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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

Vol. 52, No. 3—June 2024

THE ANNALS OF STATISTICS Vol. 52, No. 3, pp. 869–1275 June 2024

The Annals of Statistics **[ISSN 0090-5364 (print); ISSN 2168-8966 (online)],** Volume 52, Number 3, June 2024. Published bimonthly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, OH 44094, USA. Periodicals postage paid at Cleveland, Ohio, and at additional mailing offices.

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

The Annals of Statistics
2024, Vol. 52, No. 3, 869–894
https://doi.org/10.1214/24-AOS2367

DIMENSION-FREE MIXING TIMES OF GIBBS SAMPLERS FOR BAYESIAN HIERARCHICAL MODELS

BY FILIPPO ASCOLANI[1,a] AND GIACOMO ZANELLA[2,b]

[1]*Department of Statistical Science, Duke University,* [a]*filippo.ascolani@duke.edu*
[2]*Department of Decision Sciences, Bocconi University,* [b]*giacomo.zanella@unibocconi.it*

Gibbs samplers are popular algorithms to approximate posterior distributions arising from Bayesian hierarchical models. Despite their popularity and good empirical performance, however, there are still relatively few quantitative results on their convergence properties, for example, much less than for gradient-based sampling methods. In this work, we analyse the behaviour of total variation mixing times of Gibbs samplers targeting hierarchical models using tools from Bayesian asymptotics. We obtain dimension-free convergence results under random data-generating assumptions for a broad class of two-level models with generic likelihood function. Specific examples with Gaussian, binomial and categorical likelihoods are discussed.

REFERENCES

[1] AMIT, Y. (1991). On rates of convergence of stochastic relaxation for Gaussian and non-Gaussian distributions. *J. Multivariate Anal.* **38** 82–99. MR1128938 https://doi.org/10.1016/0047-259X(91)90033-X

[2] ANDRIEU, C., LEE, A., POWER, S. and WANG, A. Q. (2022). Explicit convergence bounds for Metropolis Markov chains: Isoperimetry, spectral gaps and profiles. arXiv preprint. Available at arXiv:2211.08959.

[3] ASCOLANI, F. and ZANELLA, G. (2024). Supplement to "Dimension-free mixing times of Gibbs samplers for Bayesian hierarchical models." https://doi.org/10.1214/24-AOS2367SUPP

[4] ATCHADÉ, Y. F. (2021). Approximate spectral gaps for Markov chain mixing times in high dimensions. *SIAM J. Math. Data Sci.* **3** 854–872. MR4303259 https://doi.org/10.1137/19M1283082

[5] BASS, M. R. and SAHU, S. K. (2017). A comparison of centring parameterisations of Gaussian process-based models for Bayesian computation using MCMC. *Stat. Comput.* **27** 1491–1512. MR3687322 https://doi.org/10.1007/s11222-016-9700-z

[6] BELLONI, A. and CHERNOZHUKOV, V. (2009). On the computational complexity of MCMC-based estimators in large samples. *Ann. Statist.* **37** 2011–2055. MR2533478 https://doi.org/10.1214/08-AOS634

[7] BESKOS, A., PILLAI, N., ROBERTS, G., SANZ-SERNA, J.-M. and STUART, A. (2013). Optimal tuning of the hybrid Monte Carlo algorithm. *Bernoulli* **19** 1501–1534. MR3129023 https://doi.org/10.3150/12-BEJ414

[8] BROOKS, S., GELMAN, A., JONES, G. L. and MENG, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. CRC Press, Boca Raton, FL. MR2742422 https://doi.org/10.1201/b10905

[9] CAPRIO, R. and JOHANSEN, A. M. (2023). A calculus for Markov chain Monte Carlo: Studying approximations in algorithms. arXiv preprint. Available at arXiv:2310.03853.

[10] CASELLA, G. and GEORGE, E. I. (1992). Explaining the Gibbs sampler. *Amer. Statist.* **46** 167–174. MR1183069 https://doi.org/10.2307/2685208

[11] CHLEBICKA, I., LATUSZYNSKI, K. and MIASOJEDOW, B. (2023). Solidarity of Gibbs samplers: The spectral gap. arXiv preprint. Available at arXiv:2304.02109.

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

[13] DIEBOLT, J. and ROBERT, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *J. Roy. Statist. Soc. Ser. B* **56** 363–375. MR1281940

[14] DURANTE, D. (2019). Conjugate Bayes for probit regression via unified skew-normal distributions. *Biometrika* **106** 765–779. MR4031198 https://doi.org/10.1093/biomet/asz034

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

[16] DWIVEDI, R., CHEN, Y., WAINWRIGHT, M. J. and YU, B. (2019). Log-concave sampling: Metropolis–Hastings algorithms are fast. *J. Mach. Learn. Res.* **20** Paper No. 183, 42. MR4048994

[17] FLEGAL, J. M., HUGHES, J., VATS, D., GUPTA, K. and MAJI, U. (2021). mcmcse: Monte Carlo standard errors for MCMC. R package.

[18] GELFAND, A. E., KIM, H.-J., SIRMANS, C. F. and BANERJEE, S. (2003). Spatial modeling with spatially varying coefficient processes. *J. Amer. Statist. Assoc.* **98** 387–396. MR1995715 https://doi.org/10.1198/016214503000170

[19] GELFAND, A. E., SAHU, S. K. and CARLIN, B. P. (1995). Efficient parameterisations for normal linear mixed models. *Biometrika* **82** 479–488. MR1366275 https://doi.org/10.1093/biomet/82.3.479

[20] GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. and RUBIN, D. B. (2014). *Bayesian Data Analysis*, 3rd ed. *Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. MR3235677

[21] GELMAN, A. and HILL, J. L. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge Univ. Press, Cambridge.

[22] GILKS, W. R. and WILD, P. (1992). Adaptive rejection sampling for Gibbs sampling. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **41** 337–348.

[23] GONG, L. and FLEGAL, J. M. (2016). A practical sequential stopping rule for high-dimensional Markov chain Monte Carlo. *J. Comput. Graph. Statist.* **25** 684–700. MR3533633 https://doi.org/10.1080/10618600.2015.1044092

[24] GREEN, P. J., ŁATUSZYŃSKI, K., PEREYRA, M. and ROBERT, C. P. (2015). Bayesian computation: A summary of the current state, and samples backwards and forwards. *Stat. Comput.* **25** 835–862. MR3360496 https://doi.org/10.1007/s11222-015-9574-5

[25] HOBERT, J. P. (2011). The data augmentation algorithm: Theory and methodology. In *Handbook of Markov Chain Monte Carlo*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 253–293. CRC Press, Boca Raton, FL. MR2858452

[26] JIN, Z. and HOBERT, J. P. (2022). Dimension free convergence rates for Gibbs samplers for Bayesian linear mixed models. *Stochastic Process. Appl.* **148** 25–67. MR4393342 https://doi.org/10.1016/j.spa.2022.02.003

[27] JOHNDROW, J. E., SMITH, A., PILLAI, N. and DUNSON, D. B. (2019). MCMC for imbalanced categorical data. *J. Amer. Statist. Assoc.* **114** 1394–1403. MR4011787 https://doi.org/10.1080/01621459.2018.1505626

[28] KAMATANI, K. (2014). Local consistency of Markov chain Monte Carlo methods. *Ann. Inst. Statist. Math.* **66** 63–74. MR3147545 https://doi.org/10.1007/s10463-013-0403-3

[29] KHARE, K. and ZHOU, H. (2009). Rates of convergence of some multivariate Markov chains with polynomial eigenfunctions. *Ann. Appl. Probab.* **19** 737–777. MR2521887 https://doi.org/10.1214/08-AAP562

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

[31] LIU, J. S. (1994). Fraction of missing information and convergence rate for data augmentation. In *Computationally Intensive Statistical Methods*: *Proceedings of the* 26*th Symposium Interface*.

[32] LOVÁSZ, L. and SIMONOVITS, M. (1993). Random walks in a convex body and an improved volume algorithm. *Random Structures Algorithms* **4** 359–412. MR1238906 https://doi.org/10.1002/rsa.3240040402

[33] MARTIN, G. M., FRAZIER, D. T. and ROBERT, C. P. (2024). Computing Bayes: From then 'til now. *Statist. Sci.* **39** 3–19. MR4718524 https://doi.org/10.1214/22-sts876

[34] NEAL, R. M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 113–162. CRC Press, Boca Raton, FL. MR2858447

[35] NEGREA, J., YANG, J., FENG, H., ROY, D. M. and HUGGINS, J. H. (2022). Statistical inference with stochastic gradient algorithms. arXiv preprint. Available at arXiv:2207.12395.

[36] NGUYEN, X. (2013). Convergence of latent mixing measures in finite and infinite mixture models. *Ann. Statist.* **41** 370–400. MR3059422 https://doi.org/10.1214/12-AOS1065

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

[38] PAPASPILIOPOULOS, O., ROBERTS, G. O. and SKÖLD, M. (2003). Non-centered parameterizations for hierarchical models and data augmentation. In *Bayesian Statistics*, 7 (*Tenerife*, 2002) (J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West, eds.) 307–326. Oxford Univ. Press, New York. With a discussion by Alan E. Gelfand, Ole F. Christensen and Darren J. Wilkinson, and a reply by the authors. MR2003180

[39] PAPASPILIOPOULOS, O., ROBERTS, G. O. and SKÖLD, M. (2007). A general framework for the parametrization of hierarchical models. *Statist. Sci.* **22** 59–73. MR2408661 https://doi.org/10.1214/088342307000000014

[40] PAPASPILIOPOULOS, O., STUMPF-FÉTIZON, T. and ZANELLA, G. (2023). Scalable computation for Bayesian hierarchical models. arXiv preprint. Available at arXiv:2103.10875.

[41] QIN, Q. and HOBERT, J. P. (2019). Convergence complexity analysis of Albert and Chib's algorithm for Bayesian probit regression. *Ann. Statist.* **47** 2320–2347. MR3953453 https://doi.org/10.1214/18-AOS1749

[42] QIN, Q. and HOBERT, J. P. (2022). Wasserstein-based methods for convergence complexity analysis of MCMC with applications. *Ann. Appl. Probab.* **32** 124–166. MR4386523 https://doi.org/10.1214/21-aap1673

[43] RAJARATNAM, B. and SPARKS, D. (2015). MCMC-based inference in the era of big data: A fundamental analysis of the convergence complexity of high-dimensional chains. arXiv preprint. Available at arXiv:1508.00947.

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

[45] ROBERTS, G. O. and ROSENTHAL, J. S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 255–268. MR1625691 https://doi.org/10.1111/1467-9868.00123

[46] ROBERTS, G. O. and ROSENTHAL, J. S. (2001). Markov chains and de-initializing processes. *Scand. J. Stat.* **28** 489–504. MR1858413 https://doi.org/10.1111/1467-9469.00250

[47] ROBERTS, G. O. and ROSENTHAL, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv.* **1** 20–71. MR2095565 https://doi.org/10.1214/154957804100000024

[48] ROBERTS, G. O. and SAHU, S. K. (1997). Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *J. Roy. Statist. Soc. Ser. B* **59** 291–317. MR1440584 https://doi.org/10.1111/1467-9868.00070

[49] ROBERTS, G. O. and SAHU, S. K. (2001). Approximate predetermined convergence properties of the Gibbs sampler. *J. Comput. Graph. Statist.* **10** 216–229. MR1939698 https://doi.org/10.1198/10618600152627915

[50] ROBERTS, G. O. and SMITH, A. F. M. (1994). Simple conditions for the convergence of the Gibbs sampler and Metropolis–Hastings algorithms. *Stochastic Process. Appl.* **49** 207–216. MR1260190 https://doi.org/10.1016/0304-4149(94)90134-1

[51] ROBERTS, G. O. and TWEEDIE, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** 341–363. MR1440273 https://doi.org/10.2307/3318418

[52] ROSENTHAL, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **90** 558–566. MR1340509

[53] ROSENTHAL, J. S. and ROSENTHAL, P. (2015). Spectral bounds for certain two-factor non-reversible MCMC algorithms. *Electron. Commun. Probab.* **20** no. 91, 10. MR3434208 https://doi.org/10.1214/ECP.v20-4528

[54] TANG, R. and YANG, Y. (2022). Computational complexity of Metropolis-adjusted Langevin algorithms for Bayesian posterior sampling. arXiv preprint. Available at arXiv:2206.06491.

[55] THOMPSON, M. A comparison of methods for computing autocorrelation time. Technical report no. 1007, Department of Statistics, University of Toronto.

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

[57] WU, K., SCHMIDLER, S. and CHEN, Y. (2022). Minimax mixing time of the Metropolis-adjusted Langevin algorithm for log-concave sampling. *J. Mach. Learn. Res.* **23** Paper No. [270], 63. MR4577709

[58] YANG, J. and ROSENTHAL, J. S. (2023). Complexity results for MCMC derived from quantitative bounds. *Ann. Appl. Probab.* **33** 1259–1300. MR4564431 https://doi.org/10.1214/22-aap1846

[59] YANG, Y., WAINWRIGHT, M. J. and JORDAN, M. I. (2016). On the computational complexity of high-dimensional Bayesian variable selection. *Ann. Statist.* **44** 2497–2532. MR3576552 https://doi.org/10.1214/15-AOS1417

[60] ZHOU, Q., YANG, J., VATS, D., ROBERTS, G. O. and ROSENTHAL, J. S. (2022). Dimension-free mixing for high-dimensional Bayesian variable selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 1751–1784. MR4515557 https://doi.org/10.1111/rssb.12546

The Annals of Statistics
2024, Vol. 52, No. 3, 895–921
https://doi.org/10.1214/24-AOS2372

RECONCILING MODEL-X AND DOUBLY ROBUST APPROACHES TO CONDITIONAL INDEPENDENCE TESTING

BY ZIANG NIU[1,a], ABHINAV CHAKRABORTY[1,b], OLIVER DUKES[2,d] AND EUGENE KATSEVICH[1,c]

[1]*Department of Statistics and Data Science, University of Pennsylvania,* [a]*ziangniu@wharton.upenn.edu,* [b]*abch@wharton.upenn.edu,* [c]*ekatsevi@wharton.upenn.edu*

[2]*Department of Applied Mathematics, Computer Science and Statistics, Ghent University,* [d]*oliver.dukes@ugent.be*

Model-X approaches to testing conditional independence between a predictor and an outcome variable given a vector of covariates usually assume exact knowledge of the conditional distribution of the predictor given the covariates. Nevertheless, model-X methodologies are often deployed with this conditional distribution learned in sample. We investigate the consequences of this choice through the lens of the distilled conditional randomization test (dCRT). We find that Type-I error control is still possible, but only if the mean of the outcome variable given the covariates is estimated well enough. This demonstrates that the dCRT is doubly robust, and motivates a comparison to the generalized covariance measure (GCM) test, another doubly robust conditional independence test. We prove that these two tests are asymptotically equivalent, and show that the GCM test is optimal against (generalized) partially linear alternatives by leveraging semiparametric efficiency theory. In an extensive simulation study, we compare the dCRT to the GCM test. These two tests have broadly similar Type-I error and power, though dCRT can have somewhat better Type-I error control but somewhat worse power in small samples or when the response is discrete. We also find that post-lasso based test statistics (as compared to lasso based statistics) can dramatically improve Type-I error control for both methods.

REFERENCES

AUFIERO, M. and JANSON, L. (2022). Surrogate-based global sensitivity analysis with statistical guarantees via floodgate. arXiv.

BARBER, R. F., CANDÈS, E. J. and SAMWORTH, R. J. (2020). Robust inference with knockoffs. *Ann. Statist.* **48** 1409–1431. MR4124328 https://doi.org/10.1214/19-AOS1852

BARBER, R. F. and JANSON, L. (2022). Testing goodness-of-fit and conditional independence with approximate co-sufficient sampling. *Ann. Statist.* **50** 2514–2544. MR4500617 https://doi.org/10.1214/22-aos2187

BARRY, T., WANG, X., MORRIS, J. A., ROEDER, K. and KATSEVICH, E. (2021). SCEPTRE improves calibration and sensitivity in single-cell CRISPR screen analysis. *Genome Biol.* **22** 1–19.

BATES, S., SESIA, M., SABATTI, C. and CANDÈS, E. (2020). Causal inference in genetic trio studies. *Proc. Natl. Acad. Sci. USA* **117** 24117–24126. MR4250261 https://doi.org/10.1073/pnas.2007743117

BAYATI, M. and MONTANARI, A. (2011). The LASSO risk for Gaussian matrices. *IEEE Trans. Inf. Theory* **58** 1997–2017. MR2951312 https://doi.org/10.1109/TIT.2011.2174612

BELLONI, A. and CHERNOZHUKOV, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli* **19** 521–547. MR3037163 https://doi.org/10.3150/11-BEJ410

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

BERRETT, T. B., WANG, Y., BARBER, R. F. and SAMWORTH, R. J. (2020). The conditional permutation test for independence while controlling for confounders. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 175–197. MR4060981

MSC2020 subject classifications. 62J07, 62G10, 62G09.

Key words and phrases. Model-X, conditional randomization test, conditional independence testing, double robustness.

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

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

CANONNE, C. L., DIAKONIKOLAS, I., KANE, D. M. and STEWART, A. (2018). Testing conditional independence of discrete distributions. In *STOC'*18*—Proceedings of the* 50*th Annual ACM SIGACT Symposium on Theory of Computing* 735–748. ACM, New York. MR3826290 https://doi.org/10.1145/3188745.3188756

CELENTANO, M. and MONTANARI, A. (2021). CAD: Debiasing the Lasso with inaccurate covariate model. arXiv.

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

CHERNOZHUKOV, V., ESCANCIANO, J. C., ICHIMURA, H., NEWEY, W. K. and ROBINS, J. M. (2022). Locally robust semiparametric estimation. *Econometrica* **90** 1501–1535. MR4467437 https://doi.org/10.3982/ecta16294

CHOI, S., HALL, W. J. and SCHICK, A. (1996). Asymptotically uniformly most powerful tests in parametric and semiparametric models. *Ann. Statist.* **24** 841–861. MR1394992 https://doi.org/10.1214/aos/1032894469

DONALD, S. G. and NEWEY, W. K. (1994). Series estimation of semilinear models. *J. Multivariate Anal.* **50** 30–40. MR1292606 https://doi.org/10.1006/jmva.1994.1032

DUKES, O. and VANSTEELANDT, S. (2020). How to obtain valid tests and confidence intervals after propensity score variable selection? *Stat. Methods Med. Res.* **29** 677–694. MR4078242 https://doi.org/10.1177/0962280219862005

FAN, Y., DEMIRKAYA, E., LI, G. and LV, J. (2020b). RANK: Large-scale inference with graphical nonlinear knockoffs. *J. Amer. Statist. Assoc.* **115** 362–379. MR4078469 https://doi.org/10.1080/01621459.2018.1546589

FAN, Y., GAO, L. and LV, J. (2023). ARK: Robust knockoffs inference with coupling. arXiv.

FAN, Y., LV, J., SHARIFVAGHEFI, M. and UEMATSU, Y. (2020b). IPAD: Stable interpretable forecasting with knockoffs inference. *J. Amer. Statist. Assoc.* **115** 1822–1834. MR4189760 https://doi.org/10.1080/01621459.2019.1654878

HAM, D. W., IMAI, K. and JANSON, L. (2022). Using machine learning to test causal hypotheses in conjoint analysis. arXiv.

HÄRDLE, W., LIANG, H. and GAO, J. (2000). *Partially Linear Models*. *Contributions to Statistics*. Physica-Verlag, Heidelberg. MR1787637 https://doi.org/10.1007/978-3-642-57700-0

HENMI, M. and EGUCHI, S. (2004). A paradox concerning nuisance parameters and projected estimating functions. *Biometrika* **91** 929–941. MR2126042 https://doi.org/10.1093/biomet/91.4.929

HUANG, D. and JANSON, L. (2020). Relaxing the assumptions of knockoffs by conditioning. *Ann. Statist.* **48** 3021–3042. MR4152633 https://doi.org/10.1214/19-AOS1920

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

JAVANMARD, A. and MONTANARI, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909. MR3277152

KATSEVICH, E. and RAMDAS, A. (2022). On the power of conditional independence testing under model-X. *Electron. J. Stat.* **16** 6348–6394. MR4517344 https://doi.org/10.1214/22-ejs2085

KIM, I., NEYKOV, M., BALAKRISHNAN, S. and WASSERMAN, L. (2022). Local permutation tests for conditional independence. *Ann. Statist.* **50** 3388–3414. MR4524501 https://doi.org/10.1214/22-aos2233

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

LI, S. and LIU, M. (2023). Maxway CRT: Improving the robustness of the model-X inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **85** 1441–1470. MR4718544 https://doi.org/10.1093/jrsssb/qkad081

LI, S., SESIA, M., ROMANO, Y., CANDÈS, E. and SABATTI, C. (2022). Searching for robust associations with a multi-environment knockoff filter. *Biometrika* **109** 611–629. MR4472838 https://doi.org/10.1093/biomet/asab055

LIU, J. and RIGOLLET, P. (2019). Power analysis of knockoff filters for correlated designs. In 33*rd Conference on Neural Information Processing Systems*.

LIU, M., KATSEVICH, E., JANSON, L. and RAMDAS, A. (2022). Fast and powerful conditional randomization testing via distillation. *Biometrika* **109** 277–293. MR4430958 https://doi.org/10.1093/biomet/asab039

LUNDBORG, A. R., KIM, I., SHAH, R. D. and SAMWORTH, R. J. (2022). The projected covariance measure for assumption-lean variable significance testing. arXiv.

NEYKOV, M., BALAKRISHNAN, S. and WASSERMAN, L. (2021). Minimax optimal conditional independence testing. *Ann. Statist.* **49** 2151–2177. MR4319245 https://doi.org/10.1214/20-aos2030

NING, Y. and LIU, H. (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *Ann. Statist.* **45** 158–195. MR3611489 https://doi.org/10.1214/16-AOS1448

NIU, Z., CHAKRABORTY, A., DUKES, O. and KATSEVICH, E. (2024). Supplement to "Reconciling model-X and doubly robust approaches to conditional independence testing." https://doi.org/10.1214/24-AOS2372SUPP

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

ROBINS, J. M., MARK, S. D. and NEWEY, W. K. (1992). Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics* **48** 479–495. MR1173493 https://doi.org/10.2307/2532304

ROBINS, J. M. and ROTNITZKY, A. (2001). Comment on the Bickel and Kwon article, "Inference for semiparametric models: Some questions and an answer". *Statist. Sinica* **11** 920–936. MR1867326

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

ROMANO, Y., SESIA, M. and CANDÈS, E. (2019). Deep knockoffs. *J. Amer. Statist. Assoc.* **115** 1861–1872. MR4189763 https://doi.org/10.1080/01621459.2019.1660174

SESIA, M., BATES, S., CANDÈS, E., MARCHINI, J. and SABATTI, C. (2021). False discovery rate control in genome-wide association studies with population structure. *Proc. Natl. Acad. Sci. USA* **118** 1–12. https://doi.org/10.1073/pnas.2105841118

SESIA, M., KATSEVICH, E., BATES, S., CANDÈS, E. and SABATTI, C. (2020). Multi-resolution localization of causal variants across the genome. *Nat. Commun.* **11** 1093.

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

SESIA, M. and SUN, T. (2022). Individualized conditional independence testing under model-X with heterogeneous samples and interactions. arXiv.

SHAH, R. D. and PETERS, J. (2020). The hardness of conditional independence testing and the generalised covariance measure. *Ann. Statist.* **48** 1514–1538. MR4124333 https://doi.org/10.1214/19-AOS1857

SMUCLER, E., ROTNITZKY, A. and ROBINS, J. M. (2019). A unifying approach for doubly-robust L1 regularized estimation of causal contrasts. arXiv.

SPECTOR, A. and FITHIAN, W. (2022). Asymptotically optimal knockoff statistics via the masked likelihood ratio. arXiv.

VANSTEELANDT, S., VANDERWEELE, T. J., TCHETGEN, E. J. and ROBINS, J. M. (2008). Multiply robust inference for statistical interactions. *J. Amer. Statist. Assoc.* **103** 1693–1704. MR2510295 https://doi.org/10.1198/016214508000001084

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

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

WAINWRIGHT, M. J. (2019). *High-Dimensional Statistics: A Non-asymptotic Viewpoint. Cambridge Series in Statistical and Probabilistic Mathematics* **48**. Cambridge Univ. Press, Cambridge. MR3967104 https://doi.org/10.1017/9781108627771

WANG, W. and JANSON, L. (2022). A high-dimensional power analysis of the conditional randomization test and knockoffs. *Biometrika* **109** 631–645. MR4472839 https://doi.org/10.1093/biomet/asab052

WEINSTEIN, A., BARBER, R. and CANDES, E. (2017). A power analysis for knockoffs under Gaussian designs. arXiv.

WEINSTEIN, A., SU, W. J., BOGDAN, M., FOYGEL BARBER, R. and CANDÈS, E. J. (2023). A power analysis for model-X knockoffs with ℓ_p-regularized statistics. *Ann. Statist.* **51** 1005–1029. MR4630938 https://doi.org/10.1214/23-aos2274

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

ZHONG, Y., KUFFNER, T. and LAHIRI, S. (2021). Conditional randomization rank test. arXiv.

The Annals of Statistics
2024, Vol. 52, No. 3, 922–947
https://doi.org/10.1214/24-AOS2376

DISTRIBUTED ESTIMATION AND INFERENCE FOR SEMIPARAMETRIC BINARY RESPONSE MODELS

BY XI CHEN[1,a], WENBO JING[1,b], WEIDONG LIU[2,c] AND YICHEN ZHANG[3,d]

[1]*Stern School of Business, New York University,* [a]*xc13@stern.nyu.edu,* [b]*wj2093@nyu.edu*
[2]*School of Mathematical Sciences, Shanghai Jiao Tong University,* [c]*weidongl@sjtu.edu.cn*
[3]*Mitchell E. Daniels, Jr. School of Business, Purdue University,* [d]*zhang@purdue.edu*

The development of modern technology has enabled data collection of unprecedented size, which poses new challenges to many statistical estimation and inference problems. This paper studies the maximum score estimator of a semiparametric binary choice model under a distributed computing environment without prespecifying the noise distribution. An intuitive divide-and-conquer estimator is computationally expensive and restricted by a nonregular constraint on the number of machines, due to the highly nonsmooth nature of the objective function.

We propose (1) a one-shot divide-and-conquer estimator after smoothing the objective to relax the constraint, and (2) a multiround estimator to completely remove the constraint via iterative smoothing. We specify an adaptive choice of kernel smoother with a sequentially shrinking bandwidth to achieve the superlinear improvement of the optimization error over multiple iterations. The improved statistical accuracy per iteration is derived, and a quadratic convergence up to the optimal statistical error rate is established. We further provide two generalizations to handle the heterogeneity of data sets and high-dimensional problems where the parameter of interest is sparse.

REFERENCES

BANERJEE, M. and DUROT, C. (2019). Circumventing superefficiency: An effective strategy for distributed computing in non-standard problems. *Electron. J. Stat.* **13** 1926–1977. MR3964267 https://doi.org/10.1214/19-EJS1559

BANERJEE, M., DUROT, C. and SEN, B. (2019). Divide and conquer in nonstandard problems and the superefficiency phenomenon. *Ann. Statist.* **47** 720–757. MR3909948 https://doi.org/10.1214/17-AOS1633

BATTEY, H., FAN, J., LIU, H., LU, J. and ZHU, Z. (2018). Distributed estimation and inference with statistical guarantees. *Ann. Statist.* **46** 1352–1382.

BROCKHOFF, P. M. and MÜLLER, H.-G. (1997). Random effect threshold models for dose–response relationships with repeated measurements. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **59** 431–446.

BROWN, L. D., LOW, M. G. and ZHAO, L. H. (1997). Superefficiency in nonparametric function estimation. *Ann. Statist.* **25** 2607–2625. MR1604424 https://doi.org/10.1214/aos/1030741087

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

CHAMBERLAIN, G. (1986). Asymptotic efficiency in semiparametric models with censoring. *J. Econometrics* **32** 189–218. MR0864926 https://doi.org/10.1016/0304-4076(86)90038-2

CHEN, X., JING, W., LIU, W. and ZHANG, Y. (2024). Supplement to "Distributed estimation and inference for semiparametric binary response models." https://doi.org/10.1214/24-AOS2376SUPP

CHEN, X., LIU, W. and ZHANG, Y. (2019). Quantile regression under memory constraint. *Ann. Statist.* **47** 3244–3273. MR4025741 https://doi.org/10.1214/18-AOS1777

CHEN, X., LIU, W. and ZHANG, Y. (2022). First-order Newton-type estimator for distributed estimation and inference. *J. Amer. Statist. Assoc.* **117** 1858–1874. MR4528476 https://doi.org/10.1080/01621459.2021.1891925

CHEN, X. and XIE, M. (2014). A split-and-conquer approach for analysis of extraordinarily large data. *Statist. Sinica* **24** 1655–1684. MR3308656

MSC2020 subject classifications. Primary 62G20, 62G08; secondary 62E20.

Key words and phrases. Distributed inference, semiparametric inference, binary response model, maximum score estimator, divide and conquer.

DOBRIBAN, E. and SHENG, Y. (2020). WONDER: Weighted one-shot distributed ridge regression in high dimensions. *J. Mach. Learn. Res.* **21** Paper No. 66, 52. MR4095345

DOBRIBAN, E. and SHENG, Y. (2021). Distributed linear regression by averaging. *Ann. Statist.* **49** 918–943. MR4255113 https://doi.org/10.1214/20-aos1984

DUAN, R., NING, Y. and CHEN, Y. (2022). Heterogeneity-aware and communication-efficient distributed statistical inference. *Biometrika* **109** 67–83. MR4374641 https://doi.org/10.1093/biomet/asab007

FAN, J., GUO, Y. and WANG, K. (2023). Communication-efficient accurate statistical estimation. *J. Amer. Statist. Assoc.* **118** 1000–1010. MR4595472 https://doi.org/10.1080/01621459.2021.1969238

FAN, J., WANG, D., WANG, K. and ZHU, Z. (2019). Distributed estimation of principal eigenspaces. *Ann. Statist.* **47** 3009–3031. MR4025733 https://doi.org/10.1214/18-AOS1713

FENG, H., NING, Y. and ZHAO, J. (2022). Nonregular and minimax estimation of individualized thresholds in high dimension with binary responses. *Ann. Statist.* **50** 2284–2305. MR4474491 https://doi.org/10.1214/22-aos2188

FERNANDES, M., GUERRE, E. and HORTA, E. (2021). Smoothing quantile regressions. *J. Bus. Econom. Statist.* **39** 338–357. MR4187194 https://doi.org/10.1080/07350015.2019.1660177

FLORIOS, K. and SKOURAS, S. (2008). Exact computation of max weighted score estimators. *J. Econometrics* **146** 86–91. MR2459645 https://doi.org/10.1016/j.jeconom.2008.05.018

GAO, Y., LIU, W., WANG, H., WANG, X., YAN, Y. and ZHANG, R. (2022). A review of distributed statistical inference. *Stat. Theory Relat. Fields* **6** 89–99. MR4440050 https://doi.org/10.1080/24754269.2021.1974158

GREENE, W. (2009). Discrete choice modeling. In *Palgrave Handbook of Econometrics* 473–556. Springer, Berlin.

HÄRDLE, W., MÜLLER, M., SPERLICH, S. and WERWATZ, A. (2004). *Nonparametric and Semiparametric Models. Springer Series in Statistics*. Springer, New York. MR2061786 https://doi.org/10.1007/978-3-642-17146-8

HE, X., PAN, X., TAN, K. M. and ZHOU, W.-X. (2022). Scalable estimation and inference for censored quantile regression process. *Ann. Statist.* **50** 2899–2924. MR4500629 https://doi.org/10.1214/22-aos2214

HE, X., PAN, X., TAN, K. M. and ZHOU, W.-X. (2023). Smoothed quantile regression with large-scale inference. *J. Econometrics* **232** 367–388. MR4539491 https://doi.org/10.1016/j.jeconom.2021.07.010

HOROWITZ, J. L. (1992). A smoothed maximum score estimator for the binary response model. *Econometrica* **60** 505–531. MR1162997 https://doi.org/10.2307/2951582

HOROWITZ, J. L. (1993). Semiparametric and nonparametric estimation of quantal response models. In *Handbook of Statistics*, *Vol.* 11 45–72. North-Holland, Amsterdam. MR1247239 https://doi.org/10.1016/S0169-7161(05)80037-9

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

HUANG, C. and HUO, X. (2019). A distributed one-step estimator. *Math. Program.* **174** 41–76. MR3935072 https://doi.org/10.1007/s10107-019-01369-0

JORDAN, M. I., LEE, J. D. and YANG, Y. (2019). Communication-efficient distributed statistical inference. *J. Amer. Statist. Assoc.* **114** 668–681. MR3963171 https://doi.org/10.1080/01621459.2018.1429274

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

LEE, J. D., LIU, Q., SUN, Y. and TAYLOR, J. E. (2017). Communication-efficient sparse regression. *J. Mach. Learn. Res.* **18** Paper No. 5, 30. MR3625709

LI, R., LIN, D. K. J. and LI, B. (2013). Statistical inference in massive data sets. *Appl. Stoch. Models Bus. Ind.* **29** 399–409. MR3117826 https://doi.org/10.1002/asmb.1927

LUO, J., SUN, Q. and ZHOU, W.-X. (2022). Distributed adaptive Huber regression. *Comput. Statist. Data Anal.* **169** Paper No. 107419, 23. MR4366282 https://doi.org/10.1016/j.csda.2021.107419

MANSKI, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *J. Econometrics* **3** 205–228. MR0436905 https://doi.org/10.1016/0304-4076(75)90032-9

MANSKI, C. F. (1985). Semiparametric analysis of discrete response. Asymptotic properties of the maximum score estimator. *J. Econometrics* **27** 313–333. MR0788628 https://doi.org/10.1016/0304-4076(85)90009-0

MUKHERJEE, D., BANERJEE, M., MUKHERJEE, D. and RITOV, Y. (2023). Asymptotic normality of a change plane estimator in fixed dimension with near-optimal rate. *Electron. J. Stat.* **17** 2289–2316. MR4649982 https://doi.org/10.1214/23-ejs2144

MUKHERJEE, D., BANERJEE, M. and RITOV, Y. (2019). Non-standard asymptotics in high dimensions: Manski's maximum score estimator revisited. arXiv preprint. Available at arXiv:1903.10063.

ŞENTÜRK, D. and MÜLLER, H.-G. (2009). Covariate-adjusted generalized linear models. *Biometrika* **96** 357–370. MR2507148 https://doi.org/10.1093/biomet/asp012

SHAMIR, O., SREBRO, N. and ZHANG, T. (2014). Communication-efficient distributed optimization using an approximate Newton-type method. In *International Conference on Machine Learning*.

SHI, C., LU, W. and SONG, R. (2018). A massive data framework for M-estimators with cubic-rate. *J. Amer. Statist. Assoc.* **113** 1698–1709. MR3902239 https://doi.org/10.1080/01621459.2017.1360779

TAN, K. M., BATTEY, H. and ZHOU, W.-X. (2022). Communication-constrained distributed quantile regression with optimal statistical guarantees. *J. Mach. Learn. Res.* **23** Paper No. [272], 61. MR4577711

TAN, K. M., WANG, L. and ZHOU, W.-X. (2022). High-dimensional quantile regression: Convolution smoothing and concave regularization. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 205–233. MR4400395 https://doi.org/10.1111/rssb.12485

TU, J., LIU, W., MAO, X. and CHEN, X. (2021). Variance reduced median-of-means estimator for Byzantine-robust distributed inference. *J. Mach. Learn. Res.* **22** Paper No. 84, 67. MR4253777

VOLGUSHEV, S., CHAO, S.-K. and CHENG, G. (2019). Distributed inference for quantile regression processes. *Ann. Statist.* **47** 1634–1662. MR3911125 https://doi.org/10.1214/18-AOS1730

WANG, J., KOLAR, M., SREBRO, N. and ZHANG, T. (2017). Efficient distributed learning with sparsity. In *International Conference on Machine Learning*.

WANG, X., YANG, Z., CHEN, X. and LIU, W. (2019). Distributed inference for linear support vector machine. *J. Mach. Learn. Res.* **20** Paper No. 113, 41. MR3990467

WANG, Y. and ZHU, Z. (2022). ReBoot: Distributed statistical learning via refitting Bootstrap samples. arXiv preprint. Available at arXiv:2207.09098.

WHITE, H. (1981). Consequences and detection of misspecified nonlinear regression models. *J. Amer. Statist. Assoc.* **76** 419–433. MR0624344

WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50** 1–25. MR0640163 https://doi.org/10.2307/1912526

XU, T., WANG, J. and FANG, Y. (2014). A model-free estimation for the covariate-adjusted Youden index and its associated cut-point. *Stat. Med.* **33** 4963–4974. MR3276512 https://doi.org/10.1002/sim.6290

YU, Y., CHAO, S.-K. and CHENG, G. (2020). Simultaneous inference for massive data: Distributed Bootstrap. In *International Conference on Machine Learning*.

YU, Y., CHAO, S.-K. and CHENG, G. (2022). Distributed bootstrap for simultaneous inference under high dimensionality. *J. Mach. Learn. Res.* **23** Paper No. [195], 77. MR4577148 https://doi.org/10.1515/bejte-2021-0025

ZHANG, Y., DUCHI, J. and WAINWRIGHT, M. (2015). Divide and conquer kernel ridge regression: A distributed algorithm with minimax optimal rates. *J. Mach. Learn. Res.* **16** 3299–3340. MR3450540

ZHANG, Y., DUCHI, J. C. and WAINWRIGHT, M. J. (2013). Communication-efficient algorithms for statistical optimization. *J. Mach. Learn. Res.* **14** 3321–3363. MR3144464

ZHAO, T., CHENG, G. and LIU, H. (2016). A partially linear framework for massive heterogeneous data. *Ann. Statist.* **44** 1400–1437. MR3519928 https://doi.org/10.1214/15-AOS1410

ZHOU, H. H., SINGH, V., JOHNSON, S. C., WAHBA, G. and INITIATIVE, A. D. N. (2018). Statistical tests and identifiability conditions for pooling and analyzing multisite datasets. *Proc. Natl. Acad. Sci. USA* **115** 1481–1486. MR3767073 https://doi.org/10.1073/pnas.1719747115

The Annals of Statistics
2024, Vol. 52, No. 3, 948–965
https://doi.org/10.1214/24-AOS2378

ON BLOCKWISE AND REFERENCE PANEL-BASED ESTIMATORS FOR GENETIC DATA PREDICTION IN HIGH DIMENSIONS

BY BINGXIN ZHAO[1,a], SHURONG ZHENG[2,b] AND HONGTU ZHU[3,c]

[1]*Department of Statistics and Data Science, University of Pennsylvania,* [a]*bxzhao@wharton.upenn.edu*
[2]*School of Mathematics and Statistics, Northeast Normal University,* [b]*zhengsr@nenu.edu.cn*
[3]*Department of Biostatistics, University of North Carolina at Chapel Hill,* [c]*htzhu@email.unc.edu*

Genetic prediction holds immense promise for translating genetic discoveries into medical advances. As the high-dimensional covariance matrix (or the linkage disequilibrium (LD) pattern) of genetic variants often presents a block-diagonal structure, numerous methods account for the dependence among variants in predetermined local LD blocks. Moreover, due to privacy considerations and data protection concerns, genetic variant dependence in each LD block is typically estimated from external reference panels rather than the original training data set. This paper presents a unified analysis of blockwise and reference panel-based estimators in a high-dimensional prediction framework without sparsity restrictions. We find that, surprisingly, even when the covariance matrix has a block-diagonal structure with well-defined boundaries, blockwise estimation methods adjusting for local dependence can be substantially less accurate than methods controlling for the whole covariance matrix. Further, estimation methods built on the original training data set and external reference panels are likely to have varying performance in high dimensions, which may reflect the cost of having only access to summary level data from the training data set. This analysis is based on novel results in random matrix theory for block-diagonal covariance matrix. We numerically evaluate our results using extensive simulations and real data analysis in the UK Biobank.

REFERENCES

[1] 1000-GENOMES-CONSORTIUM (2015). A global reference for human genetic variation. *Nature* **526** 68–74.
[2] BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2567175 https://doi.org/10.1007/978-1-4419-0661-8
[3] BAI, Z. and ZHOU, W. (2008). Large sample covariance matrices without independence structures in columns. *Statist. Sinica* **18** 425–442. MR2411613
[4] BAI, Z. D. and SILVERSTEIN, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** 553–605. MR2040792 https://doi.org/10.1214/aop/1078415845
[5] BENNER, C., HAVULINNA, A. S., JÄRVELIN, M.-R., SALOMAA, V., RIPATTI, S. and PIRINEN, M. (2017). Prospects of fine-mapping trait-associated genomic regions by using summary statistics from genome-wide association studies. *Amer. J. Hum. Genet.* **101** 539–551. https://doi.org/10.1016/j.ajhg.2017.08.012
[6] BERISA, T. and PICKRELL, J. K. (2016). Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* **32** 283–285. https://doi.org/10.1093/bioinformatics/btv546
[7] BULIK-SULLIVAN, B. K., LOH, P.-R., FINUCANE, H. K., RIPKE, S., YANG, J., PATTERSON, N., DALY, M. J., PRICE, A. L., NEALE, B. M. and SCHIZOPHRENIA WORKING GROUP OF THE PSYCHIATRIC GENOMICS CONSORTIUM (2015). LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47** 291–295.
[8] BYCROFT, C., FREEMAN, C., PETKOVA, D., BAND, G., ELLIOTT, L., SHARP, K., MOTYER, A., VUKCEVIC, D., DELANEAU, O. et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562** 203–209.

MSC2020 subject classifications. Primary 62J05; secondary 60B20.
Key words and phrases. Block-diagonal covariance matrix, high-dimensional prediction, linkage disequilibrium, random matrix theory, reference panel.

[9] DENG, Y. and PAN, W. (2018). Improved use of small reference panels for conditional and joint analysis with GWAS summary statistics. *Genetics* **209** 401–408. https://doi.org/10.1534/genetics.118.300813

[10] DICKER, L. H. (2011). Dense signals, linear estimators, and out-of-sample prediction for high-dimensional linear models. arXiv preprint. Available at arXiv:1102.2952.

[11] DING, Y., HOU, K., BURCH, K. S., LAPINSKA, S., PRIVÉ, F., VILHJÁLMSSON, B., SANKARARAMAN, S. and PASANIUC, B. (2022). Large uncertainty in individual polygenic risk score estimation impacts PRS-based risk stratification. *Nat. Genet.* **54** 30–39. https://doi.org/10.1038/s41588-021-00961-5

[12] 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

[13] GE, T., CHEN, C.-Y., NI, Y., FENG, Y.-C. A. and SMOLLER, J. W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10** 1–10.

[14] HOU, K., XU, Z., DING, Y., HARPAK, A. and PASANIUC, B. (2023). Calibrated prediction intervals for polygenic scores across diverse contexts. medRxiv 2023-07.

[15] HU, Y., LU, Q., POWLES, R., YAO, X., YANG, C., FANG, F., XU, X. and ZHAO, H. (2017). Leveraging functional annotations in genetic risk prediction for human complex diseases. *PLoS Comput. Biol.* **13** e1005589.

[16] JIANG, J., LI, C., PAUL, D., YANG, C. and ZHAO, H. (2016). On high-dimensional misspecified mixed model analysis in genome-wide association study. *Ann. Statist.* **44** 2127–2160. MR3546446 https://doi.org/10.1214/15-AOS1421

[17] LIU, Y., LI, Z. and LIN, X. (2022). A minimax optimal ridge-type set test for global hypothesis with applications in whole genome sequencing association studies. *J. Amer. Statist. Assoc.* **117** 897–908. MR4436321 https://doi.org/10.1080/01621459.2020.1831926

[18] MAK, T. S. H., PORSCH, R. M., CHOI, S. W., ZHOU, X. and SHAM, P. C. (2017). Polygenic scores via penalized regression on summary statistics. *Genet. Epidemiol.* **41** 469–480. https://doi.org/10.1002/gepi.22050

[19] MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb.* (*N.S.*) **114** 507–536. MR0208649

[20] O'CONNOR, L. J., SCHOECH, A. P., HORMOZDIARI, F., GAZAL, S., PATTERSON, N. and PRICE, A. L. (2019). Extreme polygenicity of complex traits is explained by negative selection. *Amer. J. Hum. Genet.* **105** 456–476. https://doi.org/10.1016/j.ajhg.2019.07.003

[21] PAIN, O., GLANVILLE, K. P., HAGENAARS, S. P., SELZAM, S., FÜRTJES, A. E., GASPAR, H. A., COLEMAN, J. R., RIMFELD, K., BREEN, G. et al. (2021). Evaluation of polygenic prediction methodology within a reference-standardized framework. *PLoS Genet.* **17** e1009021.

[22] PASANIUC, B. and PRICE, A. L. (2017). Dissecting the genetics of complex traits using summary association statistics. *Nat. Rev. Genet.* **18** 117–127. https://doi.org/10.1038/nrg.2016.142

[23] PATTEE, J. and PAN, W. (2020). Penalized regression and model selection methods for polygenic scores on summary statistics. *PLoS Comput. Biol.* **16** e1008271. https://doi.org/10.1371/journal.pcbi.1008271

[24] QIAN, J., TANIGAWA, Y., DU, W., AGUIRRE, M., CHANG, C., TIBSHIRANI, R., RIVAS, M. A. and HASTIE, T. (2020). A fast and scalable framework for large-scale and ultrahigh-dimensional sparse regression with application to the UK Biobank. *PLoS Genet.* **16** e1009141. MR4455904 https://doi.org/10.1214/21-aoas1575

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

[26] SONG, S., JIANG, W., HOU, L. and ZHAO, H. (2020). Leveraging effect size distributions to improve polygenic risk scores derived from summary statistics of genome-wide association studies. *PLoS Comput. Biol.* **16** e1007565.

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

[28] TIMPSON, N. J., GREENWOOD, C. M., SORANZO, N., LAWSON, D. J. and RICHARDS, J. B. (2018). Genetic architecture: The shape of the genetic contribution to human traits and disease. *Nat. Rev. Genet.* **19** 110–125.

[29] TORKAMANI, A., WINEINGER, N. E. and TOPOL, E. J. (2018). The personal and clinical utility of polygenic risk scores. *Nat. Rev. Genet.* **19** 581–590.

[30] VAN HILTEN, A., KUSHNER, S. A., KAYSER, M., ARFAN IKRAM, M., ADAMS, H. H., KLAVER, C. C., NIESSEN, W. J. and ROSHCHUPKIN, G. V. (2021). GenNet framework: Interpretable deep learning for predicting phenotypes from genetic data. *Commun. Biol.* **4** 1–9.

[31] VILHJÁLMSSON, B. J., YANG, J., FINUCANE, H. K., GUSEV, A., LINDSTRÖM, S., RIPKE, S., GENOVESE, G., LOH, P.-R., BHATIA, G. et al. (2015). Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Amer. J. Hum. Genet.* **97** 576–592.

[32] WANG, J., WANG, W. and LI, H. (2022). Sparse block signal detection and identification for shared cross-trait association analysis. *Ann*. *Appl*. *Stat*. **16** 866–886. MR4438815 https://doi.org/10.1214/21-aoas1523

[33] YANG, J., LEE, S. H., GODDARD, M. E. and VISSCHER, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *Amer. J*. *Hum*. *Genet*. **88** 76–82.

[34] YANG, S. and ZHOU, X. (2020). Accurate and scalable construction of polygenic scores in large biobank data sets. *Amer. J*. *Hum*. *Genet*. **106** 679–693.

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

[36] ZHAO, B., ZHENG, S. and ZHU, H. (2024). Supplement to "On blockwise and reference panel-based estimators for genetic data prediction in high dimensions." https://doi.org/10.1214/24-AOS2378SUPP

The Annals of Statistics
2024, Vol. 52, No. 3, 966–998
https://doi.org/10.1214/24-AOS2379

PLUGIN ESTIMATION OF SMOOTH OPTIMAL TRANSPORT MAPS

BY TUDOR MANOLE[1,a], SIVARAMAN BALAKRISHNAN[1,b], JONATHAN NILES-WEED[2,d] AND LARRY WASSERMAN[1,c]

[1]*Department of Statistics and Data Science, Carnegie Mellon University,* [a]*tmanole@andrew.cmu.edu,* [b]*siva@stat.cmu.edu,* [c]*larry@stat.cmu.edu*

[2]*Courant Institute of Mathematical Sciences and Center for Data Science, New York University,* [d]*jnw@cims.nyu.edu*

We analyze a number of natural estimators for the optimal transport map between two distributions and show that they are minimax optimal. We adopt the plugin approach: our estimators are simply optimal couplings between measures derived from our observations, appropriately extended so that they define functions on $\mathbb{R}^d$. When the underlying map is assumed to be Lipschitz, we show that computing the optimal coupling between the empirical measures, and extending it using linear smoothers, already gives a minimax optimal estimator. When the underlying map enjoys higher regularity, we show that the optimal coupling between appropriate nonparametric density estimates yields faster rates. Our work also provides new bounds on the risk of corresponding plugin estimators for the quadratic Wasserstein distance, and we show how this problem relates to that of estimating optimal transport maps using stability arguments for smooth and strongly convex Brenier potentials. As an application of our results, we derive central limit theorems for plugin estimators of the squared Wasserstein distance, which are centered at their population counterpart when the underlying distributions have sufficiently smooth densities. In contrast to known central limit theorems for empirical estimators, this result easily lends itself to statistical inference for the quadratic Wasserstein distance.

REFERENCES

AJTAI, M., KOMLÓS, J. and TUSNÁDY, G. (1984). On optimal matchings. *Combinatorica* **4** 259–264. MR0779885 https://doi.org/10.1007/BF02579135

AMBROSIO, L., COLOMBO, M., DE PHILIPPIS, G. and FIGALLI, A. (2012). Existence of Eulerian solutions to the semigeostrophic equations in physical space: The 2-dimensional periodic case. *Comm. Partial Differential Equations* **37** 2209–2227. MR3005541 https://doi.org/10.1080/03605302.2012.669443

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

BLACK, E., YEOM, S. and FREDRIKSON, M. (2020). Fliptest: Fairness testing via optimal transport. In *Proceedings of the* 2020 *Conference on Fairness, Accountability, and Transparency* 111–121.

BOBKOV, S. and LEDOUX, M. (2019). One-dimensional empirical measures, order statistics, and Kantorovich transport distances. *Mem. Amer. Math. Soc.* **261** v+126. MR4028181 https://doi.org/10.1090/memo/1259

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

BONNOTTE, N. (2013). From Knothe's rearrangement to Brenier's optimal transport map. *SIAM J. Math. Anal.* **45** 64–87. MR3032969 https://doi.org/10.1137/120874850

BRENIER, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.* **44** 375–417. MR1100809 https://doi.org/10.1002/cpa.3160440402

CAFFARELLI, L. A. (1991). Some regularity properties of solutions of Monge Ampère equation. *Comm. Pure Appl. Math.* **44** 965–969. MR1127042 https://doi.org/10.1002/cpa.3160440809

CAFFARELLI, L. A. (1992a). Boundary regularity of maps with convex potentials. *Comm. Pure Appl. Math.* **45** 1141–1151. MR1177479 https://doi.org/10.1002/cpa.3160450905

MSC2020 subject classifications. Primary 62G05, 62G20; secondary 62G07, 62C20.

Key words and phrases. Optimal transport map, Wasserstein distance, Brenier potential, minimax estimation, density estimation, central limit theorem, semiparametric efficiency.

CAFFARELLI, L. A. (1992b). The regularity of mappings with a convex potential. *J. Amer. Math. Soc.* **5** 99–104. MR1124980 https://doi.org/10.2307/2152752

CAFFARELLI, L. A. (1996). Boundary regularity of maps with convex potentials. II. *Ann. of Math.* (2) **144** 453–496. MR1426885 https://doi.org/10.2307/2118564

CAFFARELLI, L. A. (2000). Monotonicity properties of optimal transportation and the FKG and related inequalities. *Comm. Math. Phys.* **214** 547–563. MR1800860 https://doi.org/10.1007/s002200000257

CHERNOZHUKOV, V., GALICHON, A., HALLIN, M. and HENRY, M. (2017). Monge–Kantorovich depth, quantiles, ranks and signs. *Ann. Statist.* **45** 223–256. MR3611491 https://doi.org/10.1214/16-AOS1450

CHIZAT, L., ROUSSILLON, P., LÉGER, F., VIALARD, F.-X. and PEYRÉ, G. (2020). Faster Wasserstein distance estimation with the Sinkhorn divergence. *Adv. Neural Inf. Process. Syst.* **33** 2257–2269.

CLEANTHOUS, G., GEORGIADIS, A. G., KERKYACHARIAN, G., PETRUSHEV, P. and PICARD, D. (2020). Kernel and wavelet density estimators on manifolds and more general metric spaces. *Bernoulli* **26** 1832–1862. MR4091093 https://doi.org/10.3150/19-BEJ1171

COHEN, A., DAUBECHIES, I. and VIAL, P. (1993). Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.* **1** 54–81. MR1256527 https://doi.org/10.1006/acha.1993.1005

CORDERO-ERAUSQUIN, D. (1999). Sur le transport de mesures périodiques. *C. R. Acad. Sci. Paris Sér. I Math.* **329** 199–202. MR1711060 https://doi.org/10.1016/S0764-4442(00)88593-6

COURTY, N., FLAMARY, R., TUIA, D. and RAKOTOMAMONJY, A. (2016). Optimal transport for domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39** 1853–1865. MR3721792 https://doi.org/10.1109/tnnls.2016.2600243

COVER, T. (1968). Estimation by the nearest neighbor rule. *IEEE Trans. Inf. Theory* **14** 50–55.

CUEVAS, A. (2009). Set estimation: Another bridge between statistics and geometry. *Bol. Estad. Investig. Oper.* **25** 71–85. MR2750781

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

CUTURI, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. *Adv. Neural Inf. Process. Syst.* 2292–2300.

DE LARA, L., GONZÁLEZ-SANZ, A. and LOUBES, J.-M. (2021). A consistent extension of discrete optimal transport maps for machine learning applications. arXiv preprint. Available at arXiv:2102.08644.

DE PHILIPPIS, G. and FIGALLI, A. (2014). The Monge–Ampère equation and its link to optimal transportation. *Bull. Amer. Math. Soc.* (*N.S.*) **51** 527–580. MR3237759 https://doi.org/10.1090/S0273-0979-2014-01459-4

DEB, N., BHATTACHARYA, B. B. and SEN, B. (2021). Efficiency lower bounds for distribution-free hotelling-type two-sample tests based on optimal transport. arXiv preprint. Available at arXiv:2104.01986.

DEB, N., GHOSAL, P. and SEN, B. (2021). Rates of estimation of optimal transport maps using plug-in estimators via barycentric projections. *Adv. Neural Inf. Process. Syst.* **34**.

DEB, N. and SEN, B. (2023). Multivariate rank-based distribution-free nonparametric testing using measure transportation. *J. Amer. Statist. Assoc.* **118** 192–207. MR4571116 https://doi.org/10.1080/01621459.2021.1923508

DEL BARRIO, E., GINÉ, E. and MATRÁN, C. (1999). Central limit theorems for the Wasserstein distance between the empirical and the true distributions. *Ann. Probab.* **27** 1009–1071. MR1698999 https://doi.org/10.1214/aop/1022677394

DEL BARRIO, E., GINÉ, E. and UTZET, F. (2005). Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli* **11** 131–189. MR2121458 https://doi.org/10.3150/bj/1110228245

DEL BARRIO, E., GORDALIZA, P. and LOUBES, J.-M. (2019). A central limit theorem for L_p transportation cost on the real line with application to fairness assessment in machine learning. *Inf. Inference* **8** 817–849. MR4045479 https://doi.org/10.1093/imaiai/iaz016

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

DELALANDE, A. and MÉRIGOT, Q. (2023). Quantitative stability of optimal transport maps under variations of the target measure. *Duke Math. J.* **172** 3321–3357. MR4688680 https://doi.org/10.1215/00127094-2022-0106

DIVOL, V. (2021). A short proof on the rate of convergence of the empirical measure for the Wasserstein distance. arXiv preprint. Available at arXiv:2101.08126.

DIVOL, V. (2022). Measure estimation on manifolds: An optimal transport approach. *Probab. Theory Related Fields* **183** 581–647. MR4421180 https://doi.org/10.1007/s00440-022-01118-z

DIVOL, V., NILES-WEED, J. and POOLADIAN, A.-A. (2022). Optimal transport map estimation in general function spaces. arXiv preprint. Available at arXiv:2212.03722.

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

DUNLOP, M. M., SLEPČEV, D., STUART, A. M. and THORPE, M. (2020). Large data and zero noise limits of graph-based semi-supervised learning algorithms. *Appl. Comput. Harmon. Anal.* **49** 655–697. MR4117856 https://doi.org/10.1016/j.acha.2019.03.005

EFROMOVICH, S. (1999). *Nonparametric Curve Estimation: Methods, Theory, and Applications. Springer Series in Statistics*. Springer, New York. MR1705298

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

FAN, J. and HU, T. C. (1992). Bias correction and higher order kernel functions. *Statist. Probab. Lett.* **13** 235–243. MR1158864 https://doi.org/10.1016/0167-7152(92)90053-8

FIGALLI, A. (2017). *The Monge–Ampère Equation and Its Applications. Zurich Lectures in Advanced Mathematics*. Eur. Math. Soc., Zürich. MR3617963 https://doi.org/10.4171/170

FINLAY, C., GEROLIN, A., OBERMAN, A. M. and POOLADIAN, A.-A. (2020). Learning normalizing flows from Entropy–Kantorovich potentials. arXiv preprint. Available at arXiv:2006.06033.

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

FREITAG, G. and MUNK, A. (2005). On Hadamard differentiability in k-sample semiparametric models—with applications to the assessment of structural relationships. *J. Multivariate Anal.* **94** 123–158. MR2161214 https://doi.org/10.1016/j.jmva.2004.03.006

GHODRATI, L. and PANARETOS, V. M. (2022). Distribution-on-distribution regression via optimal transport maps. *Biometrika* **109** 957–974. MR4519110 https://doi.org/10.1093/biomet/asac005

GHOSAL, P. and SEN, B. (2022). Multivariate ranks and quantiles using optimal transport: Consistency, rates and nonparametric testing. *Ann. Statist.* **50** 1012–1037. MR4404927 https://doi.org/10.1214/21-aos2136

GIGLI, N. (2011). On Hölder continuity-in-time of the optimal transport map towards measures along a curve. *Proc. Edinb. Math. Soc.* (2) **54** 401–409. MR2794662 https://doi.org/10.1017/S001309150800117X

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

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

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

GONZÁLEZ-DELGADO, J., GONZÁLEZ-SANZ, A., CORTÉS, J. and NEUVIAL, P. (2023). Two-sample goodness-of-fit tests on the flat torus based on Wasserstein distance and their relevance to structural biology. *Electron. J. Stat.* **17** 1547–1586. MR4598874 https://doi.org/10.1214/23-ejs2135

GORDALIZA, P., DEL BARRIO, E., FABRICE, G. and LOUBES, J.-M. (2019). Obtaining fairness using optimal transport theory. In *International Conference on Machine Learning* 2357–2365.

GUITTET, K. (2003). On the time-continuous mass transport problem and its approximation by augmented Lagrangian techniques. *SIAM J. Numer. Anal.* **41** 382–399. MR1974507 https://doi.org/10.1137/S0036142901386069

GUNSILIUS, F. F. (2022). On the convergence rate of potentials of Brenier maps. *Econometric Theory* **38** 381–417. MR4407051 https://doi.org/10.1017/S0266466621000037

GYÖRFI, L., KOHLER, M., KRZYŻAK, A. and WALK, H. (2002). *A Distribution-Free Theory of Nonparametric Regression. Springer Series in Statistics*. Springer, New York. MR1920390 https://doi.org/10.1007/b97848

HALLIN, M., DEL BARRIO, E., CUESTA-ALBERTOS, J. and MATRÁN, C. (2021). Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach. *Ann. Statist.* **49** 1139–1165. MR4255122 https://doi.org/10.1214/20-aos1996

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

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

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

HÜTTER, J.-C. and RIGOLLET, P. (2021). Minimax estimation of smooth optimal transport maps. *Ann. Statist.* **49** 1166–1194. MR4255123 https://doi.org/10.1214/20-aos1997

KANTOROVICH, L. V. (1948). On a problem of Monge. *C. R. (Dokl.) Acad. Sci. URSS* **3** 225–226.

KANTOROVITCH, L. (1942). On the translocation of masses. *C. R. (Dokl.) Acad. Sci. URSS* **37** 199–201. MR0009619

KERKYACHARIAN, G. and PICARD, D. (1992). Density estimation in Besov spaces. *Statist. Probab. Lett.* **13** 15–24. MR1147634 https://doi.org/10.1016/0167-7152(92)90231-S

KLEIN, N., ORELLANA, J., BRINCAT, S. L., MILLER, E. K. and KASS, R. E. (2020). Torus graphs for multivariate phase coupling analysis. *Ann. Appl. Stat.* **14** 635–660. MR4117823 https://doi.org/10.1214/19-AOAS1300
KNOTT, M. and SMITH, C. S. (1984). On the optimal mapping of distributions. *J. Optim. Theory Appl.* **43** 39–49. MR0745785 https://doi.org/10.1007/BF00934745
KOLOURI, S., PARK, S. R., THORPE, M., SLEPCEV, D. and ROHDE, G. K. (2017). Optimal mass transport: Signal processing and machine-learning applications. *IEEE Signal Process. Mag.* **34** 43–59.
KOMISKE, P. T., MASTANDREA, R., METODIEV, E. M., NAIK, P. and THALER, J. (2020a). Exploring the space of jets with CMS open data. *Phys. Rev. D* **101** 034009.
KOMISKE, P. T., METODIEV, E. M. and THALER, J. (2019). Metric space of collider events. *Phys. Rev. Lett.* **123** 041801. https://doi.org/10.1103/PhysRevLett.123.041801
KRISHNAMURTHY, A., KANDASAMY, K., POCZOS, B. and WASSERMAN, L. (2014). Nonparametric estimation of Renyi divergence and friends. In *International Conference on Machine Learning* 919–927.
LEDOUX, M. (2019). On optimal matching of Gaussian samples. *J. Math. Sci.* **238** 495–522. MR3723584
LEI, J. (2020). Convergence and concentration of empirical measures under Wasserstein distance in unbounded functional spaces. *Bernoulli* **26** 767–798. MR4036051 https://doi.org/10.3150/19-BEJ1151
LEVY, B. and SCHWINDT, E. (2018). Notions of optimal transport theory and how to implement them on a computer. *Comput. Graph.* **72** 135–148.
LIANG, T. (2019). On the minimax optimality of estimating the Wasserstein metric. arXiv preprint. Available at arXiv:1908.10324.
LIN, T., CUTURI, M. and JORDAN, M. I. (2023). A specialized semismooth Newton method for kernel-based optimal transport. arXiv preprint. Available at arXiv:2310.14087.
LOEPER, G. and RAPETTI, F. (2005). Numerical solution of the Monge–Ampère equation by a Newton's algorithm. *C. R. Math. Acad. Sci. Paris* **340** 319–324. MR2121899 https://doi.org/10.1016/j.crma.2004.12.018
MA, X.-N., TRUDINGER, N. S. and WANG, X.-J. (2005). Regularity of potential functions of the optimal transportation problem. *Arch. Ration. Mech. Anal.* **177** 151–183. MR2188047 https://doi.org/10.1007/s00205-005-0362-9
MAKKUVA, A., TAGHVAEI, A., OH, S. and LEE, J. (2020). Optimal transport mapping via input convex neural networks. In *The 37th International Conference on Machine Learning* 6672–6681. PMLR.
MANOLE, T., BALAKRISHNAN, S., NILES-WEED, J. and WASSERMAN, L. (2024). Supplement to "Plugin estimation of smooth optimal transport maps." https://doi.org/10.1214/24-AOS2379SUPP
MANOLE, T., BALAKRISHNAN, S. and WASSERMAN, L. (2022). Minimax confidence intervals for the sliced Wasserstein distance. *Electron. J. Stat.* **16** 2252–2345. MR4402565 https://doi.org/10.1214/22-ejs2001
MANOLE, T. and NILES-WEED, J. (2024). Sharp convergence rates for empirical optimal transport with smooth costs. *Ann. Appl. Probab.* **34** 1108–1135. MR4700254 https://doi.org/10.1214/23-aap1986
MASRY, E. (1997). Multivariate probability density estimation by wavelet methods: Strong consistency and rates for stationary time series. *Stochastic Process. Appl.* **67** 177–193. MR1449830 https://doi.org/10.1016/S0304-4149(96)00005-1
MAZUMDER, R., CHOUDHURY, A., IYENGAR, G. and SEN, B. (2019). A computational framework for multivariate convex regression and its variants. *J. Amer. Statist. Assoc.* **114** 318–331. MR3941257 https://doi.org/10.1080/01621459.2017.1407771
MÉRIGOT, Q. (2011). A multiscale approach to optimal transport. *Comput. Graph. Forum* **30** 1583–1592.
MÉRIGOT, Q., DELALANDE, A. and CHAZAL, F. (2019). Quantitative stability of optimal transport maps and linearization of the 2-Wasserstein space. arXiv preprint. Available at arXiv:1910.05954.
MONGE, G. (1781). Mémoire Sur La Théorie Des Déblais et Des Remblais. Histoire de L'Académie Royale des Sciences de Paris.
MUNK, A. and CZADO, C. (1998). Nonparametric validation of similar distributions and assessment of goodness of fit. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 223–241. MR1625620 https://doi.org/10.1111/1467-9868.00121
MUZELLEC, B., VACHER, A., BACH, F., VIALARD, F.-X. and RUDI, A. (2021). Near-optimal estimation of smooth transport maps with kernel sums-of-squares. arXiv preprint. Available at arXiv:2112.01907.
NATH, J. S. and JAWANPURIA, P. (2020). Statistical optimal transport posed as learning kernel embedding. arXiv preprint. Available at arXiv:2002.03179.
NILES-WEED, J. and BERTHET, Q. (2022). Minimax estimation of smooth densities in Wasserstein distance. *Ann. Statist.* **50** 1519–1540. MR4441130 https://doi.org/10.1214/21-aos2161
NILES-WEED, J. and RIGOLLET, P. (2022). Estimation of Wasserstein distances in the spiked transport model. *Bernoulli* **28** 2663–2688. MR4474558 https://doi.org/10.3150/21-bej1433
ONKEN, D., FUNG, S. W., LI, X. and RUTHOTTO, L. (2021). OT-flow: Fast and accurate continuous normalizing flows via optimal transport. arXiv preprint. Available at arXiv:2006.00104.
PANARETOS, V. M. and ZEMEL, Y. (2019). Statistical aspects of Wasserstein distances. *Annu. Rev. Stat. Appl.* **6** 405–431. MR3939527 https://doi.org/10.1146/annurev-statistics-030718-104938

PERROT, M., COURTY, N., FLAMARY, R. and HABRARD, A. (2016). Mapping estimation for discrete optimal transport. *Adv. Neural Inf. Process. Syst.* **29** 4197–4205.

PEYRÉ, G. and CUTURI, M. (2019). Computational optimal transport: With applications to data science. *Found. Trends Mach. Learn.* **11** 355–607.

PEYRE, R. (2018). Comparison between W_2 distance and $\dot{H}^{-1}$ norm, and localization of Wasserstein distance. *ESAIM Control Optim. Calc. Var.* **24** 1489–1501. MR3922440 https://doi.org/10.1051/cocv/2017050

POOLADIAN, A.-A. and NILES-WEED, J. (2021). Entropic estimation of optimal transport maps. arXiv preprint. Available at arXiv:2109.12004.

RACHEV, S. T. and RÜSCHENDORF, L. (1998). *Mass Transportation Problems. Vol. I: Theory. Probability and Its Applications (New York)*. Springer, New York. MR1619170

RAKOTOMAMONJY, A., FLAMARY, R., GASSO, G., EL ALAYA, M., BERAR, M. and COURTY, N. (2022). Optimal transport for conditional domain matching and label shift. *Mach. Learn.* **111** 1651–1670. MR4426353 https://doi.org/10.1007/s10994-021-06088-2

READ, A. L. (1999). Linear interpolation of histograms. *Nucl. Instrum. Methods Phys. Res., Sect. A, Accel. Spectrom. Detect. Assoc. Equip.* **425** 357–360.

REDKO, I., COURTY, N., FLAMARY, R. and TUIA, D. (2019). Optimal transport for multi-source domain adaptation under target shift. In *The 22nd International Conference on Artificial Intelligence and Statistics* 849–858. PMLR.

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

SCHIEBINGER, G., SHU, J., TABAKA, M., CLEARY, B., SUBRAMANIAN, V., SOLOMON, A., GOULD, J., LIU, S., LIN, S. et al. (2019). Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell* **176** 928–943.e22.

SEIJO, E. and SEN, B. (2011). Nonparametric least squares estimation of a multivariate convex regression function. *Ann. Statist.* **39** 1633–1657. MR2850215 https://doi.org/10.1214/10-AOS852

SHI, H., DRTON, M. and HAN, F. (2022). Distribution-free consistent independence tests via center-outward ranks and signs. *J. Amer. Statist. Assoc.* **117** 395–410. MR4399094 https://doi.org/10.1080/01621459.2020.1782223

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

STAUDT, T., HUNDRIESER, S. and MUNK, A. (2022). On the uniqueness of Kantorovich potentials. arXiv preprint. Available at arXiv:2201.08316.

TALAGRAND, M. (1992). The Ajtai–Komlós–Tusnády matching theorem for general measures. In *Probability in Banach Spaces*, 8 (*Brunswick, ME*, 1991). *Progress in Probability* **30** 39–54. Birkhäuser, Boston, MA. MR1227608

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

TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation. Springer Series in Statistics*. Springer, New York. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats. MR2724359 https://doi.org/10.1007/b13794

URBAS, J. (1997). On the second boundary value problem for equations of Monge–Ampère type. *J. Reine Angew. Math.* **487** 115–124. MR1454261 https://doi.org/10.1515/crll.1997.487.115

VACHER, A., MUZELLEC, B., RUDI, A., BACH, F. and VIALARD, F.-X. (2021). A dimension-free computational upper-bound for smooth optimal transport estimation. arXiv preprint. Available at arXiv:2101.05380.

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. (2002). Semiparametric statistics. In *Lectures on Probability Theory and Statistics: École DÉté de Probabilités de Saint-Flour XXIX*—1999 (P. Bernard, ed.) Springer, Berlin.

VILLANI, C. (2003). *Topics in Optimal Transportation. Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI. MR1964483 https://doi.org/10.1090/gsm/058

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

WANG, W., OZOLEK, J. A., SLEPČEV, D., LEE, A. B., CHEN, C. and ROHDE, G. K. (2011). An optimal transportation approach for nuclear structure-based pathology. *IEEE Trans. Med. Imag.* **30** 621–631.

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

WIECHERS, H., ELTZNER, B., MARDIA, K. V. and HUCKEMANN, S. F. (2023). Learning torus PCA-based classification for multiscale RNA correction with application to SARS-CoV-2. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **72** 271–293. MR4719276 https://doi.org/10.1093/jrsssc/qlad004

ZHU, J., GUHA, A., XU, M., MA, Y., LEI, R., LOFFREDO, V., NGUYEN, X. and ZHAO, D. (2021). Functional optimal transport: Mapping estimation and domain adaptation for functional data. arXiv preprint. Available at arXiv:2102.03895.

The Annals of Statistics
2024, Vol. 52, No. 3, 999–1026
https://doi.org/10.1214/24-AOS2380

CHANGE-POINT INFERENCE IN HIGH-DIMENSIONAL REGRESSION MODELS UNDER TEMPORAL DEPENDENCE

BY HAOTIAN XU[1,a], DAREN WANG[2,c], ZIFENG ZHAO[3,d] AND YI YU[1,b]

[1]*Department of Statistics, University of Warwick,* [a]*haotian.xu.1@warwick.ac.uk,* [b]*yi.yu.2@warwick.ac.uk*
[2]*Department of Statistics, University of Notre Dame,* [c]*dwang24@nd.edu*
[3]*Mendoza College of Business, University of Notre Dame,* [d]*zzhao2@nd.edu*

This paper concerns the limiting distributions of change-point estimators, in a high-dimensional linear regression time-series context, where a regression object $(y_t, X_t) \in \mathbb{R} \times \mathbb{R}^p$ is observed at every time point $t \in \{1, \dots, n\}$. At unknown time points, called change points, the regression coefficients change, with the jump sizes measured in ℓ_2-norm. We provide limiting distributions of the change-point estimators in the regimes where the minimal jump size vanishes and where it remains a constant. We allow for both the covariate and noise sequences to be temporally dependent, in the functional dependence framework, which is the first time seen in the change-point inference literature. We show that a block-type long-run variance estimator is consistent under the functional dependence, which facilitates the practical implementation of our derived limiting distributions. We also present a few important byproducts of our analysis, which are of their own interest. These include a novel variant of the dynamic programming algorithm to boost the computational efficiency, consistent change-point localization rates under temporal dependence and a new Bernstein inequality for data possessing functional dependence. Extensive numerical results are provided to support our theoretical results. The proposed methods are implemented in the R package `changepoints` (Xu et al. (2022)).

REFERENCES

AUE, A., GABRYS, R., HORVÁTH, L. and KOKOSZKA, P. (2009). Estimation of a change-point in the mean function of functional data. *J. Multivariate Anal.* **100** 2254–2269. MR2560367 https://doi.org/10.1016/j.jmva.2009.04.001

AUE, A. and HORVÁTH, L. (2013). Structural breaks in time series. *J. Time Series Anal.* **34** 1–16. MR3008012 https://doi.org/10.1111/j.1467-9892.2012.00819.x

AUE, A., RICE, G. and SÖNMEZ, O. (2018). Detecting and dating structural breaks in functional data without dimension reduction. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 509–529. MR3798876 https://doi.org/10.1111/rssb.12257

BAI, J. (1994). Least squares estimation of a shift in linear processes. *J. Time Series Anal.* **15** 453–472. MR1292161 https://doi.org/10.1111/j.1467-9892.1994.tb00204.x

BAI, J. (2010). Common breaks in means and variances for panel data. *J. Econometrics* **157** 78–92. MR2652280 https://doi.org/10.1016/j.jeconom.2009.10.020

BERKES, I., GABRYS, R., HORVÁTH, L. and KOKOSZKA, P. (2009). Detecting changes in the mean of functional observations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 927–946. MR2750251 https://doi.org/10.1111/j.1467-9868.2009.00713.x

BHATTACHARJEE, M., BANERJEE, M. and MICHAILIDIS, G. (2020). Change point estimation in a dynamic stochastic block model. *J. Mach. Learn. Res.* **21** 1–59. MR4119175

BRODSKY, B. E. and DARKHOVSKY, B. S. (1993). *Nonparametric Methods in Change-Point Problems. Mathematics and Its Applications* **243**. Kluwer Academic, Dordrecht. MR1228205 https://doi.org/10.1007/978-94-015-8163-9

MSC2020 subject classifications. Primary 62F30; secondary 62F35.
Key words and phrases. Change-point inference, confidence interval, high-dimensional linear regression, long-run variance, temporal dependence.

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

CAO, H. and WU, W. B. (2015). Changepoint estimation: Another look at multiple testing problems. *Biometrika* **102** 974–980. MR3431567 https://doi.org/10.1093/biomet/asv031

CASINI, A. and PERRON, P. (2019). *Structural Breaks in Time Series*. *Oxford Research Encyclopedia of Economics and Finance* (forthcoming). Oxford Univ. Press, London.

CHEN, L., WANG, W. and WU, W. B. (2022). Inference of breakpoints in high-dimensional time series. *J. Amer. Statist. Assoc.* **117** 1951–1963. MR4528482 https://doi.org/10.1080/01621459.2021.1893178

CHEN, Y. B., NG, S. and BAI, J. (2022). fbi: Factor-Based Imputation and FRED-MD/QD Data Set R package version 0.6.0.

CHU, N. (2016). *Air Quality in Northern Taiwan*. Available at https://www.kaggle.com/datasets/nelsonchu/air-quality-in-northern-taiwan [Accessed: April 16, 2023].

CSÖRGŐ, M. and HORVÁTH, L. (1997). *Limit Theorems in Change-Point Analysis*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR2743035

DAHLHAUS, R., RICHTER, S. and WU, W. B. (2019). Towards a general theory for nonlinear locally stationary processes. *Bernoulli* **25** 1013–1044. MR3920364 https://doi.org/10.3150/17-bej1011

EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499. MR2060166 https://doi.org/10.1214/009053604000000067

FRIEDRICH, F., KEMPE, A., LIEBSCHER, V. and WINKLER, G. (2008). Complexity penalized M-estimation: Fast computation. *J. Comput. Graph. Statist.* **17** 201–224. MR2424802 https://doi.org/10.1198/106186008X285591

HAN, Y. and TSAY, R. S. (2020). High-dimensional linear regression for dependent data with applications to nowcasting. *Statist. Sinica* **30** 1797–1827. MR4260745 https://doi.org/10.5705/ss.202018.0044

KAUL, A., FOTOPOULOS, S. B., JANDHYALA, V. K. and SAFIKHANI, A. (2021). Inference on the change point under a high dimensional sparse mean shift. *Electron. J. Stat.* **15** 71–134. MR4195770 https://doi.org/10.1214/20-EJS1791

KAUL, A., JANDHYALA, V. K. and FOTOPOULOS, S. B. (2019). An efficient two step algorithm for high dimensional change point regression models without grid search. *J. Mach. Learn. Res.* **20** 1–40. MR3990465

KAUL, A. and MICHAILIDIS, G. (2021). Inference for change points in high dimensional mean shift models. arXiv preprint. Available at arXiv:2107.09150.

KILLICK, R., FEARNHEAD, P. and ECKLEY, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *J. Amer. Statist. Assoc.* **107** 1590–1598. MR3036418 https://doi.org/10.1080/01621459.2012.737745

KUCHIBHOTLA, A. K., BROWN, L. D., BUJA, A., GEORGE, E. I. and ZHAO, L. (2023). Uniform-in-submodel bounds for linear regression in a model-free framework. *Econometric Theory* **39** 1202–1248. MR4678102 https://doi.org/10.1017/s0266466621000219

LEE, S., SEO, M. H. and SHIN, Y. (2016). The lasso for high dimensional regression with a possible change point. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 193–210. MR3453652 https://doi.org/10.1111/rssb.12108

LEONARDI, F. BÜHLMANN, P. (2016). Computationally efficient change point detection for high-dimensional regression. arXiv preprint. Available at arXiv:1601.03704.

MADRID PADILLA, C. M., WANG, D., ZHAO, Z. and YU, Y. Change-point detection for sparse and dense functional data in general dimensions. arXiv e-prints arXiv–2205.

MCCRACKEN, M. W. and NG, S. (2016). FRED-MD: A monthly database for macroeconomic research. *J. Bus. Econom. Statist.* **34** 574–589. MR3547997 https://doi.org/10.1080/07350015.2015.1086655

MERLEVÈDE, F., PELIGRAD, M. and RIO, E. (2009). Bernstein inequality and moderate deviations under strong mixing conditions. In *High Dimensional Probability V: The Luminy Volume*. *Inst. Math. Stat.* (*IMS*) *Collect.* **5** 273–292. IMS, Beachwood, OH. MR2797953 https://doi.org/10.1214/09-IMSCOLL518

MUKHERJEE, D., BANERJEE, M. and RITOV, Y. (2022). On robust learning in the canonical change point problem under heavy tailed errors in finite and growing dimensions. *Electron. J. Stat.* **16** 1153–1252. MR4381059 https://doi.org/10.1214/21-ejs1927

NATIONAL RESEARCH COUNCIL (2013). *Frontiers in Massive Data Analysis*. The National Academies Press, Washington, DC.

PILLIAT, E., CARPENTIER, A. and VERZELEN, N. (2023). Optimal multiple change-point detection for high-dimensional data. *Electron. J. Stat.* **17** 1240–1315. MR4576243 https://doi.org/10.1214/23-ejs2126

RICHTER, S., WANG, W. and WU, W. B. (2023). Testing for parameter change epochs in GARCH time series. *Econom. J.* **26** 467–491. MR4643834 https://doi.org/10.1093/ectj/utad006

RINALDO, A., WANG, D., WEN, Q., WILLETT, R. and YU, Y. (2021). Localizing changes in high-dimensional regression models. In *International Conference on Artificial Intelligence and Statistics* 2089–2097. PMLR.

SHAO, X. and ZHANG, X. (2010). Testing for change points in time series. *J. Amer. Statist. Assoc.* **105** 1228–1240. MR2752617 https://doi.org/10.1198/jasa.2010.tm10103

VAN DE GEER, S. (2018). On tight bounds for the Lasso. *J. Mach. Learn. Res.* **19** 1–48. MR3874154

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

VERSHYNIN, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science. Cambridge Series in Statistical and Probabilistic Mathematics* **47**. Cambridge Univ. Press, Cambridge. MR3837109 https://doi.org/10.1017/9781108231596

VERZELEN, N., FROMONT, M., LERASLE, M. and REYNAUD-BOURET, P. (2023). Optimal change-point detection and localization. *Ann. Statist.* **51** 1586–1610. MR4658569 https://doi.org/10.1214/23-aos2297

WANG, D., YU, Y. and RINALDO, A. (2020). Univariate mean change point detection: Penalization, CUSUM and optimality. *Electron. J. Stat.* **14** 1917–1961. MR4091859 https://doi.org/10.1214/20-EJS1710

WANG, D., YU, Y. and RINALDO, A. (2021). Optimal change point detection and localization in sparse dynamic networks. *Ann. Statist.* **49** 203–232. MR4206675 https://doi.org/10.1214/20-AOS1953

WANG, D. and ZHAO, Z. (2022). Optimal change-point testing for high-dimensional linear models with temporal dependence. arXiv preprint. Available at arXiv:2205.03880.

WANG, D., ZHAO, Z., LIN, K. Z. and WILLETT, R. (2021). Statistically and computationally efficient change point localization in regression settings. *J. Mach. Learn. Res.* **22** 1–46.

WANG, R. and SHAO, X. (2023). Dating the break in high-dimensional data. *Bernoulli* **29** 2879–2901. MR4632124 https://doi.org/10.3150/22-bej1567

WANG, R., ZHU, C., VOLGUSHEV, S. and SHAO, X. (2022). Inference for change points in high-dimensional data via selfnormalization. *Ann. Statist.* **50** 781–806. MR4405366 https://doi.org/10.1214/21-aos2127

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

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

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

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

XU, H., PADILLA, O., WANG, D. and LI, M. (2022). changepoints: A collection of change-point detection methods R package version 1.1.0.

XU, H., WANG, D., ZHAO, Z. and YU, Y. (2024). Supplement to "Change-point inference in high-dimensional regression models under temporal dependence." https://doi.org/10.1214/24-AOS2380SUPP

YAO, Y.-C. (1987). Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. *Ann. Statist.* **15** 1321–1328. MR0902262 https://doi.org/10.1214/aos/1176350509

YAO, Y.-C. (1988). Estimating the number of change-points via Schwarz' criterion. *Statist. Probab. Lett.* **6** 181–189. MR0919373 https://doi.org/10.1016/0167-7152(88)90118-6

YAO, Y.-C. and AU, S. T. (1989). Least-squares estimation of a step function. *Sankhyā Ser. A* **51** 370–381. MR1175613

YU, M. and CHEN, X. (2021). Finite sample change point inference and identification for high-dimensional mean vectors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 247–270. MR4250275 https://doi.org/10.1111/rssb.12406

YU, Y., CHATTERJEE, S. and XU, H. (2022). Localising change points in piecewise polynomials of general degrees. *Electron. J. Stat.* **16** 1855–1890. MR4396490 https://doi.org/10.1214/21-ejs1963

ZHANG, B., GENG, J. and LAI, L. (2015). Multiple change-points estimation in linear regression models via sparse group lasso. *IEEE Trans. Signal Process.* **63** 2209–2224. MR3331995 https://doi.org/10.1109/TSP.2015.2411220

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

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, Y., WANG, R. and SHAO, X. (2022). Adaptive inference for change points in high-dimensional data. *J. Amer. Statist. Assoc.* **117** 1751–1762. MR4528468 https://doi.org/10.1080/01621459.2021.1884562

The Annals of Statistics
2024, Vol. 52, No. 3, 1027–1049
https://doi.org/10.1214/24-AOS2381

HIGH-DIMENSIONAL COVARIANCE MATRICES UNDER DYNAMIC VOLATILITY MODELS: ASYMPTOTICS AND SHRINKAGE ESTIMATION

BY YI DING[1,a] AND XINGHUA ZHENG[2,b]

[1]*Faculty of Business Administration, University of Macau,* [a]*yiding@um.edu.mo*
[2]*Department of ISOM, Hong Kong University of Science and Technology,* [b]*xhzheng@ust.hk*

We study the estimation of high-dimensional covariance matrices and their empirical spectral distributions under dynamic volatility models. Data under such models have nonlinear dependency both cross-sectionally and temporally. We establish the condition under which the limiting spectral distribution (LSD) of the sample covariance matrix under scalar BEKK models is different from the i.i.d. case. We then propose a time-variation adjusted (TV-adj) sample covariance matrix and prove that its LSD follows the Marčenko–Pastur law. Based on the asymptotics of the TV-adj sample covariance matrix, we develop a consistent population spectrum estimator and an asymptotically optimal nonlinear shrinkage estimator of the unconditional covariance matrix.

REFERENCES

AO, M., LI, Y. and ZHENG, X. (2019). Approaching mean-variance efficiency for large portfolios. *Rev. Financ. Stud.* **32** 2890–2919.

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

BANNA, M. and MERLEVÈDE, F. (2015). Limiting spectral distribution of large sample covariance matrices associated with a class of stationary processes. *J. Theoret. Probab.* **28** 745–783. MR3370674 https://doi.org/10.1007/s10959-013-0508-x

BHATTACHARJEE, M. and BOSE, A. (2016). Large sample behaviour of high dimensional autocovariance matrices. *Ann. Statist.* **44** 598–628. MR3476611 https://doi.org/10.1214/15-AOS1378

BOLLERSLEV, T., ENGLE, R. F. and WOOLDRIDGE, J. M. (1988). A capital asset pricing model with time-varying covariances. *J. Polit. Econ.* **96** 116–131.

DE NARD, G., LEDOIT, O. and WOLF, M. (2021). Factor models for portfolio selection in large dimensions: The good, the better and the ugly. *J. Financ. Econom.* **19** 236–257.

DING, Y., LI, Y. and ZHENG, X. (2021). High dimensional minimum variance portfolio estimation under statistical factor models. *J. Econometrics* **222** 502–515. MR4234830 https://doi.org/10.1016/j.jeconom.2020.07.013

DING, Y. and ZHENG, X. (2024). Supplement to "High-dimensional covariance matrices under dynamic volatility models: Asymptotics and shrinkage estimation." https://doi.org/10.1214/24-AOS2381SUPP

DING, Z. and ENGLE, R. F. (2001). Large scale conditional covariance matrix modeling, estimation and testing. *Academia Economic Papers* **29** 157–184.

EL KAROUI, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Ann. Statist.* **36** 2757–2790. MR2485012 https://doi.org/10.1214/07-AOS581

ENGLE, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *J. Bus. Econom. Statist.* **20** 339–350. MR1939905 https://doi.org/10.1198/073500102288618487

ENGLE, R. F., GRANGER, C. W. J. and KRAFT, D. (1984). Combining competing forecasts of inflation using a bivariate ARCH model. *J. Econom. Dynam. Control* **8** 151–165. MR0781644 https://doi.org/10.1016/0165-1889(84)90031-9

ENGLE, R. F. and KRONER, K. F. (1995). Multivariate simultaneous generalized arch. *Econometric Theory* **11** 122–150. MR1325104 https://doi.org/10.1017/S0266466600009063

ENGLE, R. F., LEDOIT, O. and WOLF, M. (2019). Large dynamic covariance matrices. *J. Bus. Econom. Statist.* **37** 363–375. MR3948411 https://doi.org/10.1080/07350015.2017.1345683

MSC2020 subject classifications. Primary 62B20; secondary 62H12, 62M10.

Key words and phrases. High-dimension, dynamic volatility model, sample covariance matrix, spectral distribution, nonlinear shrinkage.

FRANCQ, C. and ZAKOÏAN, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA–GARCH processes. *Bernoulli* **10** 605–637. MR2076065 https://doi.org/10.3150/bj/1093265632

FRANCQ, C. and ZAKOIAN, J.-M. (2007). Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero. *Stochastic Process*. *Appl*. **117** 1265–1284. MR2343939 https://doi.org/10.1016/j.spa.2007.01.001

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

LEDOIT, O. and PÉCHÉ, S. (2011). Eigenvectors of some large sample covariance matrix ensembles. *Probab. Theory Related Fields* **151** 233–264. MR2834718 https://doi.org/10.1007/s00440-010-0298-3

LEDOIT, O. and WOLF, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Ann*. *Statist*. **40** 1024–1060. MR2985942 https://doi.org/10.1214/12-AOS989

LEDOIT, O. and WOLF, M. (2015). Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. *J. Multivariate Anal*. **139** 360–384. MR3349498 https://doi.org/10.1016/j.jmva.2015.04.006

LEDOIT, O. and WOLF, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *Rev. Financ*. *Stud*. **30** 4349–4388.

LEDOIT, O. and WOLF, M. (2020). Analytical nonlinear shrinkage of large-dimensional covariance matrices. *Ann*. *Statist*. **48** 3043–3065. MR4152634 https://doi.org/10.1214/19-AOS1921

LIU, H., AUE, A. and PAUL, D. (2015). On the Marčenko–Pastur law for linear time series. *Ann*. *Statist*. **43** 675–712. MR3319140 https://doi.org/10.1214/14-AOS1294

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

MERLEVÈDE, F. and PELIGRAD, M. (2016). On the empirical spectral distribution for matrices with long memory and independent rows. *Stochastic Process*. *Appl*. **126** 2734–2760. MR3522299 https://doi.org/10.1016/j.spa.2016.02.016

PAKEL, C., SHEPHARD, N., SHEPPARD, K. and ENGLE, R. F. (2021). Fitting vast dimensional time-varying covariance models. *J. Bus*. *Econom*. *Statist*. **39** 652–668. MR4272926 https://doi.org/10.1080/07350015.2020.1713795

PEDERSEN, R. S. and RAHBEK, A. (2014). Multivariate variance targeting in the BEKK–GARCH model. *Econom*. *J*. **17** 24–55. MR3171211 https://doi.org/10.1111/ectj.12019

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

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

YANG, X., ZHENG, X. and CHEN, J. (2021). Testing high-dimensional covariance matrices under the elliptical distribution and beyond. *J. Econometrics* **221** 409–423. MR4215033 https://doi.org/10.1016/j.jeconom.2020.05.017

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

YASKOV, P. A. (2017). On the spectrum of sample covariance matrices for time series. *Teor. Veroyatn*. *Primen*. **62** 542–555. MR3684648 https://doi.org/10.1137/S0040585X97T988721

YIN, Y. Q. (1986). Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal*. **20** 50–68. MR0862241 https://doi.org/10.1016/0047-259X(86)90019-9

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

The Annals of Statistics
2024, Vol. 52, No. 3, 1050–1075
https://doi.org/10.1214/24-AOS2382

CHANGE ACCELERATION AND DETECTION

BY YANGLEI SONG[1,a] AND GEORGIOS FELLOURIS[2,b]

[1]*Department of Mathematics and Statistics, Queen's University,* [a]*yanglei.song@queensu.ca*
[2]*Department of Statistics, University of Illinois Urbana-Champaign,* [b]*fellouri@illinois.edu*

A novel sequential change detection problem is proposed, in which the goal is to not only detect but also accelerate the change. Specifically, it is assumed that the sequentially collected observations are responses to treatments selected in real time. The assigned treatments determine the pre-change and post-change distributions of the responses and also influence when the change happens. The goal is to find a treatment assignment rule and a stopping rule that minimize the expected total number of observations subject to a user-specified bound on the false alarm probability. The optimal solution is obtained under a general Markovian change-point model. Moreover, an alternative procedure is proposed, whose applicability is not restricted to Markovian change-point models and whose design requires minimal computation. For a large class of change-point models, the proposed procedure is shown to achieve the optimal performance in an asymptotic sense. Finally, its performance is found in simulation studies to be comparable to the optimal, uniformly with respect to the error probability.

REFERENCES

[1] BAKER, R. S. and INVENTADO, P. S. (2014). Educational data mining and learning analytics. In *Learning Analytics* 61–75. Springer, Berlin.
[2] BERTSEKAS, D. P. (2001). *Dynamic Programming and Optimal Control*. *Vol*. *II*, 2nd ed. Athena Scientific, Belmont, MA. MR2182753
[3] BERTSEKAS, D. P. (2022). *Abstract Dynamic Programming*. *Athena Scientific Optimization and Computation Series*. Athena Scientific, Belmont, MA. Third edition [of 3204932]. MR4496006
[4] BERTSEKAS, D. P. and TSITSIKLIS, J. N. (1991). An analysis of stochastic shortest path problems. *Math*. *Oper*. *Res*. **16** 580–595. MR1120471 https://doi.org/10.1287/moor.16.3.580
[5] CHAUDHURI, A., FELLOURIS, G. and TAJER, A. (2024). Round Robin active sequential change detection for dependent multi-channel data. arXiv preprint. Available at arXiv:2403.16297.
[6] CHEN, Y., CULPEPPER, S. A., WANG, S. and DOUGLAS, J. (2018). A hidden Markov model for learning trajectories in cognitive diagnosis with application to spatial rotation skills. *Appl*. *Psychol*. *Meas*. **42** 5–23.
[7] CHERNOFF, H. (1959). Sequential design of experiments. *Ann*. *Math*. *Stat*. **30** 755–770. MR0108874 https://doi.org/10.1214/aoms/1177706205
[8] COLLINS, L. M. and LANZA, S. T. (2009). *Latent Class and Latent Transition Analysis*: *With Applications in the Social*, *Behavioral*, *and Health Sciences*, *Vol*. 718. Wiley, New York.
[9] HERNÁNDEZ-LERMA, O. and LASSERRE, J. B. (1996). *Discrete-Time Markov Control Processes*: *Basic Optimality Criteria*. *Applications of Mathematics* (*New York*) **30**. Springer, New York. MR1363487 https://doi.org/10.1007/978-1-4612-0729-0
[10] KAYA, Y. and LEITE, W. L. (2017). Assessing change in latent skills across time with longitudinal cognitive diagnosis modeling: An evaluation of model performance. *Educ*. *Psychol*. *Meas*. **77** 369–388.
[11] KEENER, R. (1984). Second order efficiency in the sequential design of experiments. *Ann*. *Statist*. **12** 510–532. MR0740909 https://doi.org/10.1214/aos/1176346503
[12] KIEFER, J. and SACKS, J. (1963). Asymptotically optimum sequential inference and design. *Ann*. *Math*. *Stat*. **34** 705–750. MR0150907 https://doi.org/10.1214/aoms/1177704000
[13] LAI, T. L. (1998). Information bounds and quick detection of parameter changes in stochastic systems. *IEEE Trans*. *Inf*. *Theory* **44** 2917–2929. MR1672051 https://doi.org/10.1109/18.737522

MSC2020 subject classifications. 62L05, 62L10.
Key words and phrases. Sequential design of experiments, change-point detection, asymptotic optimality.

[14] LEVINE, S., KUMAR, A., TUCKER, G. and FU, J. (2020). Offline reinforcement learning: Tutorial, review, and perspectives on open problems. arXiv preprint. Available at arXiv:2005.01643.

[15] LI, F., COHEN, A., BOTTGE, B. and TEMPLIN, J. (2016). A latent transition analysis model for assessing change in cognitive skills. *Educ*. *Psychol*. *Meas*. **76** 181–204. https://doi.org/10.1177/0013164415588946

[16] LIANG, Q., LA TORRE, J. D. and LAW, N. (2023). Latent transition cognitive diagnosis model with covariates: A three-step approach. *J*. *Educ*. *Behav*. *Stat*. **10769986231163320**.

[17] LIU, K., MEI, Y. and SHI, J. (2015). An adaptive sampling strategy for online high-dimensional process monitoring. *Technometrics* **57** 305–319. MR3384946 https://doi.org/10.1080/00401706.2014.947005

[18] LORDEN, G. (1971). Procedures for reacting to a change in distribution. *Ann*. *Math*. *Stat*. **42** 1897–1908. MR0309251 https://doi.org/10.1214/aoms/1177693055

[19] MOUSTAKIDES, G. V. (2008). Sequential change detection revisited. *Ann*. *Statist*. **36** 787–807. MR2396815 https://doi.org/10.1214/009053607000000938

[20] NAGHSHVAR, M. and JAVIDI, T. (2013). Active sequential hypothesis testing. *Ann*. *Statist*. **41** 2703–2738. MR3161445 https://doi.org/10.1214/13-AOS1144

[21] NITINAWARAT, S., ATIA, G. K. and VEERAVALLI, V. V. (2013). Controlled sensing for multihypothesis testing. *IEEE Trans*. *Automat*. *Control* **58** 2451–2464. MR3106054 https://doi.org/10.1109/TAC.2013.2261188

[22] PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–115. MR0088850 https://doi.org/10.1093/biomet/41.1-2.100

[23] PATEK, S. D. (2001). On partially observed stochastic shortest path problems. In *Proceedings of the* 40*th IEEE Conference on Decision and Control* (*Cat*. *No*. 01*CH*37228), *Vol*. 5. 5050–5055. https://doi.org/10.1109/.2001.981011

[24] PATEK, S. D. (2007). Partially observed stochastic shortest path problems with approximate solution by neurodynamic programming. *IEEE Trans*. *Syst*. *Man Cybern*., *Part A*, *Syst*. *Humans* **37** 710–720.

[25] POLLAK, M. (1985). Optimal detection of a change in distribution. *Ann*. *Statist*. **13** 206–227. MR0773162 https://doi.org/10.1214/aos/1176346587

[26] PRUDENCIO, R. F., MAXIMO, M. R. and COLOMBINI, E. L. (2023). A survey on offline reinforcement learning: Taxonomy, review, and open problems. *IEEE Trans*. *Neural Netw*. *Learn*. *Syst*. MR4753658 https://doi.org/10.1007/s10994-023-06353-6

[27] SHEWHART, W. A. (2001). *Economic Control of Quality of Manufactured Product*. Macmillan & Co., London.

[28] SHIRYAEV, A. N. (2008). *Optimal Stopping Rules*. *Stochastic Modelling and Applied Probability* **8**. Springer, Berlin. Translated from the 1976 Russian second edition by A. B. Aries, Reprint of the 1978 translation. MR2374974

[29] ŠIRJAEV, A. N. (1963). Optimal methods in quickest detection problems. *Teor*. *Veroyatn*. *Primen*. **8** 26–51. MR0155708

[30] SONG, Y. and FELLOURIS, G. (2024). Supplement to "Change acceleration and detection." https://doi.org/10.1214/24-AOS2382SUPP

[31] TAJER, A., HEYDARI, J. and POOR, H. V. (2022). Active sampling for the quickest detection of Markov networks. *IEEE Trans*. *Inf*. *Theory* **68** 2479–2508. MR4413566 https://doi.org/10.1109/tit.2021.3124166

[32] TARTAKOVSKY, A., NIKIFOROV, I. and BASSEVILLE, M. (2015). *Sequential Analysis*: *Hypothesis Testing and Changepoint Detection*. *Monographs on Statistics and Applied Probability* **136**. CRC Press, Boca Raton, FL. MR3241619

[33] TARTAKOVSKY, A. G. (2017). On asymptotic optimality in sequential changepoint detection: Non-iid case. *IEEE Trans*. *Inf*. *Theory* **63** 3433–3450. MR3658534 https://doi.org/10.1109/TIT.2017.2683496

[34] TARTAKOVSKY, A. G. and VEERAVALLI, V. V. (2004). General asymptotic Bayesian theory of quickest change detection. *Teor*. *Veroyatn*. *Primen*. **49** 538–582. MR2144868 https://doi.org/10.1137/S0040585X97981202

[35] TEMPLIN, J., HENSON, R. A. et al. (2010). *Diagnostic Measurement*: *Theory*, *Methods*, *and Applications*. Guilford, New York.

[36] TEMPLIN, J. L. and HENSON, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychol*. *Methods* **11** 287–305.

[37] UEHARA, M., SHI, C. and KALLUS, N. (2022). A review of off-policy evaluation in reinforcement learning. arXiv preprint. Available at arXiv:2212.06355.

[38] VEERAVALLI, V. V. and BANERJEE, T. (2014). Quickest change detection. In *Academic Press Library in Signal Processing* **3** 209–255. Elsevier, Amsterdam.

[39] VEERAVALLI, V. V., FELLOURIS, G. and MOUSTAKIDES, G. V. (2024). Quickest change detection with controlled sensing. *IEEE J*. *Sel*. *Areas Inf*. *Theory* **5** 1–11. https://doi.org/10.1109/JSAIT.2024.3362324

[40] WANG, S. and CHEN, Y. (2020). Using response times and response accuracy to measure fluency within cognitive diagnosis models. *Psychometrika* **85** 600–629. MR4169490 https://doi.org/10.1007/s11336-020-09717-2

[41] WANG, S., YANG, Y., CULPEPPER, S. A. and DOUGLAS, J. A. (2018). Tracking skill acquisition with cognitive diagnosis models: A higher-order, hidden Markov model with covariates. *J. Educ. Behav. Stat.* **43** 57–87.

[42] XU, Q. and MEI, Y. (2023). Asymptotic optimality theory for active quickest detection with unknown postchange parameters. *Sequential Anal.* **42** 150–181. MR4595431 https://doi.org/10.1080/07474946.2023.2187417

[43] XU, Q., MEI, Y. and MOUSTAKIDES, G. V. (2021). Optimum multi-stream sequential change-point detection with sampling control. *IEEE Trans. Inf. Theory* **67** 7627–7636. MR4345143 https://doi.org/10.1109/tit.2021.3074961

[44] ZHANG, S. and CHANG, H.-H. (2016). From smart testing to smart learning: How testing technology can assist the new generation of education. *Int. J. Smart Technol. Learn.* **1** 67–92.

[45] ZHANG, S., LIU, J. and YING, Z. (2023). Statistical applications to cognitive diagnostic testing. *Annu. Rev. Stat. Appl.* **10** 651–675. MR4567809 https://doi.org/10.1146/annurev-statistics-033021-111803

The Annals of Statistics
2024, Vol. 52, No. 3, 1076–1101
https://doi.org/10.1214/24-AOS2383

SPECTRAL REGULARIZED KERNEL TWO-SAMPLE TESTS

BY OMAR HAGRASS[a], BHARATH SRIPERUMBUDUR[b] AND BING LI[c]

Department of Statistics, Pennsylvania State University, [a]*oih3@psu.edu,* [b]*bks18@psu.edu,* [c]*bxl9@psu.edu*

Over the last decade, an approach that has gained a lot of popularity to tackle nonparametric testing problems on general (i.e., non-Euclidean) domains is based on the notion of reproducing kernel Hilbert space (RKHS) embedding of probability distributions. The main goal of our work is to understand the optimality of two-sample tests constructed based on this approach. First, we show the popular MMD (maximum mean discrepancy) two-sample test to be not optimal in terms of the separation boundary measured in Hellinger distance. Second, we propose a modification to the MMD test based on spectral regularization by taking into account the covariance information (which is not captured by the MMD test) and prove the proposed test to be minimax optimal with a smaller separation boundary than that achieved by the MMD test. Third, we propose an adaptive version of the above test which involves a data-driven strategy to choose the regularization parameter and show the adaptive test to be almost minimax optimal up to a logarithmic factor. Moreover, our results hold for the permutation variant of the test where the test threshold is chosen elegantly through the permutation of the samples. Through numerical experiments on synthetic and real data, we demonstrate the superior performance of the proposed test in comparison to the MMD test and other popular tests in the literature.

REFERENCES

ADAMS, R. A. and FOURNIER, J. J. F. (2003). *Sobolev Spaces*, 2nd ed. *Pure and Applied Mathematics* (*Amsterdam*) **140**. Elsevier/Academic Press, Amsterdam. MR2424078

ALBERT, M., LAURENT, B., MARREL, A. and MEYNAOUI, A. (2022). Adaptive test of independence based on HSIC measures. *Ann*. *Statist*. **50** 858–879. MR4404921 https://doi.org/10.1214/21-aos2129

ARONSZAJN, N. (1950). Theory of reproducing kernels. *Trans*. *Amer*. *Math*. *Soc*. **68** 337–404. MR0051437 https://doi.org/10.2307/1990404

BALASUBRAMANIAN, K., LI, T. and YUAN, M. (2021). On the optimality of kernel-embedding based goodness-of-fit tests. *J*. *Mach*. *Learn*. *Res*. **22** Paper No. 1, 45. MR4253694

BAUER, F., PEREVERZEV, S. and ROSASCO, L. (2007). On regularization algorithms in learning theory. *J*. *Complexity* **23** 52–72. MR2297015 https://doi.org/10.1016/j.jco.2006.07.001

BURNAŠEV, M. V. (1979). Minimax detection of an imperfectly known signal against a background of Gaussian white noise. *Teor*. *Veroyatn*. *Primen*. **24** 106–118. MR0522240

CAPONNETTO, A. and DE VITO, E. (2007). Optimal rates for the regularized least-squares algorithm. *Found*. *Comput*. *Math*. **7** 331–368. MR2335249 https://doi.org/10.1007/s10208-006-0196-8

CUCKER, F. and ZHOU, D.-X. (2007). *Learning Theory*: *An Approximation Theory Viewpoint*. *Cambridge Monographs on Applied and Computational Mathematics* **24**. Cambridge Univ. Press, Cambridge. MR2354721 https://doi.org/10.1017/CBO9780511618796

DINCULEANU, N. (2000). *Vector Integration and Stochastic Integration in Banach Spaces*. *Pure and Applied Mathematics* (*New York*). Wiley Interscience, New York. MR1782432 https://doi.org/10.1002/9781118033012

DRINEAS, P. and MAHONEY, M. W. (2005). On the Nyström method for approximating a Gram matrix for improved kernel-based learning. *J*. *Mach*. *Learn*. *Res*. **6** 2153–2175. MR2249884

DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann*. *Math*. *Stat*. **27** 642–669. MR0083864 https://doi.org/10.1214/aoms/1177728174

MSC2020 subject classifications. Primary 62G10; secondary 65J20, 65J22, 46E22, 47A52.

Key words and phrases. Two-sample test, maximum mean discrepancy, reproducing kernel Hilbert space, permutation test, U-statistics, Bernstein's inequality, spectral regularization, adaptivity, covariance operator.

ENGL, H. W., HANKE, M. and NEUBAUER, A. (1996). *Regularization of Inverse Problems*. *Mathematics and Its Applications* **375**. Kluwer Academic, Dordrecht. MR1408680

FASANO, G. and FRANCESCHINI, A. (1987). A multidimensional version of the Kolmogorov–Smirnov test. *Mon. Not. R. Astron. Soc.* **225** 155–170.

GRETTON, A., BORGWARDT, K., RASCH, M., SCHÖLKOPF, B. and SMOLA, A. (2006). A kernel method for the two-sample problem. In *Advances in Neural Information Processing Systems* (B. Schölkopf, J. Platt and T. Hoffman, eds.) **19** 513–520. MIT Press, Cambridge.

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

GRETTON, A., FUKUMIZU, K., HARCHAOUI, Z. and SRIPERUMBUDUR, B. K. (2009). A fast, consistent kernel two-sample test. In *Advances in Neural Information Processing Systems* (Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams and A. Culotta, eds.) **22**. Curran Associates, Red Hook.

HAGRASS, O., SRIPERUMBUDUR, B. and LI, B. (2024). Supplement to "Spectral regularized kernel two-sample tests." https://doi.org/10.1214/24-AOS2383SUPP

HARCHAOUI, Z., BACH, F. R. and MOULINES, E. (2007). Testing for homogeneity with kernel Fisher discriminant analysis. In *Advances in Neural Information Processing Systems* (J. Platt, D. Koller, Y. Singer and S. Roweis, eds.) **20**. Curran Associates, Red Hook.

HOEFFDING, W. (1992). A class of statistics with asymptotically normal distribution. In *Breakthroughs in Statistics* 308–334.

INGSTER, YU. I. (1986). A minimax test of nonparametric hypotheses on the density of a distribution in L_p metrics. *Teor. Veroyatn. Primen.* **31** 384–389. MR0851000

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

INGSTER, YU. I. (2000). Adaptive chi-square tests. *J. Math. Sci.* **99** 1110–1119.

KIM, I., BALAKRISHNAN, S. and WASSERMAN, L. (2022). Minimax optimality of permutation tests. *Ann. Statist.* **50** 225–251. MR4382015 https://doi.org/10.1214/21-aos2103

LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. *Springer Series in Statistics*. Springer, New York. MR0856411 https://doi.org/10.1007/978-1-4612-4946-7

LECUN, Y., CORTES, C. and BURGES, C. (2010). MNIST handwritten digit database. AT &T Labs.

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

LI, T. and YUAN, M. (2019). On the optimality of Gaussian kernel based nonparametric tests against smooth alternatives.

MENDELSON, S. and NEEMAN, J. (2010). Regularization in kernel learning. *Ann. Statist.* **38** 526–565. MR2590050 https://doi.org/10.1214/09-AOS728

MINH, H. Q., NIYOGI, P. and YAO, Y. (2006). Mercer's theorem, feature maps, and smoothing. In *Learning Theory* (G. Lugosi and H. U. Simon, eds.). *Lecture Notes in Computer Science* **4005** 154–168. Springer, Berlin. MR2280604 https://doi.org/10.1007/11776420_14

MUANDET, K., FUKUMIZU, K., SRIPERUMBUDUR, B. and SCHÖLKOPF, B. (2017). Kernel mean embedding of distributions: A review and beyond. *Found. Trends Mach. Learn.* **10** 1–141.

PESARIN, F. and SALMASO, L. (2010). *Permutation Tests for Complex Data*: *Theory*, *Applications and Software*. Wiley, New York.

PURITZ, C., NESS-COHN, E. and BRAUN, R. (2022). fasano.franceschini.test: An Implementation of a Multidimensional KS Test in R.

RAHIMI, A. and RECHT, B. (2008). Random features for large-scale kernel machines. In *Advances in Neural Information Processing Systems* (J. C. Platt, D. Koller, Y. Singer and S. T. Roweis, eds.) **20** 1177–1184. Curran Associates, Red Hook.

REED, M. and SIMON, B. (1980). *Methods of Modern Mathematical Physics*. *I*: *Functional Analysis*, 2nd ed. Academic Press, New York. MR0751959

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

SCHRAB, A., KIM, I., ALBERT, M., LAURENT, B., GUEDJ, B. and GRETTON, A. (2023). MMD aggregated two-sample test. *J. Mach. Learn. Res.* **24** Paper No. [194], 81. MR4633583

SIMON-GABRIEL, C.-J. and SCHÖLKOPF, B. (2018). Kernel distribution embeddings: Universal kernels, characteristic kernels and kernel metrics on distributions. *J. Mach. Learn. Res.* **19** Paper No. 44, 29. MR3874152

SMOLA, A. J., GRETTON, A., SONG, L. and SCHÖLKOPF, B. (2007). A Hilbert space embedding for distributions. In *Algorithmic Learning Theory* (M. Hutter, R. A. Servedio and E. Takimoto, eds.) 13–31. Springer-Verlag, Berlin, Germany.

SRIPERUMBUDUR, B. (2016). On the optimal estimation of probability measures in weak and strong topologies. *Bernoulli* **22** 1839–1893. MR3474835 https://doi.org/10.3150/15-BEJ713

SRIPERUMBUDUR, B. K., FUKUMIZU, K., GRETTON, A., LANCKRIET, G. R. G. and SCHÖLKOPF, B. (2009). Kernel choice and classifiability for RKHS embeddings of probability distributions. In *Advances in Neural Information Processing Systems* (Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams and A. Culotta, eds.) **22** 1750–1758. MIT Press, Cambridge, MA.

SRIPERUMBUDUR, B. K., FUKUMIZU, K. and LANCKRIET, G. R. G. (2011). Universality, characteristic kernels and RKHS embedding of measures. *J. Mach. Learn. Res.* **12** 2389–2410. MR2825431

SRIPERUMBUDUR, B. K., GRETTON, A., FUKUMIZU, K., SCHÖLKOPF, B. and LANCKRIET, G. R. G. (2010). Hilbert space embeddings and metrics on probability measures. *J. Mach. Learn. Res.* **11** 1517–1561. MR2645460

SRIPERUMBUDUR, B. K. and STERGE, N. (2022). Approximate kernel PCA: Computational versus statistical trade-off. *Ann. Statist.* **50** 2713–2736. MR4500622 https://doi.org/10.1214/22-aos2204

STEINWART, I. and CHRISTMANN, A. (2008). *Support Vector Machines*. Springer, New York.

STEINWART, I., HUSH, D. and SCOVEL, C. (2006). An explicit description of the reproducing kernel Hilbert spaces of Gaussian RBF kernels. *IEEE Trans. Inf. Theory* **52** 4635–4643. MR2300845 https://doi.org/10.1109/TIT.2006.881713

STEINWART, I. and SCOVEL, C. (2012). Mercer's theorem on general domains: On the interaction between measures, kernels, and RKHSs. *Constr. Approx.* **35** 363–417. MR2914365 https://doi.org/10.1007/s00365-012-9153-3

SZEKELY, G. and RIZZO, M. (2004). Testing for equal distributions in high dimension. *InterStat* **5**.

WILLIAMS, C. K. I. and SEEGER, M. (2001). Using the Nyström method to speed up kernel machines. In *Advances in Neural Information Processing Systems* (V. Tresp, T. K. Leen and T. G. Diettrich, eds.) **13** 682–688. MIT Press, Cambridge, MA.

YANG, Y., PILANCI, M. and WAINWRIGHT, M. J. (2017). Randomized sketches for kernels: Fast and optimal nonparametric regression. *Ann. Statist.* **45** 991–1023. MR3662446 https://doi.org/10.1214/16-AOS1472

The Annals of Statistics
2024, Vol. 52, No. 3, 1102–1126
https://doi.org/10.1214/24-AOS2384

MARS VIA LASSO

BY DOHYEONG KI[1,a], BILLY FANG[2,c] AND ADITYANAND GUNTUBOYINA[1,b]

[1]*Department of Statistics, University of California, Berkeley,* [a]*dohyeong_ki@berkeley.edu,* [b]*aditya@stat.berkeley.edu*
[2]*Google LLC,* [c]*blfang@berkeley.edu*

Multivariate adaptive regression splines (MARS) is a popular method for nonparametric regression introduced by Friedman in 1991. MARS fits simple nonlinear and non-additive functions to regression data. We propose and study a natural lasso variant of the MARS method. Our method is based on least squares estimation over a convex class of functions obtained by considering infinite-dimensional linear combinations of functions in the MARS basis and imposing a variation based complexity constraint. Our estimator can be computed via finite-dimensional convex optimization, although it is defined as a solution to an infinite-dimensional optimization problem. Under a few standard design assumptions, we prove that our estimator achieves a rate of convergence that depends only logarithmically on dimension and thus avoids the usual curse of dimensionality to some extent. We also show that our method is naturally connected to nonparametric estimation techniques based on smoothness constraints. We implement our method with a cross-validation scheme for the selection of the involved tuning parameter and compare it to the usual MARS method in various simulation and real data settings.

REFERENCES

AISTLEITNER, C. and DICK, J. (2015). Functions of bounded variation, signed measures, and a general Koksma–Hlawka inequality. *Acta Arith.* **167** 143–171. MR3312093 https://doi.org/10.4064/aa167-2-4

ARTSTEIN, S., MILMAN, V., SZAREK, S. and TOMCZAK-JAEGERMANN, N. (2004). On convexified packing and entropy duality. *Geom. Funct. Anal.* **14** 1134–1141. MR2105957 https://doi.org/10.1007/s00039-004-0486-3

BLEI, R., GAO, F. and LI, W. V. (2007). Metric entropy of high dimensional distributions. *Proc. Amer. Math. Soc.* **135** 4009–4018. MR2341952 https://doi.org/10.1090/S0002-9939-07-08935-6

BREDIES, K. and PIKKARAINEN, H. K. (2013). Inverse problems in spaces of measures. *ESAIM Control Optim. Calc. Var.* **19** 190–218. MR3023066 https://doi.org/10.1051/cocv/2011205

BUNGARTZ, H.-J. and GRIEBEL, M. (2004). Sparse grids. *Acta Numer.* **13** 147–269. MR2249147 https://doi.org/10.1017/S0962492904000182

CANDÈS, E. J. and FERNANDEZ-GRANDA, C. (2014). Towards a mathematical theory of super-resolution. *Comm. Pure Appl. Math.* **67** 906–956. MR3193963 https://doi.org/10.1002/cpa.21455

CHEN, X. and LI, W. V. (2003). Quadratic functionals and small ball probabilities for the m-fold integrated Brownian motion. *Ann. Probab.* **31** 1052–1077. MR1964958 https://doi.org/10.1214/aop/1048516545

CONDAT, L. (2020). Atomic norm minimization for decomposition into complex exponentials and optimal transport in Fourier domain. *J. Approx. Theory* **258** 105456, 24. MR4127283 https://doi.org/10.1016/j.jat.2020.105456

DE CASTRO, Y. and GAMBOA, F. (2012). Exact reconstruction using Beurling minimal extrapolation. *J. Math. Anal. Appl.* **395** 336–354. MR2943626 https://doi.org/10.1016/j.jmaa.2012.05.011

DE CASTRO, Y., GAMBOA, F., HENRION, D. and LASSERRE, J.-B. (2017). Exact solutions to super resolution on semi-algebraic domains in higher dimensions. *IEEE Trans. Inf. Theory* **63** 621–630. MR3599963 https://doi.org/10.1109/TIT.2016.2619368

MSC2020 subject classifications. 62G08.

Key words and phrases. Bracketing entropy bounds, constrained least squares estimation, curse of dimensionality, Hardy–Krause variation, infinite-dimensional optimization, integrated Brownian sheet, locally adaptive regression spline, L1 penalty, metric entropy bounds, mixed derivatives, nonparametric regression, piecewise linear function estimation, small ball probability, tensor products, total variation regularization, trend filtering.

DENOYELLE, Q., DUVAL, V., PEYRÉ, G. and SOUBIES, E. (2020). The sliding Frank–Wolfe algorithm and its application to super-resolution microscopy. *Inverse Probl*. **36** 014001, 42. MR4040984 https://doi.org/10.1088/1361-6420/ab2a29

DŨNG, D., TEMLYAKOV, V. and ULLRICH, T. (2018). *Hyperbolic Cross Approximation*. *Advanced Courses in Mathematics*. *CRM Barcelona*. Birkhäuser, Cham. Edited and with a foreword by Sergey Tikhonov. MR3887571 https://doi.org/10.1007/978-3-319-92240-9

DUNKER, T., LINDE, W., KÜHN, T. and LIFSHITS, M. A. (1999). Metric entropy of integration operators and small ball probabilities for the Brownian sheet. *J. Approx. Theory* **101** 63–77. MR1724026 https://doi.org/10.1006/jath.1999.3354

DUVAL, V. and PEYRÉ, G. (2015). Exact support recovery for sparse spikes deconvolution. *Found. Comput. Math*. **15** 1315–1355. MR3394712 https://doi.org/10.1007/s10208-014-9228-6

FANG, B., GUNTUBOYINA, A. and SEN, B. (2021). Multivariate extensions of isotonic regression and total variation denoising via entire monotonicity and Hardy–Krause variation. *Ann. Statist*. **49** 769–792. MR4255107 https://doi.org/10.1214/20-aos1977

FRIEDMAN, J. H. (1991). Multivariate adaptive regression splines. *Ann. Statist*. **19** 1–67. MR1091842 https://doi.org/10.1214/aos/1176347963

FRIEDMAN, J. H. (1993). Fast MARS Technical Report No. LCS110, Stanford Univ. Press, Stanford.

GAO, F. (2008). Entropy estimate for k-monotone functions via small ball probability of integrated Brownian motion. *Electron. Commun. Probab*. **13** 121–130. MR2386068 https://doi.org/10.1214/ECP.v13-1357

GUNTUBOYINA, A., LIEU, D., CHATTERJEE, S. and SEN, B. (2020). Adaptive risk bounds in univariate total variation denoising and trend filtering. *Ann. Statist*. **48** 205–229. MR4065159 https://doi.org/10.1214/18-AOS1799

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

KI, D., FANG, B. and GUNTUBOYINA, A. (2024). Supplement to "MARS via LASSO." https://doi.org/10.1214/24-AOS2384SUPPA, https://doi.org/10.1214/24-AOS2384SUPPB

KIM, S.-J., KOH, K., BOYD, S. and GORINEVSKY, D. (2009). l_1 trend filtering. *SIAM Rev*. **51** 339–360. MR2505584 https://doi.org/10.1137/070690274

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

LI, W. V. and LINDE, W. (1999). Approximation, metric entropy and small ball estimates for Gaussian measures. *Ann. Probab*. **27** 1556–1578. MR1733160 https://doi.org/10.1214/aop/1022677459

LIN, Y. (1998). *Tensor Product Space ANOVA Models in Multivariate Function Estimation*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—University of Pennsylvania. MR2697355

LIN, Y. (2000). Tensor product space ANOVA models. *Ann. Statist*. **28** 734–755. MR1792785 https://doi.org/10.1214/aos/1015951996

MAMMEN, E. and VAN DE GEER, S. (1997). Locally adaptive regression splines. *Ann. Statist*. **25** 387–413. MR1429931 https://doi.org/10.1214/aos/1034276635

MEYER, D., LEISCH, F. and HORNIK, K. (2003). The support vector machine under test. *Neurocomputing* **55** 169–186.

NEMIROVSKI, A. S. (2000). Topics in nonparametric statistics. In *Lectures on Probability Theory and Statistics*: *École D'Été de Probabilités de Saint-Flour XXVIII*-1998. *Lecture Notes in Mathematics* **1738**. Springer, Berlin, Heidelberg.

ORTELLI, F. and VAN DE GEER, S. (2021). Tensor denoising with trend filtering. *Math. Stat. Learn*. **4** 87–142. MR4383732 https://doi.org/10.4171/msl/26

OWEN, A. B. (2005). Multidimensional variation for quasi-Monte Carlo. In *Contemporary Multivariate Analysis and Design of Experiments*. *Ser. Biostat*. **2** 49–74. World Scientific, Hackensack, NJ. MR2271076

PARHI, R. and NOWAK, R. D. (2021). Banach space representer theorems for neural networks and ridge splines. *J. Mach. Learn. Res*. **22** 1–40. MR4253736

PARHI, R. and NOWAK, R. D. (2023). Near-minimax optimal estimation with shallow ReLU neural networks. *IEEE Trans. Inf. Theory* **69** 1125–1140. MR4564646 https://doi.org/10.1109/tit.2022.3208653

PETERSEN, A., WITTEN, D. and SIMON, N. (2016). Fused lasso additive model. *J. Comput. Graph. Statist*. **25** 1005–1025. MR3572026 https://doi.org/10.1080/10618600.2015.1073155

POTTS, D. and SCHMISCHKE, M. (2021). Interpretable approximation of high-dimensional data. *SIAM J. Math. Data Sci*. **3** 1301–1323. MR4344888 https://doi.org/10.1137/21M1407707

POTTS, D. and SCHMISCHKE, M. (2022). Learning multivariate functions with low-dimensional structures using polynomial bases. *J. Comput. Appl. Math*. **403** Paper No. 113821, 19. MR4321625 https://doi.org/10.1016/j.cam.2021.113821

ROSSET, S., SWIRSZCZ, G., SREBRO, N. and ZHU, J. (2007). ℓ_1 regularization in infinite dimensional feature spaces. In *Learning Theory*. *Lecture Notes in Computer Science* **4539** 544–558. Springer, Berlin. MR2397611 https://doi.org/10.1007/978-3-540-72927-3_39

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

SADHANALA, V. and TIBSHIRANI, R. J. (2019). Additive models with trend filtering. *Ann*. *Statist*. **47** 3032–3068. MR4025734 https://doi.org/10.1214/19-AOS1833

SADHANALA, V., WANG, Y.-X., HU, A. J. and TIBSHIRANI, R. J. (2021). Multivariate trend filtering for lattice data. arXiv preprint. Available at arXiv:2112.14758.

STEIDL, G., DIDAS, S. and NEUMANN, J. (2006). Splines in higher order TV regularization. *Int*. *J*. *Comput*. *Vis*. **70** 241–255.

TEMLYAKOV, V. (2018). *Multivariate Approximation*. *Cambridge Monographs on Applied and Computational Mathematics* **32**. Cambridge Univ. Press, Cambridge. MR3837133 https://doi.org/10.1017/9781108689687

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

TIBSHIRANI, R. J. (2014). Adaptive piecewise polynomial estimation via trend filtering. *Ann*. *Statist*. **42** 285–323. MR3189487 https://doi.org/10.1214/13-AOS1189

VAN DER LAAN, M. J., BENKESER, D. and CAI, W. (2023). Efficient estimation of pathwise differentiable target parameters with the undersmoothed highly adaptive lasso. *Int*. *J*. *Biostat*. **19** 261–289.

YANG, T. and TAN, Z. (2018). Backfitting algorithms for total-variation and empirical-norm penalized additive modelling with high-dimensional data. *Stat* **7** e198, 19. MR3905854 https://doi.org/10.1002/sta4.198

YANG, T. and TAN, Z. (2021). Hierarchical total variations and doubly penalized ANOVA modeling for multivariate nonparametric regression. *J*. *Comput*. *Graph*. *Statist*. **30** 848–862. MR4356590 https://doi.org/10.1080/10618600.2021.1923513

The Annals of Statistics
2024, Vol. 52, No. 3, 1127–1151
https://doi.org/10.1214/24-AOS2386

SHARP ADAPTIVE AND PATHWISE STABLE SIMILARITY TESTING FOR SCALAR ERGODIC DIFFUSIONS

BY JOHANNES BRUTSCHE[a] AND ANGELIKA ROHDE[b]

Mathematical Institute, University of Freiburg, [a]*johannes.brutsche@stochastik.uni-freiburg.de,* [b]*angelika.rohde@stochastik.uni-freiburg.de*

Within the nonparametric diffusion model, we develop a multiple test to infer about *similarity* of an unknown drift b to some reference drift b_0: At prescribed significance, we simultaneously identify those regions where violation from similarity occurs, without a priori knowledge of their number, size and location. This test is shown to be minimax-optimal and adaptive. At the same time, the procedure is robust under small deviation from Brownian motion as the driving noise process. A detailed investigation for fractional driving noise, which is neither a semimartingale nor a Markov process, is provided for Hurst indices close to the Brownian motion case.

REFERENCES

[1] ALTMAN, C. G. and BLAND, J. M. (1995). Statistics notes: Absence of evidence is not evidence of absence. *Br. Med. J.* **311** 485.

[2] BERKSON, J. (1938). Some difficulties of interpretation encountered in the application of the chi-square test. *J. Amer. Statist. Assoc.* **33** 526–536.

[3] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York. MR0233396

[4] BRUTSCHE, J. and ROHDE, A. (2024). Supplement to "Sharp adaptive and pathwise stable similarity testing for scalar ergodic diffusions." https://doi.org/10.1214/24-AOS2386SUPP

[5] BÜCHER, A., DETTE, H. and HEINRICHS, F. (2021). Are deviations in a gradually varying mean relevant? A testing approach based on sup-norm estimators. *Ann. Statist.* **49** 3583–3617. MR4352542 https://doi.org/10.1214/21-aos2098

[6] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564–1597. MR3262461 https://doi.org/10.1214/14-AOS1230

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

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

[9] DATTA, P. and SEN, B. (2021). Optimal inference with a multidimensional multiscale statistic. *Electron. J. Stat.* **15** 5203–5244. MR4349258 https://doi.org/10.1214/21-ejs1914

[10] DETTE, H., KOKOT, K. and AUE, A. (2020). Functional data analysis in the Banach space of continuous functions. *Ann. Statist.* **48** 1168–1192. MR4102692 https://doi.org/10.1214/19-AOS1842

[11] DIEHL, J., FRIZ, P. and MAI, H. (2016). Pathwise stability of likelihood estimators for diffusions via rough paths. *Ann. Appl. Probab.* **26** 2169–2192. MR3543893 https://doi.org/10.1214/15-AAP1143

[12] DONOHO, D. L. (1994). Statistical estimation and optimal recovery. *Ann. Statist.* **22** 238–270. MR1272082 https://doi.org/10.1214/aos/1176325367

[13] DÜMBGEN, L., PITERBARG, V. I. and ZHOLUD, D. (2006). On the limit distribution of multiscale test statistics for nonparametric curve estimation. *Math. Methods Statist.* **15** 20–25. MR2225428

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

[15] DÜMBGEN, L. and WALTHER, G. (2008). Multiscale inference about a density. *Ann. Statist.* **36** 1758–1785. MR2435455 https://doi.org/10.1214/07-AOS521

[16] FOGARTY, C. B. and SMALL, D. S. (2014). Equivalence testing for functional data with an application to comparing pulmonary function devices. *Ann. Appl. Stat.* **8** 2002–2026. MR3292487 https://doi.org/10.1214/14-AOAS763

MSC2020 subject classifications. Primary 62G10, 62G20, 62M02; secondary 60G22.
Key words and phrases. Similarity, sharp adaptivity, empirical processes, diffusions.

[17] INGSTER, Y. I. (1982). Minimax nonparametric detection of signals in white Gaussian noise. *Probl. Inf. Transm.* **18** 130–140. MR0689340

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

[19] KALLENBERG, O. (2021). *Foundations of Modern Probability*, 3rd ed. *Probability Theory and Stochastic Modelling* **99**. Springer, Cham. MR4226142 https://doi.org/10.1007/978-3-030-61871-1

[20] KARATZAS, I. and SHREVE, S. E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed. *Graduate Texts in Mathematics* **113**. Springer, New York. MR1121940 https://doi.org/10.1007/978-1-4612-0949-2

[21] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131. MR0375412 https://doi.org/10.1007/BF00533093

[22] KÖNIG, C., MUNK, A. and WERNER, F. (2020). Multidimensional multiscale scanning in exponential families: Limit theory and statistical consequences. *Ann. Statist.* **48** 655–678. MR4102671 https://doi.org/10.1214/18-AOS1806

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

[24] LEONOV, S. L. (1999). Remarks on extremal problems in nonparametric curve estimation. *Statist. Probab. Lett.* **43** 169–178. MR1693289 https://doi.org/10.1016/S0167-7152(98)00256-9

[25] MCBRIDE, G. B. (1999). Equivalence tests can enhance environmental science and management. *Aust. N. Z. J. Stat.* **41** 19–29.

[26] NUALART, D. and OUKNINE, Y. (2002). Regularization of differential equations by fractional noise. *Stochastic Process. Appl.* **102** 103–116. MR1934157 https://doi.org/10.1016/S0304-4149(02)00155-2

[27] PROKSCH, K., WERNER, F. and MUNK, A. (2018). Multiscale scanning in inverse problems. *Ann. Statist.* **46** 3569–3602. MR3852662 https://doi.org/10.1214/17-AOS1669

[28] ROHDE, A. (2008). Adaptive goodness-of-fit tests based on signed ranks. *Ann. Statist.* **36** 1346–1374. MR2418660 https://doi.org/10.1214/009053607000000992

[29] ROHDE, A. (2011). Optimal calibration for multiple testing against local inhomogeneity in higher dimension. *Probab. Theory Related Fields* **149** 515–559. MR2776625 https://doi.org/10.1007/s00440-010-0263-1

[30] ROMANO, J. P. (2005). Optimal testing of equivalence hypotheses. *Ann. Statist.* **33** 1036–1047. MR2195627 https://doi.org/10.1214/009053605000000048

[31] SAMKO, S. G., KILBAS, A. A. and MARICHEV, O. I. (1993). *Fractional Integrals and Derivatives*. Gordon and Breach Science Publishers, Yverdon. MR1347689

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

[33] STRAUCH, C. (2018). Adaptive invariant density estimation for ergodic diffusions over anisotropic classes. *Ann. Statist.* **46** 3451–3480. MR3852658 https://doi.org/10.1214/17-AOS1664

[34] TUDOR, C. A. and VIENS, F. G. (2007). Statistical aspects of the fractional stochastic calculus. *Ann. Statist.* **35** 1183–1212. MR2341703 https://doi.org/10.1214/009053606000001541

[35] WELLEK, S. (2010). *Testing Statistical Hypotheses of Equivalence and Noninferiority*, 2nd ed. CRC Press, Boca Raton, FL. MR2676002 https://doi.org/10.1201/EBK1439808184

The Annals of Statistics
2024, Vol. 52, No. 3, 1152–1177
https://doi.org/10.1214/24-AOS2388

A BLOCKWISE EMPIRICAL LIKELIHOOD METHOD FOR TIME SERIES IN FREQUENCY DOMAIN INFERENCE

BY HAIHAN YU[1,a], MARK S. KAISER[2,b] AND DANIEL J. NORDMAN[2,c]

[1]*Department of Computer Science and Statistics, University of Rhode Island,* [a]*haihan.yu@uri.edu*
[2]*Department of Statistics, Iowa State University,* [b]*mskaiser@iastate.edu,* [c]*dnordman@iastate.edu*

Frequency domain analysis of time series is often difficult, as periodogram-based statistics involve non-linear averages with complicated variances. Due to the latter, nonparametric approximations from resampling or empirical likelihood (EL) are useful. However, current versions of periodogram-based EL for time series are highly restricted: these are valid only for linear processes and for special parameters (i.e., ratios). For general frequency domain inference with stationary, weakly dependent time series, we develop a spectral EL (SEL) method by combining two previously separate EL frameworks for time series: block-based EL and periodogram-based EL. This hybridization strategy is new and theoretically non-trivial, particularly as existing block-based EL relies on time domain averages that differ substantially from frequency domain counterparts. We formulate SEL statistics for parameters based on spectral estimating functions and periodogram subsamples. Under mild conditions, SEL log-ratio statistics are shown to be well-defined, admitting chi-square limits. Further, we formally establish an effective bootstrap procedure coupled with SEL. As a result, the SEL method can be used for nonparametric, asymptotically correct confidence regions and tests for frequency domain inference without explicit estimation of intricate variances of periodogram-based statistics. This broadly extends the applicability of EL for time series in three directions: (i) SEL can treat any spectral mean parameters; (ii) SEL is valid for both linear and non-linear processes; and (iii) SEL has a provable bootstrap development, which is rare for time series EL, and provides a novel alternative to other resampling approximations in the frequency domain. Simulation studies suggest the proposed method performs well compared to other non-EL approaches. A real data example demonstrates that SEL has application and extension to complicated scenarios.

REFERENCES

[1] AKASHI, F., LIU, Y. and TANIGUCHI, M. (2015). An empirical likelihood approach for symmetric α-stable processes. *Bernoulli* **21** 2093–2119.

[2] ANDERSON, T. W. (1993). Goodness of fit tests for spectral distributions. *Ann. Statist.* **21** 830–847. MR1232521 https://doi.org/10.1214/aos/1176349153

[3] ATHREYA, K. B. and LAHIRI, S. N. (2006). *Measure Theory and Probability Theory. Springer Texts in Statistics*. Springer, New York. MR2247694

[4] BAI, S. and TAQQU, M. S. (2017). On the validity of resampling methods under long memory. *Ann. Statist.* **45** 2365–2399. MR3737895 https://doi.org/10.1214/16-AOS1524

[5] BANDYOPADHYAY, S., LAHIRI, S. N. and NORDMAN, D. J. (2015). A frequency domain empirical likelihood method for irregularly spaced spatial data. *Ann. Statist.* **43** 519–545. MR3316189 https://doi.org/10.1214/14-AOS1291

[6] BRAVO, F. (2004). Empirical likelihood based inference with applications to some econometric models. *Econometric Theory* **20** 231–264. MR2044271 https://doi.org/10.1017/S0266466604202018

[7] BRAVO, F. (2005). Blockwise empirical entropy tests for time series regressions. *J. Time Series Anal.* **26** 185–210. MR2122895 https://doi.org/10.1111/j.1467-9892.2005.00398.x

---

MSC2020 subject classifications. Primary 62M15, 62G09; secondary 62M10.
Key words and phrases. Bootstrap, periodogram, resampling, spectral mean statistic, subsampling, Whittle estimation.

[8] BRAVO, F. (2009). Blockwise generalized empirical likelihood inference for non-linear dynamic moment conditions models. *Econom. J.* **12** 208–231. MR2562384 https://doi.org/10.1111/j.1368-423X.2009.00286.x

[9] BRILLINGER, D. R. (2001). *Time Series: Data Analysis and Theory. Classics in Applied Mathematics* **36**. SIAM, Philadelphia, PA. MR1853554 https://doi.org/10.1137/1.9780898719246

[10] BROCKWELL, P. J. and DAVIS, R. A. (1996). *Time Series: Theory and Methods*, 2nd ed. *Springer Series in Statistics*. Springer, New York.

[11] CHAN, N. H., CHEN, K. and YAU, C. Y. (2014). On the Bartlett correction of empirical likelihood for Gaussian long-memory time series. *Electron. J. Stat.* **8** 1460–1490. MR3545163

[12] CHEN, K. and HUANG, R. (2021). Robust empirical likelihood for time series. *J. Time Series Anal.* **42** 4–18. MR4192925 https://doi.org/10.1111/jtsa.12552

[13] DAHLHAUS, R. (1985). Asymptotic normality of spectral estimates. *J. Multivariate Anal.* **16** 412–431. MR0793500 https://doi.org/10.1016/0047-259X(85)90028-4

[14] DAHLHAUS, R. (1985). On the asymptotic distribution of Bartlett's U_p-statistic. *J. Time Series Anal.* **6** 213–227. MR0824665 https://doi.org/10.1111/j.1467-9892.1985.tb00411.x

[15] DAHLHAUS, R. and JANAS, D. (1996). A frequency domain bootstrap for ratio statistics in time series analysis. *Ann. Statist.* **24** 1934–1963. MR1421155 https://doi.org/10.1214/aos/1069362304

[16] DAHLHAUS, R. and WEFELMEYER, W. (1996). Asymptotically optimal estimation in misspecified time series models. *Ann. Statist.* **24** 952–974. MR1401832 https://doi.org/10.1214/aos/1032526951

[17] 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

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

[19] HALL, P. and LA SCALA, B. (1990). Methodology and algorithms of empirical likelihood. *Int. Stat. Rev.* **58** 109–127.

[20] HATHAWAY, D. H. (2015). The solar cycle. *Living Rev. Sol. Phys.* **12** 4.

[21] HJORT, N. L., MCKEAGUE, I. W. and VAN KEILEGOM, I. (2009). Extending the scope of empirical likelihood. *Ann. Statist.* **37** 1079–1111. MR2509068 https://doi.org/10.1214/07-AOS555

[22] HOSOYA, Y. (1997). A limit theory for long-range dependence and statistical inference on related models. *Ann. Statist.* **25** 105–137. MR1429919 https://doi.org/10.1214/aos/1034276623

[23] HOSOYA, Y. and TANIGUCHI, M. (1982). A central limit theorem for stationary processes and the parameter estimation of linear processes. *Ann. Statist.* **10** 132–153. MR0642725

[24] KAKIZAWA, Y. (2013). Frequency domain generalized empirical likelihood method. *J. Time Series Anal.* **34** 691–716. MR3127214 https://doi.org/10.1111/jtsa.12043

[25] KITAMURA, Y. (1997). Empirical likelihood methods with weakly dependent processes. *Ann. Statist.* **25** 2084–2102. MR1474084 https://doi.org/10.1214/aos/1069362388

[26] KREISS, J.-P. and PAPARODITIS, E. (2003). Autoregressive-aided periodogram bootstrap for time series. *Ann. Statist.* **31** 1923–1955. MR2036395 https://doi.org/10.1214/aos/1074290332

[27] KREISS, J.-P. and PAPARODITIS, E. (2012). The hybrid wild bootstrap for time series. *J. Amer. Statist. Assoc.* **107** 1073–1084. MR3010895 https://doi.org/10.1080/01621459.2012.695664

[28] KREISS, J.-P. and PAPARODITIS, E. (2023). Bootstrapping Whittle estimators. *Biometrika* **110** 499–518. MR4588361 https://doi.org/10.1093/biomet/asac044

[29] KREISS, J.-P., PAPARODITIS, E. and POLITIS, D. N. (2011). On the range of validity of the autoregressive sieve bootstrap. *Ann. Statist.* **39** 2103–2130. MR2893863 https://doi.org/10.1214/11-AOS900

[30] 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

[31] LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data. Springer Series in Statistics*. Springer, New York. MR2001447 https://doi.org/10.1007/978-1-4757-3803-2

[32] LAHIRI, S. N. (2007). Asymptotic expansions for sums of block-variables under weak dependence. *Ann. Statist.* **35** 1324–1350. MR2341707 https://doi.org/10.1214/009053607000000190

[33] MEYER, M., PAPARODITIS, E. and KREISS, J.-P. (2020). Extending the validity of frequency domain bootstrap methods to general stationary processes. *Ann. Statist.* **48** 2404–2427. MR4134800 https://doi.org/10.1214/19-AOS1892

[34] MONTI, A. C. (1997). Empirical likelihood confidence regions in time series models. *Biometrika* **84** 395–405. MR1467055 https://doi.org/10.1093/biomet/84.2.395

[35] NORDMAN, D. J. (2009). Tapered empirical likelihood for time series data in time and frequency domains. *Biometrika* **96** 119–132. MR2482139 https://doi.org/10.1093/biomet/asn071

[36] NORDMAN, D. J. and LAHIRI, S. N. (2006). A frequency domain empirical likelihood for short- and long-range dependence. *Ann. Statist.* **34** 3019–3050. MR2329476 https://doi.org/10.1214/009053606000000902

[37] NORDMAN, D. J. and LAHIRI, S. N. (2014). A review of empirical likelihood methods for time series. *J. Statist. Plann. Inference* **155** 1–18. MR3264452 https://doi.org/10.1016/j.jspi.2013.10.001

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

[39] OWEN, A. B. (2001). *Empirical Likelihood. Monographs on Statistics and Applied Probability* **92**. CRC Press/CRC, Boca Raton.

[40] OWEN, A. B. (2013). Self-concordance for empirical likelihood. *Canad. J. Statist.* **41** 387–397. MR3101590 https://doi.org/10.1002/cjs.11183

[41] PIYADI GAMAGE, R. D., NING, W. and GUPTA, A. K. (2017). Adjusted empirical likelihood for time series models. *Sankhya B* **79** 336–360. MR3749275 https://doi.org/10.1007/s13571-017-0137-y

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

[43] PRIESTLEY, M. B. (1981). *Spectral Analysis and Time Series. Probability and Mathematical Statistics: A Series of Monographs and Texbooks*. Academic Press, San Diego.

[44] QIN, J. and LAWLESS, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22** 300–325. MR1272085 https://doi.org/10.1214/aos/1176325370

[45] ROMANO, J. P. and THOMBS, L. A. (1996). Inference for autocorrelations under weak assumptions. *J. Amer. Statist. Assoc.* **91** 590–600. MR1395728 https://doi.org/10.2307/2291655

[46] SHAO, X. (2009). Confidence intervals for spectral mean and ratio statistics. *Biometrika* **96** 107–117. MR2482138 https://doi.org/10.1093/biomet/asn067

[47] TANIGUCHI, M. (1979). On estimation of parameters of Gaussian stationary processes. *J. Appl. Probab.* **16** 575–591. MR0540794 https://doi.org/10.2307/3213086

[48] TEWES, J., POLITIS, D. N. and NORDMAN, D. J. (2019). Convolved subsampling estimation with applications to block bootstrap. *Ann. Statist.* **47** 468–496. MR3909939 https://doi.org/10.1214/18-AOS1695

[49] WU, R. and CAO, J. (2011). Blockwise empirical likelihood for time series of counts. *J. Multivariate Anal.* **102** 661–673. MR2755022 https://doi.org/10.1016/j.jmva.2010.11.008

[50] YAU, C. Y. (2012). Empirical likelihood in long-memory time series models. *J. Time Series Anal.* **33** 269–275. MR2902463 https://doi.org/10.1111/j.1467-9892.2011.00756.x

[51] YU, H., KAISER, M. S. and NORDMAN, D. J. (2023). A subsampling perspective for extending the validity of state-of-the-art bootstraps in the frequency domain. *Biometrika* **110** 1099–1115. MR4667442 https://doi.org/10.1093/biomet/asad006

[52] YU, H., KAISER, M. S and NORDMAN, D. J (2024). Supplement to "A blockwise empirical likelihood method for time series in frequency domain inference." https://doi.org/10.1214/24-AOS2388SUPP

[53] ZHANG, R., PENG, L. and QI, Y. (2012). Jackknife-blockwise empirical likelihood methods under dependence. *J. Multivariate Anal.* **104** 56–72. MR2832186 https://doi.org/10.1016/j.jmva.2011.06.009

The Annals of Statistics
2024, Vol. 52, No. 3, 1178–1200
https://doi.org/10.1214/24-AOS2389
This research was funded, in whole or in part, by EPSRC, EP/V002694/1, EP/V002694/1, EP/W016117/1. A CC BY 4.0 license is applied to this article arising from this submission, in accordance with the grant's open access conditions

NONPARAMETRIC CLASSIFICATION WITH MISSING DATA

BY TORBEN SELL[1,a], THOMAS B. BERRETT[2,c] AND TIMOTHY I. CANNINGS[1,b]

[1]*School of Mathematics and Maxwell Institute for Mathematical Sciences, University of Edinburgh,* [a]*torben.sell@ed.ac.uk,* [b]*timothy.cannings@ed.ac.uk*

[2]*Department of Statistics, University of Warwick,* [c]*tom.berrett@warwick.ac.uk*

We introduce a new nonparametric framework for classification problems in the presence of missing data. The key aspect of our framework is that the regression function decomposes into an anova-type sum of orthogonal functions, of which some (or even many) may be zero. Working under a general missingness setting, which allows features to be missing not at random, our main goal is to derive the minimax rate for the excess risk in this problem. In addition to the decomposition property, the rate depends on parameters that control the tail behaviour of the marginal feature distributions, the smoothness of the regression function and a margin condition. The ambient data dimension does not appear in the minimax rate, which can therefore be faster than in the classical nonparametric setting. We further propose a new method, called the *Hard-thresholding Anova Missing data (HAM)* classifier, based on a careful combination of a k-nearest neighbour algorithm and a thresholding step. The HAM classifier attains the minimax rate up to polylogarithmic factors and numerical experiments further illustrate its utility.

REFERENCES

AHFOCK, D. and MCLACHLAN, G. J. (2023). Semi-supervised learning of classifiers from a statistical perspective: A brief review. *Econom. Stat.* **26** 124–138. MR4569121 https://doi.org/10.1016/j.ecosta.2022.03.007

AUDIBERT, J.-Y. and TSYBAKOV, A. B. (2007). Fast learning rates for plug-in classifiers. *Ann. Statist.* **35** 608–633. MR2336861 https://doi.org/10.1214/009053606000001217

AYME, A., BOYER, C., DIEULEVEUT, A. and SCORNET, E. (2022). Near-optimal rate of consistency for linear models with missing values. In *International Conference on Machine Learning* 1211–1243. PMLR, Cambridge.

BERRETT, T. B. and SAMWORTH, R. J. (2023). Optimal nonparametric testing of missing completely at random and its connections to compatibility. *Ann. Statist.* **51** 2170–2193. MR4678800 https://doi.org/10.1214/23-aos2326

BORDINO, A. and BERRETT, T. B. (2024). Tests of missing completely at random based on sample covariance matrices. arXiv preprint. Available at arXiv:2401.05256.

BOUCHERON, S., BOUSQUET, O. and LUGOSI, G. (2005). Theory of classification: A survey of some recent advances. *ESAIM Probab. Stat.* **9** 323–375. MR2182250 https://doi.org/10.1051/ps:2005018

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

CAI, T. T. and ZHANG, A. (2016). Minimax rate-optimal estimation of high-dimensional covariance matrices with incomplete data. *J. Multivariate Anal.* **150** 55–74. MR3534902 https://doi.org/10.1016/j.jmva.2016.05.002

CAI, T. T. and ZHANG, L. (2019). High dimensional linear discriminant analysis: Optimality, adaptive algorithm and missing data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 675–705. MR3997097

CANNINGS, T. I. and FAN, Y. (2022). The correlation-assisted missing data estimator. *J. Mach. Learn. Res.* **23** Paper No. [41], 49. MR4420766

CANNINGS, T. I., FAN, Y. and SAMWORTH, R. J. (2020). Classification with imperfect training labels. *Biometrika* **107** 311–330. MR4108933 https://doi.org/10.1093/biomet/asaa011

CHANDRASEKHER, K. A., ALAOUI, A. E. and MONTANARI, A. (2020). Imputation for high-dimensional linear regression. arXiv preprint. Available at arXiv:2001.09180.

MSC2020 subject classifications. 62D10, 62G05.
Key words and phrases. Missing data, classification, minimax.

CHAPELLE, O., SCHÖLKOPF, B. and ZIEN, A. (2006). *Semi-Supervised Learning*. The MIT Press, Cambridge, MA.

DEVROYE, L., GYÖRFI, L. and LUGOSI, G. (1996). *A Probabilistic Theory of Pattern Recognition*. *Applications of Mathematics* (*New York*) **31**. Springer, New York. MR1383093 https://doi.org/10.1007/978-1-4612-0711-5

EFRON, B. and STEIN, C. (1981). The jackknife estimate of variance. *Ann. Statist.* **9** 586–596. MR0615434

ELSENER, A. and VAN DE GEER, S. (2019). Sparse spectral estimation with missing and corrupted measurements. *Stat* **8** e229, 11. MR3978409 https://doi.org/10.1002/sta4.229

ELTER, M. (2007). Mammographic mass. UCI Machine Learning Repository. https://doi.org/10.24432/C53K6Z

FIX, E. and HODGES, J. L. (1952). Discriminatory analysis-nonparametric discrimination: Small sample performance. Technical report number 4, USAF School of Aviation Medicine, Randolph Field, Texas.

FIX, E. and HODGES, J. L. (1989). Discriminatory analysis-nonparametric discrimination: Small sample performance. *Int. Stat. Rev.* **57** 238–247.

FOLLAIN, B., WANG, T. and SAMWORTH, R. J. (2022). High-dimensional changepoint estimation with heterogeneous missingness. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 1023–1055. MR4460584

FRÉNAY, B., KABÁN, A. et al. (2014). A comprehensive introduction to label noise. In *ESANN*. Citeseer.

FRÉNAY, B. and VERLEYSEN, M. (2013). Classification in the presence of label noise: A survey. *IEEE Trans. Neural Netw. Learn. Syst.* **25** 845–869.

HASTIE, T. and TIBSHIRANI, R. (1986). Generalized additive models. *Statist. Sci.* **1** 297–318. With discussion. MR0858512

JOSSE, J., PROST, N., SCORNET, E. and VAROQUAUX, G. (2019). On the consistency of supervised learning with missing values. arXiv preprint. Available at arXiv:1902.06931.

JOSSE, J. and REITER, J. P. (2018). Introduction to the special section on missing data. *Statist. Sci.* **33** 139–141. MR3797706 https://doi.org/10.1214/18-STS332IN

LE MORVAN, M., JOSSE, J., MOREAU, T., SCORNET, E. and VAROQUAUX, G. (2020). Neumiss networks: Differentiable programming for supervised learning with missing values. *Adv. Neural Inf. Process. Syst.* **33** 5980–5990.

LE MORVAN, M., JOSSE, J., SCORNET, E. and VAROQUAUX, G. (2021). What's a good imputation to predict with missing values? *Adv. Neural Inf. Process. Syst.* **34** 11530–11540.

LEE, Y. and FOYGEL BARBER, R. (2022). Binary classification with corrupted labels. *Electron. J. Stat.* **16** 1367–1392. MR4387845 https://doi.org/10.1214/22-ejs1987

LITTLE, R. J. A. and RUBIN, D. B. (2002). *Statistical Analysis with Missing Data*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley-Interscience, Hoboken, NJ. MR1925014 https://doi.org/10.1002/9781119013563

LOH, P.-L. and TAN, X. L. (2018). High-dimensional robust precision matrix estimation: Cellwise corruption under ϵ-contamination. *Electron. J. Stat.* **12** 1429–1467. MR3804842 https://doi.org/10.1214/18-EJS1427

LOH, P.-L. and WAINWRIGHT, M. J. (2012). High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *Ann. Statist.* **40** 1637–1664. MR3015038 https://doi.org/10.1214/12-AOS1018

MAMMEN, E. and TSYBAKOV, A. B. (1999). Smooth discrimination analysis. *Ann. Statist.* **27** 1808–1829. MR1765618 https://doi.org/10.1214/aos/1017939240

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

REEVE, H. W. J., CANNINGS, T. I. and SAMWORTH, R. J. (2021). Adaptive transfer learning. *Ann. Statist.* **49** 3618–3649. MR4352543 https://doi.org/10.1214/21-aos2102

SELL, T., BERRETT, T. B. and CANNINGS, T. I. (2024). Supplement to "Nonparametric classification with missing data." https://doi.org/10.1214/24-AOS2389SUPP

SPORTISSE, A., SCHMUTZ, H., HUMBERT, O., BOUVEYRON, C. and MATTEI, P.-A. (2023). Are labels informative in semi-supervised learning?–estimating and leveraging the missing-data mechanism. arXiv preprint. Available at arXiv:2302.07540.

STEKHOVEN, D. J. and BÜHLMANN, P. (2012). Missforest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* **28** 112–118.

WEISS, K., KHOSHGOFTAAR, T. M. and WANG, D. (2016). A survey of transfer learning. *J. Big Data* **3** 1–40.

ZHANG, Q., YUAN, Q., ZENG, C., LI, X. and WEI, Y. (2018). Missing data reconstruction in remote sensing image with a unified spatial–temporal–spectral deep convolutional neural network. *IEEE Trans. Geosci. Remote Sens.* **56** 4274–4288.

ZHU, Z., WANG, T. and SAMWORTH, R. J. (2022). High-dimensional principal component analysis with heterogeneous missingness. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 2000–2031. MR4515564

The Annals of Statistics
2024, Vol. 52, No. 3, 1201–1226
https://doi.org/10.1214/24-AOS2390

DEEP NONLINEAR SUFFICIENT DIMENSION REDUCTION

BY YINFENG CHEN[1,a], YULING JIAO[2,d], RUI QIU[1,b] AND ZHOU YU[1,c]

[1]*School of Statistics, KLATASDS-MOE, East China Normal University,* [a]*52194404001@stu.ecnu.edu.cn,* [b]*rqiu_stat@outlook.com,* [c]*zyu@stat.ecnu.edu.cn*

[2]*School of Mathematics and Statistics, Hubei Key Laboratory of Computational Science, Wuhan University,* [d]*yulingjiaomath@whu.edu.cn*

Linear sufficient dimension reduction, as exemplified by sliced inverse regression, has seen substantial development in the past thirty years. However, with the advent of more complex scenarios, nonlinear dimension reduction has gained considerable interest recently. This paper introduces a novel method for nonlinear sufficient dimension reduction, utilizing the generalized martingale difference divergence measure in conjunction with deep neural networks. The optimal solution of the proposed objective function is shown to be unbiased at the general level of σ-fields. And two optimization schemes, based on the fascinating deep neural networks, exhibit higher efficiency and flexibility compared to the classical eigendecomposition of linear operators. Moreover, we systematically investigate the slow rate and fast rate for the estimation error based on advanced U-process theory. Remarkably, the fast rate almost coincides with the minimax rate of nonparametric regression. The validity of our deep nonlinear sufficient dimension reduction methods is demonstrated through simulations and real data analysis.

REFERENCES

[1] BANIJAMALI, E., KARIMI, A.-H. and GHODSI, A. (2018). Deep variational sufficient dimensionality reduction. Preprint. Available at arXiv:1812.07641.

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

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

[4] BÖTTCHER, B., KELLER-RESSEL, M. and SCHILLING, R. L. (2018). Detecting independence of random vectors: Generalized distance covariance and Gaussian covariance. *Mod. Stoch. Theory Appl.* **5** 353–383. MR3868546 https://doi.org/10.15559/18-vmsta116

[5] BURA, E. and COOK, R. D. (2001). Estimating the structural dimension of regressions via parametric inverse regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 393–410. MR1841422 https://doi.org/10.1111/1467-9868.00292

[6] CHEN, Y., JIAO, Y., QIU, R. and YU, Z. (2024). Supplement to "Deep nonlinear sufficient dimension reduction." https://doi.org/10.1214/24-AOS2390SUPP

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

[8] COOK, R. D. and WEISBERG, S. (1991). Discussion of sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.* **86** 328–332.

[9] DE LA PEÑA, V. H. and GINÉ, E. (1999). *Decoupling*: *From Dependence to Independence*. *Probability and Its Applications* (*New York*). Springer, New York. MR1666908 https://doi.org/10.1007/978-1-4612-0537-1

[10] DONG, Y. (2021). Sufficient dimension reduction through independence and conditional mean independence measures. In *Festschrift in Honor of R. Dennis Cook—Fifty Years of Contribution to Statistical Science* 167–180. Springer, Cham. MR4299383

MSC2020 subject classifications. 62M45, 60G25, 62G08, 62H12.

Key words and phrases. Sufficient dimension reduction, generalized martingale difference divergence, deep neural networks, U-process.

[11] FERRÉ, L. (1998). Determining the dimension in sliced inverse regression and related methods. *J. Amer. Statist. Assoc.* **93** 132–140. MR1614604 https://doi.org/10.2307/2669610

[12] FERTL, L. and BURA, E. (2022). Conditional variance estimator for sufficient dimension reduction. *Bernoulli* **28** 1862–1891. MR4411514 https://doi.org/10.3150/21-bej1402

[13] GHOSH, T. and KIRBY, M. (2022). Supervised dimensionality reduction and visualization using centroid-encoder. *J. Mach. Learn. Res.* **23** 1–34. MR4420745

[14] GRETTON, A., BOUSQUET, O., SMOLA, A. and SCHÖLKOPF, B. (2005). Measuring statistical dependence with Hilbert–Schmidt norms. In *Algorithmic Learning Theory. Lecture Notes in Computer Science* **3734** 63–77. Springer, Berlin. MR2255909 https://doi.org/10.1007/11564089_7

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

[16] HINTON, G. E. and SALAKHUTDINOV, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science* **313** 504–507. MR2242509 https://doi.org/10.1126/science.1127647

[17] HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* **19** 293–325. MR0026294 https://doi.org/10.1214/aoms/1177730196

[18] HSING, T. and REN, H. (2009). An RKHS formulation of the inverse regression dimension-reduction problem. *Ann. Statist.* **37** 726–755. MR2502649 https://doi.org/10.1214/07-AOS589

[19] HUANG, G., LIU, Z., VAN DER MAATEN, L. and WEINBERGER, K. Q. (2017). Densely connected convolutional networks. In 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*) 2261–2269.

[20] HUANG, J., JIAO, Y., LIAO, X., LIU, J. and YU, Z. (2024). Deep dimension reduction for supervised representation learning. *IEEE Trans. Inf. Theory* **70** 3583–3598. MR4740798 https://doi.org/10.1109/tit.2023.3340658

[21] KAPLA, D., FERTL, L. and BURA, E. (2022). Fusing sufficient dimension reduction with neural networks. *Comput. Statist. Data Anal.* **168** Paper No. 107390, 20. MR4343643 https://doi.org/10.1016/j.csda.2021.107390

[22] KE, C. and YIN, X. (2020). Expected conditional characteristic function-based measures for testing independence. *J. Amer. Statist. Assoc.* **115** 985–996. MR4107694 https://doi.org/10.1080/01621459.2019.1604364

[23] LAI, T., ZHANG, Z. and WANG, Y. (2021). A kernel-based measure for conditional mean dependence. *Comput. Statist. Data Anal.* **160** Paper No. 107246, 22. MR4242933 https://doi.org/10.1016/j.csda.2021.107246

[24] LE, L., PATTERSON, A. and WHITE, M. (2018). Supervised autoencoders: Improving generalization performance with unsupervised regularizers. *Adv. Neural Inf. Process. Syst.* **31** 107–117.

[25] LECUN, Y., BOTTOU, L., BENGIO, Y. and HAFFNER, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* **86** 2278–2324. https://doi.org/10.1109/5.726791

[26] LEE, C. E. and SHAO, X. (2018). Martingale difference divergence matrix and its application to dimension reduction for stationary multivariate time series. *J. Amer. Statist. Assoc.* **113** 216–229. MR3803459 https://doi.org/10.1080/01621459.2016.1240083

[27] LEE, K.-Y., LI, B. and CHIAROMONTE, F. (2013). A general theory for nonlinear sufficient dimension reduction: Formulation and estimation. *Ann. Statist.* **41** 221–249. MR3059416 https://doi.org/10.1214/12-AOS1071

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

[29] LI, B. and DONG, Y. (2009). Dimension reduction for nonelliptically distributed predictors. *Ann. Statist.* **37** 1272–1298. MR2509074 https://doi.org/10.1214/08-AOS598

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

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

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

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

[34] LI, K.-C. and DUAN, N. (1989). Regression analysis under link violation. *Ann. Statist.* **17** 1009–1052. MR1015136 https://doi.org/10.1214/aos/1176347254

[35] LI, L., KE, C., YIN, X. and YU, Z. (2023). Generalized martingale difference divergence: Detecting conditional mean independence with applications in variable screening. *Comput. Statist. Data Anal.* **180** Paper No. 107618, 26. MR4533821 https://doi.org/10.1016/j.csda.2022.107618

[36] LOYAL, J. D., ZHU, R., CUI, Y. and ZHANG, X. (2022). Dimension reduction forests: Local variable importance using structured random forests. *J. Comput. Graph. Statist.* **31** 1104–1113. MR4513373 https://doi.org/10.1080/10618600.2022.2069777

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

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

[39] MASSART, P. (2007). *Concentration Inequalities and Model Selection. Lecture Notes in Math.* **1896**. Springer, Berlin. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, with a foreword by Jean Picard. MR2319879

[40] 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

[41] SCHOTT, J. R. (1994). Determining the dimensionality in sliced inverse regression. *J. Amer. Statist. Assoc.* **89** 141–148. MR1266291

[42] SETODJI, C. M. and COOK, R. D. (2004). K-means inverse regression. *Technometrics* **46** 421–429. MR2101510 https://doi.org/10.1198/004017004000000437

[43] SHAO, X. and ZHANG, J. (2014). Martingale difference correlation and its use in high-dimensional variable screening. *J. Amer. Statist. Assoc.* **109** 1302–1318. MR3265698 https://doi.org/10.1080/01621459.2014.887012

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

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

[46] SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. MR2382665 https://doi.org/10.1214/009053607000000505

[47] 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

[48] WAINWRIGHT, M. J. (2019). *High-Dimensional Statistics: A Non-asymptotic Viewpoint. Cambridge Series in Statistical and Probabilistic Mathematics* **48**. Cambridge Univ. Press, Cambridge. MR3967104 https://doi.org/10.1017/9781108627771

[49] WU, H.-M. (2008). Kernel sliced inverse regression with applications to classification. *J. Comput. Graph. Statist.* **17** 590–610. MR2528238 https://doi.org/10.1198/106186008X345161

[50] WU, Q., LIANG, F. and MUKHERJEE, S. (2008). Regularized sliced inverse regression for kernel models Technical Report, Duke Univ. Press, Durham NC.

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

[52] YEH, Y.-R., HUANG, S.-Y. and LEE, Y.-J. (2009). Nonlinear dimension reduction with kernel sliced inverse regression. *IEEE Trans. Knowl. Data Eng.* **21** 1590–1603.

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

[54] YIN, X. and YUAN, Q. (2020). A new class of measures for testing independence. *Statist. Sinica* **30** 2131–2154. MR4260758 https://doi.org/10.5705/ss.20

[55] YING, C. and YU, Z. (2022). Fréchet sufficient dimension reduction for random objects. *Biometrika* **109** 975–992. MR4519111 https://doi.org/10.1093/biomet/asac012

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

The Annals of Statistics
2024, Vol. 52, No. 3, 1227–1253
https://doi.org/10.1214/24-AOS2391

LOCALLY SIMULTANEOUS INFERENCE

BY TIJANA ZRNIC[1,a] AND WILLIAM FITHIAN[2,b]

[1]*Department of Electrical Engineering and Computer Sciences, University of California, Berkeley,* [a]*tijana.zrnic@berkeley.edu*
[2]*Department of Statistics, University of California, Berkeley,* [b]*wfithian@berkeley.edu*

Selective inference is the problem of giving valid answers to statistical questions chosen in a data-driven manner. A standard solution to selective inference is *simultaneous inference*, which delivers valid answers to the set of all questions that could possibly have been asked. However, simultaneous inference can be unnecessarily conservative if this set includes many questions that were unlikely to be asked in the first place. We introduce a less conservative solution to selective inference that we call *locally simultaneous inference*, which only answers those questions that could *plausibly* have been asked in light of the observed data, all the while preserving rigorous type I error guarantees. For example, if the objective is to construct a confidence interval for the "winning" treatment effect in a clinical trial with multiple treatments, and it is obvious in hindsight that only one treatment had a chance to win, then our approach will return an interval that is nearly the same as the uncorrected, standard interval. Locally simultaneous inference is implemented by refining any method for simultaneous inference of interest. Under mild conditions satisfied by common confidence intervals, locally simultaneous inference *strictly dominates* its underlying simultaneous inference method, meaning it can never yield less statistical power but only more. Compared to conditional selective inference, which demands stronger guarantees, locally simultaneous inference is more easily applicable in nonparametric settings and is more numerically stable.

REFERENCES

[1] ANDREWS, I., KITAGAWA, T. and MCCLOSKEY, A. (2024). Inference on winners. *Q. J. Econ.* **139** 305–358.

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

[3] BENJAMINI, Y., HECHTLINGER, Y. and STARK, P. B. (2019). Confidence intervals for selected parameters. arXiv preprint. Available at arXiv:1906.00505.

[4] BENNETT, G. (1962). Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.

[5] BENTKUS, V. (2004). On Hoeffding's inequalities. *Ann. Probab.* **32** 1650–1673. MR2060313 https://doi.org/10.1214/009117904000000360

[6] BERK, R., BROWN, L., BUJA, A., ZHANG, K. and ZHAO, L. (2013). Valid post-selection inference. *Ann. Statist.* **41** 802–837. MR3099122 https://doi.org/10.1214/12-AOS1077

[7] BERNSTEIN, S. (1924). On a modification of Chebyshev's inequality and of the error formula of Laplace. *Ann. Sci. Inst. Sav. Ukr., Sect. Math.* **1** 38–49.

[8] BRETZ, F., GENZ, A. and HOTHORN, L. A. (2001). On the numerical availability of multiple comparison procedures. *Biom. J.* **43** 645–656. MR1863491 https://doi.org/10.1002/1521-4036(200109)43:5<645::AID-BIMJ645>3.0.CO;2-F

[9] DICKHAUS, T. (2014). *Simultaneous Statistical Inference: With Applications in the Life Sciences*. Springer, Heidelberg. MR3184277 https://doi.org/10.1007/978-3-642-45182-9

[10] EL GHAOUI, L., VIALLON, V. and RABBANI, T. (2012). Safe feature elimination in sparse supervised learning. *Pac. J. Optim.* **8** 667–698. MR3026449

[11] FITHIAN, W., SUN, D. and TAYLOR, J. (2014). Optimal inference after model selection. arXiv preprint. Available at arXiv:1410.2597.

MSC2020 subject classifications. Primary 62J15; secondary 62F25, 62G15.
Key words and phrases. Selective inference, post-selection inference, simultaneous inference.

[12] FUENTES, C., CASELLA, G. and WELLS, M. T. (2018). Confidence intervals for the means of the selected populations. *Electron. J. Stat.* **12** 58–79. MR3743737 https://doi.org/10.1214/17-EJS1374

[13] GENZ, A. (1992). Numerical computation of multivariate normal probabilities. *J. Comput. Graph. Statist.* **1** 141–149.

[14] GENZ, A. and BRETZ, F. (1999). Numerical computation of multivariate t-probabilities with application to power calculation of multiple contrasts. *J. Stat. Comput. Simul.* **63** 361–378. MR1718625 https://doi.org/10.1080/00949659908811962

[15] GOEMAN, J. J. and SOLARI, A. (2024). On selection and conditioning in multiple testing and selective inference. *Biometrika* **111** 393–416. MR4745573 https://doi.org/10.1093/biomet/asad078

[16] HOCHBERG, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* **75** 800–802. MR0995126 https://doi.org/10.1093/biomet/75.4.800

[17] HOLM, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6** 65–70. MR0538597

[18] HOMMEL, G. (1990). A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika* **75** 383–386.

[19] HOTHORN, T., BRETZ, F. and WESTFALL, P. (2008). Simultaneous inference in general parametric models. *Biom. J.* **50** 346–363. MR2521547 https://doi.org/10.1002/bimj.200810425

[20] KIVARANOVIC, D. and LEEB, H. (2021). On the length of post-model-selection confidence intervals conditional on polyhedral constraints. *J. Amer. Statist. Assoc.* **116** 845–857. MR4270029 https://doi.org/10.1080/01621459.2020.1732989

[21] KOLTCHINSKII, V. (2006). Local Rademacher complexities and oracle inequalities in risk minimization. *Ann. Statist.* **34** 2593–2656. MR2329442 https://doi.org/10.1214/009053606000001019

[22] KOLTCHINSKII, V. (2011). *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems: École D'Été de Probabilités de Saint-Flour XXXVIII-2008. Lecture Notes in Math.* **2033**. Springer, Heidelberg. Lectures from the 38th Probability Summer School held in Saint-Flour [Saint-Flour Probability Summer School]. MR2829871 https://doi.org/10.1007/978-3-642-22147-7

[23] KOLTCHINSKII, V. and PANCHENKO, D. (2000). Rademacher processes and bounding the risk of function learning. In *High Dimensional Probability, II (Seattle, WA*, 1999). *Progress in Probability* **47** 443–457. Birkhäuser, Boston, MA. MR1857339

[24] 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

[25] LEEB, H. and PÖTSCHER, B. M. (2017). Testing in the presence of nuisance parameters: Some comments on tests post-model-selection and random critical values. In *Big and Complex Data Analysis. Contrib. Stat.* 69–82. Springer, Cham. MR3644121

[26] MCCLOSKEY, A. (2024). Hybrid confidence intervals for informative uniform asymptotic inference after model selection. *Biometrika* **111** 109–127. MR4704561 https://doi.org/10.1093/biomet/asad023

[27] RASP, S., DUEBEN, P. D., SCHER, S., WEYN, J. A., MOUATADID, S. and THUEREY, N. (2020). WeatherBench: A benchmark data set for data-driven weather forecasting. *J. Adv. Model. Earth Syst.* **12** e2020MS002203.

[28] ROMANO, J. P., SHAIKH, A. M. and WOLF, M. (2014). A practical two-step method for testing moment inequalities. *Econometrica* **82** 1979–2002. MR3268401 https://doi.org/10.3982/ECTA11011

[29] STRASSBURGER, K. and BRETZ, F. (2008). Compatible simultaneous lower confidence bounds for the Holm procedure and other Bonferroni-based closed tests. *Stat. Med.* **27** 4914–4927. MR2528773 https://doi.org/10.1002/sim.3338

[30] TIAN, X. and TAYLOR, J. (2018). Selective inference with a randomized response. *Ann. Statist.* **46** 679–710. MR3782381 https://doi.org/10.1214/17-AOS1564

[31] TIBSHIRANI, R., BIEN, J., FRIEDMAN, J., HASTIE, T., SIMON, N., TAYLOR, J. and TIBSHIRANI, R. J. (2012). Strong rules for discarding predictors in lasso-type problems. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 245–266. MR2899862 https://doi.org/10.1111/j.1467-9868.2011.01004.x

[32] TIBSHIRANI, R. J., TAYLOR, J., LOCKHART, R. and TIBSHIRANI, R. (2016). Exact post-selection inference for sequential regression procedures. *J. Amer. Statist. Assoc.* **111** 600–620. MR3538689 https://doi.org/10.1080/01621459.2015.1108848

[33] VENTER, J. (1988). Confidence bounds based on the largest treatment mean. *South Afr. J. Sci.* **84** 340.

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

[35] 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

[36] ZRNIC, T. and FITHIAN, W. (2024). Supplement to "Locally simultaneous inference." https://doi.org/10.1214/24-AOS2391SUPP

[37] ZRNIC, T. and JORDAN, M. I. (2023). Post-selection inference via algorithmic stability. *Ann. Statist.* **51** 1666–1691. MR4658572 https://doi.org/10.1214/23-aos2303

The Annals of Statistics
2024, Vol. 52, No. 3, 1254–1275
https://doi.org/10.1214/24-AOS2392

SPECTRAL ANALYSIS OF GRAM MATRICES WITH MISSING AT RANDOM OBSERVATIONS: CONVERGENCE, CENTRAL LIMIT THEOREMS, AND APPLICATIONS IN STATISTICAL INFERENCE

BY HUIQIN LI[1,a], GUANGMING PAN[2,c], YANQING YIN[1,b] AND WANG ZHOU[3,d]

[1]*School of Mathematics and Statistics, Chongqing University,* [a]*lihq118@nenu.edu.cn,* [b]*yinyq@cqu.edu.cn*
[2]*Division of Mathematical Sciences, Nanyang Technological University,* [c]*gmpan@ntu.edu.sg*
[3]*Department of Statistics and Data Science, National University of Singapore,* [d]*wangzhou@nus.edu.sg*

Motivated by the statistical inference using the Gram matrix in the context of missing at random observations, this paper investigates the spectral properties of the random matrices $\mathbf{S}_n = \frac{1}{n}\mathbf{Z}\mathbf{Z}^*$, where $\mathbf{Z} = \mathbf{D} \circ (\mathbf{\Sigma}^{1/2}\mathbf{X})$ represents a Hadamard random matrix with entries determined by independent Bernoulli variables $\mathbf{D}$. Operating within the high-dimensional framework, we establish the convergence of the empirical spectral distribution of $\mathbf{S}_n$ to a well-defined limiting distribution. In addition, we explore the impact of the missing mechanism on the second-order properties of the spectral distribution of the Gram matrix $\mathbf{S}_n$. We establish the central limit theorem for the linear spectral statistics of $\mathbf{S}_n$, shedding light on their fluctuations. Surprisingly, our analysis reveals that even in the ideal Gaussian distribution scenario, the fluctuations of statistics generated by eigenvalues are influenced by the eigenvectors of the population covariance matrix in the missing-at-random case. This discovery uncovers a remarkable phenomenon that starkly contrasts with the classical case. Subsequently, we demonstrate the practical application of our central limit theorem in hypothesis testing for the population covariance matrix.

REFERENCES

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

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

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

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

[5] CAI, T. T. and ZHANG, L. (2019). High dimensional linear discriminant analysis: Optimality, adaptive algorithm and missing data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 675–705. MR3997097

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

[7] CHEN, Y., FAN, J., MA, C. and YAN, Y. (2021). Bridging convex and nonconvex optimization in robust PCA: Noise, outliers and missing data. *Ann. Statist.* **49** 2948–2971. MR4338899 https://doi.org/10.1214/21-aos2066

[8] EL KAROUI, N. (2009). Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond. *Ann. Appl. Probab.* **19** 2362–2405. MR2588248 https://doi.org/10.1214/08-AAP548

[9] FAN, J., GUO, J. and ZHENG, S. (2022). Estimating number of factors by adjusted eigenvalues thresholding. *J. Amer. Statist. Assoc.* **117** 852–861. MR4436317 https://doi.org/10.1080/01621459.2020.1825448

MSC2020 subject classifications. Primary 62H15, 62B20; secondary 62D10.
Key words and phrases. Sample covariance matrix, missing observations, high-dimensionality, limiting spectral distribution, central limit theorem, random matrix theory.

[10] HU, J., LI, W., LIU, Z. and ZHOU, W. (2019). High-dimensional covariance matrices in elliptical distributions with application to spherical test. *Ann. Statist.* **47** 527–555. MR3909941 https://doi.org/10.1214/18-AOS1699

[11] JURCZAK, K. and ROHDE, A. (2017). Spectral analysis of high-dimensional sample covariance matrices with missing observations. *Bernoulli* **23** 2466–2532. MR3648036 https://doi.org/10.3150/16-BEJ815

[12] LI, H., PAN, G., YIN, Y. and ZHOU, W. (2024). Supplement to "Spectral analysis of gram matrices with missing at random observations: Convergence, central limit theorems, and applications in statistical inference." https://doi.org/10.1214/24-AOS2392SUPP

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

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

[15] PAN, G. M. and ZHOU, W. (2008). Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. *Ann. Appl. Probab.* **18** 1232–1270. MR2418244 https://doi.org/10.1214/07-AAP477

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

[17] VERSHYNIN, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. *Cambridge Series in Statistical and Probabilistic Mathematics* **47**. Cambridge Univ. Press, Cambridge. MR3837109 https://doi.org/10.1017/9781108231596

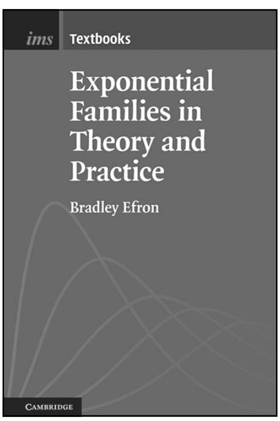
[18] WACHTER, K. W. (1978). The strong limits of random matrix spectra for sample matrices of independent elements. *Ann. Probab.* **6** 1–18. MR0467894 https://doi.org/10.1214/aop/1176995607

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

NEW BOOK

The Institute of Mathematical Statistics presents

IMS TEXTBOOKS



Exponential Families in Theory and Practice

Bradley Efron, Stanford University

During the past half-century, exponential families have attained a position at the center of parametric statistical inference. Theoretical advances have been matched, and more than matched, in the world of applications, where logistic regression by itself has become the go-to methodology in medical statistics, computer-based prediction algorithms, and the social sciences. This book is based on a one-semester graduate course for first year Ph.D. and advanced master's students. After presenting the basic structure of univariate and multivariate exponential families, their application to generalized linear models including logistic and Poisson regression is described in detail, emphasizing geometrical ideas, computational practice, and the analogy with ordinary linear regression. Connections are made with a variety of current statistical methodologies: missing data, survival analysis and proportional hazards, false discovery rates, bootstrapping, and empirical Bayes analysis. The book connects exponential family theory with its applications in a way that doesn't require advanced mathematical preparation.

Hardback $105.00

Paperback $39.99

IMS members are entitled to a 40% discount: email ims@imstat.org to request your code

www.imstat.org/cup/

Cambridge University Press, with the Institute of Mathematical Statistics, established the *IMS Monographs* and *IMS Textbooks* series of high-quality books. The series editors are Mark Handcock, Ramon van Handel, Arnaud Doucet, and John Aston.