediction intervals. Extensive numerical simulations further show that the debiased and thresholded ridge regression has favorable finite sample performance and may be preferable in some settings.

REFERENCES

- BAI, Z. D. and YIN, Y. Q. (1993). Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.* **21** 1275–1294. [MR1235416](#)
- BASU, S. and MICHAILIDIS, G. (2015). Regularized estimation in sparse high-dimensional time series models. *Ann. Statist.* **43** 1535–1567. [MR3357870](#) <https://doi.org/10.1214/15-AOS1315>
- BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. [MR2533469](#) <https://doi.org/10.1214/08-AOS620>
- BÜHLMANN, P. (2013). Statistical significance in high-dimensional linear models. *Bernoulli* **19** 1212–1242. [MR3102549](#) <https://doi.org/10.3150/12-BEJSP11>
- BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data. Springer Series in Statistics.* Springer, Heidelberg. [MR2807761](#) <https://doi.org/10.1007/978-3-642-20192-9>
- CHATTERJEE, A. and LAHIRI, S. N. (2010). Asymptotic properties of the residual bootstrap for Lasso estimators. *Proc. Amer. Math. Soc.* **138** 4497–4509. [MR2680074](#) <https://doi.org/10.1090/S0002-9939-2010-10474-4>
- CHATTERJEE, A. and LAHIRI, S. N. (2011). Bootstrapping lasso estimators. *J. Amer. Statist. Assoc.* **106** 608–625. [MR2847974](#) <https://doi.org/10.1198/jasa.2011.tm10159>
- CHEN, X. and ZHOU, W.-X. (2020). Robust inference via multiplier bootstrap. *Ann. Statist.* **48** 1665–1691. [MR4124339](#) <https://doi.org/10.1214/19-AOS1863>
- 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., WÜTHRICH, K. and ZHU, Y. (2019). Distributional conformal prediction. Preprint. Available at [arXiv:1909.07889](https://arxiv.org/abs/1909.07889).
- DAI, L., CHEN, K., SUN, Z., LIU, Z. and LI, G. (2018). Broken adaptive ridge regression and its asymptotic properties. *J. Multivariate Anal.* **168** 334–351. [MR3858369](#) <https://doi.org/10.1016/j.jmva.2018.08.007>

MSC2020 subject classifications. 62J05, 62F40.

Key words and phrases. Gaussian approximation, high-dimensional data, Lasso, prediction, regression, resampling.

- DALALYAN, A. S., HEBIRI, M. and LEDERER, J. (2017). On the prediction performance of the Lasso. *Bernoulli* **23** 552–581. [MR3556784](#) <https://doi.org/10.3150/15-BEJ756>
- DEZEURE, R., BÜHLMANN, P. and ZHANG, C.-H. (2017). High-dimensional simultaneous inference with the bootstrap. *TEST* **26** 685–719. [MR3713586](#) <https://doi.org/10.1007/s11749-017-0554-2>
- DOBRIBAN, E. and WAGER, S. (2018). High-dimensional asymptotics of prediction: Ridge regression and classification. *Ann. Statist.* **46** 247–279. [MR3766952](#) <https://doi.org/10.1214/17-AOS1549>
- DOLADO, J. J. and LÜTKEPOHL, H. (1996). Making Wald tests work for cointegrated VAR systems. *Econometric Rev.* **15** 369–386.
- 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>
- FITHIAN, W., SUN, D. and TAYLOR, J. (2017). Optimal inference after model selection. Preprint. Available at [arXiv:1410.2597](#).
- GONÇALVES, S. and VOGELSANG, T. J. (2011). Block bootstrap HAC robust tests: The sophistication of the naive bootstrap. *Econometric Theory* **27** 745–791. [MR2822364](#) <https://doi.org/10.1017/S0266466610000496>
- GREENSHEIN, E. and RITOV, Y. (2004). Persistence in high-dimensional linear predictor selection and the virtue of overparametrization. *Bernoulli* **10** 971–988. [MR2108039](#) <https://doi.org/10.3150/bj/1106314846>
- HORN, R. A. and JOHNSON, C. R. (2013). *Matrix Analysis*, 2nd ed. Cambridge Univ. Press, Cambridge. [MR2978290](#)
- JAVANMARD, A. and JAVADI, H. (2019). False discovery rate control via debiased lasso. *Electron. J. Stat.* **13** 1212–1253. [MR3935848](#) <https://doi.org/10.1214/19-ejs1554>
- JAVANMARD, A. and MONTANARI, A. (2018). Debiasing the Lasso: Optimal sample size for Gaussian designs. *Ann. Statist.* **46** 2593–2622. [MR3851749](#) <https://doi.org/10.1214/17-AOS1630>
- LEE, J. D., SUN, D. L., SUN, Y. and TAYLOR, J. E. (2016). Exact post-selection inference, with application to the lasso. *Ann. Statist.* **44** 907–927. [MR3485948](#) <https://doi.org/10.1214/15-AOS1371>
- LEI, J. and WASSERMAN, L. (2014). Distribution-free prediction bands for non-parametric regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 71–96. [MR3153934](#) <https://doi.org/10.1111/rssb.12021>
- 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>
- LIU, H. and YU, B. (2013). Asymptotic properties of Lasso+mLS and Lasso+Ridge in sparse high-dimensional linear regression. *Electron. J. Stat.* **7** 3124–3169. [MR3151764](#) <https://doi.org/10.1214/14-EJS875>
- LOPES, M. (2014). A residual bootstrap for high-dimensional regression with near low-rank designs. In *Advances in Neural Information Processing Systems* 27 3239–3247.
- 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>
- MAMMEN, E. (1996). Empirical process of residuals for high-dimensional linear models. *Ann. Statist.* **24** 307–335. [MR1389892](#) <https://doi.org/10.1214/aos/1033066211>
- MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. [MR2278363](#) <https://doi.org/10.1214/009053606000000281>
- MEINSHAUSEN, N. and YU, B. (2009). Lasso-type recovery of sparse representations for high-dimensional data. *Ann. Statist.* **37** 246–270. [MR2488351](#) <https://doi.org/10.1214/07-AOS582>
- POLITIS, D. N. (2015). *Model-Free Prediction and Regression*. *Frontiers in Probability and the Statistical Sciences*. Springer, Cham. [MR3442999](#) <https://doi.org/10.1007/978-3-319-21347-7>
- POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling*. *Springer Series in Statistics*. Springer, New York. [MR1707286](#) <https://doi.org/10.1007/978-1-4612-1554-7>
- ROMANO, Y., PATTERSON, E. and CANDÈS, E. (2019). Conformalized quantile regression. In *Advances in Neural Information Processing Systems* **32** 3543–3553. Curran Associates, Red Hook.
- ROMANO, Y., SESIA, M. and CANDÈS, E. (2020). Deep knockoffs. *J. Amer. Statist. Assoc.* **115** 1861–1872. [MR4189763](#) <https://doi.org/10.1080/01621459.2019.1660174>
- SHAO, J. (2003). *Mathematical Statistics*, 2nd ed. *Springer Texts in Statistics*. Springer, New York. [MR2002723](#) <https://doi.org/10.1007/b97553>
- SHAO, J. and DENG, X. (2012). Estimation in high-dimensional linear models with deterministic design matrices. *Ann. Statist.* **40** 812–831. [MR2933667](#) <https://doi.org/10.1214/12-AOS982>
- STINE, R. A. (1985). Bootstrap prediction intervals for regression. *J. Amer. Statist. Assoc.* **80** 1026–1031. [MR0819610](#)
- SUN, Y. (2011). Robust trend inference with series variance estimator and testing-optimal smoothing parameter. *J. Econometrics* **164** 345–366. [MR2826775](#) <https://doi.org/10.1016/j.jeconom.2011.06.017>
- SUN, Y. (2013). A heteroskedasticity and autocorrelation robust F test using an orthonormal series variance estimator. *Econom. J.* **16** 1–26. [MR3030064](#) <https://doi.org/10.1111/j.1368-423X.2012.00390.x>
- SUN, T. and ZHANG, C.-H. (2012). Scaled sparse linear regression. *Biometrika* **99** 879–898. [MR2999166](#) <https://doi.org/10.1093/biomet/ass043>

- TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. [MR1379242](#)
- TIBSHIRANI, R. J., RINALDO, A., TIBSHIRANI, R. and WASSERMAN, L. (2018). Uniform asymptotic inference and the bootstrap after model selection. *Ann. Statist.* **46** 1255–1287. [MR3798003](#) <https://doi.org/10.1214/17-AOS1584>
- VAN DE GEER, S. A. (2008). High-dimensional generalized linear models and the lasso. *Ann. Statist.* **36** 614–645. [MR2396809](#) <https://doi.org/10.1214/009053607000000929>
- VAN DE GEER, S. (2019). On the asymptotic variance of the debiased Lasso. *Electron. J. Stat.* **13** 2970–3008. [MR4010589](#) <https://doi.org/10.1214/19-ejs1599>
- VAN DE GEER, S., BÜHLMANN, P. and ZHOU, S. (2011). The adaptive and the thresholded Lasso for potentially misspecified models (and a lower bound for the Lasso). *Electron. J. Stat.* **5** 688–749. [MR2820636](#) <https://doi.org/10.1214/11-EJS624>
- WU, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* **14** 1261–1350. [MR0868303](#) <https://doi.org/10.1214/aos/1176350142>
- ZHANG, X. and CHENG, G. (2017). Simultaneous inference for high-dimensional linear models. *J. Amer. Statist. Assoc.* **112** 757–768. [MR3671768](#) <https://doi.org/10.1080/01621459.2016.1166114>
- ZHANG, Y. and POLITIS, D. N. (2022). Supplement to “Ridge regression revisited: Debiasing, thresholding and bootstrap.” <https://doi.org/10.1214/21-AOS2156SUPP>
- ZHANG, Y. and POLITIS, D. N. (2021b). Bootstrap prediction intervals with asymptotic conditional validity and unconditional guarantees. Preprint. Available at [arXiv:2005.09145](https://arxiv.org/abs/2005.09145).
- 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>
- ZHANG, C.-H. and ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 217–242. [MR3153940](#) <https://doi.org/10.1111/rssb.12026>
- ZHAO, P. and YU, B. (2006). On model selection consistency of Lasso. *J. Mach. Learn. Res.* **7** 2541–2563. [MR2274449](#)

CUBE ROOT WEAK CONVERGENCE OF EMPIRICAL ESTIMATORS OF A DENSITY LEVEL SET

BY PHILIPPE BERTHET^{1,a} AND JOHN H. J. EINMAHL^{2,b}

¹*Institut de Mathématiques de Toulouse; UMR5219, Université de Toulouse; CNRS, ^aphilippe.berthet@math.univ-toulouse.fr*

²*Department of Econometrics and OR and CentER, Tilburg University, ^bj.h.j.einmahl@uvt.nl*

Given n independent random vectors with common density f on $\mathbb{R}^d$, we study the weak convergence of three empirical-measure based estimators of the convex λ -level set L_λ of f , namely the excess mass set, the minimum volume set and the maximum probability set, all selected from a class of convex sets $\mathcal{A}$ that contains L_λ . Since these set-valued estimators approach L_λ , even the formulation of their weak convergence is nonstandard. We identify the joint limiting distribution of the symmetric difference of L_λ and each of the three estimators, at rate $n^{-1/3}$. It turns out that the minimum volume set and the maximum probability set estimators are asymptotically indistinguishable, whereas the excess mass set estimator exhibits “richer” limit behavior. Arguments rely on the boundary local empirical process, its cylinder representation, dimension-free concentration around the boundary of L_λ , and the set-valued argmax of a drifted Wiener process.

REFERENCES

- ALEXANDER, K. S. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.* **12** 1041–1067. [MR0757769](#)
- BERTHET, P. and EL-NOUTY, C. (2006). Almost sure asymptotic behaviour of the shorth estimators. *Math. Methods Statist.* **14** 379–403. [MR2210538](#)
- BRUNEL, V.-E. (2018). Methods for estimation of convex sets. *Statist. Sci.* **33** 615–632. [MR3881211](#) <https://doi.org/10.1214/18-STS669>
- CADRE, B. (2006). Kernel estimation of density level sets. *J. Multivariate Anal.* **97** 999–1023. [MR2256570](#) <https://doi.org/10.1016/j.jmva.2005.05.004>
- CAI, J.-J., EINMAHL, J. H. J. and DE HAAN, L. (2011). Estimation of extreme risk regions under multivariate regular variation. *Ann. Statist.* **39** 1803–1826. [MR2850221](#) <https://doi.org/10.1214/11-AOS891>
- CHEN, Y.-C., GENOVESE, C. R. and WASSERMAN, L. (2017). Density level sets: Asymptotics, inference, and visualization. *J. Amer. Statist. Assoc.* **112** 1684–1696. [MR3750891](#) <https://doi.org/10.1080/01621459.2016.1228536>
- CLÉMENTON, S., GOIX, N. and SABOURIN, A. (2015). On anomaly ranking and excess-mass curves. In *Proceedings of the AISTATS 2015* 287–295. San Diego.
- DAVIES, L. (1992). The asymptotics of Rousseeuw’s minimum volume ellipsoid estimator. *Ann. Statist.* **20** 1828–1843. [MR1193314](#) <https://doi.org/10.1214/aos/1176348891>
- DI BUCCHIANICO, A., EINMAHL, J. H. J. and MUSHKUDIANI, N. A. (2001). Smallest nonparametric tolerance regions. *Ann. Statist.* **29** 1320–1343. [MR1873333](#) <https://doi.org/10.1214/aos/1013203456>
- DYKSTRA, R. and CAROLAN, C. (1999). The distribution of the argmax of two-sided Brownian motion with quadratic drift. *J. Statist. Comput. Simulation* **63** 47–58. [MR1703044](#) <https://doi.org/10.1080/00949659908811948>
- EINMAHL, J. H. J. and KHMALADZE, E. V. (2011). Central limit theorems for local empirical processes near boundaries of sets. *Bernoulli* **17** 545–561. [MR2787604](#) <https://doi.org/10.3150/10-BEJ283>
- GROENEBOOM, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probab. Theory Related Fields* **81** 79–109. [MR0981568](#) <https://doi.org/10.1007/BF00343738>
- HARTIGAN, J. A. (1987). Estimation of a convex density contour in two dimensions. *J. Amer. Statist. Assoc.* **82** 267–270. [MR0883354](#)

MSC2020 subject classifications. Primary 62G05, 62G20; secondary 60F05, 60F17.

Key words and phrases. Argmax drifted Wiener process, cube root asymptotics, density level set, excess mass, local empirical process, minimum volume set, set-valued estimator.

- KHMALADZE, E. V. (2007). Differentiation of sets in measure. *J. Math. Anal. Appl.* **334** 1055–1072. [MR2338647](#) <https://doi.org/10.1016/j.jmaa.2006.12.086>
- KHMALADZE, E. and WEIL, W. (2008). Local empirical processes near boundaries of convex bodies. *Ann. Inst. Statist. Math.* **60** 813–842. [MR2453573](#) <https://doi.org/10.1007/s10463-007-0123-7>
- KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219. [MR1041391](#) <https://doi.org/10.1214/aos/1176347498>
- MAMMEN, E. and POLONIK, W. (2013). Confidence regions for level sets. *J. Multivariate Anal.* **122** 202–214. [MR3189318](#) <https://doi.org/10.1016/j.jmva.2013.07.017>
- MASON, D. M. and POLONIK, W. (2009). Asymptotic normality of plug-in level set estimates. *Ann. Appl. Probab.* **19** 1108–1142. [MR2537201](#) <https://doi.org/10.1214/08-AAP569>
- MÜLLER, D. W. (1992). *The Excess Mass Approach in Statistics*. Beiträge zur Statistik, Univ. Heidelberg.
- MÜLLER, D. W. and SAWITZKI, G. (1991). Excess mass estimates and tests for multimodality. *J. Amer. Statist. Assoc.* **86** 738–746. [MR1147099](#)
- NOLAN, D. (1991). The excess-mass ellipsoid. *J. Multivariate Anal.* **39** 348–371. [MR1147127](#) [https://doi.org/10.1016/0047-259X\(91\)90106-C](https://doi.org/10.1016/0047-259X(91)90106-C)
- POLONIK, W. (1995). Measuring mass concentrations and estimating density contour clusters—an excess mass approach. *Ann. Statist.* **23** 855–881. [MR1345204](#) <https://doi.org/10.1214/aos/1176324626>
- POLONIK, W. (1997). Minimum volume sets and generalized quantile processes. *Stochastic Process. Appl.* **69** 1–24. [MR1464172](#) [https://doi.org/10.1016/S0304-4149\(97\)00028-8](https://doi.org/10.1016/S0304-4149(97)00028-8)
- POLONIK, W. (1998). The silhouette, concentration functions and ML-density estimation under order restrictions. *Ann. Statist.* **26** 1857–1877. [MR1673281](#) <https://doi.org/10.1214/aos/1024691360>
- QIAO, W. and POLONIK, W. (2019). Nonparametric confidence regions for level sets: Statistical properties and geometry. *Electron. J. Stat.* **13** 985–1030. [MR3934621](#) <https://doi.org/10.1214/19-EJS1543>
- RODRÍGUEZ-CASAL, A. and SAAVEDRA-NIEVES, P. (2019). Minimax Hausdorff estimation of density level sets. Preprint. Available at [arXiv:1905.02897](https://arxiv.org/abs/1905.02897).
- ROUSSEEUW, P. (1985). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications, Vol. B (Bad Tatzmannsdorf, 1983)* 283–297. Reidel, Dordrecht. [MR0851060](#)
- SCHNEIDER, R. (1993). *Convex Bodies: The Brunn-Minkowski Theory*. *Encyclopedia of Mathematics and Its Applications* **44**. Cambridge Univ. Press, Cambridge. [MR1216521](#) <https://doi.org/10.1017/CBO9780511526282>
- SCHNEIDER, R. and WEIL, W. (2008). *Stochastic and Integral Geometry. Probability and Its Applications (New York)*. Springer, Berlin. [MR2455326](#) <https://doi.org/10.1007/978-3-540-78859-1>
- SCOTT, C. D. and NOWAK, R. D. (2006). Learning minimum volume sets. *J. Mach. Learn. Res.* **7** 665–704. [MR2274383](#)
- SILVERMAN, B. W. and TITTERINGTON, D. M. (1980). Minimum covering ellipses. *SIAM J. Sci. Statist. Comput.* **1** 401–409. [MR0610752](#) <https://doi.org/10.1137/0901028>
- TSYBAKOV, A. B. (1997). On nonparametric estimation of density level sets. *Ann. Statist.* **25** 948–969. [MR1447735](#) <https://doi.org/10.1214/aos/1069362732>
- VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. [MR1652247](#) <https://doi.org/10.1017/CBO9780511802256>
- VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes. With Applications to Statistics*. *Springer Series in Statistics*. Springer, New York. [MR1385671](#) <https://doi.org/10.1007/978-1-4757-2545-2>
- XU, M. and SAMWORTH, R. J. (2021). High-dimensional nonparametric density estimation via symmetry and shape constraints. *Ann. Statist.* **49** 650–672. [MR4255102](#) <https://doi.org/10.1214/20-aos1972>

HIGHER CRITICISM TO COMPARE TWO LARGE FREQUENCY TABLES, WITH SENSITIVITY TO POSSIBLE RARE AND WEAK DIFFERENCES

BY DAVID L. DONOHO^{1,a} AND ALON KIPNIS^{2,b}

¹Department of Statistics, Stanford University, ^adonoho@stanford.edu

²School of Computer Science, Reichman University, ^balon.kipnis@idc.ac.il

We adapt Higher Criticism (HC) to the comparison of two frequency tables which may—or may not—exhibit moderate differences between the tables in some unknown, relatively small subset out of a large number of categories. Our analysis of the power of the proposed HC test quantifies the rarity and size of assumed differences and applies moderate deviations-analysis to determine the asymptotic powerfulness/powerlessness of our proposed HC procedure.

Our analysis considers the null hypothesis of no difference in underlying generative model against a rare/weak perturbation alternative, in which the frequencies of $N^{1-\beta}$ out of the N categories are perturbed by $r(\log N)/2n$ in the Hellinger distance; here, n is the size of each sample. Our proposed Higher Criticism (HC) test for this setting uses P-values obtained from N exact binomial tests. We characterize the asymptotic performance of the HC-based test in terms of the sparsity parameter β and the perturbation intensity parameter r . Specifically, we derive a region in the (β, r) -plane where the test asymptotically has maximal power, while having asymptotically no power outside this region. Our analysis distinguishes between cases in which the counts in both tables are low, versus cases in which counts are high, corresponding to the cases of sparse and dense frequency tables. The phase transition curve of HC in the high-counts regime matches formally the curve delivered by HC in a two-sample normal means model.

REFERENCES

- [1] ACHARYA, J., DAS, H., JAFARPOUR, A., ORLITSKY, A., PAN, S. and SURESH, A. (2012). Competitive classification and closeness testing. In *Conference on Learning Theory* 22–1.
- [2] ANSCOMBE, F. J. (1948). The transformation of Poisson, binomial and negative-binomial data. *Biometrika* **35** 246–254. [MR0028556](#) <https://doi.org/10.1093/biomet/35.3-4.246>
- [3] 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>
- [4] ARIAS-CASTRO, E. and WANG, M. (2015). The sparse Poisson means model. *Electron. J. Stat.* **9** 2170–2201. [MR3406276](#) <https://doi.org/10.1214/15-EJS1066>
- [5] BRAVATA, D. M., McDONALD, K. M., SMITH, W. M., RYDZAK, C., SZETO, H., BUCKERIDGE, D. L., HABERLAND, C. and OWENS, D. K. (2004). Systematic review: Surveillance systems for early detection of bioterrorism-related diseases. *Ann. Intern. Med.* **140** 910–922.
- [6] BROWN, L. D., ZHANG, R. and ZHAO, L. (2001). Root un-root methodology for nonparametric density estimation. Technical Report, The Wharton School, Univ. Pennsylvania.
- [7] CAI, T. T., JENG, X. J. and JIN, J. (2011). Optimal detection of heterogeneous and heteroscedastic mixtures. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 629–662. [MR2867452](#) <https://doi.org/10.1111/j.1467-9868.2011.00778.x>
- [8] CAI, T. T. and WU, Y. (2014). Optimal detection of sparse mixtures against a given null distribution. *IEEE Trans. Inf. Theory* **60** 2217–2232. [MR3181520](#) <https://doi.org/10.1109/TIT.2014.2304295>

- [9] CHAN, S.-O., DIAKONIKOLAS, I., VALIANT, G. and VALIANT, P. (2014). Optimal algorithms for testing closeness of discrete distributions. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms* 1193–1203. ACM, New York. MR3376448 <https://doi.org/10.1137/1.9781611973402.88>
- [10] CLOPPER, C. J. and PEARSON, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26** 404–413.
- [11] CORLESS, R. M., GONNET, G. H., HARE, D. E. G., JEFFREY, D. J. and KNUTH, D. E. (1996). On the Lambert W function. *Adv. Comput. Math.* **5** 329–359. MR1414285 <https://doi.org/10.1007/BF02124750>
- [12] DEMBO, A. and ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. *Applications of Mathematics (New York)* **38**. Springer, New York. MR1619036 <https://doi.org/10.1007/978-1-4612-5320-4>
- [13] DIAKONIKOLAS, I., GOULEAKIS, T., PEEBLES, J. and PRICE, E. (2019). Collision-based testers are optimal for uniformity and closeness. *Chic. J. Theoret. Comput. Sci. Art.* **1**, 21. MR4008852 <https://doi.org/10.4086/cjcs.2019.001>
- [14] DIAKONIKOLAS, I. and KANE, D. M. (2016). A new approach for testing properties of discrete distributions. In *57th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2016* 685–694. IEEE Computer Soc., Los Alamitos, CA. MR3631031
- [15] 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>
- [16] DONOHO, D. and JIN, J. (2009). Feature selection by higher criticism thresholding achieves the optimal phase diagram. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **367** 4449–4470. MR2546396 <https://doi.org/10.1098/rsta.2009.0129>
- [17] DONOHO, D. and JIN, J. (2015). Higher criticism for large-scale inference, especially for rare and weak effects. *Statist. Sci.* **30** 1–25. MR3317751 <https://doi.org/10.1214/14-ST506>
- [18] EFRON, B. (1982). Transformation theory: How normal is a family of distributions? *Ann. Statist.* **10** 323–339. MR0653511
- [19] GONTSCHARUK, V., LANDWEHR, S. and FINNER, H. (2015). The intermediates take it all: Asymptotics of higher criticism statistics and a powerful alternative based on equal local levels. *Biom. J.* **57** 159–180. MR3298224 <https://doi.org/10.1002/bimj.201300255>
- [20] Hoeffding, W. (1994). *The Collected Works of Wassily Hoeffding. Springer Series in Statistics. Perspectives in Statistics*. Springer, New York. MR1307621
- [21] INGSTER, J. I. (1996). *On Some Problems of Hypothesis Testing Leading to Infinitely Divisible Distributions. Weierstrass-Institut Für Angewandte Analysis und Stochastik Berlin: Preprint*. Weierstraß-Inst. für Angewandte Analysis und Stochastik, Berlin.
- [22] 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>
- [23] JAGER, L. and WELLNER, J. A. (2007). Goodness-of-fit tests via phi-divergences. *Ann. Statist.* **35** 2018–2053. MR2363962 <https://doi.org/10.1214/00090536070000000244>
- [24] JANSSEN, A. (2000). Global power functions of goodness of fit tests. *Ann. Statist.* **28** 239–253. MR1762910 <https://doi.org/10.1214/aos/1016120371>
- [25] JIN, J. (2003). *Detecting and Estimating Sparse Mixtures*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—Stanford University. MR2704645
- [26] DONOHO, D. L. and JIN, J. (2008). Higher criticism thresholding: Optimal feature selection when useful features are rare and weak. *Proc. Natl. Acad. Sci. USA* **105** 14790–14795.
- [27] KIPNIS, A. (2022). Higher criticism for discriminating word-frequency tables and authorship attribution. *Ann. Appl. Statist.* To appear.
- [28] KIPNIS, A. (2021). Unification of rare/weak detection models using moderate deviations analysis and log-chisquared P-values. Available at [arXiv:2103.03999](https://arxiv.org/abs/2103.03999).
- [29] KIPNIS, A. and DONOHO, D. L. (2021). Two-sample testing of discrete distributions under rare/weak perturbations. In *2021 IEEE International Symposium on Information Theory (ISIT)* 3314–3319. IEEE, New York.
- [30] KIPNIS, A. and DONOHO, D. L. (2021). On the inability of the higher criticism to detect rare/weak departures. Preprint. Available at [arXiv:2103.03218](https://arxiv.org/abs/2103.03218).
- [31] LE CAM, L. (2012). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.
- [32] MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models*. 2nd ed. *Monographs on Statistics and Applied Probability*. CRC Press, London. MR3223057 <https://doi.org/10.1007/978-1-4899-3242-6>

- [33] MILLER, B., KASSENBERG, H., DUNSMUIR, W., GRIFFITH, J., HADIDI, M., NORDIN, J. D. and DANILA, R. (2004). Syndromic surveillance for influenzalike illness in ambulatory care setting. *Emerg. Infect. Dis.* **10** 1806.
- [34] MITZENMACHER, M. and UPFAL, E. (2017). *Probability and Computing. Randomization and Probabilistic Techniques in Algorithms and Data Analysis*, 2nd ed. Cambridge Univ. Press, Cambridge. [MR3674428](#)
- [35] MOSTELLER, F. and WALLACE, D. L. (1963). Inference in an authorship problem. *J. Amer. Statist. Assoc.* **58** 275–309.
- [36] MUKHERJEE, R., PILLAI, N. S. and LIN, X. (2015). Hypothesis testing for high-dimensional sparse binary regression. *Ann. Statist.* **43** 352–381. [MR3311863](#) <https://doi.org/10.1214/14-AOS1279>
- [37] NUSSBAUM, M. and KLEMELÄ, J. (2006). *Constructive Asymptotic Equivalence of Density Estimation and Gaussian White Noise*. Humboldt-Universität zu Berlin, Wirtschaftswissenschaftliche Fakultät, Berlin.
- [38] RUBIN, H. and SETHURAMAN, J. (1965). Probabilities of moderate deviations. *Sankhyā Ser. A* **27** 325–346. [MR0203783](#)
- [39] SHORACK, G. R. and WELLNER, J. A. (2009). *Empirical Processes with Applications to Statistics. Classics in Applied Mathematics* **59**. SIAM, Philadelphia, PA. [MR3396731](#) <https://doi.org/10.1137/1.9780898719017.ch1>
- [40] STÉPHANE, B., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, London.

A NEW AND FLEXIBLE DESIGN CONSTRUCTION FOR ORTHOGONAL ARRAYS FOR MODERN APPLICATIONS

BY YUANZHEN HE^{1,a} C. DEVON LIN^{2,b} AND FASHENG SUN^{3,c}

¹*School of Statistics, Beijing Normal University, heyuanzhen@bnu.edu.cn*

²*Department of Mathematics and Statistics, Queen's University, devon.lin@queensu.ca*

³*KLAS and School of Mathematics and Statistics, Northeast Normal University, sunfs359@nenu.edu.cn*

Orthogonal array, a classical and effective tool for collecting data, has been flourished with its applications in modern computer experiments and engineering statistics. Driven by the wide use of computer experiments with both qualitative and quantitative factors, multiple computer experiments, multifidelity computer experiments, cross-validation and stochastic optimization, orthogonal arrays with certain structures have been introduced. Sliced orthogonal arrays and nested orthogonal arrays are examples of such arrays. This article introduces a flexible, fresh construction method, which uses smaller arrays and a special structure. The method uncovers the hidden structure of many existing fixed-level orthogonal arrays of given run sizes, possibly with more columns. It also allows fixed-level orthogonal arrays of nearly strength three to be constructed, which are useful as there are not many construction methods for fixed-level orthogonal arrays of strength three, and also helpful for generating Latin hypercube designs with desirable low-dimensional projections. Theoretical properties of the proposed method are explored. As by-products, several theoretical results on orthogonal arrays are obtained.

REFERENCES

- [1] AI, M., JIANG, B. and LI, K. (2014). Construction of sliced space-filling designs based on balanced sliced orthogonal arrays. *Statist. Sinica* **24** 1685–1702. [MR3308657](#)
- [2] BOSE, R. C. and BUSH, K. A. (1952). Orthogonal arrays of strength two and three. *Ann. Math. Stat.* **23** 508–524. [MR0051204](#) <https://doi.org/10.1214/aoms/1177729331>
- [3] DENG, X., HUNG, Y. and LIN, C. D. (2015). Design for computer experiments with qualitative and quantitative factors. *Statist. Sinica* **25** 1567–1581. [MR3409081](#)
- [4] DENG, X., LIN, C. D., LIU, K. W. and ROWE, R. K. (2017). Additive Gaussian process for computer models with qualitative and quantitative factors. *Technometrics* **25** 1567–1581.
- [5] DEY, A. and MUKERJEE, R. (1999). *Fractional Factorial Plans*. *Wiley Series in Probability and Statistics: Probability and Statistics*. Wiley, New York. [MR1679441](#) <https://doi.org/10.1002/9780470316986>
- [6] EENDEBAK, P. and SCHOEN, E. (2017). Complete series of non-isomorphic orthogonal arrays. Available at <http://www.pietereendebak.nl/oapackage/series.html>.
- [7] GOH, J., BINGHAM, D., HOLLOWAY, J. P., GROSSKOPF, M. J., KURANZ, C. C. and RUTTER, E. (2013). Prediction and computer model calibration using outputs from multifidelity simulators. *Technometrics* **55** 501–512. [MR3176554](#) <https://doi.org/10.1080/00401706.2013.838910>
- [8] GOPALAKRISHNAN, K. and STINSON, D. R. (2006). Applications of orthogonal arrays to computer science. In *Proceeding of ICDM* 149–164.
- [9] GUPTA, V. K., NIGAM, A. K. and DEY, A. (1982). Orthogonal main-effect plans for asymmetrical factorials. *Technometrics* **24** 135–137. [MR0655577](#) <https://doi.org/10.2307/1268491>
- [10] HAALAND, B. and QIAN, P. Z. G. (2011). Accurate emulators for large-scale computer experiments. *Ann. Statist.* **39** 2974–3002. [MR3012398](#) <https://doi.org/10.1214/11-AOS929>
- [11] HE, X. and QIAN, P. Z. G. (2014). A central limit theorem for general orthogonal array based space-filling designs. *Ann. Statist.* **42** 1725–1750. [MR3262466](#) <https://doi.org/10.1214/14-AOS1231>

MSC2020 subject classifications. Primary 62K15; secondary 62K05.

Key words and phrases. Completely resolvable, difference scheme, fractional factorial design, computer experiment, nested orthogonal array, sliced orthogonal array.

- [12] HE, X. and QIAN, P. Z. G. (2016). A central limit theorem for nested or sliced Latin hypercube designs. *Statist. Sinica* **26** 1117–1128. [MR3559945](#)
- [13] HE, Y., LIN, C. D. and SUN, F. (2017). On construction of marginally coupled designs. *Statist. Sinica* **27** 665–683. [MR3674691](#)
- [14] HE, Y., LIN, C. D. and SUN, F. (2019). Construction of marginally coupled designs by subspace theory. *Bernoulli* **25** 2163–2182. [MR3961244](#) <https://doi.org/10.3150/18-BEJ1049>
- [15] HE, Y., LIN, C. D. and SUN, F. (2022). Supplement to “A new and flexible design construction for orthogonal arrays for modern applications.” <https://doi.org/10.1214/21-AOS2159SUPP>
- [16] HE, Y., LIN, C. D., SUN, F. and LV, B. (2017). Marginally coupled designs for two-level qualitative factors. *J. Statist. Plann. Inference* **187** 103–108. [MR3638045](#) <https://doi.org/10.1016/j.jspi.2017.02.010>
- [17] 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>
- [18] HEDAYAT, A. S. (1986). Orthogonal arrays of strength $t+$ and their statistical applications. Technical report, Univ. Illinois, Chicago, IL.
- [19] HEDAYAT, A. S. (1990). New properties of orthogonal arrays and their statistical applications. In *Statistical Design and Analysis of Industrial Experiments* (S. Ghosh, ed.) 407–422. Dekker, New York.
- [20] HEDAYAT, A. S. (2017). A scientific tour on orthogonal arrays. In *Mathematics Across Contemporary Sciences* (T. Abualrub, A. Jarrah, S. Kallel and H. Sulieman, eds.). *Springer Proc. Math. Stat.* **190** 111–135. Springer, Cham. [MR3630785](#) https://doi.org/10.1007/978-3-319-46310-0_7
- [21] HEDAYAT, A. S., SLOANE, N. J. A. and STUFKEN, J. (1999). *Orthogonal Arrays: Theory and Applications*. *Springer Series in Statistics*. Springer, New York. [MR1693498](#) <https://doi.org/10.1007/978-1-4612-1478-6>
- [22] HUANG, H., LIN, D. K. J., LIU, M.-Q. and YANG, J.-F. (2016). Computer experiments with both qualitative and quantitative variables. *Technometrics* **58** 495–507. [MR3556617](#) <https://doi.org/10.1080/00401706.2015.1094416>
- [23] HWANG, Y., QIAN, P. Z. G. and HE, X. (2016). Sliced orthogonal array-based Latin hypercube designs. *Technometrics* **58** 50–61. [MR3463156](#) <https://doi.org/10.1080/00401706.2014.993092>
- [24] KONG, X., AI, M. and TSUI, K. L. (2018). Flexible sliced designs for computer experiments. *Ann. Inst. Statist. Math.* **70** 631–646. [MR3785710](#) <https://doi.org/10.1007/s10463-017-0603-3>
- [25] LEKIVETZ, R., SITTER, R., BINGHAM, D., HAMADA, M. S., MOORE, L. M. and WENDELBERGER, J. R. (2015). On algorithms for obtaining orthogonal and near-orthogonal arrays for main-effects screening. *J. Qual. Technol.* **47** 2–13.
- [26] LIN, C. D., SITTER, R. R. and TANG, B. (2012). Creating catalogues of two-level nonregular fractional factorial designs based on the criteria of generalized aberration. *J. Statist. Plann. Inference* **142** 445–456. [MR2843048](#) <https://doi.org/10.1016/j.jspi.2011.08.003>
- [27] 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>
- [28] MUKERJEE, R. and WU, C. F. J. (2006). *A Modern Theory of Factorial Designs*. *Springer Series in Statistics*. Springer, New York. [MR2230487](#)
- [29] OWEN, A. B. (1992). Orthogonal arrays for computer experiments, integration and visualization. *Statist. Sinica* **2** 439–452. [MR1187952](#)
- [30] QIAN, P. Z. G., AI, M. and WU, C. F. J. (2009). Construction of nested space-filling designs. *Ann. Statist.* **37** 3616–3643. [MR2549572](#) <https://doi.org/10.1214/09-AOS690>
- [31] QIAN, P. Z. G., TANG, B. and WU, C. F. J. (2009). Nested space-filling designs for computer experiments with two levels of accuracy. *Statist. Sinica* **19** 287–300. [MR2487890](#)
- [32] QIAN, P. Z. G. and WU, C. F. J. (2009). Sliced space-filling designs. *Biometrika* **96** 945–956. [MR2767280](#) <https://doi.org/10.1093/biomet/asp044>
- [33] QIAN, P. Z. G., WU, H. and WU, C. F. J. (2008). Gaussian process models for computer experiments with qualitative and quantitative factors. *Technometrics* **50** 383–396. [MR2457574](#) <https://doi.org/10.1198/004017008000000262>
- [34] SACKS, J., WELCH, W. J., MITCHELL, T. J. and WYNN, H. P. (1989). Design and analysis of computer experiments. *Statist. Sci.* **4** 409–435. [MR1041765](#)
- [35] 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. [MR3887662](#)
- [36] SLOANE, N. J. A. (2004). A library of orthogonal arrays. Available at <http://neilsloane.com/oadir/>.
- [37] SUEN, C.-Y. (1989). Some resolvable orthogonal arrays with two symbols. *Comm. Statist. Theory Methods* **18** 3875–3881. [MR1040681](#) <https://doi.org/10.1080/03610928908830128>
- [38] SUN, F., LIU, M.-Q. and QIAN, P. Z. G. (2014). On the construction of nested space-filling designs. *Ann. Statist.* **42** 162–193. [MR3226161](#) <https://doi.org/10.1214/14-AOS1229>

- [39] SUN, F., WANG, Y. and XU, H. (2019). Uniform projection designs. *Ann. Statist.* **47** 641–661. [MR3909945](#)
<https://doi.org/10.1214/18-AOS1705>
- [40] TANG, B. (1993). Orthogonal array-based Latin hypercubes. *J. Amer. Statist. Assoc.* **88** 1392–1397.
[MR1245375](#)
- [41] WANG, J. C. and WU, C. F. J. (1992). Nearly orthogonal arrays with mixed levels and small runs. *Technometrics* **34** 409–422. [MRMR1939683](#)
- [42] WEI, Y. and XIONG, S. (2019). Bayesian integrative analysis for multi-fidelity computer experiments. *J. Appl. Stat.* **46** 1973–1987. [MR3963046](#) <https://doi.org/10.1080/02664763.2019.1575340>
- [43] WU, C. F. J. and HAMADA, M. S. (2009). *Experiments: Planning, Analysis, and Optimization*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. [MR2583259](#)
- [44] XU, H. (2002). An algorithm for constructing orthogonal and nearly-orthogonal arrays with mixed levels and small runs. *Technometrics* **44** 356–368. [MR1939683](#) <https://doi.org/10.1198/004017002188618554>
- [45] XU, H., CHENG, S.-W. and WU, C. F. J. (2004). Optimal projective three-level designs for factor screening and interaction detection. *Technometrics* **46** 280–292. [MR2082498](#) <https://doi.org/10.1198/004017004000000310>
- [46] XU, H., PHOA, F. K. H. and WONG, W. K. (2009). Recent developments in nonregular fractional factorial designs. *Stat. Surv.* **3** 18–46. [MR2520978](#) <https://doi.org/10.1214/08-SS040>
- [47] YANG, F., LIN, C. D., ZHOU, Y. D. and HE, Y. (2021). Doubly coupled designs for computer experiments with both qualitative and quantitative factors. *Statist. Sinica*. <https://doi.org/0.5705/ss.202020.0317>
- [48] YANG, J.-F., LIN, C. D., QIAN, P. Z. G. and LIN, D. K. J. (2013). Construction of sliced orthogonal Latin hypercube designs. *Statist. Sinica* **23** 1117–1130. [MR3114707](#)
- [49] YANG, X., CHEN, H. and LIU, M.-Q. (2014). Resolvable orthogonal array-based uniform sliced Latin hypercube designs. *Statist. Probab. Lett.* **93** 108–115. [MR3244562](#) <https://doi.org/10.1016/j.spl.2014.06.021>
- [50] ZHANG, Y., TAO, S., CHEN, W. and APLEY, D. W. (2020). A latent variable approach to Gaussian process modeling with qualitative and quantitative factors. *Technometrics* **62** 291–302. [MR4125497](#)
<https://doi.org/10.1080/00401706.2019.1638834>
- [51] ZHOU, Y. and TANG, B. (2019). Column-orthogonal strong orthogonal arrays of strength two plus and three minus. *Biometrika* **106** 997–1004. [MR4031213](#) <https://doi.org/10.1093/biomet/asz043>

CONSISTENCY OF BAYESIAN INFERENCE FOR MULTIVARIATE MAX-STABLE DISTRIBUTIONS

BY SIMONE A. PADOAN^{1,a} AND STEFANO RIZZELLI^{2,b}

¹Department of Decision Sciences, Bocconi University, ^asimone.padoan@unibocconi.it

²Department of Statistical Sciences, Catholic University, ^bstefano.rizzelli@unicatt.it

Predicting extreme events is important in many applications in risk analysis. Extreme-value theory suggests modelling extremes by max-stable distributions. The Bayesian approach provides a natural framework for statistical prediction. Although various Bayesian inferential procedures have been proposed in the literature of univariate extremes and some for multivariate extremes, the study of their asymptotic properties has been left largely untouched. In this paper we focus on a semiparametric Bayesian method for estimating max-stable distributions in arbitrary dimension. We establish consistency of the pertaining posterior distributions for fairly general, well-specified max-stable models, whose margins can be short-, light- or heavy-tailed. We then extend our consistency results to the case where data are samples of block maxima whose distribution is only approximately a max-stable one, which represents the most realistic inferential setting.

REFERENCES

- [1] ASADI, P., DAVISON, A. C. and ENGELKE, S. (2015). Extremes on river networks. *Ann. Appl. Stat.* **9** 2023–2050. [MR3456363](#) <https://doi.org/10.1214/15-AOAS863>
- [2] BARRIENTOS, A. F., JARA, A. and QUINTANA, F. A. (2015). Bayesian density estimation for compositional data using random Bernstein polynomials. *J. Statist. Plann. Inference* **166** 116–125. [MR3390138](#) <https://doi.org/10.1016/j.jspi.2015.01.006>
- [3] BEIRLANT, J., GÖEGBEUR, Y., TEUGELS, J. and SEGERS, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley Series in Probability and Statistics. Wiley, Chichester. [MR2108013](#) <https://doi.org/10.1002/0470012382>
- [4] BERANGER, B. and PADOAN, S. A. (2016). Extreme dependence models. In *Extreme Value Modeling and Risk Analysis* 325–352. CRC Press, Boca Raton, FL. [MR3644320](#) <https://doi.org/10.1201/b19721-20>
- [5] BOLDI, M.-O. and DAVISON, A. C. (2007). A mixture model for multivariate extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 217–229. [MR2325273](#) <https://doi.org/10.1111/j.1467-9868.2007.00585.x>
- [6] BÜCHER, A. and SEGERS, J. (2017). On the maximum likelihood estimator for the generalized extreme-value distribution. *Extremes* **20** 839–872. [MR3737387](#) <https://doi.org/10.1007/s10687-017-0292-6>
- [7] BÜCHER, A. and SEGERS, J. (2018). Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series. *Bernoulli* **24** 1427–1462. [MR3706798](#) <https://doi.org/10.3150/16-BEJ903>
- [8] BUHL, S. and KLÜPPELBERG, C. (2016). Anisotropic Brown–Resnick space–time processes: Estimation and model assessment. *Extremes* **19** 627–660. [MR3558348](#) <https://doi.org/10.1007/s10687-016-0257-1>
- [9] CASTILLO, E., HADI, A. S., BALAKRISHNAN, N. and SARABIA, J. M. (2005). *Extreme Value and Related Models with Applications in Engineering and Science*. Wiley Series in Probability and Statistics. Wiley Interscience, Hoboken, NJ. [MR2191401](#)
- [10] CHOUDHURI, N., GHOSAL, S. and ROY, A. (2004). Bayesian estimation of the spectral density of a time series. *J. Amer. Statist. Assoc.* **99** 1050–1059. [MR2109494](#) <https://doi.org/10.1198/016214504000000557>
- [11] COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. Springer London, Ltd., London. [MR1932132](#) <https://doi.org/10.1007/978-1-4471-3675-0>

MSC2020 subject classifications. Primary 62G20, 62G32; secondary 60G70, 62C10.

Key words and phrases. Bernstein polynomials, extreme-value copula, multivariate max-stable distribution, nonparametric estimation, Pickands dependence function, angular measure, posterior consistency.

- [12] COLES, S. and PERICCHI, L. (2003). Anticipating catastrophes through extreme value modelling. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **52** 405–416. [MR2012566](#) <https://doi.org/10.1111/1467-9876.00413>
- [13] CORMIER, E., GENEST, C. and NEŠLEHOVÁ, J. G. (2014). Using B-splines for nonparametric inference on bivariate extreme-value copulas. *Extremes* **17** 633–659. [MR3282690](#) <https://doi.org/10.1007/s10687-014-0199-4>
- [14] DAVISON, A. C., PADOAN, S. A. and RIBATET, M. (2012). Statistical modeling of spatial extremes. *Statist. Sci.* **27** 199–201. [MR2963980](#) <https://doi.org/10.1214/11-STS376>
- [15] DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer, New York. [MR2234156](#) <https://doi.org/10.1007/0-387-34471-3>
- [16] DELLACHERIE, C. and MEYER, P.-A. (1982). *Probabilities and Potential. B*. North-Holland Mathematics Studies **72**. North-Holland, Amsterdam. [MR0745449](#) [https://doi.org/10.1016/s0304-0208\(08\)x7175-0](https://doi.org/10.1016/s0304-0208(08)x7175-0)
- [17] DEY, D. and YAN, J. (2016). *Extreme Value Modeling and Risk Analysis*. CRC Press, Boca Raton, FL. [MR3644305](#) <https://doi.org/10.1201/b19721>
- [18] DOMBRY, C. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli* **21** 420–436. [MR3322325](#) <https://doi.org/10.3150/13-BEJ573>
- [19] DOMBRY, C., ENGELKE, S. and OESTING, M. (2017). Asymptotic properties of the maximum likelihood estimator for multivariate extreme value distributions. ArXiv e-prints. Available at <https://arxiv.org/abs/1612.05178>.
- [20] DOMBRY, C., ENGELKE, S. and OESTING, M. (2017). Bayesian inference for multivariate extreme value distributions. *Electron. J. Stat.* **11** 4813–4844. [MR3729660](#) <https://doi.org/10.1214/17-EJS1367>
- [21] DOMBRY, C. and EYI-MINKO, F. (2013). Regular conditional distributions of continuous max-infinitely divisible random fields. *Electron. J. Probab.* **18** 7–21. [MR3024101](#) <https://doi.org/10.1214/EJP.v18-1991>
- [22] EINMAHL, J. H. J. and MASON, D. M. (1992). Generalized quantile processes. *Ann. Statist.* **20** 1062–1078. [MR1165606](#) <https://doi.org/10.1214/aos/1176348670>
- [23] FALK, M., HÜSLER, J. and REISS, R.-D. (2011). *Laws of Small Numbers: Extremes and Rare Events*, extended ed. Birkhäuser/Springer Basel AG, Basel. [MR2732365](#) <https://doi.org/10.1007/978-3-0348-0009-9>
- [24] FALK, M., PADOAN, S. A. and RIZZELLI, S. (2020). Strong convergence of multivariate maxima. *J. Appl. Probab.* **57** 314–331. [MR4094397](#) <https://doi.org/10.1017/jpr.2019.100>
- [25] FOURNIER, N. and GUILLIN, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* **162** 707–738. [MR3383341](#) <https://doi.org/10.1007/s00440-014-0583-7>
- [26] GHOSAL, S. (2001). Convergence rates for density estimation with Bernstein polynomials. *Ann. Statist.* **29** 1264–1280. [MR1873330](#) <https://doi.org/10.1214/aos/1013203453>
- [27] GHOSAL, S. and VAN DER VAART, A. W. (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>
- [28] GNEITING, T., BALABDAOUI, F. and RAFTERY, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 243–268. [MR2325275](#) <https://doi.org/10.1111/j.1467-9868.2007.00587.x>
- [29] GUILLOTTE, S. and PERRON, F. (2016). Polynomial Pickands functions. *Bernoulli* **22** 213–241. [MR3449781](#) <https://doi.org/10.3150/14-BEJ656>
- [30] GUILLOTTE, S., PERRON, F. and SEGERS, J. (2011). Non-parametric Bayesian inference on bivariate extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 377–406. [MR2815781](#) <https://doi.org/10.1111/j.1467-9868.2010.00770.x>
- [31] HANSON, T. E., DE CARVALHO, M. and CHEN, Y. (2017). Bernstein polynomial angular densities of multivariate extreme value distributions. *Statist. Probab. Lett.* **128** 60–66. [MR3656376](#) <https://doi.org/10.1016/j.spl.2017.03.030>
- [32] HUSER, R., DOMBRY, C., RIBATET, M. and GENTON, M. G. (2019). Full likelihood inference for max-stable data. *Stat* **8** e218. [MR3938281](#) <https://doi.org/10.1002/sta4.218>
- [33] JOE, H. (2015). *Dependence Modeling with Copulas*. Monographs on Statistics and Applied Probability **134**. CRC Press, Boca Raton, FL. [MR3328438](#)
- [34] KLEIJN, B. J. K. (2021). Frequentist validity of Bayesian limits. *Ann. Statist.* **49** 182–202. [MR4206674](#) <https://doi.org/10.1214/20-AOS1952>
- [35] KLEIJN, B. J. K. and VAN DER VAART, A. W. (2006). Misspecification in infinite-dimensional Bayesian statistics. *Ann. Statist.* **34** 837–877. [MR2283395](#) <https://doi.org/10.1214/009053606000000029>

- [36] KLÜPPELBERG, C. and MAY, A. (2006). Bivariate extreme value distributions based on polynomial dependence functions. *Math. Methods Appl. Sci.* **29** 1467–1480. MR2247312 <https://doi.org/10.1002/mma.525>
- [37] KRUIJER, W. and VAN DER VAART, A. W. (2008). Posterior convergence rates for Dirichlet mixtures of beta densities. *J. Statist. Plann. Inference* **138** 1981–1992. MR2406419 <https://doi.org/10.1016/j.jspi.2007.07.012>
- [38] LONGIN, F., ed. (2017). *Extreme Events in Finance: A Handbook of Extreme Value Theory and Its Applications. Wiley Handbook in Financial Engineering and Econometrics*. Wiley, Hoboken, NJ. MR3523908 <https://doi.org/10.1002/9781118650318>
- [39] MARCON, G., PADOAN, S. A. and ANTONIANO–VILLALOBOS, I. (2016). Bayesian inference for the extremal dependence. *Electron. J. Stat.* **10** 3310–3337. MR3572851 <https://doi.org/10.1214/16-EJS1162>
- [40] MARCON, G., PADOAN, S. A., NAVEAU, P., MULIERE, P. and SEGERS, J. (2017). Multivariate nonparametric estimation of the Pickands dependence function using Bernstein polynomials. *J. Statist. Plann. Inference* **183** 1–17. MR3589492 <https://doi.org/10.1016/j.jspi.2016.10.004>
- [41] MARDIA, K. V. (1970). *Families of Bivariate Distributions. Griffin's Statistical Monographs & Courses, No. 27*. Hafner Publishing Co., Darien, CT. MR0319318
- [42] NG, K. W., TIAN, G.-L. and TANG, M.-L. (2011). *Dirichlet and Related Distributions: Theory, Methods and Applications. Wiley Series in Probability and Statistics*. Wiley, Chichester. MR2830563 <https://doi.org/10.1002/9781119995784>
- [43] 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>
- [44] NORTHROP, P. J. and ATTALIDES, N. (2016). Posterior propriety in Bayesian extreme value analyses using reference priors. *Statist. Sinica* **26** 721–743. MR3497768 <https://doi.org/10.5705/ss.2014.034>
- [45] PADOAN, S. A., RIBATET, M. and SISSON, S. A. (2010). Likelihood-based inference for max-stable processes. *J. Amer. Statist. Assoc.* **105** 263–277. MR2757202 <https://doi.org/10.1198/jasa.2009.tm08577>
- [46] PADOAN, S. A. and RIZZELLI, S. (2022). Supplement to “Consistency of Bayesian inference for multivariate max-stable distributions.” <https://doi.org/10.1214/21-AOS2160SUPP>
- [47] PENG, L. and QI, Y. (2017). *Inference for Heavy-Tailed Data: Applications in Insurance and Finance*. Academic Press, London. <https://doi.org/10.1016/C2015-0-01492-7>
- [48] PETRONE, S., RIZZELLI, S., ROUSSEAU, J. and SCRICCILO, C. (2014). Empirical Bayes methods in classical and Bayesian inference. *Metron* **72** 201–215. MR3233149 <https://doi.org/10.1007/s40300-014-0044-1>
- [49] PETRONE, S. and WASSERMAN, L. (2002). Consistency of Bernstein polynomial posteriors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 79–100. MR1881846 <https://doi.org/10.1111/1467-9868.00326>
- [50] RESNICK, S. I. (2008). *Extreme Values, Regular Variation and Point Processes. Springer Series in Operations Research and Financial Engineering*. Springer, New York. MR2364939 <https://doi.org/10.1007/978-0-387-75953-1>
- [51] RITOV, Y., BICKEL, P. J., GAMST, A. C. and KLEIJN, B. J. K. (2014). The Bayesian analysis of complex, high-dimensional models: Can it be CODA? *Statist. Sci.* **29** 619–639. MR3300362 <https://doi.org/10.1214/14-STS483>
- [52] SABOURIN, A. and NAVEAU, P. (2014). Bayesian Dirichlet mixture model for multivariate extremes: A reparametrization. *Comput. Statist. Data Anal.* **71** 542–567. MR3131989 <https://doi.org/10.1016/j.csda.2013.04.021>
- [53] SHIN, Y., LEE, Y., CHOI, J. and PARK, J.-S. (2019). Integration of max-stable processes and Bayesian model averaging to predict extreme climatic events in multi-model ensembles. *Stoch. Environ. Res. Risk Assess.* **33** 47–57. <https://doi.org/10.1007/s00477-018-1629-7>
- [54] SRIVASTAVA, S. M. (1998). *A Course on Borel Sets. Graduate Texts in Mathematics* **180**. Springer, New York. MR1619545 <https://doi.org/10.1007/978-3-642-85473-6>
- [55] STEPHENSON, A. and TAWN, J. (2004). Bayesian inference for extremes: Accounting for the three extremal types. *Extremes* **7** 291–307. MR2212389 <https://doi.org/10.1007/s10687-004-3479-6>
- [56] THIBAUD, E., AALTO, J., COOLEY, D. S., DAVISON, A. C. and HEIKKINEN, J. (2016). Bayesian inference for the Brown–Resnick process, with an application to extreme low temperatures. *Ann. Appl. Stat.* **10** 2303–2324. MR3592058 <https://doi.org/10.1214/16-AOAS980>
- [57] WADSWORTH, J. L. and TAWN, J. A. (2014). Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika* **101** 1–15. MR3180654 <https://doi.org/10.1093/biomet/ast042>

MINIMAX ESTIMATION OF SMOOTH DENSITIES IN WASSERSTEIN DISTANCE

BY JONATHAN NILES-WEED^{1,a} AND QUENTIN BERTHET^{2,b}

¹*Courant Institute of Mathematical Sciences, New York University, jnw@cims.nyu.edu*

²*Google Research, qberthet@google.com*

We study nonparametric density estimation problems where error is measured in the Wasserstein distance, a metric on probability distributions popular in many areas of statistics and machine learning. We give the first minimax-optimal rates for this problem for general Wasserstein distances for two classes of densities: smooth probability densities on $[0, 1]^d$ bounded away from 0, and sub-Gaussian densities lying in the Hölder class C^s , $s \in (0, 1)$. Unlike classical nonparametric density estimation, these rates depend on whether the densities in question are bounded below, even in the compactly supported case. Motivated by variational problems involving the Wasserstein distance, we also show how to construct discretely supported measures, suitable for computational purposes, which achieve the minimax rates. Our main technical tool is an inequality giving a nearly tight dual characterization of the Wasserstein distances in terms of Besov norms.

REFERENCES

- ALAUX, J., GRAVE, E., CUTURI, M. and JOULIN, A. (2019). Unsupervised hyper-alignment for multilingual word embeddings. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6–9, 2019*. OpenReview.net. Available at <https://openreview.net/forum?id=HJe62s09tX>.
- ALTSCHULER, J., WEED, J. and RIGOLLET, P. (2017). Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4–9 December 2017, Long Beach, CA, USA 1961–1971*. Available <http://papers.nips.cc/paper/6792-near-linear-time-approximation-algorithms-for-optimal-transport-via-sinkhorn-iteration>.
- ALVAREZ-MELIS, D., JAAKKOLA, T. S. and JEGELKA, S. (2018). Structured optimal transport. In *International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9–11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain* (A. J. Storkey and F. Pérez-Cruz, eds.). *Proceedings of Machine Learning Research* **84** 1771–1780. PMLR. Available at <http://proceedings.mlr.press/v84/alvarez-melis18a.html>.
- ARJOVSKY, M., CHINTALA, S. and BOTTOU, L. (2017). Wasserstein GAN. ArXiv preprint. Available at [arXiv:1701.07875](https://arxiv.org/abs/1701.07875).
- BASSETTI, F., BODINI, A. and REGAZZINI, E. (2006). On minimum Kantorovich distance estimators. *Statist. Probab. Lett.* **76** 1298–1302. <https://doi.org/10.1016/j.spl.2006.02.001>
- BENAMOU, J.-D. and BRENIER, Y. (2000). A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numer. Math.* **84** 375–393. <https://doi.org/10.1007/s002110050002>
- BERNTON, E., JACOB, P. E., GERBER, M. and ROBERT, C. P. (2019). On parameter estimation with the Wasserstein distance. *Inf. Inference* **8** 657–676.
- BOISSARD, E. and LE GOUIC, T. (2014). On the mean speed of convergence of empirical and occupation measures in Wasserstein distance. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** 539–563. <https://doi.org/10.1214/12-AIHP517>
- BRENIER, Y. (2003). Extended Monge–Kantorovich theory. In *Optimal Transportation and Applications. Lecture Notes in Math.* **1813** 91–121. Springer, Berlin. https://doi.org/10.1007/978-3-540-44857-0_4
- CHEN, G., IWEN, M. A., CHIN, S. and MAGGIONI, M. (2012). A fast multiscale framework for data in high-dimensions: Measure estimation, anomaly detection, and compressive measurements. In *2012 Visual Communications and Image Processing, VCIP 2012, San Diego, CA, USA, November 27–30, 2012* 1–6. <https://doi.org/10.1109/VCIP.2012.6410789>

MSC2020 subject classifications. Primary 62F99; secondary 62H99.

Key words and phrases. Wasserstein distance, optimal transport, high-dimensional statistics, density estimation.

- COHEN, A. (2003). *Numerical Analysis of Wavelet Methods. Studies in Mathematics and Its Applications*. **32**. North-Holland, Amsterdam.
- CUTURI, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a Meeting Held December 5–8, 2013* 2292–2300, Lake Tahoe, NV, United States. Available at <http://papers.nips.cc/paper/4927-sinkhorn-distances-lightspeed-computation-of-optimal-transport>.
- CUTURI, M. and DOUCET, A. (2014). Fast computation of Wasserstein barycenters. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21–26 June 2014. JMLR Workshop and Conference Proceedings* **32** 685–693. JMLR.org. Available at <http://proceedings.mlr.press/v32/cuturi14.html>.
- DOMINITZ, A., ANGENENT, S. and TANNENBAUM, A. (2008). On the computation of optimal transport maps using gradient flows and multiresolution analysis. In *Recent Advances in Learning and Control. Lect. Notes Control Inf. Sci.* **371** 65–78. Springer, London. https://doi.org/10.1007/978-1-84800-155-8_5
- DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24** 508–539. <https://doi.org/10.1214/aos/1032894451>
- DOUKHAN, P. and LEÓN, J. R. (1990). Déviation quadratique d’estimateurs de densité par projections orthogonales. *C. R. Acad. Sci. Paris Sér. I Math.* **310** 425–430.
- DUDLEY, R. M. (1969). The speed of mean Glivenko–Cantelli convergence. *Ann. Math. Stat.* **40** 40–50.
- EFRON, B. and TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability*. **57**. CRC Press, New York. <https://doi.org/10.1007/978-1-4899-4541-9>
- EVANS, L. C. (2010). *Partial Differential Equations*, 2nd ed. *Graduate Studies in Mathematics* **19**. Amer. Math. Soc., Providence, RI. <https://doi.org/10.1090/gsm/019>
- FLAMARY, R., CUTURI, M., COURTY, N. and RAKOTOMAMONJY, A. (2018). Wasserstein discriminant analysis. *Mach. Learn.* **107** 1923–1945. <https://doi.org/10.1007/s10994-018-5717-1>
- FOURNIER, N. and GUILLIN, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* **162** 707–738. <https://doi.org/10.1007/s00440-014-0583-7>
- GENEVAY, A., PEYRÉ, G. and CUTURI, M. (2018a). Learning generative models with Sinkhorn divergences. In *International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9–11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain* (A. J. Storkey and F. Pérez-Cruz, eds.). *Proceedings of Machine Learning Research* **84** 1608–1617. Available at <http://proceedings.mlr.press/v84/genevay18a.html>.
- GENEVAY, A., CHIZAT, L., BACH, F., CUTURI, M. and PEYRÉ, G. (2018b). Sample complexity of Sinkhorn divergences. In *The 22nd International Conference on Artificial Intelligence and Statistics, AISTATS 2019, 16–18 April 2019, Naha, Okinawa, Japan* (K. Chaudhuri and M. Sugiyama, eds.). *Proceedings of Machine Learning Research* **89** 1574–1583. PMLR. Available at <http://proceedings.mlr.press/v97/gordaliza19a.html>.
- GORDALIZA, P., DEL BARRIO, E., GAMBOA, F. and LOUBES, J.-M. (2019). Obtaining fairness using optimal transport theory. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9–15 June 2019, Long Beach, California, USA* (K. Chaudhuri and R. Salakhutdinov, eds.). *Proceedings of Machine Learning Research* **97** 2357–2365. PMLR. Available at <http://proceedings.mlr.press/v97/gordaliza19a.html>.
- GRAVE, E., JOULIN, A. and BERTHET, Q. (2019). Unsupervised alignment of embeddings with Wasserstein procrustes. In *The 22nd International Conference on Artificial Intelligence and Statistics, AISTATS 2019, 16–18 April 2019, Naha, Okinawa, Japan* (K. Chaudhuri and M. Sugiyama, eds.). *Proceedings of Machine Learning Research* **89** 1880–1890. PMLR. Available at <http://proceedings.mlr.press/v89/grave19a.html>.
- HÄRDLE, W., KERKYACHARIAN, G., PICARD, D. and TSYBAKOV, A. (1998). *Wavelets, Approximation, and Statistical Applications. Lecture Notes in Statistics* **129**. Springer, New York. <https://doi.org/10.1007/978-1-4612-2222-4>
- KANTOROVITCH, L. (1942). On the translocation of masses. *C. R. (Dokl.) Acad. Sci. URSS* **37** 199–201.
- KERKYACHARIAN, G. and PICARD, D. (1992). Density estimation in Besov spaces. *Statist. Probab. Lett.* **13** 15–24. [https://doi.org/10.1016/0167-7152\(92\)90231-S](https://doi.org/10.1016/0167-7152(92)90231-S)
- KLOECKNER, B. (2018). Empirical measures: Regularity is a counter-curse to dimensionality. ArXiv preprint. Available at [arXiv:1802.04038](https://arxiv.org/abs/1802.04038).
- LE GOUIC, T. and LOUBES, J.-M. (2017). Existence and consistency of Wasserstein barycenters. *Probab. Theory Related Fields* **168** 901–917.
- LECAM, L. (1973). Convergence of estimates under dimensionality restrictions. *Ann. Statist.* **1** 38–53.
- LEDoux, M. (2017). On optimal matching of Gaussian samples. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* **457** 226–264.
- LEEB, W. and COIFMAN, R. (2015). Hölder–lipschitz norms and their duals on spaces with semigroups, with applications to earth mover’s distance. *J. Fourier Anal. Appl.* **22** 910–953. <https://doi.org/10.1007/s00041-015-9439-5>
- LIANG, T. (2017). How well can generative adversarial networks (GAN) learn densities: A nonparametric view. ArXiv preprint. Available at [arXiv:1712.08244](https://arxiv.org/abs/1712.08244).

- LOEPER, G. (2006). Uniqueness of the solution to the Vlasov–Poisson system with bounded density. *J. Math. Pures Appl.* (9) **86** 68–79. <https://doi.org/10.1016/j.matpur.2006.01.005>
- MARTIAL, A. and CARLIER, G. (2011). Barycenters in the Wasserstein space. *SIAM J. Math. Anal.* **43** 904–924.
- MAURY, B., ROUDNEFF-CHUPIN, A. and SANTAMBROGIO, F. (2010). A macroscopic crowd motion model of gradient flow type. *Math. Models Methods Appl. Sci.* **20** 1787–1821. <https://doi.org/10.1142/S0218202510004799>
- MENA, G. and NILES-WEED, J. (2019). Statistical bounds for entropic optimal transport: Sample complexity and the central limit theorem. In *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*.
- MEYER, Y. (1990). *Ondelettes et Opérateurs. I. Actualités Mathématiques*. [Current Mathematical Topics]. Hermann, Paris.
- MONGE, G. (1781). Mémoire sur la théorie des déblais et des remblais. *Histoire de L’Académie Royale des Sciences* **1** 666–704.
- MONTAVON, G., MÜLLER, K.-R. and CUTURI, M. (2016). Wasserstein training of restricted Boltzmann machines. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5–10, 2016, Barcelona, Spain* (D. D. Lee, M. Sugiyama, U. von Luxburg, I. Guyon and R. Garnett, eds.) 3711–3719. Available at <https://proceedings.neurips.cc/paper/2016/hash/728f206c2a01bf572b5940d7d9a8fa4c-Abstract.html>.
- MOSER, J. (1965). On the volume elements on a manifold. *Trans. Amer. Math. Soc.* **120** 286–294. <https://doi.org/10.2307/1994022>
- MÜLLER, A. (1997). Integral probability metrics and their generating classes of functions. *Adv. in Appl. Probab.* **29** 429–443. <https://doi.org/10.2307/1428011>
- NILES-WEED, J. and BERTHET, Q. (2022). Supplement to “Minimax estimation of smooth densities in Wasserstein distance.” <https://doi.org/10.1214/21-AOS2161SUPP>
- NILES-WEED, J. and RIGOLLET, P. (2019). Estimation of Wasserstein distances in the spiked transport model.
- PEYRE, R. (2018). Comparison between W_2 distance and $\dot{H}^{-1}$ norm, and localization of Wasserstein distance. In *ESAIM: Control, Optimisation and Calculus of Variations*. <https://doi.org/10.1051/cocv/2017050>
- PEYRÉ, G. and CUTURI, M. (2019). Computational optimal transport: With Applications to Data Science. In *Foundations and Trends® in Machine Learning* **11** 355–607. <https://doi.org/10.1561/22000000073>
- RABIN, J., PEYRÉ, G., DELON, J. and BERNOT, M. (2011). Wasserstein barycenter and its application to texture mixing. In *International Conference on Scale Space and Variational Methods in Computer Vision* 435–446. Springer, Berlin.
- RIGOLLET, P. and WEED, J. (2018). Entropic optimal transport is maximum-likelihood deconvolution. *C. R. Math. Acad. Sci. Paris* **356** 1228–1235. [MR3907589 https://doi.org/10.1016/j.crma.2018.10.010](https://doi.org/10.1016/j.crma.2018.10.010)
- RIGOLLET, P. and WEED, J. (2019). Uncoupled isotonic regression via minimum Wasserstein deconvolution, information and inference. *Inf. Inference* **8** 691–717. [MR4045481 https://doi.org/10.1093/imaiai/iaz006](https://doi.org/10.1093/imaiai/iaz006)
- SHIRDHONKAR, S. and JACOBS, D. W. (2008). Approximate earth mover’s distance in linear time. In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008)*, 24–26 June 2008, Anchorage, Alaska, USA. <https://doi.org/10.1109/CVPR.2008.4587662>
- SINGH, S. and PÓCZOS, B. (2018). Minimax distribution estimation in Wasserstein distance. ArXiv preprint. Available at [arXiv:1802.08855](https://arxiv.org/abs/1802.08855).
- SINGH, S., UPPAL, A., LI, B., LI, C.-L., ZAHEER, M. and PÓCZOS, C.-L. (2018). Nonparametric density estimation under adversarial losses. In *Advances in Neural Information Processing Systems* (S. Bengio, H. M. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi and R. Garnett, eds.) **31** 10246–10257. Available at <https://proceedings.neurips.cc/paper/2018/hash/4996dcc43b5be197b5887a4e60817b1c-Abstract.html>.
- TRIEBEL, H. (2010). *Bases in Function Spaces, Sampling, Discrepancy, Numerical Integration*. EMS Tracts in Mathematics. European Mathematical **11**. Society (EMS), Zürich. <https://doi.org/10.4171/085>
- TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, New York. Revised and extended from the 2004 French original, translated by Vladimir Zaiats.
- UPPAL, A., SINGH, S. and PÓCZOS, B. (2019). Nonparametric density estimation & convergence rates for GANS under Besov IPM losses. In *Advances in Neural Information Processing Systems* (H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. B. Fox and R. Garnett, eds.) **32** 9086–9097. Available at <https://proceedings.neurips.cc/paper/2019/hash/fb09f481d40c4d3c0861a46bd2dc52c0-Abstract.html>.
- VILLANI, C. (2008). *Optimal Transport: Old and New* **338**. Springer, Berlin.
- WALTER, G. G. (1992). Approximation of the delta function by wavelets. *J. Approx. Theory* **71** 329–343. [https://doi.org/10.1016/0021-9045\(92\)90123-6](https://doi.org/10.1016/0021-9045(92)90123-6)
- WEED, J. and BACH, F. (2019). Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli* **25** 2620–2648. [MRMR4003560 https://doi.org/10.3150/18-BEJ1065](https://doi.org/10.3150/18-BEJ1065)
- WEED, J. and BERTHET, Q. (2019). Estimation of smooth densities in Wasserstein distance. In *Conference on Learning Theory, COLT 2019, 25–28 June 2019, Phoenix, AZ, USA* (A. Beygelzimer and D. Hsu, eds.). *Proceedings of Machine Learning Research* **99** 3118–3119. PMLR. Available at <http://proceedings.mlr.press/v99/weed19a.html>.

LARGE-SCALE INFERENCE WITH BLOCK STRUCTURE

BY JIYAO KOU^{1,a} AND GUENTHER WALTHER^{2,b}

^{1a}koujiyao@gmail.com

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

The detection of weak and rare effects in large amounts of data arises in a number of modern data analysis problems. Known results show that in this situation the potential of statistical inference is severely limited by the large-scale multiple testing that is inherent in these problems. Here, we show that fundamentally more powerful statistical inference is possible when there is some structure in the signal that can be exploited, for example, if the signal is clustered in many small blocks, as is the case in some relevant applications. We derive the detection boundary in such a situation where we allow both the number of blocks and the block length to grow polynomially with sample size. We derive these results both for the univariate and the multivariate settings as well as for the problem of detecting clusters in a network. These results recover as special cases the sparse signal detection problem (*Ann. Statist.* **32** (2004) 962–994) where there is no structure in the signal, as well as the scan problem (*Statist. Sinica* **23** (2013) 409–428) where the signal comprises a single interval. We develop methodology that allows optimal adaptive detection in the general setting, thus exploiting the structure if it is present without incurring a relevant penalty in the case where there is no structure. The advantage of this methodology can be considerable, as in the case of no structure the means need to increase at the rate $\sqrt{\log n}$ to ensure detection, while the presence of structure allows detection even if the means decrease at a polynomial rate.

REFERENCES

- ARIAS-CASTRO, E., CANDÈS, E. J. and DURAND, A. (2011). Detection of an anomalous cluster in a network. *Ann. Statist.* **39** 278–304. [MR2797847](#) <https://doi.org/10.1214/10-AOS839>
- ARIAS-CASTRO, E., DONOHO, D. L. and HUO, X. (2005). Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Inf. Theory* **51** 2402–2425. [MR2246369](#) <https://doi.org/10.1109/TIT.2005.850056>
- CAI, T. T., JENG, X. J. and JIN, J. (2011). Optimal detection of heterogeneous and heteroscedastic mixtures. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 629–662. [MR2867452](#) <https://doi.org/10.1111/j.1467-9868.2011.00778.x>
- CHAN, H. P. (2009). Detection of spatial clustering with average likelihood ratio test statistics. *Ann. Statist.* **37** 3985–4010. [MR2572450](#) <https://doi.org/10.1214/09-AOS701>
- CHAN, H. P. and WALTHER, G. (2013). Detection with the scan and the average likelihood ratio. *Statist. Sinica* **23** 409–428. [MR3076173](#)
- DELAIGLE, A. and HALL, P. (2009). Higher criticism in the context of unknown distribution, non-independence and classification. In *Perspectives in Mathematical Sciences. I. Stat. Sci. Interdiscip. Res.* **7** 109–138. World Sci. Publ., Hackensack, NJ. [MR2581742](#) https://doi.org/10.1142/9789814273633_0006
- 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>
- DUEMBGEN, L. and WELLNER, J. A. (2014). Confidence bands for distribution functions: A new look at the law of the iterated logarithm. Available at [arXiv:1402.2918](https://arxiv.org/abs/1402.2918).
- FAN, Y., JIN, J. and YAO, Z. (2013). Optimal classification in sparse Gaussian graphic model. *Ann. Statist.* **41** 2537–2571. [MR3161437](#) <https://doi.org/10.1214/13-AOS1163>

MSC2020 subject classifications. Primary 62G10, 62G32; secondary 62H15.

Key words and phrases. Block structure, heterogeneous mixture detection, sparse signal detection, structured higher criticism, structured Berk–Jones statistic, structured ϕ -divergence, tail bound for supremum of standardized Brownian bridge, tail bound for supremum of binomial log likelihood ratio process, tail bound for higher criticism statistic and Berk–Jones statistic.

- GANGNON, R. E. and CLAYTON, M. K. (2001). A weighted average likelihood ratio test for spatial clustering of disease. *Stat. Med.* **20** 2977–2987.
- HALL, P. and JIN, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.* **38** 1686–1732. [MR2662357](#) <https://doi.org/10.1214/09-AOS764>
- INGSTER, YU. I. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods Statist.* **6** 47–69. [MR1456646](#)
- INGSTER, YU. I. (1998). Minimax detection of a signal for l^n -balls. *Math. Methods Statist.* **7** 401–428. [MR1680087](#)
- INGSTER, Y. I., POUET, C. and TSYBAKOV, A. B. (2009). Classification of sparse high-dimensional vectors. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **367** 4427–4448. [MR2546395](#) <https://doi.org/10.1098/rsta.2009.0156>
- INGSTER, YU. I. and SUSLINA, I. A. (2002). On the detection of a signal with a known shape in a multichannel system. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* **294** 88–112, 261. [MR1976749](#) <https://doi.org/10.1007/s10958-005-0133-z>
- ITÔ, K. and MCKEAN, H. P. JR. (1965). *Diffusion Processes and Their Sample Paths. Die Grundlehren der Mathematischen Wissenschaften, Band 125*. Academic Press, New York; Springer, Berlin–New York. [MR0199891](#)
- JAGER, L. and WELLNER, J. A. (2007). Goodness-of-fit tests via phi-divergences. *Ann. Statist.* **35** 2018–2053. [MR2363962](#) <https://doi.org/10.1214/0009053607000000244>
- JENG, X. J., CAI, T. T. and LI, H. (2010). Optimal sparse segment identification with application in copy number variation analysis. *J. Amer. Statist. Assoc.* **105** 1156–1166. [MR2752611](#) <https://doi.org/10.1198/jasa.2010.tm10083>
- KOU, J. (2017). *Large-Scale Inference with Block Structure*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)–Stanford University. [MR4257243](#)
- KOU, J. (2021). Identifying the support of rectangular signals in Gaussian noise. *Comm. Statist. Theory Methods* 1–28.
- KULLDORFF, M. (1999). Spatial scan statistics: Models, calculations, and applications. In *Scan Statistics and Applications. Stat. Ind. Technol.* 303–322. Birkhäuser, Boston, MA. [MR1697758](#) https://doi.org/10.1007/978-1-4612-1578-3_14
- LI, J. and SIEGMUND, D. (2015). Higher criticism: p -values and criticism. *Ann. Statist.* **43** 1323–1350. [MR3346705](#) <https://doi.org/10.1214/15-AOS1312>
- MILLER, R. and SIEGMUND, D. (1982). Maximally selected chi square statistics. *Biometrics* **38** 1011–1016. [MR0695368](#) <https://doi.org/10.2307/2529881>
- RIVERA, C. and WALTHER, G. (2013). Optimal detection of a jump in the intensity of a Poisson process or in a density with likelihood ratio statistics. *Scand. J. Stat.* **40** 752–769. [MR3145116](#) <https://doi.org/10.1111/sjos.12027>
- SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. Wiley, New York. [MR0838963](#)
- VERZELEN, N. and ARIAS-CASTRO, E. (2017). Detection and feature selection in sparse mixture models. *Ann. Statist.* **45** 1920–1950. [MR3718157](#) <https://doi.org/10.1214/16-AOS1513>
- WALTHER, G. (2010). Optimal and fast detection of spatial clusters with scan statistics. *Ann. Statist.* **38** 1010–1033. [MR2604703](#) <https://doi.org/10.1214/09-AOS732>
- WALTHER, G. (2013). The average likelihood ratio for large-scale multiple testing and detecting sparse mixtures. In *From Probability to Statistics and Back: High-Dimensional Models and Processes. Inst. Math. Stat. (IMS) Collect.* **9** 317–326. IMS, Beachwood, OH. [MR3202643](#) <https://doi.org/10.1214/12-IMSCOLL923>
- ZHONG, P.-S., CHEN, S. X. and XU, M. (2013). Tests alternative to higher criticism for high-dimensional means under sparsity and column-wise dependence. *Ann. Statist.* **41** 2820–2851. [MR3161449](#) <https://doi.org/10.1214/13-AOS1168>

UNIFORM CONVERGENCE OF LOCAL FRÉCHET REGRESSION WITH APPLICATIONS TO LOCATING EXTREMA AND TIME WARPING FOR METRIC SPACE VALUED TRAJECTORIES

BY YAQING CHEN^a AND HANS-GEORG MÜLLER^b

Department of Statistics, University of California, Davis

^ayaqchen@ucdavis.edu, ^bhgmuller@ucdavis.edu

Local Fréchet regression is a nonparametric regression method for metric space valued responses and Euclidean predictors, which can be utilized to obtain estimates of smooth trajectories taking values in general metric spaces from noisy metric space valued random objects. We derive uniform rates of convergence, which so far have eluded theoretical analysis of this method, for both fixed and random target trajectories, where we utilize tools from empirical processes. These results are shown to be widely applicable in metric space valued data analysis. In addition to simulations, we provide two pertinent examples where these results are important: The consistent estimation of the location of properly defined extrema in metric space valued trajectories, which we illustrate with the problem of locating the age of minimum brain connectivity as obtained from fMRI data; and time warping for metric space valued trajectories, illustrated with yearly age-at-death distributions for different countries.

REFERENCES

- AGUEH, M. and CARLIER, G. (2011). Barycenters in the Wasserstein space. *SIAM J. Math. Anal.* **43** 904–924. [MR2801182 https://doi.org/10.1137/100805741](https://doi.org/10.1137/100805741)
- AMBROSIO, L., GIGLI, N. and SAVARÉ, G. (2004). Gradient flows with metric and differentiable structures, and applications to the Wasserstein space. *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl.* **15** 327–343. [MR2148889](https://doi.org/10.1016/S0926-6410(04)00088-9)
- BALABDAOUI, F., RUFIBACH, K. and WELLNER, J. A. (2009). Limit distribution theory for maximum likelihood estimation of a log-concave density. *Ann. Statist.* **37** 1299–1331. [MR2509075 https://doi.org/10.1214/08-AOS609](https://doi.org/10.1214/08-AOS609)
- BELITSER, E., GHOSAL, S. and VAN ZANTEN, H. (2012). Optimal two-stage procedures for estimating location and size of the maximum of a multivariate regression function. *Ann. Statist.* **40** 2850–2876. [MR3097962 https://doi.org/10.1214/12-AOS1053](https://doi.org/10.1214/12-AOS1053)
- BHATTACHARYA, R. and PATRANGENARU, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann. Statist.* **31** 1–29. [MR1962498 https://doi.org/10.1214/aos/1046294456](https://doi.org/10.1214/aos/1046294456)
- BHATTACHARYA, R. and PATRANGENARU, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *Ann. Statist.* **33** 1225–1259. [MR2195634 https://doi.org/10.1214/009053605000000093](https://doi.org/10.1214/009053605000000093)
- BUCKNER, R. L., SEPULCRE, J., TALUKDAR, T., KRIENEN, F. M., LIU, H., HEDDEN, T., ANDREWS-HANNA, J. R., SPERLING, R. A. and JOHNSON, K. A. (2009). Cortical hubs revealed by intrinsic functional connectivity: Mapping, assessment of stability, and relation to Alzheimer’s disease. *J. Neurosci.* **29** 1860–1873.
- BULLMORE, E. and SPORNS, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* **10** 186–198.
- CAI, J., LIU, A., MI, T., GARG, S., TRAPPE, W., MCKEOWN, M. J. and WANG, Z. J. (2019). Dynamic graph theoretical analysis of functional connectivity in Parkinson’s disease: The importance of Fiedler value. *IEEE J. Biomed. Health Inform.* **23** 1720–1729.
- CHEN, Y. and MÜLLER, H.-G. (2022). Supplement to “Uniform convergence of local Fréchet regression with applications to locating extrema and time warping for metric space valued trajectories.” <https://doi.org/10.1214/21-AOS2163SUPP>

MSC2020 subject classifications. Primary 62G05, 62G20; secondary 62G08.

Key words and phrases. Random objects, metric space valued random processes, rates of convergence, smoothing, fMRI, mortality distributions.

- DAVIS, B. C., FLETCHER, P. T., BULLITT, E. and JOSHI, S. (2007). Population shape regression from random design data. In 2007 *IEEE 11th International Conference on Computer Vision*.
- DE HAAN, W., VAN DER FLIER, W. M., WANG, H., VAN MIEGHEM, P. F., SCHELTENS, P. and STAM, C. J. (2012). Disruption of functional brain networks in Alzheimer's disease: What can we learn from graph spectral analysis of resting-state magnetoencephalography? *Brain Connect.* **2** 45–55.
- DENNIS, E. L. and THOMPSON, P. M. (2014). Functional brain connectivity using fMRI in aging and Alzheimer's disease. *Neuropsychol. Rev.* **24** 49–62. <https://doi.org/10.1007/s11065-014-9249-6>
- DEVROYE, L. P. (1978). The uniform convergence of nearest neighbor regression function estimators and their application in optimization. *IEEE Trans. Inf. Theory* **24** 142–151. [MR0518942 https://doi.org/10.1109/tit.1978.1055865](https://doi.org/10.1109/tit.1978.1055865)
- DRYDEN, I. L., KOLOYDENKO, A. and ZHOU, D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Ann. Appl. Stat.* **3** 1102–1123. [MR2750388 https://doi.org/10.1214/09-AOAS249](https://doi.org/10.1214/09-AOAS249)
- DUBEY, P. and MÜLLER, H.-G. (2020). Functional models for time-varying random objects. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 275–327. [MR4084166](https://doi.org/10.1111/j.1467-9868.2004.B5582.x)
- FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. Monographs on Statistics and Applied Probability **66**. CRC Press, London. [MR1383587](https://doi.org/10.1111/j.1467-9868.2004.B5582.x)
- FARAGÓ, A., LINDER, T. and LUGOSI, G. (1993). Fast nearest-neighbor search in dissimilarity spaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **15** 957–962.
- FERREIRA, L. K. and BUSATTO, G. F. (2013). Resting-state functional connectivity in normal brain aging. *Neurosci. Biobehav. Rev.* **37** 384–400.
- FIEDLER, M. (1973). Algebraic connectivity of graphs. *Czechoslovak Math. J.* **23** 298–305. [MR0318007](https://doi.org/10.1111/j.1467-9868.2004.B5582.x)
- FRÉCHET, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. Henri Poincaré* **10** 215–310. [MR0027464](https://doi.org/10.1111/j.1467-9868.2004.B5582.x)
- GASSER, T. and KNEIP, A. (1995). Searching for structure in curve samples. *J. Amer. Statist. Assoc.* **90** 1179–1188.
- GERVINI, D. and GASSER, T. (2004). Self-modelling warping functions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 959–971. [MR2102475 https://doi.org/10.1111/j.1467-9868.2004.B5582.x](https://doi.org/10.1111/j.1467-9868.2004.B5582.x)
- GONG, D. and MEDIONI, G. (2011). Dynamic manifold warping for view invariant action recognition. In *Proceedings of International Conference on Computer Vision* 571–578.
- HEIN, M. (2009). Robust nonparametric regression with metric-space valued output. In *Advances in Neural Information Processing Systems* 718–726.
- HOFFMAN, A. J. and WIELANDT, H. W. (1953). The variation of the spectrum of a normal matrix. *Duke Math. J.* **20** 37–39. [MR0052379](https://doi.org/10.1111/j.1467-9868.2004.B5582.x)
- HUCKEMANN, S. F. (2012). On the meaning of mean shape: Manifold stability, locus and the two sample test. *Ann. Inst. Statist. Math.* **64** 1227–1259. [MR2981621 https://doi.org/10.1007/s10463-012-0352-2](https://doi.org/10.1007/s10463-012-0352-2)
- HUCKEMANN, S. F. (2015). (Semi-)Intrinsic statistical analysis on non-Euclidean spaces. In *Advances in Complex Data Modeling and Computational Methods in Statistics* 103–118. Springer, New York.
- JAMES, G. M. (2007). Curve alignment by moments. *Ann. Appl. Stat.* **1** 480–501. [MR2415744 https://doi.org/10.1214/07-AOAS127](https://doi.org/10.1214/07-AOAS127)
- KLOECKNER, B. (2010). A geometric study of Wasserstein spaces: Euclidean spaces. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* **9** 297–323. [MR2731158](https://doi.org/10.1111/j.1467-9868.2004.B5582.x)
- KNEIP, A. and GASSER, T. (1992). Statistical tools to analyze data representing a sample of curves. *Ann. Statist.* **20** 1266–1305. [MR1186250 https://doi.org/10.1214/aos/1176348769](https://doi.org/10.1214/aos/1176348769)
- LE GOUIC, T. and LOUBES, J.-M. (2017). Existence and consistency of Wasserstein barycenters. *Probab. Theory Related Fields* **168** 901–917. [MR3663634 https://doi.org/10.1007/s00440-016-0727-z](https://doi.org/10.1007/s00440-016-0727-z)
- MACK, Y. P. and SILVERMAN, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrsch. Verw. Gebiete* **61** 405–415. [MR0679685 https://doi.org/10.1007/BF00539840](https://doi.org/10.1007/BF00539840)
- MARRON, J. S., RAMSAY, J. O., SANGALLI, L. M. and SRIVASTAVA, A. (2015). Functional data analysis of amplitude and phase variation. *Statist. Sci.* **30** 468–484. [MR3432837 https://doi.org/10.1214/15-STS524](https://doi.org/10.1214/15-STS524)
- MEVEL, K., LANDEAU, B., FOUQUET, M., LA JOIE, R., VILLAIN, N., MÉZENGE, F., PERROTIN, A., EUSTACHE, F., DESGRANGES, B. and CHÉTELAT, G. (2013). Age effect on the default mode network, inner thoughts, and cognitive abilities. *Neurobiol. Aging* **34** 1292–1301.
- MÜLLER, H.-G. (1989). Adaptive nonparametric peak estimation. *Ann. Statist.* **17** 1053–1069. [MR1015137 https://doi.org/10.1214/aos/1176347255](https://doi.org/10.1214/aos/1176347255)
- PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Stat.* **33** 1065–1076. [MR0143282 https://doi.org/10.1214/aoms/1177704472](https://doi.org/10.1214/aoms/1177704472)
- PELLETIER, B. (2006). Non-parametric regression estimation on closed Riemannian manifolds. *J. Nonparametr. Stat.* **18** 57–67. [MR2214065 https://doi.org/10.1080/10485250500504828](https://doi.org/10.1080/10485250500504828)

- PETERSEN, A. and MÜLLER, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *Ann. Statist.* **47** 691–719. [MR3909947 https://doi.org/10.1214/17-AOS1624](https://doi.org/10.1214/17-AOS1624)
- PETERSEN, A., ZHAO, J., CARMICHAEL, O. and MÜLLER, H.-G. (2016). Quantifying individual brain connectivity with functional principal component analysis for networks. *Brain Connect.* **6** 540–547.
- PHILLIPS, D. J., MCGLAUGHLIN, A., RUTH, D., JAGER, L. R., SOLDAN, A. and ADNI (2015). Graph theoretic analysis of structural connectivity across the spectrum of Alzheimer’s disease: The importance of graph creation methods. *NeuroImage Clin.* **7** 377–390.
- RAMSAY, J. O. and LI, X. (1998). Curve registration. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 351–363. [MR1616045 https://doi.org/10.1111/1467-9868.00129](https://doi.org/10.1111/1467-9868.00129)
- RUBINOV, M. and SPORNS, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage* **52** 1059–1069.
- RUPPERT, D. and WAND, M. P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.* **22** 1346–1370. [MR1311979 https://doi.org/10.1214/aos/1176325632](https://doi.org/10.1214/aos/1176325632)
- SAKOE, H. and CHIBA, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust. Speech Signal Process.* **26** 43–49.
- SILVERMAN, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.* **6** 177–184. [MR0471166](https://doi.org/10.1214/aos/1176325632)
- STEINKE, F. and HEIN, M. (2009). Non-parametric regression between manifolds. In *Advances in Neural Information Processing Systems* 1561–1568.
- STEINKE, F., HEIN, M. and SCHÖLKOPF, B. (2010). Nonparametric regression between general Riemannian manifolds. *SIAM J. Imaging Sci.* **3** 527–563. [MR2736019 https://doi.org/10.1137/080744189](https://doi.org/10.1137/080744189)
- STURM, K.-T. (2003). Probability measures on metric spaces of nonpositive curvature. In *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces (Paris, 2002)*. *Contemp. Math.* **338** 357–390. Amer. Math. Soc., Providence, RI. [MR2039961 https://doi.org/10.1090/conm/338/06080](https://doi.org/10.1090/conm/338/06080)
- TANG, R. and MÜLLER, H.-G. (2008). Pairwise curve synchronization for functional data. *Biometrika* **95** 875–889. [MR2461217 https://doi.org/10.1093/biomet/asn047](https://doi.org/10.1093/biomet/asn047)
- TRIGEORGIS, G., NICOLAOU, M. A., SCHULLER, B. W. and ZAFEIRIOU, S. (2018). Deep canonical time warping for simultaneous alignment and representation learning of sequences. *IEEE Trans. Pattern Anal. Mach. Intell.* **40** 1128–1138. <https://doi.org/10.1109/TPAMI.2017.2710047>
- 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)
- VIEU, P. (1996). A note on density mode estimation. *Statist. Probab. Lett.* **26** 297–307. [MR1393913 https://doi.org/10.1016/0167-7152\(95\)00024-0](https://doi.org/10.1016/0167-7152(95)00024-0)
- VU, H. T., CAREY, C. and MAHADEVAN, S. (2012). Manifold warping: Manifold alignment over time. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence* 1155–1161.
- WANG, K. and GASSER, T. (1997). Alignment of curves by dynamic time warping. *Ann. Statist.* **25** 1251–1276. [MR1447750 https://doi.org/10.1214/aos/1069362747](https://doi.org/10.1214/aos/1069362747)
- 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](https://doi.org/10.1111/j.1467-9868.2011.01022.x)
- ZONNEVELD, H. I., PRUIM, R. H. R., BOS, D., VROOMAN, H. A., MUETZEL, R. L., HOFMAN, A., ROMBOUTS, S. A. R. B., VAN DER LUGT, A., NIESSEN, W. J. IKRAM, M. A. and VERNOOIJAB, M. W. (2019). Patterns of functional connectivity in an aging population: The Rotterdam study. *NeuroImage* **189** 432–444.

COINTEGRATION IN LARGE VARS

BY ANNA BYKHOVSKAYA^{1,a} AND VADIM GORIN^{2,b}

¹*Department of Economics, University of Wisconsin-Madison, ^aanna.bykhovskaya@wisc.edu*

²*Department of Mathematics, University of Wisconsin-Madison, ^bvadicgor@gmail.com*

The paper analyzes cointegration in vector autoregressive processes (VARs) for the cases when both the number of coordinates, N , and the number of time periods, T , are large and of the same order. We propose a way to examine a VAR of order 1 for the presence of cointegration based on a modification of the Johansen likelihood ratio test. The advantage of our procedure over the original Johansen test and its finite sample corrections is that our test does not suffer from overrejection. This is achieved through novel asymptotic theorems for eigenvalues of matrices in the test statistic in the regime of proportionally growing N and T . Our theoretical findings are supported by Monte Carlo simulations and an empirical illustration. Moreover, we find a surprising connection with multivariate analysis of variance (MANOVA) and explain why it emerges.

REFERENCES

- BAI, J. and NG, S. (2008). Large dimensional factor analysis. *Found Trends Econom.* **3** 890163.
- BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2567175 <https://doi.org/10.1007/978-1-4419-0661-8>
- BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. MR2165575 <https://doi.org/10.1214/009117905000000233>
- BAO, Z., PAN, G. and ZHOU, W. (2013). Central limit theorem for partial linear eigenvalue statistics of Wigner matrices. *J. Stat. Phys.* **150** 88–129. MR3018879 <https://doi.org/10.1007/s10955-012-0663-y>
- BAO, Z., HU, J., PAN, G. and ZHOU, W. (2019). Canonical correlation coefficients of high-dimensional Gaussian vectors: Finite rank case. *Ann. Statist.* **47** 612–640. MR3909944 <https://doi.org/10.1214/18-AOS1704>
- BEN AROUS, G. and GUIONNET, A. (1997). Large deviations for Wigner’s law and Voiculescu’s non-commutative entropy. *Probab. Theory Related Fields* **108** 517–542. MR1465640 <https://doi.org/10.1007/s004400050119>
- BORODIN, A. and GORIN, V. (2015). General β -Jacobi corners process and the Gaussian free field. *Comm. Pure Appl. Math.* **68** 1774–1844. MR3385342 <https://doi.org/10.1002/cpa.21546>
- BORODIN, A. and GORIN, V. (2016). Lectures on integrable probability. In *Probability and Statistical Physics in St. Petersburg. Proc. Sympos. Pure Math.* **91** 155–214. Amer. Math. Soc., Providence, RI. MR3526828 <https://doi.org/10.1007/s00029-010-0034-y>
- BREITUNG, J. and PESARAN, M. H. (2008). Unit roots and cointegration in panels. In *The Econometrics of Panel Data* (L. Mátyás and P. Sevestre, eds.). *Advanced Studies in Theoretical and Applied Econometrics* **46** 279–322. Springer, Berlin, Heidelberg.
- BYKHOVSKAYA, A. (2021). Time series approach to the evolution of networks: Prediction and estimation. *J. Bus. Econom. Statist.*
- BYKHOVSKAYA, A. and GORIN, V. (2022a). Asymptotics of cointegration tests for high-dimensional VAR(k). Available at [arXiv:2202.07150](https://arxiv.org/abs/2202.07150).
- BYKHOVSKAYA, A. and GORIN, V. (2022b). Supplement to “Cointegration in large VARs.” <https://doi.org/10.1214/21-AOS2164SUPP>
- CAVALIERE, G., RAHBEK, A. and TAYLOR, A. M. R. (2012). Bootstrap determination of the co-integration rank in vector autoregressive models. *Econometrica* **80** 1721–1740. MR2977435 <https://doi.org/10.3982/ECTA9099>
- DUMITRIU, I. and EDELMAN, A. (2002). Matrix models for beta ensembles. *J. Math. Phys.* **43** 5830–5847. MR1936554 <https://doi.org/10.1063/1.1507823>

- ELLIOTT, G., ROTHENBERG, T. J. and STOCK, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica* **64** 813–836. [MR1399219](#) <https://doi.org/10.2307/2171846>
- ENGLE, R. F. and GRANGER, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica* **55** 251–276. [MR0882095](#) <https://doi.org/10.2307/1913236>
- ERDŐS, L. and YAU, H.-T. (2012). Universality of local spectral statistics of random matrices. *Bull. Amer. Math. Soc. (N.S.)* **49** 377–414. [MR2917064](#) <https://doi.org/10.1090/S0273-0979-2012-01372-1>
- FORRESTER, P. J. (1993). The spectrum edge of random matrix ensembles. *Nuclear Phys. B* **402** 709–728. [MR1236195](#) [https://doi.org/10.1016/0550-3213\(93\)90126-A](https://doi.org/10.1016/0550-3213(93)90126-A)
- FORRESTER, P. J. (2010). *Log-Gases and Random Matrices*. London Mathematical Society Monographs Series **34**. Princeton Univ. Press, Princeton, NJ. [MR2641363](#) <https://doi.org/10.1515/9781400835416>
- GONZALO, J. and PITARAKIS, J. Y. (1999). Dimensionality effect in cointegration analysis. In *Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive WJ Granger* 212–229. Oxford Univ. Press, Oxford. Chapter 9.
- GRANGER, C. W. (1981). Some properties of time series data and their use in econometric model specification. *J. Econometrics* **16** 121–130.
- GUIONNET, A. and NOVAK, J. (2015). Asymptotics of unitary multimatrix models: The Schwinger-Dyson lattice and topological recursion. *J. Funct. Anal.* **268** 2851–2905. [MR3331788](#) <https://doi.org/10.1016/j.jfa.2015.03.002>
- HAN, X., PAN, G. and YANG, Q. (2018). A unified matrix model including both CCA and F matrices in multivariate analysis: The largest eigenvalue and its applications. *Bernoulli* **24** 3447–3468. [MR3788178](#) <https://doi.org/10.3150/17-BEJ965>
- HAN, X., PAN, G. and ZHANG, B. (2016). The Tracy-Widom law for the largest eigenvalue of F type matrices. *Ann. Statist.* **44** 1564–1592. [MR3519933](#) <https://doi.org/10.1214/15-AOS1427>
- HARVEY, D. I., LEYBOURNE, S. J. and TAYLOR, A. M. R. (2009). Unit root testing in practice: Dealing with uncertainty over the trend and initial condition. *Econometric Theory* **25** 587–636. [MR2507521](#) <https://doi.org/10.1017/S026646660809018X>
- HO, M. S. and SØRENSEN, B. E. (1996). Finding cointegration rank in high dimensional systems using the Johansen test: An illustration using data based Monte Carlo simulations. *Rev. Econ. Stat.* **78** 726–732.
- HUA, L. K. (1963). *Harmonic Analysis of Functions of Several Complex Variables in the Classical Domains*. Amer. Math. Soc., Providence, R.I. [MR0171936](#)
- JOHANSEN, S. (1988). Statistical analysis of cointegration vectors *J. Econom. Dynam. Control* **12** 231–254. [MR0986516](#) [https://doi.org/10.1016/0165-1889\(88\)90041-3](https://doi.org/10.1016/0165-1889(88)90041-3)
- JOHANSEN, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* **59** 1551–1580. [MR1135641](#) <https://doi.org/10.2307/2938278>
- JOHANSEN, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford Univ. Press, New York. [MR1487375](#) <https://doi.org/10.1093/0198774508.001.0001>
- JOHANSEN, S. (2002). A small sample correction for the test of cointegrating rank in the vector autoregressive model. *Econometrica* **70** 1929–1961. [MR1925161](#) <https://doi.org/10.1111/1468-0262.00358>
- JOHANSSON, K. (1998). On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.* **91** 151–204. [MR1487983](#) <https://doi.org/10.1215/S0012-7094-98-09108-6>
- JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. [MR1863961](#) <https://doi.org/10.1214/aos/1009210544>
- JOHNSTONE, I. M. (2008). Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy-Widom limits and rates of convergence. *Ann. Statist.* **36** 2638–2716. [MR2485010](#) <https://doi.org/10.1214/08-AOS605>
- JOHNSTONE, I. M. and ONATSKI, A. (2020). Testing in high-dimensional spiked models. *Ann. Statist.* **48** 1231–1254. [MR4124321](#) <https://doi.org/10.1214/18-AOS1697>
- KADELL, K. W. J. (1997). The Selberg-Jack symmetric functions. *Adv. Math.* **130** 33–102. [MR1467311](#) <https://doi.org/10.1006/aima.1997.1642>
- KILLIP, R. and NENCIU, I. (2004). Matrix models for circular ensembles. *Int. Math. Res. Not.* **50** 2665–2701. [MR2127367](#) <https://doi.org/10.1155/S1073792804141597>
- KRAUSS, C. (2017). Statistical arbitrage pairs trading strategies: Review and outlook. *J. Econ. Surv.* **31** 513–545.
- LÜTKEPOHL, H., SAIKKONEN, P. and TRENKLER, C. (2001). Maximum eigenvalue versus trace tests for the cointegrating rank of a VAR process. *Econom. J.* **4** 287–310. [MR1889173](#) <https://doi.org/10.1111/1368-423X.00068>
- MADDALA, G. S. and KIM, I.-M. (1998). *Unit Roots, Cointegration and Structural Change. Themes in Modern Econometrics*. Cambridge Univ. Press, Cambridge. [MR1675977](#)
- MEHTA, M. L. (2004). *Random Matrices*, 3rd ed. *Pure and Applied Mathematics (Amsterdam)* **142**. Elsevier/Academic Press, Amsterdam. [MR2129906](#)
- MINGO, J. A. and POPA, M. (2013). Real second order freeness and Haar orthogonal matrices. *J. Math. Phys.* **54** 051701, 35. [MR3098908](#) <https://doi.org/10.1063/1.4804168>

- MUIRHEAD, R. J. (2009). *Aspects of Multivariate Statistical Theory*. Wiley, New York.
- MÜLLER, U. K. and ELLIOTT, G. (2003). Tests for unit roots and the initial condition. *Econometrica* **71** 1269–1286. [MR1995831](#) <https://doi.org/10.1111/1468-0262.00447>
- O’ROURKE, S. (2010). Gaussian fluctuations of eigenvalues in Wigner random matrices. *J. Stat. Phys.* **138** 1045–1066. [MR2601422](#) <https://doi.org/10.1007/s10955-009-9906-y>
- ONATSKI, A. and WANG, C. (2018). Alternative asymptotics for cointegration tests in large VARs. *Econometrica* **86** 1465–1478. [MR3843495](#) <https://doi.org/10.3982/ECTA14649>
- ONATSKI, A. and WANG, C. (2019). Extreme canonical correlations and high-dimensional cointegration analysis. *J. Econometrics* **212** 307–322. [MR3994019](#) <https://doi.org/10.1016/j.jeconom.2019.04.032>
- PARUOLO, P. (2001). The power of lambda max. *Oxf. Bull. Econ. Stat.* **63** 395–403.
- PHILLIPS, P. C. B. and OULIARIS, S. (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica* **58** 165–193. [MR1046923](#) <https://doi.org/10.2307/2938339>
- RAMÍREZ, J. A., RIDER, B. and VIRÁG, B. (2011). Beta ensembles, stochastic Airy spectrum, and a diffusion. *J. Amer. Math. Soc.* **24** 919–944. [MR2813333](#) <https://doi.org/10.1090/S0894-0347-2011-00703-0>
- REINSEL, G. C. and AHN, S. K. (1992). Vector autoregressive models with unit roots and reduced rank structure: Estimation, likelihood ratio test, and forecasting. *J. Time Series Anal.* **13** 353–375. [MR1173565](#) <https://doi.org/10.1111/j.1467-9892.1992.tb00113.x>
- SODIN, S. (2014). Several applications of the moment method in random matrix theory. In *Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. III* 451–475. Kyung Moon Sa, Seoul. [MR3729037](#)
- STOCK, J. H. and WATSON, M. W. (1988). Testing for common trends. *J. Amer. Statist. Assoc.* **83** 1097–1107. [MR0997588](#)
- SWENSEN, A. R. (2006). Bootstrap algorithms for testing and determining the cointegration rank in VAR models. *Econometrica* **74** 1699–1714. [MR2268413](#) <https://doi.org/10.1111/j.1468-0262.2006.00723.x>
- TAO, T. and VU, V. (2014). Random matrices: The universality phenomenon for Wigner ensembles. In *Modern Aspects of Random Matrix Theory. Proc. Sympos. Appl. Math.* **72** 121–172. Amer. Math. Soc., Providence, RI. [MR3288230](#) <https://doi.org/10.1090/psapm/072/00615>
- TRACY, C. A. and WIDOM, H. (1996). On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.* **177** 727–754. [MR1385083](#)
- XIAO, Z. and PHILLIPS, P. C. B. (1999). Efficient detrending in cointegrating regression. *Econometric Theory* **15** 519–548. [MR1717966](#) <https://doi.org/10.1017/S0266466699154033>
- ZHANG, B., PAN, G. and GAO, J. (2018). CLT for largest eigenvalues and unit root testing for high-dimensional nonstationary time series. *Ann. Statist.* **46** 2186–2215. [MR3845015](#) <https://doi.org/10.1214/17-AOS1616>

PARTIAL RECOVERY FOR TOP- k RANKING: OPTIMALITY OF MLE AND SUBOPTIMALITY OF THE SPECTRAL METHOD

BY PINHAN CHEN^{1,a}, CHAO GAO^{1,b} AND ANDERSON Y. ZHANG^{2,c}

¹Department of Statistics, University of Chicago, ^apinhanchen@uchicago.edu, ^bchaogao@uchicago.edu

²Department of Statistics, University of Pennsylvania, ^cayz@wharton.upenn.edu

Given partially observed pairwise comparison data generated by the Bradley–Terry–Luce (BTL) model, we study the problem of top- k ranking. That is, to optimally identify the set of top- k players. We derive the minimax rate with respect to a normalized Hamming loss. This provides the first result in the literature that characterizes the partial recovery error in terms of the proportion of mistakes for top- k ranking. We also derive the optimal signal to noise ratio condition for the exact recovery of the top- k set. The maximum likelihood estimator (MLE) is shown to achieve both optimal partial recovery and optimal exact recovery. On the other hand, we show another popular algorithm, the spectral method, is in general suboptimal. Our results complement the recent work (*Ann. Statist.* **47** (2019) 2204–2235) that shows both the MLE and the spectral method achieve the optimal sample complexity for exact recovery. It turns out the leading constants of the sample complexity are different for the two algorithms. Another contribution that may be of independent interest is the analysis of the MLE without any penalty or regularization for the BTL model. This closes an important gap between theory and practice in the literature of ranking.

REFERENCES

- [1] ABBE, E., FAN, J., WANG, K. and ZHONG, Y. (2020). Entrywise eigenvector analysis of random matrices with low expected rank. *Ann. Statist.* **48** 1452–1474. [MR4124330 https://doi.org/10.1214/19-AOS1854](https://doi.org/10.1214/19-AOS1854)
- [2] BORDA, J. D. (1784). Mémoire sur les élections au scrutin. Histoire de l’Academie Royale des Sciences pour 1781 (Paris, 1784).
- [3] BRADLEY, R. A. and TERRY, M. E. (1952). Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika* **39** 324–345. [MR0070925 https://doi.org/10.2307/2334029](https://doi.org/10.2307/2334029)
- [4] BUTUCEA, C., NDAOUD, M., STEPANOVA, N. A. and TSYBAKOV, A. B. (2018). Variable selection with Hamming loss. *Ann. Statist.* **46** 1837–1875. [MR3845003 https://doi.org/10.1214/17-AOS1572](https://doi.org/10.1214/17-AOS1572)
- [5] CHEN, P., GAO, C. and ZHANG, A. Y. (2022). Supplement to “Partial Recovery for Top- k Ranking: Optimality of MLE and SubOptimality of the Spectral Method.” <https://doi.org/10.1214/21-AOS2166SUPP>
- [6] CHEN, X., GOPI, S., MAO, J. and SCHNEIDER, J. (2017). Competitive analysis of the top- K ranking problem. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms* 1245–1264. SIAM, Philadelphia, PA. [MR3627810 https://doi.org/10.1137/1.9781611974782.81](https://doi.org/10.1137/1.9781611974782.81)
- [7] CHEN, Y., FAN, J., MA, C. and WANG, K. (2019). Spectral method and regularized MLE are both optimal for top- K ranking. *Ann. Statist.* **47** 2204–2235. [MR3953449 https://doi.org/10.1214/18-AOS1745](https://doi.org/10.1214/18-AOS1745)
- [8] CHEN, Y. and SUH, C. (2015). Spectral mle: Top- k rank aggregation from pairwise comparisons. In *International Conference on Machine Learning* 371–380.
- [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](https://doi.org/10.1214/13-AOS1161)
- [10] COPELAND, A. (1951). A ‘reasonable’ social welfare function, Seminar on mathematics in social sciences, University of Michigan. Cited Indirectly from Its Mention by Luce and Raiffa (1957) 358.

MSC2020 subject classifications. 62F07.

Key words and phrases. Top- k ranking, pairwise comparison, MLE, spectral method, partial recovery, exact recovery.

- [11] COSSOCK, D. and ZHANG, T. (2006). Subset ranking using regression. In *Learning Theory. Lecture Notes in Computer Science* **4005** 605–619. Springer, Berlin. [MR2280634](#) https://doi.org/10.1007/11776420_44
- [12] DWORK, C., KUMAR, R., NAOR, M. and SIVAKUMAR, D. (2001). Rank aggregation methods for the web. In *Proceedings of the 10th International Conference on World Wide Web* 613–622.
- [13] ERDŐS, P. and RÉNYI, A. (1960). On the evolution of random graphs. *Magy. Tud. Akad. Mat. Kut. Intéz. Közl.* **5** 17–61. [MR0125031](#)
- [14] JANG, M., KIM, S., SUH, C. and OH, S. (2016). Top- k ranking from pairwise comparisons: When spectral ranking is optimal. ArXiv preprint. Available at [arXiv:1603.04153](#).
- [15] JANG, M., KIM, S., SUH, C. and OH, S. (2017). Optimal sample complexity of m -wise data for top- k ranking. In *Advances in Neural Information Processing Systems* 1686–1696.
- [16] LÖFFLER, M., ZHANG, A. Y. and ZHOU, H. H. (2021). Optimality of spectral clustering in the Gaussian mixture model. *Ann. Statist.* **49** 2506–2530. [MR4338373](#) <https://doi.org/10.1214/20-aos2044>
- [17] LU, Y. and ZHOU, H. H. (2016). Statistical and computational guarantees of Lloyd’s algorithm and its variants. ArXiv preprint. Available at [arXiv:1612.02099](#).
- [18] LUCE, R. D. (2012). *Individual Choice Behavior: A Theoretical Analysis*. Courier Corporation.
- [19] MOTEGI, S. and MASUDA, N. (2012). A network-based dynamical ranking system for competitive sports. *Sci. Rep.* **2** 904. <https://doi.org/10.1038/srep00904>
- [20] NDAOUD, M. and TSYBAKOV, A. B. (2020). Optimal variable selection and adaptive noisy compressed sensing. *IEEE Trans. Inf. Theory* **66** 2517–2532. [MR4087700](#) <https://doi.org/10.1109/TIT.2020.2965738>
- [21] NEGAHBAN, S., OH, S. and SHAH, D. (2017). Rank centrality: Ranking from pairwise comparisons. *Oper. Res.* **65** 266–287. [MR3613103](#) <https://doi.org/10.1287/opre.2016.1534>
- [22] SHA, L., LUCEY, P., YUE, Y., CARR, P., ROHLF, C. and MATTHEWS, I. (2016). Chalkboarding: A new spatiotemporal query paradigm for sports play retrieval. In *Proceedings of the 21st International Conference on Intelligent User Interfaces* 336–347.
- [23] SHAH, N. B. and WAINWRIGHT, M. J. (2017). Simple, robust and optimal ranking from pairwise comparisons. *J. Mach. Learn. Res.* **18** 199. [MR3827087](#)
- [24] TROPP, J. A. (2015). An introduction to matrix concentration inequalities. *Found. Trends Mach. Learn.* **8** 1–230.
- [25] 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>

A DATA-ADAPTIVE METHOD FOR ESTIMATING DENSITY LEVEL SETS UNDER SHAPE CONDITIONS

BY ALBERTO RODRÍGUEZ-CASAL^{1,a} AND PAULA SAAVEDRA-NIEVES^{2,b}

¹*Faculdade de Matemáticas, Universidade de Santiago de Compostela, alberto.rodriguez.casal@usc.es*

²*Faculdade de Medicina e Odontoloxía, Universidade de Santiago de Compostela, paula.saaavedra@usc.es*

Given a random sample of points from some unknown density, we propose a method for estimating density level sets, for a positive threshold t , under the r -convexity assumption. This shape condition generalizes the convexity property and allows to consider level sets with more than one connected component. The main problem in practice is that r is an unknown geometric characteristic of the set related to its curvature, which may depend on t . A stochastic algorithm is proposed for selecting its value from data. The resulting reconstruction of the level set is able to achieve minimax rates for Hausdorff metric and distance in measure uniformly on the level t .

REFERENCES

- [1] ARIAS-CASTRO, E. and RODRÍGUEZ-CASAL, A. (2017). On estimating the perimeter using the alpha-shape. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 1051–1068. [MR3689960](#) <https://doi.org/10.1214/16-AIHP747>
- [2] BAÍLLO, A. (2003). Total error in a plug-in estimator of level sets. *Statist. Probab. Lett.* **65** 411–417. [MR2039885](#) <https://doi.org/10.1016/j.spl.2003.08.007>
- [3] BAÍLLO, A. and CUEVAS, A. (2006). Parametric versus nonparametric tolerance regions in detection problems. *Comput. Statist.* **21** 523–536. [MR2312631](#) <https://doi.org/10.1007/s00180-006-0010-3>
- [4] CADRE, B. (2006). Kernel estimation of density level sets. *J. Multivariate Anal.* **97** 999–1023. [MR2256570](#) <https://doi.org/10.1016/j.jmva.2005.05.004>
- [5] CADRE, B., PELLETIER, B. and PUDLO, P. (2013). Estimation of density level sets with a given probability content. *J. Nonparametr. Stat.* **25** 261–272. [MR3039981](#) <https://doi.org/10.1080/10485252.2012.750319>
- [6] CHACÓN, J. E. and DUONG, T. (2018). *Multivariate Kernel Smoothing and Its Applications. Monographs on Statistics and Applied Probability* **160**. CRC Press, Boca Raton, FL. [MR3822372](#) <https://doi.org/10.1201/9780429485572>
- [7] CHEN, Y.-C., GENOVESE, C. R. and WASSERMAN, L. (2017). Density level sets: Asymptotics, inference, and visualization. *J. Amer. Statist. Assoc.* **112** 1684–1696. [MR3750891](#) <https://doi.org/10.1080/01621459.2016.1228536>
- [8] CHEN, Y. C. and DOBRA, A. (2017). Measuring human activity spaces with density ranking based on GPS data. Preprint. Available at [arXiv:1708.05017](https://arxiv.org/abs/1708.05017).
- [9] CHERNOFF, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.* **16** 31–41. [MR0172382](#) <https://doi.org/10.1007/BF02868560>
- [10] CUEVAS, A. and FRAIMAN, R. (1997). A plug-in approach to support estimation. *Ann. Statist.* **25** 2300–2312. [MR1604449](#) <https://doi.org/10.1214/aos/1030741073>
- [11] CUEVAS, A., FRAIMAN, R. and PATEIRO-LÓPEZ, B. (2012). On statistical properties of sets fulfilling rolling-type conditions. *Adv. in Appl. Probab.* **44** 311–329. [MR2977397](#) <https://doi.org/10.1239/aap/1339878713>
- [12] DIGGLE, P. J. Index of Point Pattern Datasets. Retrieved February 12, 2014 from <http://www.lancaster.ac.uk/staff/diggle/pointpatternbook/datasets/>.
- [13] DOSS, C. R. and WENG, G. (2018). Bandwidth selection for kernel density estimators of multivariate level sets and highest density regions. *Electron. J. Stat.* **12** 4313–4376. [MR3892342](#) <https://doi.org/10.1214/18-ejs1501>
- [14] EDDY, W. F. and HARTIGAN, J. A. (1977). Uniform convergence of the empirical distribution function over convex sets. *Ann. Statist.* **5** 370–374. [MR0431344](#)

- [15] EDELSBRUNNER, H., KIRKPATRICK, D. G. and SEIDEL, R. (1983). On the shape of a set of points in the plane. *IEEE Trans. Inf. Theory* **29** 551–559. [MR0713690](#) <https://doi.org/10.1109/TIT.1983.1056714>
- [16] GAYRAUD, G. and ROUSSEAU, J. (2005). Rates of convergence for a Bayesian level set estimation. *Scand. J. Stat.* **32** 639–660. [MR2232347](#) <https://doi.org/10.1111/j.1467-9469.2005.00448.x>
- [17] GINÉ, E. and GUILLOU, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators *Ann. Inst. Henri Poincaré Probab. Stat.* **38** 907–921. [MR1955344](#) [https://doi.org/10.1016/S0246-0203\(02\)01128-7](https://doi.org/10.1016/S0246-0203(02)01128-7)
- [18] GRÜBEL, R. (1988). The length of the shorth. *Ann. Statist.* **16** 619–628. [MR0947565](#) <https://doi.org/10.1214/aos/1176350823>
- [19] HAND, D. J. (1982). *Kernel Discriminant Analysis. Electronic & Electrical Engineering Research Studies: Pattern Recognition & Image Processing Series 2*. Research Studies Press [Wiley], Chichester–New York. [MR0666869](#)
- [20] HARTIGAN, J. A. (1975). *Clustering Algorithms. Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. [MR0405726](#)
- [21] HARTIGAN, J. A. (1987). Estimation of a convex density contour in two dimensions. *J. Amer. Statist. Assoc.* **82** 267–270. [MR0883354](#)
- [22] HENDERSON, R., SHIMAKURA, S. and GORST, D. (2002). Modeling spatial variation in leukemia survival data. *J. Amer. Statist. Assoc.* **97** 965–972. [MR1973859](#) <https://doi.org/10.1198/016214502388618753>
- [23] HUO, X. and LU, J.-C. (2004). A network flow approach in finding maximum likelihood estimate of high concentration regions. *Comput. Statist. Data Anal.* **46** 33–56. [MR2056823](#) [https://doi.org/10.1016/S0167-9473\(03\)00134-8](https://doi.org/10.1016/S0167-9473(03)00134-8)
- [24] HYNDMAN, R. J. (1996). Computing and graphing highest density regions. *Amer. Statist.* **50** 120–126.
- [25] LAFARGE, T., PATEIRO-LÓPEZ, B., POSSOLO, A. and DUNKERS, J. (2014). R implementation of a polyhedral approximation to a 3D set of points using the α -shape. *J. Stat. Softw.* **56** 1–19.
- [26] MAMMEN, E. and POLONIK, W. (2013). Confidence regions for level sets. *J. Multivariate Anal.* **122** 202–214. [MR3189318](#) <https://doi.org/10.1016/j.jmva.2013.07.017>
- [27] MAMMEN, E. and POLONIK, W. (2013). Confidence regions for level sets. *J. Multivariate Anal.* **122** 202–214. [MR3189318](#) <https://doi.org/10.1016/j.jmva.2013.07.017>
- [28] MAMMEN, E. and TSYBAKOV, A. B. (1995). Asymptotical minimax recovery of sets with smooth boundaries. *Ann. Statist.* **23** 502–524. [MR1332579](#) <https://doi.org/10.1214/aos/1176324533>
- [29] MARKOU, M. and SINGH, S. (2003). Novelty detection: A reviewpart 1: Statistical approaches. *Signal Process.* **83** 2481–2497.
- [30] MASON, D. M. and POLONIK, W. (2009). Asymptotic normality of plug-in level set estimates. *Ann. Appl. Probab.* **19** 1108–1142. [MR2537201](#) <https://doi.org/10.1214/08-AAP569>
- [31] MÜLLER, D. and SAWITZKI, G. (1987). Using excess mass estimates to investigate the modality of distribution. Sonderforschungsbereich. Heidelberg University.
- [32] MÜLLER, D. W. and SAWITZKI, G. (1991). Excess mass estimates and tests for multimodality. *J. Amer. Statist. Assoc.* **86** 738–746. [MR1147099](#)
- [33] PATEIRO-LÓPEZ, B. and RODRÍGUEZ-CASAL, A. (2010). Generalizing the convex hull of a sample: The R package alphahull. *J. Stat. Softw.* **34** 1–28.
- [34] POLONIK, W. (1995). Measuring mass concentrations and estimating density contour clusters—an excess mass approach. *Ann. Statist.* **23** 855–881. [MR1345204](#) <https://doi.org/10.1214/aos/1176324626>
- [35] POLONIK, W. (1997). Minimum volume sets and generalized quantile processes. *Stochastic Process. Appl.* **69** 1–24. [MR1464172](#) [https://doi.org/10.1016/S0304-4149\(97\)00028-8](https://doi.org/10.1016/S0304-4149(97)00028-8)
- [36] QIAO, W. (2020). Asymptotics and optimal bandwidth for nonparametric estimation of density level sets. *Electron. J. Stat.* **14** 302–344. [MR4048601](#) <https://doi.org/10.1214/19-EJS1668>
- [37] RIGOLLET, P. and VERT, R. (2009). Optimal rates for plug-in estimators of density level sets. *Bernoulli* **15** 1154–1178. [MR2597587](#) <https://doi.org/10.3150/09-BEJ184>
- [38] RINALDO, A. and WASSERMAN, L. (2010). Generalized density clustering. *Ann. Statist.* **38** 2678–2722. [MR2722453](#) <https://doi.org/10.1214/10-AOS797>
- [39] RODRÍGUEZ-CASAL, A. and SAAVEDRA-NIEVES, P. (2016). A fully data-driven method for estimating the shape of a point cloud. *ESAIM Probab. Stat.* **20** 332–348. [MR3557598](#) <https://doi.org/10.1051/ps/2016015>
- [40] RODRÍGUEZ-CASAL, A. and SAAVEDRA-NIEVES, P. (2022). Supplement to “A data-adaptive method for estimating density level sets under shape conditions.” <https://doi.org/10.1214/21-AOS2168SUPP>
- [41] ROEDERER, M. and HARDY, R. R. (2001). Frequency difference gating: A multivariate method for identifying subsets that differ between samples. *Cytometry* **45** 56–64. [https://doi.org/10.1002/1097-0320\(20010901\)45:1<56::aid-cyto1144>3.0.co;2-9](https://doi.org/10.1002/1097-0320(20010901)45:1<56::aid-cyto1144>3.0.co;2-9)
- [42] SAMWORTH, R. J. and WAND, M. P. (2010). Asymptotics and optimal bandwidth selection for highest density region estimation. *Ann. Statist.* **38** 1767–1792. [MR2662359](#) <https://doi.org/10.1214/09-AOS766>

- [43] SERRA, J. (1984). *Image Analysis and Mathematical Morphology*. Academic Press, London. [MR0753649](#)
- [44] SINGH, A., SCOTT, C. and NOWAK, R. (2009). Adaptive Hausdorff estimation of density level sets. *Ann. Statist.* **37** 2760–2782. [MR2541446](#) <https://doi.org/10.1214/08-AOS661>
- [45] STEINWART, I. (2015). Fully adaptive density-based clustering. *Ann. Statist.* **43** 2132–2167. [MR3396981](#) <https://doi.org/10.1214/15-AOS1331>
- [46] TSYBAKOV, A. B. (1997). On nonparametric estimation of density level sets. *Ann. Statist.* **25** 948–969. [MR1447735](#) <https://doi.org/10.1214/aos/1069362732>
- [47] WALTHER, G. (1997). Granulometric smoothing. *Ann. Statist.* **25** 2273–2299. [MR1604445](#) <https://doi.org/10.1214/aos/1030741072>
- [48] WALTHER, G. (1999). On a generalization of Blaschke’s rolling theorem and the smoothing of surfaces. *Math. Methods Appl. Sci.* **22** 301–316. [MR1671447](#) [https://doi.org/10.1002/\(SICI\)1099-1476\(19990310\)22:4<301::AID-MMA42>3.3.CO;2-D](https://doi.org/10.1002/(SICI)1099-1476(19990310)22:4<301::AID-MMA42>3.3.CO;2-D)
- [49] WAND, M. P. and JONES, M. C. (1994). Multivariate plug-in bandwidth selection. *Comput. Statist.* **9** 97–116. [MR1280754](#)
- [50] XU, M. and SAMWORTH, R. J. (2021). High-dimensional nonparametric density estimation via symmetry and shape constraints. *Ann. Statist.* **49** 650–672. [MR4255102](#) <https://doi.org/10.1214/20-aos1972>

A PRECISE HIGH-DIMENSIONAL ASYMPTOTIC THEORY FOR BOOSTING AND MINIMUM- ℓ_1 -NORM INTERPOLATED CLASSIFIERS

BY TENGYUAN LIANG^{1,a} AND PRAGYA SUR^{*2,b}

¹Booth School of Business, University of Chicago, ^atengyuan.liang@chicagobooth.edu

²Department of Statistics, Harvard University, ^bpragya@fas.harvard.edu

This paper establishes a precise high-dimensional asymptotic theory for boosting on separable data, taking statistical and computational perspectives. We consider a high-dimensional setting where the number of features (weak learners) p scales with the sample size n , in an overparametrized regime. Under a class of statistical models, we provide an exact analysis of the generalization error of boosting when the algorithm interpolates the training data and maximizes the empirical ℓ_1 -margin. Further, we explicitly pin down the relation between the boosting test error and the optimal Bayes error, as well as the proportion of active features at interpolation (with zero initialization). In turn, these precise characterizations answer certain questions raised in (*Neural Comput.* **11** (1999) 1493–1517; *Ann. Statist.* **26** (1998) 1651–1686) surrounding boosting, under assumed data generating processes. At the heart of our theory lies an in-depth study of the maximum- ℓ_1 -margin, which can be accurately described by a new system of nonlinear equations; to analyze this margin, we rely on Gaussian comparison techniques and develop a novel uniform deviation argument. Our statistical and computational arguments can handle (1) any finite-rank spiked covariance model for the feature distribution and (2) variants of boosting corresponding to general ℓ_q -geometry, $q \in [1, 2]$. As a final component, via the Lindeberg principle, we establish a universality result showcasing that the scaled ℓ_1 -margin (asymptotically) remains the same, whether the covariates used for boosting arise from a nonlinear random feature model or an appropriately linearized model with matching moments.

REFERENCES

- [1] ADVANI, M. and GANGULI, S. (2016). Statistical mechanics of optimal convex inference in high dimensions. *Phys. Rev. A* **6** 031034.
- [2] ALBERT, A. and ANDERSON, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* **71** 1–10. [MR0738319](#) <https://doi.org/10.1093/biomet/71.1.1>
- [3] AMID, E. and WARMUTH, M. K. (2020). Winoing with gradient descent. *Proc. Mach. Learn. Res.* **125** 1–20.
- [4] BACH, F. (2017). Breaking the curse of dimensionality with convex neural networks. *J. Mach. Learn. Res.* **18** 19. [MR3634886](#)
- [5] 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>
- [6] BARTLETT, P. L. and TRASKIN, M. (2007). AdaBoost is consistent. *J. Mach. Learn. Res.* **8** 2347–2368. [MR2353835](#)
- [7] BEAN, D., BICKEL, P. J., EL KAROUI, N. and YU, B. (2013). Optimal M-estimation in high-dimensional regression. *Proc. Natl. Acad. Sci. USA* **110** 14563–14568.
- [8] BELKIN, M., HSU, D., MA, S. and MANDAL, S. (2018). Reconciling modern machine learning and the bias-variance trade-off. ArXiv preprint. Available at [arXiv:1812.11118](#).
- [9] 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>

MSC2020 subject classifications. Primary 68Q32; secondary 62H30.

Key words and phrases. Boosting, high-dimensional asymptotics, minimum-norm interpolation, over-parametrization, margin theory.

- [10] BELKIN, M., MA, S. and MANDAL, S. (2018). To understand deep learning we need to understand kernel learning. ArXiv preprint. Available at [arXiv:1802.01396](https://arxiv.org/abs/1802.01396).
- [11] BELKIN, M., RAKHLIN, A. and TSYBAKOV, A. B. (2018). Does data interpolation contradict statistical optimality? ArXiv preprint. Available at [arXiv:1806.09471](https://arxiv.org/abs/1806.09471).
- [12] BICKEL, P. J., RITOV, Y. and ZAKAI, A. (2006). Some theory for generalized Boosting Algorithms. *J. Mach. Learn. Res.* **7** 705–732. [MR2274384](https://arxiv.org/abs/2006.09471)
- [13] BLANCHARD, G., LUGOSI, G. and VAYATIS, N. (2004). On the rate of convergence of regularized boosting classifiers. *J. Mach. Learn. Res.* **4** 861–894. [MR2076000](https://arxiv.org/abs/2006.09471) <https://doi.org/10.1162/1532443041424319>
- [14] BREIMAN, L. (1996). Arcing classifiers. *Ann. Statist.* **26** 123–40.
- [15] BREIMAN, L. (1996). Bias, variance, and arcing classifiers Technical Report Tech. Rep. 460, Statistics Department, Univ. California, Berkeley.
- [16] BREIMAN, L. (1999). Prediction games and arcing algorithms. *Neural Comput.* **11** 1493–1517.
- [17] BREIMAN, L. (2004). Population theory for boosting ensembles. *Ann. Statist.* **32** 1–11. [MR2050998](https://arxiv.org/abs/2006.09471) <https://doi.org/10.1214/aos/1079120126>
- [18] BÜHLMANN, P. (2006). Boosting for high-dimensional linear models. *Ann. Statist.* **34** 559–583. [MR2281878](https://arxiv.org/abs/2006.09471) <https://doi.org/10.1214/0090536060000000092>
- [19] BÜHLMANN, P. and HOTHORN, T. (2007). Boosting Algorithms: Regularization, prediction and model fitting. *Statist. Sci.* **22** 477–505. [MR2420454](https://arxiv.org/abs/2006.09471) <https://doi.org/10.1214/07-STS242>
- [20] BÜHLMANN, P. and HOTHORN, T. (2010). Twin boosting: Improved feature selection and prediction. *Stat. Comput.* **20** 119–138. [MR2610767](https://arxiv.org/abs/2006.09471) <https://doi.org/10.1007/s11222-009-9148-5>
- [21] BÜHLMANN, P. and YU, B. (2003). Boosting with the L_2 loss: Regression and classification. *J. Amer. Statist. Assoc.* **98** 324–339. [MR1995709](https://arxiv.org/abs/2006.09471) <https://doi.org/10.1198/016214503000125>
- [22] BÜHLMANN, P. and YU, B. (2006). Sparse boosting. *J. Mach. Learn. Res.* **7** 1001–1024. [MR2274395](https://arxiv.org/abs/2006.09471)
- [23] BUNEA, F., STRIMAS-MACKEY, S. and WEGKAMP, M. (2020). Interpolation under latent factor regression models. ArXiv preprint. Available at [arXiv:2002.02525](https://arxiv.org/abs/2002.02525).
- [24] CANDÈS, E. J. and SUR, P. (2020). The phase transition for the existence of the maximum likelihood estimate in high-dimensional logistic regression. *Ann. Statist.* **48** 27–42. [MR4065151](https://arxiv.org/abs/2006.09471) <https://doi.org/10.1214/18-AOS1789>
- [25] CHATTERJI, N. S. and LONG, P. M. (2021). Finite-sample analysis of interpolating linear classifiers in the overparameterized regime. *J. Mach. Learn. Res.* **22** 129. [MR4279780](https://arxiv.org/abs/2006.09471)
- [26] CHIZAT, L. and BACH, F. (2020). Implicit bias of gradient descent for wide two-layer neural networks trained with the logistic loss. In *Conference on Learning Theory* 1305–1338. PMLR.
- [27] CHOULDECHOVA, A. and ROTH, A. (2018). The frontiers of fairness in machine learning. ArXiv preprint. Available at [arXiv:1810.08810](https://arxiv.org/abs/1810.08810).
- [28] 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.
- [29] DENG, Z., KAMMOUN, A. and THRAMOULIDIS, C. (2019). A model of double descent for high-dimensional binary linear classification. Available at [arXiv:1911.05822](https://arxiv.org/abs/1911.05822) [cs, Eess, Stat].
- [30] DHIFALLAH, O., THRAMOULIDIS, C. and LU, Y. M. (2018). Phase retrieval via polytope optimization: Geometry, phase transitions, and new algorithms. ArXiv preprint. Available at [arXiv:1805.09555](https://arxiv.org/abs/1805.09555).
- [31] DONOHO, D. and MONTANARI, A. (2016). High dimensional robust M-estimation: Asymptotic variance via approximate message passing. *Probab. Theory Related Fields* **166** 935–969. [MR3568043](https://arxiv.org/abs/2006.09471) <https://doi.org/10.1007/s00440-015-0675-z>
- [32] DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2009). Message-passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci. USA* **106** 18914–18919.
- [33] DOU, X. and LIANG, T. (2021). Training neural networks as learning data-adaptive kernels: Provable representation and approximation benefits. *J. Amer. Statist. Assoc.* **116** 1507–1520. [MR4309289](https://arxiv.org/abs/2006.09471) <https://doi.org/10.1080/01621459.2020.1745812>
- [34] DRUCKER, H. and CORTES, C. (1996). Boosting decision trees. In *Advances in Neural Information Processing Systems* 479–485.
- [35] EL KAROUI, N. (2018). On the impact of predictor geometry on the performance on high-dimensional ridge-regularized generalized robust regression estimators. *Probab. Theory Related Fields* **170** 95–175. [MR3748322](https://arxiv.org/abs/2006.09471) <https://doi.org/10.1007/s00440-016-0754-9>
- [36] EL KAROUI, N., BEAN, D., BICKEL, P. J., LIM, C. and YU, B. (2013). On robust regression with high-dimensional predictors. *Proc. Natl. Acad. Sci. USA* **110** 14557–14562.
- [37] FENG, O. Y., VENKATARAMANAN, R., RUSH, C. and SAMWORTH, R. J. (2021). A unifying tutorial on approximate message passing. ArXiv preprint. Available at [arXiv:2105.02180](https://arxiv.org/abs/2105.02180).
- [38] FREUND, R. M., GRIGAS, P. and MAZUMDER, R. (2013). Adaboost and forward stagewise regression are first-order convex optimization methods. ArXiv preprint. Available at [arXiv:1307.1192](https://arxiv.org/abs/1307.1192).

- [39] FREUND, R. M., GRIGAS, P. and MAZUMDER, R. (2017). A new perspective on boosting in linear regression via subgradient optimization and relatives. *Ann. Statist.* **45** 2328–2364. [MR3737894](#) <https://doi.org/10.1214/16-AOS1505>
- [40] FREUND, Y. (1995). Boosting a weak learning algorithm by majority. *Inform. and Comput.* **121** 256–285. [MR1348530](#) <https://doi.org/10.1006/inco.1995.1136>
- [41] FREUND, Y. and SCHAPIRE, R. E. (1995). A decision-theoretic generalization of on-line learning and an application to boosting. In *European Conference on Computational Learning Theory* 23–37. Springer, Berlin.
- [42] FREUND, Y. and SCHAPIRE, R. E. (1995). A decision-theoretic generalization of on-line learning and an application to boosting. In *European Conference on Computational Learning Theory* 23–37. Springer, Berlin.
- [43] FREUND, Y. and SCHAPIRE, R. E. (1996). Experiments with a new boosting algorithm. In *ICML* **96** 148–156. Citeseer.
- [44] FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2000). Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors). *Ann. Statist.* **28** 337–407. [MR1790002](#) <https://doi.org/10.1214/aos/1016218223>
- [45] FRIEDMAN, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Statist.* **29** 1189–1232. [MR1873328](#) <https://doi.org/10.1214/aos/1013203451>
- [46] GARDNER, E. (1988). The space of interactions in neural network models. *J. Phys. A* **21** 257–270. [MR0939730](#)
- [47] GORDON, Y. (1985). Some inequalities for Gaussian processes and applications. *Israel J. Math.* **50** 265–289. [MR0800188](#) <https://doi.org/10.1007/BF02759761>
- [48] GORDON, Y. (1988). On Milman’s inequality and random subspaces which escape through a mesh in $\mathbf{R}^n$. In *Geometric Aspects of Functional Analysis* (1986/87). *Lecture Notes in Math.* **1317** 84–106. Springer, Berlin. [MR0950977](#) <https://doi.org/10.1007/BFb0081737>
- [49] GROVE, A. J. and SCHUURMANS, D. (1998). Boosting in the limit: Maximizing the margin of learned ensembles. In *AAAI/IAAI* 692–699.
- [50] GUNASEKAR, S., LEE, J., SOUDRY, D. and SREBRO, N. (2018). Characterizing implicit bias in terms of optimization geometry. In *International Conference on Machine Learning* 1832–1841. PMLR.
- [51] HASTIE, T., MONTANARI, A., ROSSET, S. and TIBSHIRANI, R. J. (2019). Surprises in high-dimensional ridgeless least squares interpolation. ArXiv preprint. Available at [arXiv:1903.08560](#).
- [52] HU, H. and LU, Y. M. (2019). Asymptotics and optimal designs of SLOPE for sparse linear regression. In *2019 IEEE International Symposium on Information Theory (ISIT)* 375–379. IEEE, Los Alamitos.
- [53] HU, H. and LU, Y. M. (2020). Universality laws for high-dimensional learning with random features. ArXiv preprint. Available at [arXiv:2009.07669](#).
- [54] JI, Z. and TELGARSKY, M. (2021). Characterizing the implicit bias via a primal-dual analysis. In *Algorithmic Learning Theory* 772–804. PMLR.
- [55] JIANG, W. (2001). Some theoretical aspects of boosting in the presence of noisy data. In *Proceedings of the Eighteenth International Conference on Machine Learning* Citeseer.
- [56] JIANG, W. (2004). Process consistency for AdaBoost. *Ann. Statist.* **32** 13–29. [MR2050999](#) <https://doi.org/10.1214/aos/1079120128>
- [57] KLEINBERG, J. and MULLAINATHAN, S. (2019). Simplicity creates inequity: Implications for fairness, stereotypes, and interpretability. In *Proceedings of the 2019 ACM Conference on Economics and Computation* 807–808.
- [58] KOLTCHINSKII, V. and BEZNOSOVA, O. (2005). Exponential convergence rates in classification. In *Learning Theory. Lecture Notes in Computer Science* **3559** 295–307. Springer, Berlin. [MR2203269](#) https://doi.org/10.1007/11503415_20
- [59] KOLTCHINSKII, V. and PANCHENKO, D. (2002). Empirical margin distributions and bounding the generalization error of combined classifiers. *Ann. Statist.* **30** 1–50. [MR1892654](#) <https://doi.org/10.1214/aos/1015362182>
- [60] KOLTCHINSKII, V. and PANCHENKO, D. (2005). Complexities of convex combinations and bounding the generalization error in classification. *Ann. Statist.* **33** 1455–1496. [MR2166553](#) <https://doi.org/10.1214/009053605000000228>
- [61] LESAFFRE, E. and ALBERT, A. (1989). Partial separation in logistic discrimination. *J. Roy. Statist. Soc. Ser. B* **51** 109–116. [MR0984997](#)
- [62] LI, A. H. and BRADIC, J. (2018). Boosting in the presence of outliers: Adaptive classification with nonconvex loss functions. *J. Amer. Statist. Assoc.* **113** 660–674. [MR3832217](#) <https://doi.org/10.1080/01621459.2016.1273116>
- [63] 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>

- [64] 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 33rd Conference on Learning Theory* (J. Abernethy and S. Agarwal, eds.). *Proceedings of Machine Learning Research* **125** 2683–2711. PMLR.
- [65] LIANG, T. and RECHT, B. (2021). Interpolating classifiers make few mistakes. ArXiv preprint. Available at [arXiv:2101.11815](https://arxiv.org/abs/2101.11815).
- [66] LIANG, T. and SUR, P. (2022). Supplement to “A Precise High-Dimensional Asymptotic Theory for Boosting and Minimum- ℓ_1 -Norm Interpolated Classifiers.” <https://doi.org/10.1214/22-AOS2170SUPP>
- [67] LIANG, T. and TRAN-BACH, H. (2021). Mehler’s formula, branching process, and compositional kernels of deep neural networks. *J. Amer. Statist. Assoc.* **0** 1–14.
- [68] LIPTON, Z. C. (2018). The myths of model interpretability. *ACM Queue* **16** 31–57.
- [69] LUGOSI, G. and VAYATIS, N. (2004). On the Bayes-risk consistency of regularized boosting methods. *Ann. Statist.* **32** 30–55. [MR2051000 https://doi.org/10.1214/aos/1079120129](https://doi.org/10.1214/aos/1079120129)
- [70] MANNOR, S. and MEIR, R. (2001). Geometric bounds for generalization in boosting. In *Computational Learning Theory* (Amsterdam, 2001). *Lecture Notes in Computer Science* **2111** 461–472. Springer, Berlin. [MR2042053 https://doi.org/10.1007/3-540-44581-1_30](https://doi.org/10.1007/3-540-44581-1_30)
- [71] MANNOR, S. and MEIR, R. (2002). On the existence of linear weak learners and applications to boosting. *Mach. Learn.* **48** 219–251.
- [72] MANNOR, S., MEIR, R. and ZHANG, T. (2002). The consistency of greedy algorithms for classification. In *Computational Learning Theory* (Sydney, 2002). *Lecture Notes in Computer Science* **2375** 319–333. Springer, Berlin. [MR2040422 https://doi.org/10.1007/3-540-45435-7_22](https://doi.org/10.1007/3-540-45435-7_22)
- [73] MASON, L., BAXTER, J., BARTLETT, P. L. and FREAN, M. R. (2000). Boosting algorithms as gradient descent. In *Advances in Neural Information Processing Systems* 512–518.
- [74] MEI, S. and MONTANARI, A. (2019). The generalization error of random features regression: Precise asymptotics and the double descent curve. *Comm. Pure Appl. Math.*
- [75] 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. ArXiv preprint. Available at [arXiv:1911.01544](https://arxiv.org/abs/1911.01544).
- [76] MUKHERJEE, I., RUDIN, C. and SCHAPIRE, R. E. (2011). The rate of convergence of AdaBoost. In *Proceedings of the 24th Annual Conference on Learning Theory* 537–558.
- [77] QUINLAN, J. (1996). Bagging, boosting, and C4. 5. In ‘AAAI’96 Proceedings of the Thirteenth National Conference on Artificial Intelligence—Volume 1’, 4–8 August 1996, Portland, OR, USA.
- [78] RAHIMI, A. and RECHT, B. (2009). Weighted sums of random kitchen sinks: Replacing minimization with randomization in learning. In *Advances in Neural Information Processing Systems* 21 (D. Koller, D. Schuurmans, Y. Bengio and L. Bottou, eds.) 1313–1320. Curran Associates, Red Hook.
- [79] RÄTSCH, G., ONODA, T. and MÜLLER, K.-R. (2001). Soft margins for AdaBoost. *Mach. Learn.* **42** 287–320.
- [80] RÄTSCH, G. and WARMUTH, M. K. (2005). Efficient margin maximizing with boosting. *J. Mach. Learn. Res.* **6** 2131–2152. [MR2249883](https://doi.org/10.1162/jmlr.2005.6.2131)
- [81] REYZIN, L. and SCHAPIRE, R. E. (2006). How boosting the margin can also boost classifier complexity. In *Proceedings of the 23rd International Conference on Machine Learning* 753–760.
- [82] ROSSET, S., ZHU, J. and HASTIE, T. (2003/04). Boosting as a regularized path to a maximum margin classifier. *J. Mach. Learn. Res.* **5** 941–973. [MR2248005](https://doi.org/10.1162/jmlr.2004.5.941)
- [83] RUDIN, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **1** 206–215.
- [84] RUDIN, C., SCHAPIRE, R. E. and DAUBECHIES, I. (2007). Analysis of Boosting Algorithms using the smooth margin function. *Ann. Statist.* **35** 2723–2768. [MR2382664 https://doi.org/10.1214/009053607000000785](https://doi.org/10.1214/009053607000000785)
- [85] SALEHI, F., ABBASI, E. and HASSIBI, B. (2019). The impact of regularization on high-dimensional logistic regression. In *Advances in Neural Information Processing Systems* 11982–11992.
- [86] SANTNER, T. J. and DUFFY, D. E. (1986). A note on A. Albert and J. A. Anderson’s conditions for the existence of maximum likelihood estimates in logistic regression models. *Biometrika* **73** 755–758. [MR0897873 https://doi.org/10.1093/biomet/73.3.755](https://doi.org/10.1093/biomet/73.3.755)
- [87] SCHAPIRE, R. E. (1990). The strength of weak learnability. *Mach. Learn.* **5** 197–227.
- [88] SCHAPIRE, R. E., FREUND, Y., BARTLETT, P. and LEE, W. S. (1998). Boosting the margin: A new explanation for the effectiveness of voting methods. *Ann. Statist.* **26** 1651–1686. [MR1673273 https://doi.org/10.1214/aos/1024691352](https://doi.org/10.1214/aos/1024691352)
- [89] SHALEV-SHWARTZ, S. and SINGER, Y. (2010). On the equivalence of weak learnability and linear separability: New relaxations and efficient Boosting Algorithms. *Mach. Learn.* **80** 141–163. [MR3108163 https://doi.org/10.1007/s10994-010-5173-z](https://doi.org/10.1007/s10994-010-5173-z)

- [90] SHCHERBINA, M. and TIROZZI, B. (2003). Rigorous solution of the Gardner problem. *Comm. Math. Phys.* **234** 383–422. [MR1964377 https://doi.org/10.1007/s00220-002-0783-3](https://doi.org/10.1007/s00220-002-0783-3)
- [91] STOJNIC, M. (2013). A framework to characterize performance of lasso algorithms. ArXiv preprint. Available at [arXiv:1303.7291](https://arxiv.org/abs/1303.7291).
- [92] STOJNIC, M. (2013). Meshes that trap random subspaces. ArXiv preprint. Available at [arXiv:1304.0003](https://arxiv.org/abs/1304.0003).
- [93] STOJNIC, M. (2013). Upper-bounding ℓ_1 -optimization weak thresholds. Available at ArXiv.
- [94] SUR, P. (2019). *A Modern Maximum Likelihood Theory for High-Dimensional Logistic Regression*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—Stanford University. [MR4197622](https://arxiv.org/abs/1907.06222)
- [95] SUR, P. and CANDÈS, E. J. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. *Proc. Natl. Acad. Sci. USA* **116** 14516–14525. [MR3984492 https://doi.org/10.1073/pnas.1810420116](https://doi.org/10.1073/pnas.1810420116)
- [96] SUR, P., CHEN, Y. and CANDÈS, E. J. (2019). The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled chi-square. *Probab. Theory Related Fields* **175** 487–558. [MR4009715 https://doi.org/10.1007/s00440-018-00896-9](https://doi.org/10.1007/s00440-018-00896-9)
- [97] TAHERI, H., PEDARSANI, R. and THRAMPOLIDIS, C. (2020). Sharp asymptotics and optimal performance for inference in binary models. In *International Conference on Artificial Intelligence and Statistics* 3739–3749. PMLR.
- [98] TELGARSKY, M. (2013). Margins, shrinkage, and boosting. ArXiv preprint. Available at [arXiv:1303.4172](https://arxiv.org/abs/1303.4172).
- [99] THRAMPOLIDIS, C., ABBASI, E. and HASSIBI, B. (2015). Lasso with non-linear measurements is equivalent to one with linear measurements. *Adv. Neural Inf. Process. Syst.* **28** 3420–3428.
- [100] THRAMPOLIDIS, C., ABBASI, E. and HASSIBI, B. (2018). Precise error analysis of regularized M -estimators in high dimensions. *IEEE Trans. Inf. Theory* **64** 5592–5628. [MR3832326 https://doi.org/10.1109/TIT.2018.2840720](https://doi.org/10.1109/TIT.2018.2840720)
- [101] THRAMPOLIDIS, C., OYMAK, S. and HASSIBI, B. (2014). The Gaussian min-max theorem in the presence of convexity. ArXiv preprint. Available at [arXiv:1408.4837](https://arxiv.org/abs/1408.4837).
- [102] THRAMPOLIDIS, C., OYMAK, S. and HASSIBI, B. (2015). Regularized linear regression: A precise analysis of the estimation error. In *Conference on Learning Theory* 1683–1709.
- [103] VILLANI, C. (2008). *Optimal Transport: Old and New* **338**. Springer, Berlin.
- [104] WANG, S., WENG, H. and MALEKI, A. (2020). Which bridge estimator is the best for variable selection? *Ann. Statist.* **48** 2791–2823. [MR4152121 https://doi.org/10.1214/19-AOS1906](https://doi.org/10.1214/19-AOS1906)
- [105] WELLER, A. (2019). Transparency: Motivations and challenges. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* 23–40. Springer, Berlin.
- [106] ZHANG, C., BENGIO, S., HARDT, M., RECHT, B. and VINYALS, O. (2016). Understanding deep learning requires rethinking generalization. ArXiv preprint. Available at [arXiv:1611.03530](https://arxiv.org/abs/1611.03530).
- [107] ZHANG, T. (2004). Statistical behavior and consistency of classification methods based on convex risk minimization. *Ann. Statist.* **32** 56–85. [MR2051001 https://doi.org/10.1214/aos/1079120130](https://doi.org/10.1214/aos/1079120130)
- [108] ZHANG, T. and YU, B. (2005). Boosting with early stopping: Convergence and consistency. *Ann. Statist.* **33** 1538–1579. [MR2166555 https://doi.org/10.1214/00905360500000025](https://doi.org/10.1214/00905360500000025)
- [109] ZHAO, Q., SUR, P. and CANDÈS, E. J. (2020). The asymptotic distribution of the MLE in high-dimensional logistic models: Arbitrary covariance. ArXiv preprint. Available at [arXiv:2001.09351](https://arxiv.org/abs/2001.09351).

INTRINSIC RIEMANNIAN FUNCTIONAL DATA ANALYSIS FOR SPARSE LONGITUDINAL OBSERVATIONS

BY LINGXUAN SHAO^{1,a}, ZHENHUA LIN^{2,b} AND FANG YAO^{3,c}

For the Alzheimer's Disease Neuroimaging Initiative

¹*School of Mathematical Sciences, Center for Statistical Science, Peking University, ^ashao-14@pku.edu.cn*

²*Department of Statistics and Data Science, National University of Singapore, ^blinz@nus.edu.sg*

³*School of Mathematical Sciences, Center for Statistical Science, Peking University, ^cfyao@math.pku.edu.cn*

A new framework is developed to intrinsically analyze sparsely observed Riemannian functional data. It features four innovative components: a frame-independent covariance function, a smooth vector bundle termed *covariance vector bundle*, a parallel transport and a smooth bundle metric on the covariance vector bundle. The introduced intrinsic covariance function links estimation of covariance structure to smoothing problems that involve raw covariance observations derived from sparsely observed Riemannian functional data, while the covariance vector bundle provides a rigorous mathematical foundation for formulating such smoothing problems. The parallel transport and the bundle metric together make it possible to measure fidelity of fit to the covariance function. They also play a critical role in quantifying the quality of estimators for the covariance function. As an illustration, based on the proposed framework, we develop a local linear smoothing estimator for the covariance function, analyze its theoretical properties and provide numerical demonstration via simulated and real data sets. The intrinsic feature of the framework makes it applicable to not only Euclidean submanifolds but also manifolds without a canonical ambient space.

REFERENCES

- AFSARI, B. (2011). Riemannian L^p center of mass: Existence, uniqueness, and convexity. *Proc. Amer. Math. Soc.* **139** 655–673. [MR2736346 https://doi.org/10.1090/S0002-9939-2010-10541-5](https://doi.org/10.1090/S0002-9939-2010-10541-5)
- ANEIROS, G., CAO, R., FRAIMAN, R., GENEST, C. and VIEU, P. (2019). Recent advances in functional data analysis and high-dimensional statistics. *J. Multivariate Anal.* **170** 3–9. [MR3913024 https://doi.org/10.1016/j.jmva.2018.11.007](https://doi.org/10.1016/j.jmva.2018.11.007)
- ARSIGNY, V., FILLARD, P., PENNEC, X. and AYACHE, N. (2006). Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magn. Reson. Med.* **56** 411–421.
- ARSIGNY, V., FILLARD, P., PENNEC, X. and AYACHE, N. (2006/07). Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **29** 328–347. [MR2288028 https://doi.org/10.1137/050637996](https://doi.org/10.1137/050637996)
- BHATTACHARYA, R. and PATRANGENARU, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann. Statist.* **31** 1–29. [MR1962498 https://doi.org/10.1214/aos/1046294456](https://doi.org/10.1214/aos/1046294456)
- BHATTACHARYA, R. and PATRANGENARU, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *Ann. Statist.* **33** 1225–1259. [MR2195634 https://doi.org/10.1214/009053605000000093](https://doi.org/10.1214/009053605000000093)
- CAI, T. and YUAN, M. (2010). Nonparametric covariance function estimation for functional and longitudinal data Technical Report Univ. Pennsylvania.
- CAI, T. T. and YUAN, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *Ann. Statist.* **39** 2330–2355. [MR2906870 https://doi.org/10.1214/11-AOS898](https://doi.org/10.1214/11-AOS898)
- 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](https://doi.org/10.1111/rssb.12169)

MSC2020 subject classifications. Primary 62R10; secondary 62R30.

Key words and phrases. Diffusion tensor, Fréchet mean, intrinsic covariance function, parallel transport, smoothing, vector bundle.

- DAI, X., LIN, Z. and MÜLLER, H.-G. (2021). Modeling sparse longitudinal data on Riemannian manifolds. *Biometrics* **77** 1328–1341. [MR4357841](#) <https://doi.org/10.1111/biom.13385>
- DAI, X. and MÜLLER, H.-G. (2018). Principal component analysis for functional data on Riemannian manifolds and spheres. *Ann. Statist.* **46** 3334–3361. [MR3852654](#) <https://doi.org/10.1214/17-AOS1660>
- DELAIGLE, A., HALL, P., HUANG, W. and KNEIP, A. (2021). Estimating the covariance of fragmented and other related types of functional data. *J. Amer. Statist. Assoc.* **116** 1383–1401. [MR4309280](#) <https://doi.org/10.1080/01621459.2020.1723597>
- DESCARY, M.-H. and PANARETOS, V. M. (2019). Recovering covariance from functional fragments. *Biometrika* **106** 145–160. [MR3912388](#) <https://doi.org/10.1093/biomet/asy055>
- DRYDEN, I. L., KOLOYDENKO, A. and ZHOU, D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Ann. Appl. Stat.* **3** 1102–1123. [MR2750388](#) <https://doi.org/10.1214/09-AOAS249>
- DUBEY, P. and MÜLLER, H.-G. (2020). Functional models for time-varying random objects. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 275–327. [MR4084166](#)
- FARAWAY, J. J. (2014). Regression for non-Euclidean data using distance matrices. *J. Appl. Stat.* **41** 2342–2357. [MR3256391](#) <https://doi.org/10.1080/02664763.2014.909794>
- FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Series in Statistics. Springer, New York. [MR2229687](#)
- FILLARD, P., ARSIGNY, V., AYACHE, N. and PENNEC, X. (2005). A Riemannian framework for the processing of tensor-valued images. In *Deep Structure, Singularities, and Computer Vision* (O. Fogh Olsen, L. Florack and A. Kuijper, eds.) 112–123. Springer, Heidelberg.
- FLETCHER, P. T. (2013). Geodesic regression and the theory of least squares on Riemannian manifolds. *Int. J. Comput. Vis.* **105** 171–185. [MR3104017](#) <https://doi.org/10.1007/s11263-012-0591-y>
- FLETCHER, T. and JOSHI, S. (2007). Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Process.* **87** 250–262.
- HALL, P., MÜLLER, H.-G. and WANG, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34** 1493–1517. [MR2278365](#) <https://doi.org/10.1214/009053606000000272>
- HEIN, M. (2009). Robust nonparametric regression with metric-space valued output. In *Advances in Neural Information Processing Systems* 718–726.
- HINKLE, J., FLETCHER, P. T. and JOSHI, S. (2014). Intrinsic polynomials for regression on Riemannian manifolds. *J. Math. Imaging Vision* **50** 32–52. [MR3233133](#) <https://doi.org/10.1007/s10851-013-0489-5>
- HSING, T. and EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley Series in Probability and Statistics. Wiley, Chichester. [MR3379106](#) <https://doi.org/10.1002/9781118762547>
- KOKOSZKA, P. and REIMHERR, M. (2017). *Introduction to Functional Data Analysis*. Texts in Statistical Science Series. CRC Press, Boca Raton, FL. [MR3793167](#)
- LEE, J. M. (1997). *Riemannian Manifolds: An Introduction to Curvature*. Graduate Texts in Mathematics **176**. Springer, New York. [MR1468735](#) <https://doi.org/10.1007/b98852>
- LENGLET, C., ROUSSON, M., DERICHE, R. and FAUGERAS, O. (2006). Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor MRI processing. *J. Math. Imaging Vision* **25** 423–444. [MR2283616](#) <https://doi.org/10.1007/s10851-006-6897-z>
- LI, Y. and HSING, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist.* **38** 3321–3351. [MR2766854](#) <https://doi.org/10.1214/10-AOS813>
- LIN, Z. (2019). Riemannian geometry of symmetric positive definite matrices via Cholesky decomposition. *SIAM J. Matrix Anal. Appl.* **40** 1353–1370. [MR4032859](#) <https://doi.org/10.1137/18M1221084>
- LIN, Z. and MÜLLER, H.-G. (2021). Total variation regularized Fréchet regression for metric-space valued data. *Ann. Statist.* **49** 3510–3533. [MR4352539](#) <https://doi.org/10.1214/21-aos2095>
- LIN, Z., WANG, J.-L. and ZHONG, Q. (2021). Basis expansions for functional snippets. *Biometrika* **108** 709–726. [MR4298773](#) <https://doi.org/10.1093/biomet/asaa088>
- LIN, Z. and YAO, F. (2019). Intrinsic Riemannian functional data analysis. *Ann. Statist.* **47** 3533–3577. [MR4025751](#) <https://doi.org/10.1214/18-AOS1787>
- LINDBERG, O., WALTERFANG, M., LOOI, J. C. L., MALYKHIN, N., ÖSTBERG, P., ZANDBELT, B., STYNER, M., VELAKOULIS, D., ÖRNDahl, E. et al. (2012). Shape analysis of the hippocampus in Alzheimer’s disease and subtypes of frontotemporal lobar degeneration. *J. Alzheimer’s Dis.* **30** 355–365.
- MOAKHER, M. (2005). A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **26** 735–747. [MR2137480](#) <https://doi.org/10.1137/S0895479803436937>
- PELLETIER, B. (2006). Non-parametric regression estimation on closed Riemannian manifolds. *J. Nonparametr. Stat.* **18** 57–67. [MR2214065](#) <https://doi.org/10.1080/10485250500504828>

- PENNEC, X. (2019). Curvature effects on the empirical mean in Riemannian and affine manifolds: A non-asymptotic high concentration expansion in the small-sample regime. Available at [arXiv:arxiv.org/1906.07418](https://arxiv.org/1906.07418).
- PENNEC, X. (2020). Manifold-valued image processing with SPD matrices. In *Riemannian Geometric Statistics in Medical Image Analysis* 75–134. Elsevier, London. [MR4254214 https://doi.org/10.1016/B978-0-12-814725-2.00010-8](https://doi.org/10.1016/B978-0-12-814725-2.00010-8)
- PENNEC, X., FILLARD, P. and AYACHE, N. (2006). A Riemannian framework for tensor computing. *Int. J. Comput. Vis.* **66** 41–66. <https://doi.org/10.1007/s11263-005-3222-z>
- PETERSEN, A. and MÜLLER, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *Ann. Statist.* **47** 691–719. [MR3909947 https://doi.org/10.1214/17-AOS1624](https://doi.org/10.1214/17-AOS1624)
- PRÉVÔT, C. and RÖCKNER, M. (2007). *A Concise Course on Stochastic Partial Differential Equations. Lecture Notes in Math.* **1905**. Springer, Berlin. [MR2329435](https://doi.org/10.1007/978-3-540-72868-9)
- RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. [MR2168993](https://doi.org/10.1007/978-1-4939-9738-9)
- SCHÖTZ, C. (2019). Convergence rates for the generalized Fréchet mean via the quadruple inequality. *Electron. J. Stat.* **13** 4280–4345. [MR4023955 https://doi.org/10.1214/19-EJS1618](https://doi.org/10.1214/19-EJS1618)
- SHAO, L., LIN, Z. and YAO, F. (2022). Supplement to “Intrinsic Riemannian functional data analysis for sparse longitudinal observations.” <https://doi.org/10.1214/22-AOS2172SUPP>
- SHI, X., STYNER, M., LIEBERMAN, J., IBRAHIM, J. G., LIN, W. and ZHU, H. (2009). Intrinsic regression models for manifold-valued data. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI* **12** 192–199.
- STEINKE, F., HEIN, M. and SCHÖLKOPF, B. (2010). Nonparametric regression between general Riemannian manifolds. *SIAM J. Imaging Sci.* **3** 527–563. [MR2736019 https://doi.org/10.1137/080744189](https://doi.org/10.1137/080744189)
- STURM, K.-T. (2003). Probability measures on metric spaces of nonpositive curvature. In *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces (Paris, 2002)*. *Contemp. Math.* **338** 357–390. Amer. Math. Soc., Providence, RI. [MR2039961 https://doi.org/10.1090/conm/338/06080](https://doi.org/10.1090/conm/338/06080)
- SU, J., KURTEK, S., KLASSEN, E. and SRIVASTAVA, A. (2014). Statistical analysis of trajectories on Riemannian manifolds: Bird migration, hurricane tracking and video surveillance. *Ann. Appl. Stat.* **8** 530–552. [MR3192001 https://doi.org/10.1214/13-AOAS701](https://doi.org/10.1214/13-AOAS701)
- WANG, J.-L., CHIOU, J.-M. and MÜLLER, H.-G. (2016). Review of functional data analysis. *Annu. Rev. Stat. Appl.* **3** 257–295.
- YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. [MR2160561 https://doi.org/10.1198/016214504000001745](https://doi.org/10.1198/016214504000001745)
- ZHANG, Z., KLASSEN, E. and SRIVASTAVA, A. (2018). Phase-amplitude separation and modeling of spherical trajectories. *J. Comput. Graph. Statist.* **27** 85–97. [MR3788303 https://doi.org/10.1080/10618600.2017.1340892](https://doi.org/10.1080/10618600.2017.1340892)
- ZHANG, X. and WANG, J.-L. (2016). From sparse to dense functional data and beyond. *Ann. Statist.* **44** 2281–2321. [MR3546451 https://doi.org/10.1214/16-AOS1446](https://doi.org/10.1214/16-AOS1446)
- ZHANG, X. and WANG, J.-L. (2018). Optimal weighting schemes for longitudinal and functional data. *Statist. Probab. Lett.* **138** 165–170. [MR3788733 https://doi.org/10.1016/j.spl.2018.03.007](https://doi.org/10.1016/j.spl.2018.03.007)
- ZHU, H., CHEN, Y., IBRAHIM, J. G., LI, Y., HALL, C. and LIN, W. (2009). Intrinsic regression models for positive-definite matrices with applications to diffusion tensor imaging. *J. Amer. Statist. Assoc.* **104** 1203–1212. [MR2750245 https://doi.org/10.1198/jasa.2009.tm08096](https://doi.org/10.1198/jasa.2009.tm08096)

SPECTRAL ESTIMATION OF HAWKES PROCESSES FROM COUNT DATA

BY FELIX CHEYSSON^a AND GABRIEL LANG^b

UMR MIA-Paris, Université Paris-Saclay, AgroParisTech, INRAE, ^afelix@cheysson.fr, ^bgabriel.lang@agroparistech.fr

This paper presents a parametric estimation method for ill-observed linear stationary Hawkes processes. When the exact locations of points are not observed, but only counts over time intervals of fixed size, methods based on the likelihood are not feasible. We show that spectral estimation based on Whittle's method is adapted to this case and provides consistent and asymptotically normal estimators, provided a mild moment condition on the reproduction function. Simulated data sets and a case-study illustrate the performances of the estimation, notably of the reproduction function even when time intervals are relatively large.

REFERENCES

- ADAMOPOULOS, L. (1976). Cluster models for earthquakes: Regional comparisons. *J. Int. Assoc. Math. Geol.* **8** 463–475.
- BACRY, E., MASTROMATTEO, I. and MUZY, J.-F. (2015). Hawkes processes in finance. *Mark. Microstruct. Liq.* **1** 1550005.
- BOWSER, C. G. (2003). Modelling security market events in continuous time: Intensity based, multivariate point process models. *SSRN Electron. J.* 1–39.
- BRADLEY, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** 107–144. [MR2178042 https://doi.org/10.1214/154957805100000104](https://doi.org/10.1214/154957805100000104)
- CELEUX, G., CHAUVEAU, D. and DIEBOLT, J. (1995). On Stochastic Versions of the EM Algorithm. Technical Report No. RR-2514 INRIA.
- CENTERS FOR DISEASE CONTROL AND PREVENTION (2015). *Epidemiology and Prevention of Vaccine-Preventable Diseases*, 13 ed. Public Health Foundation, Washington D.C.
- CHAVEZ-DEMOULIN, V., DAVISON, A. C. and MCNEIL, A. J. (2005). Estimating value-at-risk: A point process approach. *Quant. Finance* **5** 227–234. [MR2240248 https://doi.org/10.1080/14697680500039613](https://doi.org/10.1080/14697680500039613)
- CHORNOBOY, E. S., SCHRAMM, L. P. and KARR, A. F. (1988). Maximum likelihood identification of neural point process systems. *Biol. Cybernet.* **59** 265–275. [MR0961117 https://doi.org/10.1007/BF00332915](https://doi.org/10.1007/BF00332915)
- DALEY, D. J. and VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes. Vol. I*, 2nd ed. *Probability and Its Applications (New York)*. Springer, New York. [MR1950431](https://doi.org/10.1007/978-1-4612-2642-0)
- DAVYDOV, J. A. (1974). Mixing conditions for Markov chains. *Theory Probab. Appl.* **18** 312–328.
- DELYON, B., LAVIELLE, M. and MOULINES, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *Ann. Statist.* **27** 94–128. [MR1701103 https://doi.org/10.1214/aos/1018031103](https://doi.org/10.1214/aos/1018031103)
- DOUKHAN, P. (1994). *Mixing: Properties and Examples. Lecture Notes in Statistics* **85**. Springer, New York. [MR1312160 https://doi.org/10.1007/978-1-4612-2642-0](https://doi.org/10.1007/978-1-4612-2642-0)
- DZHAPARIDZE, K. O. (1974). A new method for estimating spectral parameters of a stationary regular time series. *Theory Probab. Appl.* **19** 122–132.
- DZHAPARIDZE, K. (1986). *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series. Springer Series in Statistics*. Springer, New York. [MR0812272 https://doi.org/10.1007/978-1-4612-4842-2](https://doi.org/10.1007/978-1-4612-4842-2)
- EDDELBUETTTEL, D. and FRANÇOIS, R. (2011). Rcpp: Seamless R and C++ integration. *J. Stat. Softw.* **40** 1–18.
- GAO, X. and ZHU, L. (2018). Functional central limit theorems for stationary Hawkes processes and application to infinite-server queues. *Queueing Syst.* **90** 161–206. [MR3850052 https://doi.org/10.1007/s11134-018-9570-5](https://doi.org/10.1007/s11134-018-9570-5)
- HAWKES, A. G. (1971a). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58** 83–90. [MR0278410 https://doi.org/10.1093/biomet/58.1.83](https://doi.org/10.1093/biomet/58.1.83)
- HAWKES, A. G. (1971b). Point spectra of some mutually exciting point processes. *J. Roy. Statist. Soc. Ser. B* **33** 438–443. [MR0358976](https://doi.org/10.1093/biomet/33.3.438)

- HAWKES, A. G. and OAKES, D. (1974). A cluster process representation of a self-exciting process. *J. Appl. Probab.* **11** 493–503. [MR0378093](#) <https://doi.org/10.2307/3212693>
- HEINRICH, L. and PAWLAS, Z. (2013). Absolute regularity and Brillinger-mixing of stationary point processes. *Lith. Math. J.* **53** 293–310. [MR3097306](#) <https://doi.org/10.1007/s10986-013-9209-5>
- HOSOYA, Y. (1974). Estimation problems on stationary time series models. Ph.D. dissertation, Yale Univ.
- JOHNSON, S. G. The NLOpt nonlinear-optimization package.
- KIRCHNER, M. (2016). Hawkes and INAR(∞) processes. *Stochastic Process. Appl.* **126** 2494–2525. [MR3505235](#) <https://doi.org/10.1016/j.spa.2016.02.008>
- KIRCHNER, M. (2017). An estimation procedure for the Hawkes process. *Quant. Finance* **17** 571–595. [MR3620953](#) <https://doi.org/10.1080/14697688.2016.1211312>
- LIU, D. C. and NOCEDAL, J. (1989). On the limited memory BFGS method for large scale optimization. *Math. Program.* **45** 503–528. [MR1038245](#) <https://doi.org/10.1007/BF01589116>
- MEYER, S., ELIAS, J. and HÖHLE, M. (2012). A space–time conditional intensity model for invasive meningococcal disease occurrence. *Biometrics* **68** 607–616. [MR2959628](#) <https://doi.org/10.1111/j.1541-0420.2011.01684.x>
- NISHIURA, H., MIZUMOTO, K. and ASAI, Y. (2017). Assessing the transmission dynamics of measles in Japan, 2016. *Epidemics* **20** 67–72. <https://doi.org/10.1016/j.epidem.2017.03.005>
- NOCEDAL, J. (1980). Updating quasi-Newton matrices with limited storage. *Math. Comp.* **35** 773–782. [MR0572855](#) <https://doi.org/10.2307/2006193>
- OAKES, D. (1975). The Markovian self-exciting process. *J. Appl. Probab.* **12** 69–77. [MR0362522](#) <https://doi.org/10.1017/s0021900200033106>
- OGATA, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Ann. Inst. Statist. Math.* **30** 243–261. [MR0514494](#) <https://doi.org/10.1007/BF02480216>
- OGATA, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.* **83** 9–27.
- OLSON, J. F. and CARLEY, K. M. (2013). Exact and approximate EM estimation of mutually exciting Hawkes processes. *Stat. Inference Stoch. Processes.* **16** 63–80. [MR3029333](#) <https://doi.org/10.1007/s11203-013-9074-1>
- OZAKI, T. (1979). Maximum likelihood estimation of Hawkes’ self-exciting point processes. *Ann. Inst. Statist. Math.* **31** 145–155. [MR0541960](#) <https://doi.org/10.1007/BF02480272>
- PAPANGELOU, F. (1972). Integrability of expected increments of point processes and a related random change of scale. *Trans. Amer. Math. Soc.* **165** 483–506. [MR0314102](#) <https://doi.org/10.2307/1995899>
- PAPARODITIS, E. (2000). Spectral density based goodness-of-fit tests for time series models. *Scand. J. Stat.* **27** 143–176. [MR1774049](#) <https://doi.org/10.1111/1467-9469.00184>
- POINAS, A., DELYON, B. and LAVANCIER, F. (2019). Mixing properties and central limit theorem for associated point processes. *Bernoulli* **25** 1724–1754. [MR3961228](#) <https://doi.org/10.3150/18-BEJ1033>
- R CORE TEAM (2019). R: A Language and Environment for Statistical Computing.
- REINHART, A. (2018). A review of self-exciting spatio-temporal point processes and their applications. *Statist. Sci.* **33** 299–318. [MR3843374](#) <https://doi.org/10.1214/17-STS629>
- REYNAUD-BOURET, P. and SCHBATH, S. (2010). Adaptive estimation for Hawkes processes; application to genome analysis. *Ann. Statist.* **38** 2781–2822. [MR2722456](#) <https://doi.org/10.1214/10-AOS806>
- RIO, E. (2017). *Asymptotic Theory of Weakly Dependent Random Processes. Probability Theory and Stochastic Modelling* **80**. Springer Berlin Heidelberg, Berlin, Heidelberg.
- 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>
- ROUEFF, F., VON SACHS, R. and SANSONNET, L. (2016). Locally stationary Hawkes processes. *Stochastic Process. Appl.* **126** 1710–1743. [MR3483734](#) <https://doi.org/10.1016/j.spa.2015.12.003>
- SHAO, X. (2010). A self-normalized approach to confidence interval construction in time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 343–366. [MR2758116](#) <https://doi.org/10.1111/j.1467-9868.2009.00737.x>
- WESTCOTT, M. (1971). On existence and mixing results for cluster point processes. *J. Roy. Statist. Soc. Ser. B* **33** 290–300. [MR0317443](#)
- WESTCOTT, M. (1972). The probability generating functional. *J. Aust. Math. Soc.* **14** 448–466. [MR0324772](#)
- WHEATLEY, S., FILIMONOV, V. and SORNETTE, D. (2016). The Hawkes process with renewal immigration & its estimation with an EM algorithm. *Comput. Statist. Data Anal.* **94** 120–135. [MR3412815](#) <https://doi.org/10.1016/j.csda.2015.08.007>
- WHITTLE, P. (1952). Some results in time series analysis. *Skand. Aktuarietidskr.* **35** 48–60. [MR0049539](#) <https://doi.org/10.1080/03461238.1952.10414182>

MODEL SELECTION IN THE SPACE OF GAUSSIAN MODELS INVARIANT BY SYMMETRY

BY PIOTR GRACZYK^{1,a}, HIDEYUKI ISHI^{2,b}, BARTOSZ KOŁODZIEJEK^{3,c} AND
HÉLÈNE MASSAM^{4,d}

¹LAREMA, UFR Sciences, Université d'Angers, ^apiotr.graczyk@univ-angers.fr

²Department of Mathematics, Osaka City University, ^bhideyuki@sci.osaka-cu.ac.jp

³Faculty of Mathematics and Information Science, Warsaw University of Technology, ^cb.kolodziejek@mini.pw.edu.pl

⁴Department of Mathematics and Statistics, York University, ^dmassamh@mathstat.yorku.ca

We consider multivariate centered Gaussian models for the random variable $Z = (Z_1, \dots, Z_p)$, invariant under the action of a subgroup of the group of permutations on $\{1, \dots, p\}$. Using the representation theory of the symmetric group on the field of reals, we derive the distribution of the maximum likelihood estimate of the covariance parameter Σ and also the analytic expression of the normalizing constant of the Diaconis–Ylvisaker conjugate prior for the precision parameter $K = \Sigma^{-1}$. We can thus perform Bayesian model selection in the class of complete Gaussian models invariant by the action of a subgroup of the symmetric group, which we could also call complete RCOP models. We illustrate our results with a toy example of dimension 4 and several examples for selection within cyclic groups, including a high-dimensional example with $p = 100$.

REFERENCES

- ANDERSSON, S. (1975). Invariant normal models. *Ann. Statist.* **3** 132–154. [MR0362703](#)
- ANDERSSON, S. A., BRØNS, H. K. and JENSEN, S. T. (1983). Distribution of eigenvalues in multivariate statistical analysis. *Ann. Statist.* **11** 392–415. [MR0696055](#) <https://doi.org/10.1214/aos/1176346149>
- ANDERSSON, S. and MADSEN, J. (1998). Symmetry and lattice conditional independence in a multivariate normal distribution. *Ann. Statist.* **26** 525–572. [MR1626059](#) <https://doi.org/10.1214/aos/1028144848>
- BARNDORFF-NIELSEN, O. (2014). *Information and Exponential Families in Statistical Theory*. Wiley Series in Probability and Statistics. Wiley, Chichester. [MR3221776](#) <https://doi.org/10.1002/9781118857281>
- BURKE, C. J. and ROSENBLATT, M. (1958). A Markovian function of a Markov chain. *Ann. Math. Stat.* **29** 1112–1122. [MR0101557](#) <https://doi.org/10.1214/aoms/1177706444>
- DAVIES, I. and MARIGLIANO, O. (2021). Coloured graphical models and their symmetries. *Matematiche (Catania)* **76** 501–515. [MR4334901](#) <https://doi.org/10.4418/2021.76.2.13>
- DE MAIO, A., ORLANDO, D., SOLOVEYCHIK, I. and WIESEL, A. (2016). Invariance theory for adaptive detection in interference with group symmetric covariance matrix. *IEEE Trans. Signal Process.* **64** 6299–6312. [MR3562723](#) <https://doi.org/10.1109/TSP.2016.2591502>
- DESCATHA, A., ROQUELAURE, Y., EVANOFF, B., NIEDHAMMER, I., CHASTANG, J. F., MARIOT, C., HA, C., IMBERNON, E., GOLDBERG, M. et al. (2007). Selected questions on biomechanical exposures for surveillance of upper-limb work-related musculoskeletal disorders. *Int. Arch. Occup. Environ. Health* **81** 1–8. <https://doi.org/10.1007/s00420-007-0180-5>
- DIACONIS, P. and YLVISAKER, D. (1979). Conjugate priors for exponential families. *Ann. Statist.* **7** 269–281. [MR0520238](#)
- FARAUT, J. and KORÁNYI, A. (1994). *Analysis on Symmetric Cones*. Oxford Mathematical Monographs. The Clarendon Press, Oxford Univ. Press, New York. [MR1446489](#)
- FRETS, G. P. (1921). Heredity of head form in man. *Genetica* **41** 193–400.
- FROMMLET, F., BOGDAN, M. and RAMSEY, D. (2016). *Phenotypes and Genotypes: The Search for Influential Genes*. Computational Biology **18**. Springer, London. [MR3443801](#) <https://doi.org/10.1007/978-1-4471-5310-8>

MSC2020 subject classifications. Primary 62H99, 62F15; secondary 20C35.

Key words and phrases. Colored graph, conjugate prior, covariance selection, invariance, permutation symmetry.

- GAO, X. and MASSAM, H. (2015). Estimation of symmetry-constrained Gaussian graphical models: Application to clustered dense networks. *J. Comput. Graph. Statist.* **24** 909–929. [MR3432922](#) <https://doi.org/10.1080/10618600.2014.937811>
- GEHRMANN, H. (2011). Lattices of graphical Gaussian models with symmetries. *Symmetry* **3** 653–679. [MR2845303](#) <https://doi.org/10.3390/sym3030653>
- 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>
- 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>
- GRACZYK, P., ISHI, H., KOŁODZIEJEK, B. and MASSAM, H. (2022). Supplement to “Model selection in the space of Gaussian models invariant by symmetry.” <https://doi.org/10.1214/22-AOS2174SUPP>
- GRAHAM, R. L., GRÖTSCHEL, M. and LOVÁSZ, L., eds. (1995). *Handbook of Combinatorics*. Vols. 1, 2. Elsevier, Amsterdam; MIT Press, Cambridge, MA.
- HASSAIRI, A. and LAJMI, S. (2001). Riesz exponential families on symmetric cones. *J. Theoret. Probab.* **14** 927–948. [MR1860082](#) <https://doi.org/10.1023/A:1012592618872>
- HØJSGAARD, S. and LAURITZEN, S. L. (2008). Graphical Gaussian models with edge and vertex symmetries. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 1005–1027. [MR2530327](#) <https://doi.org/10.1111/j.1467-9868.2008.00666.x>
- HOLT, D. F. (2010). Enumerating subgroups of the symmetric group. In *Computational Group Theory and the Theory of Groups, II. Contemp. Math.* **511** 33–37. Amer. Math. Soc., Providence, RI. [MR2655292](#) <https://doi.org/10.1090/conm/511/10041>
- HUBERT, L. and ARABIE, P. (1985). Comparing partitions. *J. Classification* **2** 193–218.
- JENSEN, S. T. (1988). Covariance hypotheses which are linear in both the covariance and the inverse covariance. *Ann. Statist.* **16** 302–322. [MR0924873](#) <https://doi.org/10.1214/aos/1176350707>
- LI, S. Z. (2009). *Markov Random Field Modeling in Image Analysis*, 3rd ed. *Advances in Pattern Recognition*. Springer London, Ltd., London. [MR2493908](#)
- LI, Q., GAO, X. and MASSAM, H. (2020). Bayesian model selection approach for coloured graphical Gaussian models. *J. Stat. Comput. Simul.* **90** 2631–2654. [MR4145359](#) <https://doi.org/10.1080/00949655.2020.1784175>
- MAATHUIS, M., DRTON, M., LAURITZEN, S. and WAINWRIGHT, M., eds. (2018). *Handbook of Graphical Models*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, Boca Raton, FL. [MR3889064](#)
- MADSEN, J. (2000). Invariant normal models with recursive graphical Markov structure. *Ann. Statist.* **28** 1150–1178. [MR1810923](#) <https://doi.org/10.1214/aos/1015956711>
- MASSAM, H., LI, Q. and GAO, X. (2018). Bayesian precision and covariance matrix estimation for graphical Gaussian models with edge and vertex symmetries. *Biometrika* **105** 371–388. [MR3804408](#) <https://doi.org/10.1093/biomet/asx084>
- MICHAŁEK, M., STURMFELS, B., UHLER, C. and ZWIERNIK, P. (2016). Exponential varieties. *Proc. Lond. Math. Soc.* (3) **112** 27–56. [MR3458144](#) <https://doi.org/10.1112/plms/pdv066>
- MILLER, L. D., SMEDS, J., GEORGE, J., VEGA, V. B., VERGARA, L., PLONER, A., PAWITAN, Y., HALL, P., KLAAR, S. et al. (2005). An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc. Natl. Acad. Sci. USA* **102** 13550–13555.
- MISSIO, G., MORENO, D. H., DEMETRIO, F. N., SOEIRO-DE-SOUZA, M. G., FERNANDES, F. D. S., BARROS, V. B. and MORENO, R. A. (2019). A randomized controlled trial comparing lithium plus valproic acid versus lithium plus carbamazepine in young patients with type 1 bipolar disorder: The LICAVAL study. *Trials* **20** 608. <https://doi.org/10.1186/s13063-019-3655-2>
- OLKIN, I. and PRESS, S. J. (1969). Testing and estimation for a circular stationary model. *Ann. Math. Stat.* **40** 1358–1373. [MR0245139](#) <https://doi.org/10.1214/aoms/1177697508>
- PLESKEN, W. and SOUVIGNIER, B. (1996). Constructing rational representations of finite groups. *Exp. Math.* **5** 39–47. [MR1412953](#)
- RANCIATI, S., ROVERATO, A. and LUATI, A. (2021). Fused graphical lasso for brain networks with symmetries. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **70** 1299–1322. [MR4347714](#) <https://doi.org/10.1111/rssc.12514>
- ROVERATO, A. (2017). *Graphical Models for Categorical Data*. *SemStat Elements*. Cambridge Univ. Press, Cambridge. [MR3751385](#) <https://doi.org/10.1017/9781108277495>
- SERRE, J.-P. (1977). *Linear Representations of Finite Groups*. *Graduate Texts in Mathematics* **42**. Springer, New York. [MR0450380](#)
- SHAH, P. and CHANDRASEKARAN, V. (2012). Group symmetry and covariance regularization. *Electron. J. Stat.* **6** 1600–1640. [MR2988459](#) <https://doi.org/10.1214/12-EJS723>
- SIEMONS, J. (1982). On partitions and permutation groups on unordered sets. *Arch. Math. (Basel)* **38** 391–403. [MR0666910](#) <https://doi.org/10.1007/BF01304806>

- SIEMONS, J. (1983). Automorphism groups of graphs. *Arch. Math. (Basel)* **41** 379–384. [MR0731610](#)
<https://doi.org/10.1007/BF01371410>
- SOBCZYK, P., WILCZYNSKI, S., BOGDAN, M., GRACZYK, P., JOSSE, J., PANLOUP, F., SEEGER, V. and STANIAK, M. (2020). VARCLUST: Clustering variables using dimensionality reduction. Preprint. Available at [arXiv:2011.06501](#).
- SOLOVEYCHIK, I., TRUSHIN, D. and WIESEL, A. (2016). Group symmetric robust covariance estimation. *IEEE Trans. Signal Process.* **64** 244–257. [MR3432975](#) <https://doi.org/10.1109/TSP.2015.2486739>
- TOYODA, K., YOSHIMURA, S., NAKAI, M., KOGA, M., SASAHARA, Y., SONODA, K., KAMIYAMA, K., YAZAWA, Y., KAWADA, S. et al. (2022). Twenty-year change in severity and outcome of ischemic and hemorrhagic strokes. *JAMA Neurol.* **79** 61–69. <https://doi.org/10.1001/jamaneurol.2021.4346>
- UPMEIER, H. (1986). Jordan algebras and harmonic analysis on symmetric spaces. *Amer. J. Math.* **108** 1–25. [MR0821311](#) <https://doi.org/10.2307/2374466>
- WHITTAKER, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. Wiley, Chichester. [MR1112133](#)
- WIELANDT, H. (1969). *Permutation Groups Through Invariant Relations and Invariant Functions*. Lect. Notes Dept. Math. Ohio St. Univ. Columbus.

OPTIMAL FULL RANKING FROM PAIRWISE COMPARISONS

BY PINHAN CHEN^{1,a}, CHAO GAO^{1,b} AND ANDERSON Y. ZHANG^{2,c}

¹Department of Statistics, University of Chicago, ^apinhanchen@uchicago.edu, ^bchaogao@uchicago.edu

²Department of Statistics, University of Pennsylvania, ^cayz@wharton.upenn.edu

We consider the problem of ranking n players from partial pairwise comparison data under the Bradley–Terry–Luce model. For the first time in the literature, the minimax rate of this ranking problem is derived with respect to the Kendall’s tau distance that measures the difference between two rank vectors by counting the number of inversions. The minimax rate of ranking exhibits a transition between an exponential rate and a polynomial rate depending on the magnitude of the signal-to-noise ratio of the problem. To the best of our knowledge, this phenomenon is unique to full ranking and has not been seen in any other statistical estimation problem. To achieve the minimax rate, we propose a divide-and-conquer ranking algorithm that first divides the n players into groups of similar skills and then computes local MLE within each group. The optimality of the proposed algorithm is established by a careful approximate independence argument between the two steps.

REFERENCES

- [1] BALTRUNAS, L., MAKCINSKAS, T. and RICCI, F. (2010). Group recommendations with rank aggregation and collaborative filtering. In *Proceedings of the Fourth ACM Conference on Recommender Systems* 119–126.
- [2] BEAUDOIN, D. and SWARTZ, T. (2018). A computationally intensive ranking system for paired comparison data. *Oper. Res. Perspect.* **5** 105–112. [MR3907614 https://doi.org/10.1016/j.orp.2018.03.002](https://doi.org/10.1016/j.orp.2018.03.002)
- [3] BOUMAL, N. (2013). On intrinsic Cramér–Rao bounds for Riemannian submanifolds and quotient manifolds. *IEEE Trans. Signal Process.* **61** 1809–1821. [MR3038392 https://doi.org/10.1109/TSP.2013.2242068](https://doi.org/10.1109/TSP.2013.2242068)
- [4] BRADLEY, R. A. and TERRY, M. E. (1952). Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika* **39** 324–345. [MR0070925 https://doi.org/10.2307/2334029](https://doi.org/10.2307/2334029)
- [5] BRAVERMAN, M. and MOSSEL, E. (2008). Noisy sorting without resampling. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms* 268–276. ACM, New York. [MR2485312](https://doi.org/10.1137/072485312)
- [6] BRAVERMAN, M. and MOSSEL, E. (2009). Sorting from noisy information. arXiv preprint. Available at [arXiv:0910.1191](https://arxiv.org/abs/0910.1191).
- [7] CAO, D., HE, X., MIAO, L., AN, Y., YANG, C. and HONG, R. (2018). Attentive group recommendation. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval* 645–654.
- [8] CAO, Z., QIN, T., LIU, T.-Y., TSAI, M.-F. and LI, H. (2007). Learning to rank: From pairwise approach to listwise approach. In *Proceedings of the 24th International Conference on Machine Learning* 129–136.
- [9] CHEN, P., GAO, C. and ZHANG, A. Y. (2022). Partial recovery for top- k ranking: Optimality of MLE and sub-optimality of spectral method. *Ann. Statist.* **50** 1618–1652.
- [10] CHEN, P., GAO, C. and ZHANG, A. Y. (2022). Supplement to “Optimal full ranking from pairwise comparisons.” <https://doi.org/10.1214/22-AOS2175SUPP>
- [11] CHEN, X., GOPI, S., MAO, J. and SCHNEIDER, J. (2017). Competitive analysis of the top- K ranking problem. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms* 1245–1264. SIAM, Philadelphia, PA. [MR3627810 https://doi.org/10.1137/1.9781611974782.81](https://doi.org/10.1137/1.9781611974782.81)
- [12] CHEN, Y., FAN, J., MA, C. and WANG, K. (2019). Spectral method and regularized MLE are both optimal for top- K ranking. *Ann. Statist.* **47** 2204–2235. [MR3953449 https://doi.org/10.1214/18-AOS1745](https://doi.org/10.1214/18-AOS1745)
- [13] CHEN, Y. and SUH, C. (2015). Spectral mle: Top- k rank aggregation from pairwise comparisons. In *International Conference on Machine Learning* 371–380.

- [14] CHOO, E. U. and WEDLEY, W. C. (2004). A common framework for deriving preference values from pairwise comparison matrices. *Comput. Oper. Res.* **31** 893–908.
- [15] COLLIER, O. and DALALYAN, A. (2013). Permutation estimation and minimax matching thresholds.
- [16] COLLIER, O. and DALALYAN, A. S. (2016). Minimax rates in permutation estimation for feature matching. *J. Mach. Learn. Res.* **17** Paper No. 6, 31. [MR3482926](#)
- [17] COSSOCK, D. and ZHANG, T. (2006). Subset ranking using regression. In *Learning Theory. Lecture Notes in Computer Science* **4005** 605–619. Springer, Berlin. [MR2280634](#) https://doi.org/10.1007/11776420_44
- [18] CSATÓ, L. (2013). Ranking by pairwise comparisons for Swiss-system tournaments. *CEJOR Cent. Eur. J. Oper. Res.* **21** 783–803. [MR3103279](#) <https://doi.org/10.1007/s10100-012-0261-8>
- [19] DIACONIS, P. and GRAHAM, R. L. (1977). Spearman’s footrule as a measure of disarray. *J. Roy. Statist. Soc. Ser. B* **39** 262–268. [MR0652736](#)
- [20] DWORK, C., KUMAR, R., NAOR, M. and SIVAKUMAR, D. (2001). Rank aggregation methods for the web. In *Proceedings of the 10th International Conference on World Wide Web* 613–622.
- [21] ERDŐS, P. and RÉNYI, A. (1960). On the evolution of random graphs. *Magy. Tud. Akad. Mat. Kut. Intéz. Közl.* **5** 17–61. [MR0125031](#)
- [22] GAO, C. (2017). Phase transitions in approximate ranking. arXiv preprint. Available at [arXiv:1711.11189](#).
- [23] GAO, C. and ZHANG, A. Y. (2019). Iterative algorithm for discrete structure recovery. arXiv preprint. Available at [arXiv:1911.01018](#).
- [24] HERBRICH, R., MINKA, T. and GRAEPEL, T. (2007). TrueSkill: A Bayesian skill rating system. In *Advances in Neural Information Processing Systems* 569–576.
- [25] HUNTER, D. R. (2004). MM algorithms for generalized Bradley–Terry models. *Ann. Statist.* **32** 384–406. [MR2051012](#) <https://doi.org/10.1214/aos/1079120141>
- [26] JADBABAIE, A., MAKUR, A. and SHAH, D. (2020). Estimation of skill distributions. arXiv preprint. Available at [arXiv:2006.08189](#).
- [27] JANG, M., KIM, S., SUH, C. and OH, S. (2016). Top- k ranking from pairwise comparisons: When spectral ranking is optimal. arXiv preprint. Available at [arXiv:1603.04153](#).
- [28] JANG, M., KIM, S., SUH, C. and OH, S. (2017). Optimal sample complexity of m -wise data for top- k ranking. In *Advances in Neural Information Processing Systems* 1686–1696.
- [29] JONES, M. C., MARRON, J. S. and SHEATHER, S. J. (1996). A brief survey of bandwidth selection for density estimation. *J. Amer. Statist. Assoc.* **91** 401–407. [MR1394097](#) <https://doi.org/10.2307/2291420>
- [30] KATAJAINEN, J. and TRÄFF, J. L. (1997). A meticulous analysis of mergesort programs. In *Algorithms and Complexity (Rome, 1997). Lecture Notes in Computer Science* **1203** 217–228. Springer, Berlin. [MR1488977](#) https://doi.org/10.1007/3-540-62592-5_74
- [31] KNUTH, D. E. (1997). *The Art of Computer Programming. Vol. 1: Fundamental Algorithms*, 3rd ed. Addison-Wesley, Reading, MA. [MR3077152](#)
- [32] LIU, T.-Y. (2011). *Learning to Rank for Information Retrieval*. Springer, Berlin.
- [33] LOUVIERE, J. J., HENSHER, D. A. and SWAIT, J. D. (2000). *Stated Choice Methods: Analysis and Applications*. Cambridge Univ. Press, Cambridge. [MR1881927](#) <https://doi.org/10.1017/CBO9780511753831>
- [34] LUCE, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. Wiley, New York. [MR0108411](#)
- [35] LUCE, R. D. (1977). The choice axiom after twenty years. *J. Math. Psych.* **15** 215–233. [MR0462675](#) [https://doi.org/10.1016/0022-2496\(77\)90032-3](https://doi.org/10.1016/0022-2496(77)90032-3)
- [36] MANSKI, C. F. (1977). The structure of random utility models. *Theory and Decision* **8** 229–254. [MR0459524](#) <https://doi.org/10.1007/BF00133443>
- [37] MAO, C., WEED, J. and RIGOLLET, P. (2018). Minimax rates and efficient algorithms for noisy sorting. In *Algorithmic Learning Theory. Proc. Mach. Learn. Res. (PMLR)* **83** 27. [MR3857331](#)
- [38] MCFADDEN, D. (1973). Conditional logit analysis of qualitative choice behavior.
- [39] MCFADDEN, D. and TRAIN, K. (2000). Mixed MNL models for discrete response. *J. Appl. Econometrics* **15** 447–470.
- [40] MINKA, T., CLEVEN, R. and ZAYKOV, Y. (2018). Trueskill 2: An improved Bayesian skill rating system.
- [41] MOTEGI, S. and MASUDA, N. (2012). A network-based dynamical ranking system for competitive sports. *Sci. Rep.* **2** 904. <https://doi.org/10.1038/srep00904>
- [42] NEGAHBAN, S., OH, S. and SHAH, D. (2017). Rank centrality: Ranking from pairwise comparisons. *Oper. Res.* **65** 266–287. [MR3613103](#) <https://doi.org/10.1287/opre.2016.1534>
- [43] PANANJADY, A., MAO, C., MUTHUKUMAR, V., WAINWRIGHT, M. J. and COURTADE, T. A. (2020). Worst-case versus average-case design for estimation from partial pairwise comparisons. *Ann. Statist.* **48** 1072–1097. [MR4102688](#) <https://doi.org/10.1214/19-AOS1838>
- [44] PANANJADY, A., WAINWRIGHT, M. J. and COURTADE, T. A. (2016). Linear regression with an unknown permutation: Statistical and computational limits. In *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)* 417–424. IEEE.

- [45] PLACKETT, R. L. (1975). The analysis of permutations. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **24** 193–202. [MR0391338](#) <https://doi.org/10.2307/2346567>
- [46] SAATY, T. L. (1990). *Decision Making for Leaders: The Analytic Hierarchy Process for Decisions in a Complex World*. RWS Publications.
- [47] SEDGEWICK, R. (1978). Implementing quicksort programs. *Commun. ACM* **21** 847–857.
- [48] SHA, L., LUCEY, P., YUE, Y., CARR, P., ROHLF, C. and MATTHEWS, I. (2016). Chalkboarding: A new spatiotemporal query paradigm for sports play retrieval. In *Proceedings of the 21st International Conference on Intelligent User Interfaces* 336–347.
- [49] SHAH, N., BALAKRISHNAN, S., GUNTUBOYINA, A. and WAINWRIGHT, M. (2016). Stochastically transitive models for pairwise comparisons: Statistical and computational issues. In *International Conference on Machine Learning* 11–20.
- [50] SHAH, N. B. and WAINWRIGHT, M. J. (2017). Simple, robust and optimal ranking from pairwise comparisons. *J. Mach. Learn. Res.* **18** Paper No. 199, 38. [MR3827087](#)
- [51] THURSTONE, L. L. (1927). A law of comparative judgment. *Psychol. Rev.* **34** 273.

CONFIDENCE REGIONS NEAR SINGULAR INFORMATION AND BOUNDARY POINTS WITH APPLICATIONS TO MIXED MODELS

BY KARL OSKAR EKVALL^a AND MATTEO BOTTAI^b

*Division of Biostatistics, Institute of Environmental Medicine, Karolinska Institutet, ^akarl.oskar.ekvall@ki.se,
^bmatteo.bottai@ki.se*

We propose confidence regions with asymptotically correct uniform coverage probability of parameters whose Fisher information matrix can be singular at important points of the parameter set. Our work is motivated by the need for reliable inference on scale parameters close or equal to zero in mixed models, which is obtained as a special case. The confidence regions are constructed by inverting a continuous extension of the score test statistic standardized by expected information, which we show exists at points of singular information under regularity conditions. Similar results have previously only been obtained for scalar parameters, under conditions stronger than ours, and applications to mixed models have not been considered. In simulations our confidence regions have near-nominal coverage with as few as $n = 20$ observations, regardless of how close to the boundary the true parameter is. It is a corollary of our main results that the proposed test statistic has an asymptotic chi-square distribution with degrees of freedom equal to the number of tested parameters, even if they are on the boundary of the parameter set.

REFERENCES

- AZZALINI, A. (2014). *The Skew-Normal and Related Families*. Institute of Mathematical Statistics (IMS) Monographs **3**. Cambridge Univ. Press, Cambridge. [MR3468021](#)
- BAEY, C., COURNÈDE, P.-H. and KUHN, E. (2019). Asymptotic distribution of likelihood ratio test statistics for variance components in nonlinear mixed effects models. *Comput. Statist. Data Anal.* **135** 107–122. [MR3915362](#) <https://doi.org/10.1016/j.csda.2019.01.014>
- BAEY, C. and KUHN, E. (2019). varTestnlme: Variance components testing in mixed-effect models.
- BATES, D., MÄCHLER, M., BOLKER, B. and WALKER, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**.
- BHATIA, R. (1997). *Matrix Analysis*. Graduate Texts in Mathematics **169**. Springer, New York. [MR1477662](#) <https://doi.org/10.1007/978-1-4612-0653-8>
- BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York. [MR1623559](#)
- BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd ed. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. [MR0534323](#)
- BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. *Wiley Series in Probability and Statistics: Probability and Statistics*. Wiley, New York. [MR1700749](#) <https://doi.org/10.1002/9780470316962>
- BISCIO, C. A. N., POINAS, A. and WAAGEPETERSEN, R. (2018). A note on gaps in proofs of central limit theorems. *Statist. Probab. Lett.* **135** 7–10. [MR3758254](#) <https://doi.org/10.1016/j.spl.2017.11.009>
- BOTTAI, M. (2003). Confidence regions when the Fisher information is zero. *Biometrika* **90** 73–84. [MR1966551](#) <https://doi.org/10.1093/biomet/90.1.73>
- BOTTAI, M. and ORSINI, N. (2004). Confidence intervals for the variance component of random-effects linear models. *The Stata Journal: Promoting Communications on Statistics and Stata* **4** 429–435.
- CHEN, S. T., XIAO, L. and STAIKU, A.-M. (2019). An approximate restricted likelihood ratio test for variance components in generalized linear mixed models. Available at [arXiv:1906.03320 \[stat\]](#).
- CHESHER, A. (1984). Testing for neglected heterogeneity. *Econometrica* **52** 865–872.
- COX, D. R. and HINKLEY, D. V. (2000). *Theoretical Statistics*. Chapman & Hall/CRC, Boca Raton.

- CRAINICEANU, C. M. and RUPPERT, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 165–185. [MR2035765 https://doi.org/10.1111/j.1467-9868.2004.00438.x](https://doi.org/10.1111/j.1467-9868.2004.00438.x)
- DOCKERY, D. W., BERKEY, C. S., WARE, J. H., SPEIZER, F. E. and FERRIS JR., B. G. (1983). Distribution of forced vital capacity and forced expiratory volume in one second in children 6 to 11 years of age. *Amer. Rev. Respir. Dis.* **128** 405–412.
- DRIKVANDI, R., VERBEKE, G., KHODADADI, A. and PARTOVI NIA, V. (2013). Testing multiple variance components in linear mixed-effects models. *Biostatistics* **14** 144–159.
- EKVALL, K. O. and BOTTAI, M. (2022). Supplement to “Confidence regions near singular information and boundary points with applications to mixed models.” <https://doi.org/10.1214/22-AOS2177SUPP>
- EKVALL, K. O. and JONES, G. L. (2020). Consistent maximum likelihood estimation using subsets with applications to multivariate mixed models. *Ann. Statist.* **48** 932–952. [MR4102682 https://doi.org/10.1214/19-AOS1830](https://doi.org/10.1214/19-AOS1830)
- FEWSTER, R. M. and JUPP, P. E. (2013). Information on parameters of interest decreases under transformations. *J. Multivariate Anal.* **120** 34–39. [MR3072716 https://doi.org/10.1016/j.jmva.2013.05.010](https://doi.org/10.1016/j.jmva.2013.05.010)
- FITZMAURICE, G. M., LAIRD, N. M. and WARE, J. H. (2011). *Applied Longitudinal Analysis*, 2nd ed. Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ. [MR2830137](https://doi.org/10.1002/9781118000000)
- FITZMAURICE, G. M., LIPSITZ, S. R. and IBRAHIM, J. G. (2007). A note on permutation tests for variance components in multilevel generalized linear mixed models. *Biometrics* **63** 942–946. [MR2395815 https://doi.org/10.1111/j.1541-0420.2007.00775.x](https://doi.org/10.1111/j.1541-0420.2007.00775.x)
- FOLLAND, G. B. (2007). *Real Analysis: Modern Techniques and Their Applications*, 2nd ed. Pure and Applied Mathematics (New York). Wiley, New York. [MR1681462](https://doi.org/10.1002/9781118000000)
- FRISTEDT, B. and GRAY, L. (1997). *A Modern Approach to Probability Theory. Probability and Its Applications*. Birkhäuser, Inc., Boston, MA. [MR1422917 https://doi.org/10.1007/978-1-4899-2837-5](https://doi.org/10.1007/978-1-4899-2837-5)
- GEYER, C. J. (1994). On the asymptotics of constrained M -estimation. *Ann. Statist.* **22** 1993–2010. [MR1329179 https://doi.org/10.1214/aos/1176325768](https://doi.org/10.1214/aos/1176325768)
- GIAMPAOLI, V. and SINGER, J. M. (2009). Likelihood ratio tests for variance components in linear mixed models. *J. Statist. Plann. Inference* **139** 1435–1448. [MR2485137 https://doi.org/10.1016/j.jspi.2008.06.016](https://doi.org/10.1016/j.jspi.2008.06.016)
- GREVEN, S., CRAINICEANU, C. M., KÜCHENHOFF, H. and PETERS, A. (2008). Restricted likelihood ratio testing for zero variance components in linear mixed models. *J. Comput. Graph. Statist.* **17** 870–891. [MR2649072 https://doi.org/10.1198/106186008X386599](https://doi.org/10.1198/106186008X386599)
- HALL, D. B. and PRÆSTGAARD, J. T. (2001). Order-restricted score tests for homogeneity in generalised linear and nonlinear mixed models. *Biometrika* **88** 739–751. [MR1859406 https://doi.org/10.1093/biomet/88.3.739](https://doi.org/10.1093/biomet/88.3.739)
- JIANG, J. (2013). The subset argument and consistency of MLE in GLMM: Answer to an open problem and beyond. *Ann. Statist.* **41** 177–195. [MR3059414 https://doi.org/10.1214/13-AOS1084](https://doi.org/10.1214/13-AOS1084)
- KNUDSON, C., BENSON, S., GEYER, C. and JONES, G. (2021). Likelihood-based inference for generalized linear mixed models: Inference with the R package `glmm`. *Stat* **10** Paper No. e339. [MR4218938 https://doi.org/10.1002/sta4.339](https://doi.org/10.1002/sta4.339)
- LEE, L. F. and CHESHER, A. (1986). Specification testing when score test statistics are identically zero. *J. Econometrics* **31** 121–149. [MR0846287 https://doi.org/10.1016/0304-4076\(86\)90045-X](https://doi.org/10.1016/0304-4076(86)90045-X)
- LIN, X. (1997). Variance component testing in generalised linear models with random effects. *Biometrika* **84** 309–326. [MR1467049 https://doi.org/10.1093/biomet/84.2.309](https://doi.org/10.1093/biomet/84.2.309)
- MCCULLOCH, C. E., SEARLE, S. R. and NEUHAUS, J. M. (2008). *Generalized, Linear, and Mixed Models*, 2nd ed. Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ. [MR2431553](https://doi.org/10.1002/9781118000000)
- QU, L., GUENNEL, T. and MARSHALL, S. L. (2013). Linear score tests for variance components in linear mixed models and applications to genetic association studies. *Biometrics* **69** 883–892. [MR3146784 https://doi.org/10.1111/biom.12095](https://doi.org/10.1111/biom.12095)
- ROTNITZKY, A., COX, D. R., BOTTAI, M. and ROBINS, J. (2000). Likelihood-based inference with singular information matrix. *Bernoulli* **6** 243–284. [MR1748721 https://doi.org/10.2307/3318576](https://doi.org/10.2307/3318576)
- SAVILLE, B. R. and HERRING, A. H. (2009). Testing random effects in the linear mixed model using approximate Bayes factors. *Biometrics* **65** 369–376. [MR2751460 https://doi.org/10.1111/j.1541-0420.2008.01107.x](https://doi.org/10.1111/j.1541-0420.2008.01107.x)
- SELF, S. G. and LIANG, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Amer. Statist. Assoc.* **82** 605–610. [MR0898365](https://doi.org/10.1080/01621459.1987.10477355)
- SINHA, S. K. (2009). Bootstrap tests for variance components in generalized linear mixed models. *Canad. J. Statist.* **37** 219–234. [MR2531828 https://doi.org/10.1002/cjs.10012](https://doi.org/10.1002/cjs.10012)
- STERN, S. E. and WELSH, A. H. (2000). Likelihood inference for small variance components. *Canad. J. Statist.* **28** 517–532. [MR1793108 https://doi.org/10.2307/3315962](https://doi.org/10.2307/3315962)
- STRAM, D. O. and LEE, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics* **50** 1171–1177.

- STRAM, D. O. and LEE, J. W. (1995). Corrections: Variance component testing in the longitudinal mixed effects model. *Biometrics* **51** 1196.
- VERBEKE, G. and MOLENBERGHS, G. (2003). The use of score tests for inference on variance components. *Biometrics* **59** 254–262. [MR1987392](#) <https://doi.org/10.1111/1541-0420.00032>
- WIENCIERZ, A., GREVEN, S. and KÜCHENHOFF, H. (2011). Restricted likelihood ratio testing in linear mixed models with general error covariance structure. *Electron. J. Stat.* **5** 1718–1734. [MR2870148](#) <https://doi.org/10.1214/11-EJS654>
- WOOD, S. N. (2013). A simple test for random effects in regression models. *Biometrika* **100** 1005–1010. [MR3142347](#) <https://doi.org/10.1093/biomet/ast038>
- ZHU, H. and ZHANG, H. (2006). Generalized score test of homogeneity for mixed effects models. *Ann. Statist.* **34** 1545–1569. [MR2278367](#) <https://doi.org/10.1214/009053606000000380>

COMPUTATIONAL BARRIERS TO ESTIMATION FROM LOW-DEGREE POLYNOMIALS

BY TSELIL SCHRAMM^{1,a} AND ALEXANDER S. WEIN^{2,b}

¹*Department of Statistics, Stanford University, tselil@stanford.edu*

²*Department of Mathematics, Courant Institute of Mathematical Sciences, New York University, awein@cims.nyu.edu*

One fundamental goal of high-dimensional statistics is to detect or recover planted structure (such as a low-rank matrix) hidden in noisy data. A growing body of work studies low-degree polynomials as a restricted model of computation for such problems: it has been demonstrated in various settings that low-degree polynomials of the data can match the statistical performance of the best known polynomial-time algorithms. Prior work has studied the power of low-degree polynomials for the task of *detecting* the presence of hidden structures. In this work, we extend these methods to address problems of estimation and recovery (instead of detection). For a large class of “signal plus noise” problems, we give a user-friendly lower bound for the best possible mean squared error achievable by any degree- D polynomial. To our knowledge, these are the first results to establish low-degree hardness of recovery problems for which the associated detection problem is easy. As applications, we give a tight characterization of the low-degree minimum mean squared error for the planted submatrix and planted dense subgraph problems, resolving (in the low-degree framework) open problems about the computational complexity of recovery in both cases.

REFERENCES

- [1] ACHLIOPTAS, D. and COJA-OGHLAN, A. (2008). Algorithmic barriers from phase transitions. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science* 793–802. IEEE, New York.
- [2] AKAVIA, A., GOLDREICH, O., GOLDWASSER, S. and MOSHKOVITZ, D. (2006). On basing one-way functions on NP-hardness. In *STOC’06: Proceedings of the 38th Annual ACM Symposium on Theory of Computing* 701–710. ACM, New York. [MR2277195](#) <https://doi.org/10.1145/1132516.1132614>
- [3] AMES, B. P. (2013). Robust convex relaxation for the planted clique and densest k-subgraph problems. ArXiv preprint. Available at [arXiv:1305.4891](https://arxiv.org/abs/1305.4891).
- [4] AMINI, A. A. and WAINWRIGHT, M. J. (2008). High-dimensional analysis of semidefinite relaxations for sparse principal components. In *2008 IEEE International Symposium on Information Theory* 2454–2458. IEEE.
- [5] ARIAS-CASTRO, E., CANDÈS, E. J. and DURAND, A. (2011). Detection of an anomalous cluster in a network. *Ann. Statist.* **39** 278–304. [MR2797847](#) <https://doi.org/10.1214/10-AOS839>
- [6] ARIAS-CASTRO, E. and VERZELEN, N. (2014). Community detection in dense random networks. *Ann. Statist.* **42** 940–969. [MR3210992](#) <https://doi.org/10.1214/14-AOS1208>
- [7] BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. [MR2165575](#) <https://doi.org/10.1214/009117905000000233>
- [8] BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. [MR2279680](#) <https://doi.org/10.1016/j.jmva.2005.08.003>
- [9] BALAKRISHNAN, S., KOLAR, M., RINALDO, A., SINGH, A. and WASSERMAN, L. (2011). Statistical and computational tradeoffs in biclustering. In *NeurIPS 2011 Workshop on Computational Trade-Offs in Statistical Learning* 4.

MSC2020 subject classifications. Primary 62F15; secondary 68Q17.

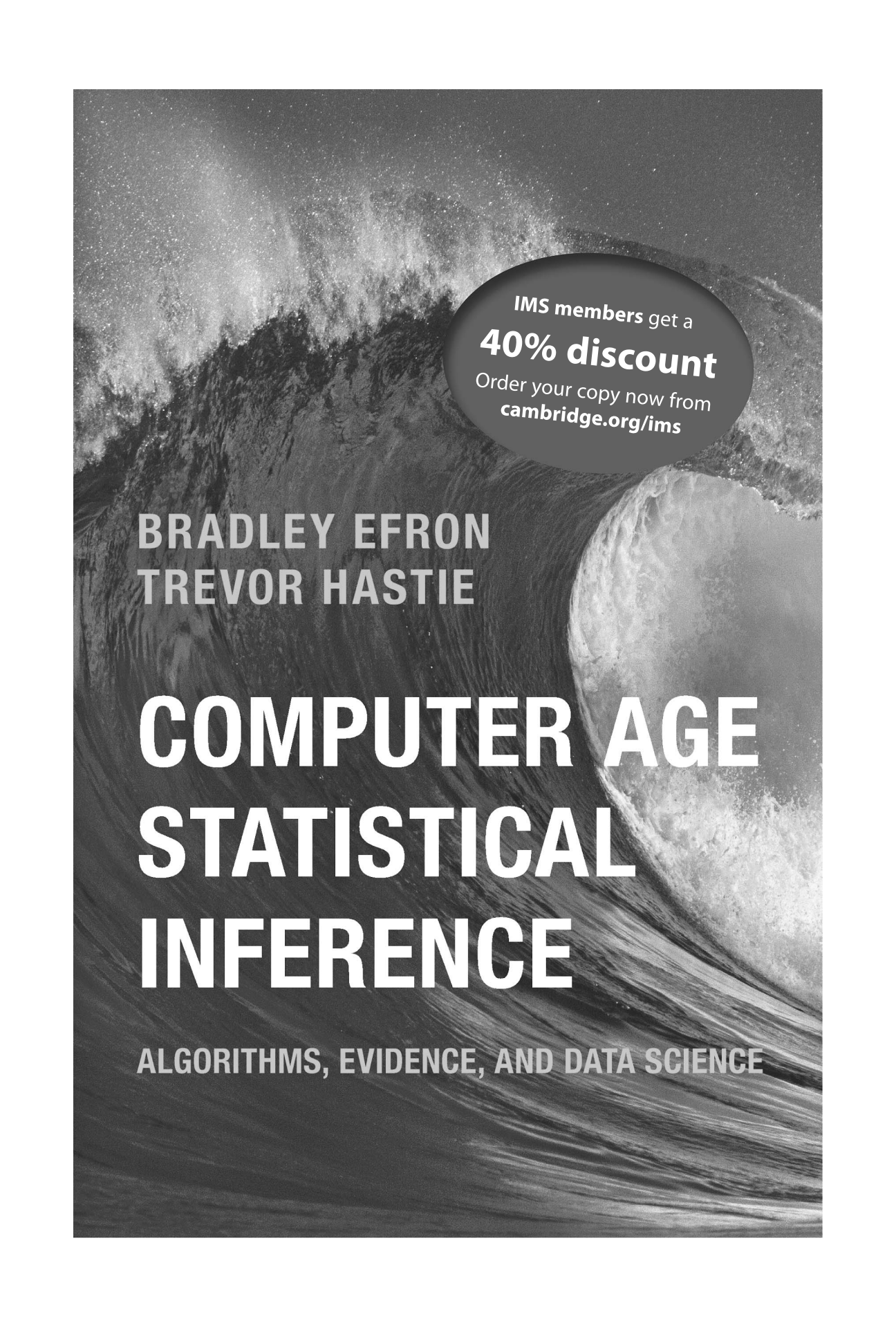
Key words and phrases. Low-degree polynomials, computational complexity, planted submatrix, planted dense subgraph.

- [10] BANDEIRA, A. S., BANKS, J., KUNISKY, D., MOORE, C. and WEIN, A. (2021). Spectral planting and the hardness of refuting cuts, colorability, and communities in random graphs. In *Conference on Learning Theory* 410–473. PMLR.
- [11] BANDEIRA, A. S., KUNISKY, D. and WEIN, A. S. (2020). Computational hardness of certifying bounds on constrained PCA problems. In *11th Innovations in Theoretical Computer Science Conference (ITCS 2020)* Schloss Dagstuhl-Leibniz-Zentrum für Informatik.
- [12] BANKS, J., KLEINBERG, R. and MOORE, C. (2019). The Lovász theta function for random regular graphs and community detection in the hard regime. *SIAM J. Comput.* **48** 1098–1119. [MR3952684](https://doi.org/10.1137/18M1180396) <https://doi.org/10.1137/18M1180396>
- [13] BANKS, J., MOHANTY, S. and RAGHAVENDRA, P. (2019). Local statistics, semidefinite programming, and community detection. ArXiv preprint. Available at [arXiv:1911.01960](https://arxiv.org/abs/1911.01960).
- [14] BARAK, B., CHOU, C.-N., LEI, Z., SCHRAMM, T. and SHENG, Y. (2019). (Nearly) efficient algorithms for the graph matching problem on correlated random graphs. In *Advances in Neural Information Processing Systems* 9190–9198.
- [15] BARAK, B., HOPKINS, S., KELNER, J., KOTHARI, P. K., MOITRA, A. and POTECHIN, A. (2019). A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM J. Comput.* **48** 687–735. [MR3945259](https://doi.org/10.1137/17M1138236) <https://doi.org/10.1137/17M1138236>
- [16] BARBIER, J., MACRIS, N. and RUSH, C. (2020). All-or-nothing statistical and computational phase transitions in sparse spiked matrix estimation. ArXiv preprint. Available at [arXiv:2006.07971](https://arxiv.org/abs/2006.07971).
- [17] BAYATI, M. and MONTANARI, A. (2011). The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inf. Theory* **57** 764–785. [MR2810285](https://doi.org/10.1109/TIT.2010.2094817) <https://doi.org/10.1109/TIT.2010.2094817>
- [18] BEN AROUS, G., GHEISSARI, R. and JAGANNATH, A. (2020). Algorithmic thresholds for tensor PCA. *Ann. Probab.* **48** 2052–2087. [MR4124533](https://doi.org/10.1214/19-AOP1415) <https://doi.org/10.1214/19-AOP1415>
- [19] BEN AROUS, G., WEIN, A. S. and ZADIK, I. (2020). Free energy wells and overlap gap property in sparse PCA. ArXiv preprint. Available at [arXiv:2006.10689](https://arxiv.org/abs/2006.10689).
- [20] BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** 494–521. [MR2782201](https://doi.org/10.1016/j.aim.2011.02.007) <https://doi.org/10.1016/j.aim.2011.02.007>
- [21] BERTHET, Q. and RIGOLLET, P. (2013). Complexity theoretic lower bounds for sparse principal component detection. In *Conference on Learning Theory* 1046–1066.
- [22] BHASKARA, A., CHARIKAR, M., CHLAMTAC, E., FEIGE, U. and VIJAYARAGHAVAN, A. (2010). Detecting high log-densities—An $O(n^{1/4})$ approximation for densest k -subgraph. In *STOC’10—Proceedings of the 2010 ACM International Symposium on Theory of Computing* 201–210. ACM, New York. [MR2743268](https://doi.org/10.1145/174003)
- [23] BIROLI, G., CAMMAROTA, C. and RICCI-TERSENGHI, F. (2020). How to iron out rough landscapes and get optimal performances: Averaged gradient descent and its application to tensor PCA. *J. Phys. A* **53** 174003. [MR4084297](https://doi.org/10.1088/1751-8121/ab7b1f) <https://doi.org/10.1088/1751-8121/ab7b1f>
- [24] BOGDANOV, A. and TREVISAN, L. (2006). On worst-case to average-case reductions for NP problems. *SIAM J. Comput.* **36** 1119–1159. [MR2272273](https://doi.org/10.1137/S0097539705446974) <https://doi.org/10.1137/S0097539705446974>
- [25] BOLTHAUSEN, E. (2014). An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model. *Comm. Math. Phys.* **325** 333–366. [MR3147441](https://doi.org/10.1007/s00220-013-1862-3) <https://doi.org/10.1007/s00220-013-1862-3>
- [26] BRENNAN, M. and BRESLER, G. (2019). Optimal average-case reductions to sparse PCA: From weak assumptions to strong hardness. In *Conference on Learning Theory* 469–470.
- [27] BRENNAN, M. and BRESLER, G. (2020). Reducibility and statistical-computational gaps from secret leakage. ArXiv preprint. Available at [arXiv:2005.08099](https://arxiv.org/abs/2005.08099).
- [28] BRENNAN, M., BRESLER, G., HOPKINS, S. B., LI, J. and SCHRAMM, T. (2021). Statistical query algorithms and low-degree tests are almost equivalent. In *Conference on Learning Theory*.
- [29] BRENNAN, M., BRESLER, G. and HULEIHEL, W. (2018). Reducibility and computational lower bounds for problems with planted sparse structure. In *Conference on Learning Theory* 48–166.
- [30] BRESLER, G. and HUANG, B. (2021). The algorithmic phase transition of random k -SAT for low degree polynomials. ArXiv preprint. Available at [arXiv:2106.02129](https://arxiv.org/abs/2106.02129).
- [31] BUTUCEA, C. and INGSTER, Y. I. (2013). Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli* **19** 2652–2688. [MR3160567](https://doi.org/10.3150/12-BEJ470) <https://doi.org/10.3150/12-BEJ470>
- [32] BUTUCEA, C., INGSTER, Y. I. and SUSLINA, I. A. (2015). Sharp variable selection of a sparse submatrix in a high-dimensional noisy matrix. *ESAIM Probab. Stat.* **19** 115–134. [MR3374872](https://doi.org/10.1051/ps/2014017) <https://doi.org/10.1051/ps/2014017>

- [33] 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>
- [34] CAI, T. T. and WU, Y. (2020). Statistical and computational limits for sparse matrix detection. *Ann. Statist.* **48** 1593–1614. MR4124336 <https://doi.org/10.1214/19-AOS1860>
- [35] CAPITAINE, M., DONATI-MARTIN, C. and FÉRAL, D. (2009). The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations. *Ann. Probab.* **37** 1–47. MR2489158 <https://doi.org/10.1214/08-AOP394>
- [36] CHEN, W.-K., GAMARNIK, D., PANCHENKO, D. and RAHMAN, M. (2019). Suboptimality of local algorithms for a class of max-cut problems. *Ann. Probab.* **47** 1587–1618. MR3945754 <https://doi.org/10.1214/18-AOP1291>
- [37] CHEN, Y. (2015). Incoherence-optimal matrix completion. *IEEE Trans. Inf. Theory* **61** 2909–2923. MR3342311 <https://doi.org/10.1109/TIT.2015.2415195>
- [38] 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. MR3491121
- [39] CHLAMTÁČ, E. and MANURANGSI, P. (2018). Sherali-adams integrality gaps matching the log-density threshold. ArXiv preprint. Available at [arXiv:1804.07842](https://arxiv.org/abs/1804.07842).
- [40] COJA-OGHLAN, A., HAQSHENAS, A. and HETTERICH, S. (2017). Walksat stalls well below satisfiability. *SIAM J. Discrete Math.* **31** 1160–1173. MR3656499 <https://doi.org/10.1137/16M1084158>
- [41] DESHPANDE, Y., ABBE, E. and MONTANARI, A. (2017). Asymptotic mutual information for the balanced binary stochastic block model. *Inf. Inference* **6** 125–170. MR3671474 <https://doi.org/10.1093/imaia/iaw017>
- [42] DESHPANDE, Y. and MONTANARI, A. (2014). Information-theoretically optimal sparse PCA. In 2014 *IEEE International Symposium on Information Theory* 2197–2201. IEEE, New York.
- [43] DESHPANDE, Y. and MONTANARI, A. (2015). Finding hidden cliques of size $\sqrt{N/e}$ in nearly linear time. *Found. Comput. Math.* **15** 1069–1128. MR3371378 <https://doi.org/10.1007/s10208-014-9215-y>
- [44] DING, Y., KUNISKY, D., WEIN, A. S. and BANDEIRA, A. S. (2019). Subexponential-time algorithms for sparse PCA. ArXiv preprint. Available at [arXiv:1907.11635](https://arxiv.org/abs/1907.11635).
- [45] 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>
- [46] DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2009). Message-passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci. USA* **106** 18914–18919.
- [47] DUDEJA, R. and HSU, D. (2021). Statistical query lower bounds for tensor PCA. *J. Mach. Learn. Res.* **22** Paper No. 83. MR4253776
- [48] FEIGENBAUM, J. and FORTNOW, L. (1993). Random-self-reducibility of complete sets. *SIAM J. Comput.* **22** 994–1005. MR1237159 <https://doi.org/10.1137/0222061>
- [49] FELDMAN, V., GRIGORESCU, E., REYZIN, L., VEMPALA, S. S. and XIAO, Y. (2017). Statistical algorithms and a lower bound for detecting planted cliques. *J. ACM* **64** Art. 8. MR3664576 <https://doi.org/10.1145/3046674>
- [50] FÉRAL, D. and PÉCHÉ, S. (2007). The largest eigenvalue of rank one deformation of large Wigner matrices. *Comm. Math. Phys.* **272** 185–228. MR2291807 <https://doi.org/10.1007/s00220-007-0209-3>
- [51] GAMARNIK, D. and JAGANNATH, A. (2021). The overlap gap property and approximate message passing algorithms for p -spin models. *Ann. Probab.* **49** 180–205. MR4203336 <https://doi.org/10.1214/20-AOP1448>
- [52] GAMARNIK, D., JAGANNATH, A. and SEN, S. (2021). The overlap gap property in principal submatrix recovery. *Probab. Theory Related Fields* **181** 757–814. MR4344133 <https://doi.org/10.1007/s00440-021-01089-7>
- [53] GAMARNIK, D., JAGANNATH, A. and WEIN, A. S. (2020). Low-degree hardness of random optimization problems. In 2020 *IEEE 61st Annual Symposium on Foundations of Computer Science* 131–140. IEEE Computer Soc., Los Alamitos, CA. MR4232029 <https://doi.org/10.1109/FOCS46700.2020.00021>
- [54] GAMARNIK, D. and SUDAN, M. (2014). Limits of local algorithms over sparse random graphs [extended abstract]. In *ITCS'14—Proceedings of the 2014 Conference on Innovations in Theoretical Computer Science* 369–375. ACM, New York. MR3359490
- [55] GAMARNIK, D. and ZADIK, I. (2017). High dimensional regression with binary coefficients. Estimating squared error and a phase transition. In *Conference on Learning Theory* 948–953.
- [56] GAMARNIK, D. and ZADIK, I. (2019). The landscape of the planted clique problem: Dense subgraphs and the overlap gap property. ArXiv preprint. Available at [arXiv:1904.07174](https://arxiv.org/abs/1904.07174).
- [57] 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>

- [58] GHOSH, M., JERONIMO, F. G., JONES, C., POTECHIN, A. and RAJENDRAN, G. (2020). Sum-of-squares lower bounds for Sherrington–Kirkpatrick via planted affine planes. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science* 954–965. IEEE Computer Soc., Los Alamitos, CA. [MR4232101](#)
- [59] HAJEK, B., WU, Y. and XU, J. (2015). Computational lower bounds for community detection on random graphs. In *Conference on Learning Theory* 899–928.
- [60] HAJEK, B., WU, Y. and XU, J. (2017). Submatrix localization via message passing. *J. Mach. Learn. Res.* **18** Paper No. 186. [MR3827074](#)
- [61] HOLMGREN, J. and WEIN, A. S. (2021). Counterexamples to the low-degree conjecture. In *12th Innovations in Theoretical Computer Science Conference. LIPIcs. Leibniz Int. Proc. Inform.* **185** Art. No. 75. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. [MR4214319](#)
- [62] HOLTZMAN, G., SOFFER, A. and VILENCHIK, D. (2020). A greedy anytime algorithm for sparse PCA. In *Conference on Learning Theory* 1939–1956.
- [63] HOPKINS, S. (2018). Statistical inference and the sum of squares method. Ph.D. thesis, Cornell University.
- [64] HOPKINS, S. B., KOTHARI, P. K., POTECHIN, A., RAGHAVENDRA, P., SCHRAMM, T. and STEURER, D. (2017). The power of sum-of-squares for detecting hidden structures. In *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2017* 720–731. IEEE Computer Soc., Los Alamitos, CA. [MR3734275](#) <https://doi.org/10.1109/FOCS.2017.72>
- [65] HOPKINS, S. B., SCHRAMM, T. and SHI, J. (2019). A robust spectral algorithm for overcomplete tensor decomposition. In *Conference on Learning Theory* 1683–1722.
- [66] HOPKINS, S. B., SHI, J., SCHRAMM, T. and STEURER, D. (2016). Fast spectral algorithms from sum-of-squares proofs: Tensor decomposition and planted sparse vectors. In *STOC’16—Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing* 178–191. ACM, New York. [MR3536564](#) <https://doi.org/10.1145/2897518.2897529>
- [67] HOPKINS, S. B., SHI, J. and STEURER, D. (2015). Tensor principal component analysis via sum-of-squares proofs. In *Conference on Learning Theory* 956–1006.
- [68] HOPKINS, S. B. and STEURER, D. (2017). Efficient Bayesian estimation from few samples: Community detection and related problems. In *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2017* 379–390. IEEE Computer Soc., Los Alamitos, CA. [MR3734245](#) <https://doi.org/10.1109/FOCS.2017.42>
- [69] JAVANMARD, A. and MONTANARI, A. (2013). State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Inf. Inference* **2** 115–144. [MR3311445](#) <https://doi.org/10.1093/imaiai/iat004>
- [70] JOHNSTONE, I. M. and LU, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* **104** 682–693. [MR2751448](#) <https://doi.org/10.1198/jasa.2009.0121>
- [71] KEARNS, M. (1998). Efficient noise-tolerant learning from statistical queries. *J. ACM* **45** 983–1006. [MR1678849](#) <https://doi.org/10.1145/293347.293351>
- [72] KOLAR, M., BALAKRISHNAN, S., RINALDO, A. and SINGH, A. (2011). Minimax localization of structural information in large noisy matrices. In *Advances in Neural Information Processing Systems* 909–917.
- [73] KOTHARI, P. K., MORI, R., O’DONNELL, R. and WITMER, D. (2017). Sum of squares lower bounds for refuting any CSP. In *STOC’17—Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing* 132–145. ACM, New York. [MR3678177](#) <https://doi.org/10.1145/3055399.3055485>
- [74] KUNISKY, D., WEIN, A. S. and BANDEIRA, A. S. (2019). Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. ArXiv preprint. Available at [arXiv:1907.11636](#).
- [75] LESIEUR, T., KRZAKALA, F. and ZDEBOROVÁ, L. (2015). Phase transitions in sparse PCA. In *2015 IEEE International Symposium on Information Theory (ISIT)* 1635–1639. IEEE.
- [76] LESIEUR, T., KRZAKALA, F. and ZDEBOROVÁ, L. (2015). MMSE of probabilistic low-rank matrix estimation: Universality with respect to the output channel. In *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)* 680–687. IEEE.
- [77] LÖFFLER, M., WEIN, A. S. and BANDEIRA, A. S. (2020). Computationally efficient sparse clustering. ArXiv preprint. Available at [arXiv:2005.10817](#).
- [78] 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>
- [79] MAGNUS, W., OBERHETTINGER, F. and SONI, R. P. (2013). *Formulas and Theorems for the Special Functions of Mathematical Physics, Vol. 52*. Springer Science & Business Media.
- [80] MASSOULIÉ, L., STEPHAN, L. and TOWSLEY, D. (2019). Planting trees in graphs, and finding them back. In *Conference on Learning Theory* 2341–2371. PMLR.

- [81] MOHANTY, S., RAGHAVENDRA, P. and XU, J. (2020). Lifting sum-of-squares lower bounds: Degree-2 to degree-4. In *STOC '20—Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing* 840–853. ACM, New York. [MR4141804](#) <https://doi.org/10.1145/3357713.3384319>
- [82] MONTANARI, A. (2019). Optimization of the Sherrington–Kirkpatrick Hamiltonian. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science* 1417–1433. IEEE Comput. Soc. Press, Los Alamitos, CA. [MR4228234](#) <https://doi.org/10.1109/FOCS.2019.00087>
- [83] MONTANARI, A. and VENKATARAMANAN, R. (2021). Estimation of low-rank matrices via approximate message passing. *Ann. Statist.* **49** 321–345. [MR4206680](#) <https://doi.org/10.1214/20-AOS1958>
- [84] NISAN, N. and SZEGEDY, M. (1994). On the degree of Boolean functions as real polynomials. *Comput. Complexity* **4** 301–313.
- [85] NOVAK, J. (2014). Three lectures on free probability. *Random matrix theory, interacting particle systems, and integrable systems* **65** 13.
- [86] PATURI, R. (1992). On the degree of polynomials that approximate symmetric Boolean functions (preliminary version). In *Proceedings of the Twenty-Fourth Annual ACM Symposium on Theory of Computing* 468–474.
- [87] PERRY, A., WEIN, A. S., BANDEIRA, A. S. and MOITRA, A. (2018). Optimality and sub-optimality of PCA I: Spiked random matrix models. *Ann. Statist.* **46** 2416–2451. [MR3845022](#) <https://doi.org/10.1214/17-AOS1625>
- [88] RAGHAVENDRA, P., SCHRAMM, T. and STEURER, D. (2018). High-dimensional estimation via sum-of-squares proofs. ArXiv preprint. Available at [arXiv:1807.11419](#).
- [89] RAHMAN, M. and VIRÁG, B. (2017). Local algorithms for independent sets are half-optimal. *Ann. Probab.* **45** 1543–1577. [MR3650409](#) <https://doi.org/10.1214/16-AOP1094>
- [90] RANGAN, S. and FLETCHER, A. K. (2012). Iterative estimation of constrained rank-one matrices in noise. In *2012 IEEE International Symposium on Information Theory Proceedings* 1246–1250. IEEE.
- [91] RICHARD, E. and MONTANARI, A. (2014). A statistical model for tensor PCA. In *Advances in Neural Information Processing Systems* 2897–2905.
- [92] SCHRAMM, T. and WEIN, A. S. (2022). Supplement to “Computational barriers to estimation from low-degree polynomials.” <https://doi.org/10.1214/22-AOS2179SUPP>
- [93] SHABALIN, A. A., WEIGMAN, V. J., PEROU, C. M. and NOBEL, A. B. (2009). Finding large average submatrices in high dimensional data. *Ann. Appl. Stat.* **3** 985–1012. [MR2750383](#) <https://doi.org/10.1214/09-AOAS239>
- [94] SZEGÖ, G. (1939). *Orthogonal Polynomials*. American Mathematical Society Colloquium Publications, Vol. 23. Amer. Math. Soc., New York. [MR0000077](#)
- [95] VERZELEN, N. and ARIAS-CASTRO, E. (2015). Community detection in sparse random networks. *Ann. Appl. Probab.* **25** 3465–3510. [MR3404642](#) <https://doi.org/10.1214/14-AAP1080>
- [96] WANG, T., BERTHET, Q. and PLAN, Y. (2016). Average-case hardness of RIP certification. In *Advances in Neural Information Processing Systems* 3819–3827.
- [97] 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>
- [98] WEIN, A. S. (2021). Optimal low-degree hardness of maximum independent set. *Math. Stat. Learn.* **4** 221–251. [MR4383734](#)
- [99] WEIN, A. S., EL ALAOUI, A. and MOORE, C. (2019). The Kikuchi hierarchy and tensor PCA. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science* 1446–1468. IEEE Comput. Soc. Press, Los Alamitos, CA. [MR4228236](#)
- [100] ZHANG, A. and XIA, D. (2018). Tensor SVD: Statistical and computational limits. *IEEE Trans. Inf. Theory* **64** 7311–7338. [MR3876445](#) <https://doi.org/10.1109/TIT.2018.2841377>



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