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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

Vol. 51, No. 4—August 2023

THE ANNALS OF STATISTICS Vol. 51, No. 4, pp. 1453–1894 August 2023

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

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

IMS OFFICERS

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

President-Elect: Tony Cai, Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, PA 19104-6304, USA

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

Executive Secretary: Peter Hoff, Department of Statistical Science, Duke University, Durham, NC 27708-0251, USA

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

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

IMS EDITORS

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

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

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

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

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

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

***The Annals of Statistics* [ISSN 0090-5364 (print); ISSN 2168-8966 (online)],** Volume 51, Number 4, August 2023. Published bimonthly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, OH 44094, USA. Periodicals postage paid at Cleveland, Ohio, and at additional mailing offices.

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

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

The Annals of Statistics
2023, Vol. 51, No. 4, 1453–1483
https://doi.org/10.1214/22-AOS2205
© Institute of Mathematical Statistics, 2023

GRAPHICAL MODELS FOR NONSTATIONARY TIME SERIES

BY SUMANTA BASU[1,a] AND SUHASINI SUBBA RAO[2,b]

[1]*Department of Statistics and Data Science, Cornell University,* [a]*sumbose@cornell.edu*
[2]*Department of Statistics, Texas A&M University,* [b]*suhasini@stat.tamu.edu*

We propose NonStGM, a general nonparametric graphical modeling framework, for studying dynamic associations among the components of a nonstationary multivariate time series. It builds on the framework of Gaussian graphical models (GGM) and stationary time series graphical models (StGM) and complements existing works on parametric graphical models based on change point vector autoregressions (VAR). Analogous to StGM, the proposed framework captures conditional noncorrelations (both intertemporal and contemporaneous) in the form of an undirected graph. In addition, to describe the more nuanced nonstationary relationships among the components of the time series, we introduce the new notion of conditional nonstationarity/stationarity and incorporate it within the graph. This can be used to search for small subnetworks that serve as the "source" of nonstationarity in a large system.

We explicitly connect conditional noncorrelation and stationarity between and within components of the multivariate time series to zero and Toeplitz embeddings of an infinite-dimensional inverse covariance operator. In the Fourier domain, conditional stationarity and noncorrelation relationships in the inverse covariance operator are encoded with a specific sparsity structure of its integral kernel operator. We show that these sparsity patterns can be recovered from finite-length time series by nodewise regression of discrete Fourier transforms (DFT) across different Fourier frequencies. We demonstrate the feasibility of learning NonStGM structure from data using simulation studies.

REFERENCES

[1] ASTON, J., DEHAY, D., DUDEK, A., FREYERMUTH, J.-M., SZUCS, D. and COLLING, L. (2019). (dual-frequency)-dependent dynamic functional connectivity analysis of vidual working memory capacity. Preprint: Hal-021335535.

[2] AUE, A. and VAN DELFT, A. (2020). Testing for stationarity of functional time series in the frequency domain. *Ann. Statist.* **48** 2505–2547. MR4152111 https://doi.org/10.1214/19-AOS1895

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

[4] BASU, S. and SUBBA RAO, S. (2023). Supplement to "Graphical models for nonstationary time series." https://doi.org/10.1214/22-AOS2205SUPP

[5] BIRR, S., DETTE, H., HALLIN, M., KLEY, T. and VOLGUSHEV, S. (2018). On Wigner–Ville spectra and the uniqueness of time-varying copula-based spectral densities. *J. Time Series Anal.* **39** 242–250. MR3796518 https://doi.org/10.1111/jtsa.12252

[6] BÖHM, H. and VON SACHS, R. (2009). Shrinkage estimation in the frequency domain of multivariate time series. *J. Multivariate Anal.* **100** 913–935. MR2498723 https://doi.org/10.1016/j.jmva.2008.09.009

[7] BRILLINGER, D. R. (1996). Remarks concerning graphical models for time series and point processes. *Braz. Rev. Econom.* **16** 1–23.

[8] BRILLINGER, D. R. (2001). *Time Series*: *Data Analysis and Theory*. *Classics in Applied Mathematics* **36**. SIAM, Philadelphia, PA. Reprint of the 1981 edition. MR1853554 https://doi.org/10.1137/1.9780898719246

[9] BROCKWELL, P. J. and DAVIS, R. A. (2006). *Time Series*: *Theory and Methods*. *Springer Series in Statistics*. Springer, New York. Reprint of the second (1991) edition. MR2839251

MSC2020 subject classifications. Primary 62M10, 62M15; secondary 62J07.
Key words and phrases. Graphical models, locally stationary time series, partial covariance, spectral analysis.

[10] CARDINALI, A. and NASON, G. P. (2010). Costationarity of locally stationary time series. *J. Time Ser. Econom.* **2** Art. 1, 33. MR2915633 https://doi.org/10.2202/1941-1928.1074

[11] CHAU, J. and VON SACHS, R. (2021). Intrinsic wavelet regression for curves of Hermitian positive definite matrices. *J. Amer. Statist. Assoc.* **116** 819–832. MR4270027 https://doi.org/10.1080/01621459.2019.1700129

[12] CONWAY, J. B. (1990). *A Course in Functional Analysis*, 2nd ed. *Graduate Texts in Mathematics* **96**. Springer, New York. MR1070713

[13] DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37. MR1429916 https://doi.org/10.1214/aos/1034276620

[14] DAHLHAUS, R. (2000). Graphical interaction models for multivariate time series. *Metrika* **51** 157–172. MR1790930 https://doi.org/10.1007/s001840000055

[15] DAHLHAUS, R. (2000). A likelihood approximation for locally stationary processes. *Ann. Statist.* **28** 1762–1794. MR1835040 https://doi.org/10.1214/aos/1015957480

[16] DAHLHAUS, R. (2012). Locally stationary processes. In *Handbook of Statistics* **30** 351–413. Elsevier, Amsterdam.

[17] DAHLHAUS, R. and EICHLER, M. (2003). Causality and graphical models in time series analysis. In *Highly Structured Stochastic Systems*. *Oxford Statist. Sci. Ser.* **27** 115–144. Oxford Univ. Press, Oxford. MR2082408

[18] DAHLHAUS, R., EICHLER, M. and SANDÜHLER, J. (1997). *J. Neurosci. Methods* **77** 93–107.

[19] DAHLHAUS, R. and GIRAITIS, L. (1998). On the optimal segment length for parameter estimates for locally stationary time series. *J. Time Series Anal.* **19** 629–655. MR1665941 https://doi.org/10.1111/1467-9892.00114

[20] DAHLHAUS, R. and POLONIK, W. (2006). Nonparametric quasi-maximum likelihood estimation for Gaussian locally stationary processes. *Ann. Statist.* **34** 2790–2824. MR2329468 https://doi.org/10.1214/009053606000000867

[21] 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

[22] DAHLHAUS, R. and SUBBA RAO, S. (2006). Statistical inference for time-varying ARCH processes. *Ann. Statist.* **34** 1075–1114. MR2278352 https://doi.org/10.1214/009053606000000227

[23] DAVIS, R. A., ZANG, P. and ZHENG, T. (2016). Sparse vector autoregressive modeling. *J. Comput. Graph. Statist.* **25** 1077–1096. MR3572029 https://doi.org/10.1080/10618600.2015.1092978

[24] DEMPSTER, A. P. (1972). Covariance selection. *Biometrics* **28** 157–175. MR3931974

[25] DIEBOLD, F. X. and YILMAZ, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *J. Econometrics* **182** 119–134. MR3212765 https://doi.org/10.1016/j.jeconom.2014.04.012

[26] DING, X. and ZHOU, Z. (2020). Estimation and inference for precision matrices of nonstationary time series. *Ann. Statist.* **48** 2455–2477. MR4134802 https://doi.org/10.1214/19-AOS1894

[27] DWIVEDI, Y. and SUBBA RAO, S. (2011). A test for second-order stationarity of a time series based on the discrete Fourier transform. *J. Time Series Anal.* **32** 68–91. MR2790673 https://doi.org/10.1111/j.1467-9892.2010.00685.x

[28] EICHLER, M. (2007). Granger causality and path diagrams for multivariate time series. *J. Econometrics* **137** 334–353. MR2354948 https://doi.org/10.1016/j.jeconom.2005.06.032

[29] EICHLER, M. (2008). Testing nonparametric and semiparametric hypotheses in vector stationary processes. *J. Multivariate Anal.* **99** 968–1009. MR2405101 https://doi.org/10.1016/j.jmva.2007.06.003

[30] EPHARTY, A., TABRIKIAN, J. and MESSER, H. (2001). Underwater source detection using a spatial stationary test. *J. Acoust. Soc. Am.* **109** 1053–1063.

[31] FIECAS, M., LENG, C., LIU, W. and YU, Y. (2019). Spectral analysis of high-dimensional time series. *Electron. J. Stat.* **13** 4079–4101. MR4017528 https://doi.org/10.1214/19-EJS1621

[32] FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441. https://doi.org/10.1093/biostatistics/kxm045

[33] GLADYŠEV, E. G. (1963). Periodically and semi-periodically correlated random processes with continuous time. *Teor. Veroyatn. Primen.* **8** 184–189. MR0152005

[34] GORROSTIETA, C., OMBAO, H. and VON SACHS, R. (2019). Time-dependent dual-frequency coherence in multivariate non-stationary time series. *J. Time Series Anal.* **40** 3–22. MR3891477 https://doi.org/10.1111/jtsa.12408

[35] HINDBERG, H. and OLHEDE, S. C. (2010). Estimation of ambiguity functions with limited spread. *IEEE Trans. Signal Process.* **58** 2383–2388. MR2752078 https://doi.org/10.1109/TSP.2009.2037663

[36] JENSEN, O. and COLGIN, L. L. (2007). Cross-frequency coupling between neuronal oscillations. *Trends Cogn. Sci.* **11** 267–269. https://doi.org/10.1016/j.tics.2007.05.003

[37] JENTSCH, C. and SUBBA RAO, S. (2015). A test for second order stationarity of a multivariate time series. *J. Econometrics* **185** 124–161. MR3300340 https://doi.org/10.1016/j.jeconom.2014.09.010

[38] JUNG, A., HANNAK, G. and GOERTZ, N. (2015). Graphical lasso based model selection for time series. *IEEE Signal Process. Lett.* **22** 1781–1785.

[39] KLEY, T., PREUSS, P. and FRYZLEWICZ, P. (2019). Predictive, finite-sample model choice for time series under stationarity and non-stationarity. *Electron. J. Stat.* **13** 3710–3774. MR4013749 https://doi.org/10.1214/19-ejs1606

[40] KRAMPE, J. and SUBBA RAO, S. (2022). Inverse covariance operators of nonstationary multivariate time series. Available at arXiv:2202.00933.

[41] LII, K.-S. and ROSENBLATT, M. (2002). Spectral analysis for harmonizable processes. *Ann. Statist.* **30** 258–297. MR1892664 https://doi.org/10.1214/aos/1015362193

[42] LOUBATON, P., ROSUEL, A. and VALLET, P. (2023). On the asymptotic distribution of the maximum sample spectral coherence of Gaussian time series in the high dimensional regime. *J. Multivariate Anal.* **194** Paper No. 105124, 22. MR4516398 https://doi.org/10.1016/j.jmva.2022.105124

[43] LUND, R., HURD, H., BLOOMFIELD, P. and SMITH, R. (1995). Climatological time series with periodic correlation. *J. Climate* **11** 2787–2809.

[44] LURIE, D. J., KESSLER, D., BASSETT, D. S., BETZEL, R. F., BREAKSPEAR, M., KHEILHOLZ, S., KUCYI, A., LIÉGEOIS, R., LINDQUIST, M. A. et al. (2020). Questions and controversies in the study of time-varying functional connectivity in resting fMRI. *Netw. Neurosci.* **4** 30–69.

[45] MALEKI, A., ANITORI, L., YANG, Z. and BARANIUK, R. G. (2013). Asymptotic analysis of complex LASSO via complex approximate message passing (CAMP). *IEEE Trans. Inf. Theory* **59** 4290–4308. MR3071330 https://doi.org/10.1109/TIT.2013.2252232

[46] MARTIN, W. and FLANDRIN, P. (1985). Wigner–Ville spectral analysis of nonstationary processes. *IEEE Trans. Acoust. Speech Signal Process.* **33** 1461–1470.

[47] MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. MR2278363 https://doi.org/10.1214/009053606000000281

[48] NASON, G. (2013). A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 879–904. MR3124795 https://doi.org/10.1111/rssb.12015

[49] OLHEDE, S. C. (2011). Ambiguity sparse processes. arXiv preprint. Available at arXiv:1103.3932v2.

[50] OLHEDE, S. C. and OMBAO, H. (2013). Modeling and estimation of covariance of replicated modulated cyclical time series. *IEEE Trans. Signal Process.* **61** 1944–1957. MR3038192 https://doi.org/10.1109/TSP.2012.2237168

[51] OMBAO, H. and PINTO, M. (2021). Spectral dependence. Available at arXiv:2103.17240.

[52] PAPARODITIS, E. (2009). Testing temporal constancy of the spectral structure of a time series. *Bernoulli* **15** 1190–1221. MR2597589 https://doi.org/10.3150/08-BEJ179

[53] PRETI, M. G., BOLTON, T. A. and VILLE, D. V. D. (2017). The dynamic functional connectome: State-of-the-art and perspectives. *NeuroImage* **160** 41–54. https://doi.org/10.1016/j.neuroimage.2016.12.061

[54] PRIESTLEY, M. B. (1965). Evolutionary spectra and non-stationary processes. *J. Roy. Statist. Soc. Ser. B* **27** 204–237. MR0199886

[55] PRIESTLEY, M. B. (1981). *Spectral Analysis and Time Series: Multivariate Series, Prediction and Control, Vol. 2. Probability and Mathematical Statistics*. Academic Press, London. MR0628736

[56] PRIESTLEY, M. B. and SUBBA RAO, T. (1969). A test for non-stationarity of time-series. *J. Roy. Statist. Soc. Ser. B* **31** 140–149. MR0269062

[57] QIU, H., HAN, F., LIU, H. and CAFFO, B. (2016). Joint estimation of multiple graphical models from high dimensional time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 487–504. MR3454206 https://doi.org/10.1111/rssb.12123

[58] SAFIKHANI, A. and SHOJAIE, A. (2022). Joint structural break detection and parameter estimation in high-dimensional nonstationary VAR models. *J. Amer. Statist. Assoc.* **117** 251–264. MR4399083 https://doi.org/10.1080/01621459.2020.1770097

[59] SUBBA RAO, S. (2006). On some nonstationary, nonlinear random processes and their stationary approximations. *Adv. in Appl. Probab.* **38** 1155–1172. MR2285698 https://doi.org/10.1239/aap/1165414596

[60] SUBBA RAO, T. (1970). The fitting of non-stationary time-series models with time-dependent parameters. *J. Roy. Statist. Soc. Ser. B* **32** 312–322. MR0269065

[61] SUN, Y., LI, Y., KUCEYESKI, A. and BASU, S. (2018). Large spectral density matrix estimation by thresholding. arXiv preprint. Available at arXiv:1812.00532.

[62] SUNDARARAJAN, R. R. and POURAHMADI, M. (2018). Stationary subspace analysis of nonstationary processes. *J. Time Series Anal.* **39** 338–355. MR3796523 https://doi.org/10.1111/jtsa.12274

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

[64] TOEPLITZ, O. (1911). Zur Theorie der quadratischen und bilinearen Formen von unendlichvielen Veränderlichen. *Math*. *Ann*. **70** 351–376. MR1511625 https://doi.org/10.1007/BF01564502

[65] WANG, D., YU, Y., RINALDO, A. and WILLETT, R. (2019). Localizing changes in high-dimensional vector autoregressive processes. arXiv preprint. Available at arXiv:1909.06359.

[66] WOODROOFE, M. B. and VAN NESS, J. W. (1967). The maximum deviation of sample spectral densities. *Ann*. *Math*. *Stat*. **38** 1558–1569. MR0216717 https://doi.org/10.1214/aoms/1177698710

[67] WU, W. B. and ZAFFARONI, P. (2018). Asymptotic theory for spectral density estimates of general multivariate time series. *Econometric Theory* **34** 1–22. MR3739234 https://doi.org/10.1017/S0266466617000068

[68] 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

[69] ZHANG, D. and WU, W. B. (2021). Convergence of covariance and spectral density estimates for high-dimensional locally stationary processes. *Ann*. *Statist*. **49** 233–254. MR4206676 https://doi.org/10.1214/20-AOS1954

[70] ZHOU, Z. and WU, W. B. (2009). Local linear quantile estimation for nonstationary time series. *Ann*. *Statist*. **37** 2696–2729. MR2541444 https://doi.org/10.1214/08-AOS636

The Annals of Statistics
2023, Vol. 51, No. 4, 1484–1509
https://doi.org/10.1214/23-AOS2277
© Institute of Mathematical Statistics, 2023

BOOTSTRAPPING PERSISTENT BETTI NUMBERS AND OTHER STABILIZING STATISTICS

BY BENJAMIN ROYCRAFT[1,a], JOHANNES KREBS[2,c] AND WOLFGANG POLONIK[1,b]

[1]*Department of Statistics, University of California, Davis,* [a]*btroycraft@ucdavis.edu,* [b]*wpolonik@ucdavis.edu*
[2]*Department of Mathematics, KU Eichstätt-Ingolstadt,* [c]*johannes.krebs@ku.de*

We investigate multivariate bootstrap procedures for general stabilizing statistics, with specific application to topological data analysis. The work relates to other general results in the area of stabilizing statistics, including central limit theorems for geometric and topological functionals of Poisson and binomial processes in the critical regime, where limit theorems prove difficult to use in practice, motivating the use of a bootstrap approach. A smoothed bootstrap procedure is shown to give consistent estimation in these settings. Specific statistics considered include the persistent Betti numbers of Čech and Vietoris–Rips complexes over point sets in $\mathbb{R}^d$, along with Euler characteristics, and the total edge length of the k-nearest neighbor graph. Special emphasis is given to weakening the necessary conditions needed to establish bootstrap consistency. In particular, the assumption of a continuous underlying density is not required. Numerical studies illustrate the performance of the proposed method.

REFERENCES

[1] ADLER, R. J., AGAMI, S. and PRANAV, P. (2017). Modeling and replicating statistical topology and evidence for CMB nonhomogeneity. *Proc. Natl. Acad. Sci. USA* **114** 11878–11883. MR3725115 https://doi.org/10.1073/pnas.1706885114

[2] ARSUAGA, J., BORRMAN, T., CAVALCANTE, R., GONZALEZ, G. and PARK, C. (2015). Identification of copy number aberrations in breast cancer subtypes using persistence topology. *Microarrays* **4** 339–369.

[3] BISCIO, C. A. N., CHENAVIER, N., HIRSCH, C. and SVANE, A. M. (2020). Testing goodness of fit for point processes via topological data analysis. *Electron. J. Stat.* **14** 1024–1074. MR4067816 https://doi.org/10.1214/20-EJS1683

[4] BLANTON, M. R., BERSHADY, M. A., ABOLFATHI, B., ALBARETI, F. D., ALLENDE PRIETO, C., ALMEIDA, A., ALONSO-GARCÍA, J., ANDERS, F., ANDERSON, S. F. et al. (2017). Sloan digital sky survey IV: Mapping the milky way, nearby galaxies, and the distant universe. *Astron. J.* **154** 28.

[5] BOBROWSKI, O. and MUKHERJEE, S. (2015). The topology of probability distributions on manifolds. *Probab. Theory Related Fields* **161** 651–686. MR3334278 https://doi.org/10.1007/s00440-014-0556-x

[6] BOISSONNAT, J.-D., CHAZAL, F. and YVINEC, M. (2018). *Geometric and Topological Inference*. *Cambridge Texts in Applied Mathematics*. Cambridge Univ. Press, Cambridge. MR3837127 https://doi.org/10.1017/9781108297806

[7] BUBENIK, P. (2015). Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.* **16** 77–102. MR3317230

[8] BUBENIK, P. and KIM, P. T. (2007). A statistical approach to persistent homology. *Homology, Homotopy Appl.* **9** 337–362. MR2366953

[9] CAMARA, P. G., ROSENBLOOM, D. I. S., EMMETT, K. J., LEVINE, A. J. and RABADAN, R. (2016). Topological data analysis generates high-resolution, genome-wide maps of human recombination. *Cell Syst.* **3** 83–94. https://doi.org/10.1016/j.cels.2016.05.008

[10] CHAZAL, F. and DIVOL, V. (2019). The density of expected persistence diagrams and its kernel based estimation. *J. Comput. Geom.* **10** 127–153. MR4088069 https://doi.org/10.20382/jocg.v10i2a7

[11] CHAZAL, F., FASY, B., LECCI, F., MICHEL, B., RINALDO, A. and WASSERMAN, L. (2015). Subsampling methods for persistent homology. *Proc.* 32*nd Int. Conf. Mach. Learn.* **37** 2143–2151.

MSC2020 subject classifications. Primary 62F40; secondary 55N31, 62R40, 62H10, 62G05.
Key words and phrases. Betti numbers, bootstrap, Euler characteristic, random geometric complexes, stabilizing statistics, stochastic geometry, topological data analysis, persistent homology.

[12] CHAZAL, F., FASY, B. T., LECCI, F., RINALDO, A., SINGH, A. and WASSERMAN, L. (2015). On the bootstrap for persistence diagrams and landscapes. *Model. Anal. Inf. Syst.* **20** 111–120.

[13] CHAZAL, F., FASY, B. T., LECCI, F., RINALDO, A. and WASSERMAN, L. (2015). Stochastic convergence of persistence landscapes and silhouettes. *J. Comput. Geom.* **6** 140–161. MR3323391 https://doi.org/10.20382/jocg.v6i2a8

[14] CHAZAL, F. and MICHEL, B. (2021). An introduction to topological data analysis: Fundamental and practical aspects for data scientists. *Front. Artif. Intell.* **4** 667963. https://doi.org/10.3389/frai.2021.667963

[15] CHEN, Y.-C., WANG, D., RINALDO, A. and WASSERMAN, L. (2015). Statistical analysis of persistence intensity functions. ArXiv Preprint. Available at arXiv:1510.02502.

[16] CHUNG, Y.-M. and LAWSON, A. (2022). Persistence curves: A canonical framework for summarizing persistence diagrams. *Adv. Comput. Math.* **48** 6. MR4368950 https://doi.org/10.1007/s10444-021-09893-4

[17] CRAWFORD, L., MONOD, A., CHEN, A. X., MUKHERJEE, S. and RABADÁN, R. (2020). Predicting clinical outcomes in glioblastoma: An application of topological and functional data analysis. *J. Amer. Statist. Assoc.* **115** 1139–1150. MR4143455 https://doi.org/10.1080/01621459.2019.1671198

[18] DE SILVA, V. and GHRIST, R. (2007). Coverage in sensor networks via persistent homology. *Algebr. Geom. Topol.* **7** 339–358. MR2308949 https://doi.org/10.2140/agt.2007.7.339

[19] DEWOSKIN, D., CLIMENT, J., CRUZ-WHITE, I., VAZQUEZ, M., PARK, C. and ARSUAGA, J. (2010). Applications of computational homology to the analysis of treatment response in breast cancer patients. *Topology Appl.* **157** 157–164. MR2556091 https://doi.org/10.1016/j.topol.2009.04.036

[20] EDELSBRUNNER, LETSCHER and ZOMORODIAN (2002). Topological persistence and simplification. *Discrete Comput. Geom.* **28** 511–533.

[21] FASY, B. T., LECCI, F., RINALDO, A., WASSERMAN, L., BALAKRISHNAN, S. and SINGH, A. (2014). Confidence sets for persistence diagrams. *Ann. Statist.* **42** 2301–2339. MR3269981 https://doi.org/10.1214/14-AOS1252

[22] HIRAOKA, Y., SHIRAI, T. and TRINH, K. D. (2018). Limit theorems for persistence diagrams. *Ann. Appl. Probab.* **28** 2740–2780. MR3847972 https://doi.org/10.1214/17-AAP1371

[23] KIM, J., SHIN, J., RINALDO, A. and WASSERMAN, L. (2019). Uniform convergence rate of the kernel density estimator adaptive to intrinsic volume dimension. In *International Conference on Machine Learning* 3398–3407. PMLR.

[24] KRAMAR, M., GOULLET, A., KONDIC, L. and MISCHAIKOW, K. (2013). Persistence of force networks in compressed granular media. *Phys. Rev., E* **87**.

[25] KRAMÁR, M., LEVANGER, R., TITHOF, J., SURI, B., XU, M., PAUL, M., SCHATZ, M. F. and MISCHAIKOW, K. (2016). Analysis of Kolmogorov flow and Rayleigh-Bénard convection using persistent homology. *Phys. D* **334** 82–98. MR3545971 https://doi.org/10.1016/j.physd.2016.02.003

[26] KREBS, J., ROYCRAFT, B. and POLONIK, W. (2021). On approximation theorems for the Euler characteristic with applications to the bootstrap. *Electron. J. Stat.* **15** 4462–4509. MR4312855 https://doi.org/10.1214/21-ejs1898

[27] KREBS, J. T. and POLONIK, W. (2019). On the asymptotic normality of persistent Betti numbers. ArXiv Preprint. Available at arXiv:1903.03280.

[28] LACHIÈZE-REY, R., PECCATI, G. and YANG, X. (2022). Quantitative two-scale stabilization on the Poisson space. *Ann. Appl. Probab.* **32** 3085–3145. MR4474528 https://doi.org/10.1214/21-aap1768

[29] LACHIÈZE-REY, R., SCHULTE, M. and YUKICH, J. E. (2019). Normal approximation for stabilizing functionals. *Ann. Appl. Probab.* **29** 931–993. MR3910021 https://doi.org/10.1214/18-AAP1405

[30] LAST, G., PECCATI, G. and SCHULTE, M. (2016). Normal approximation on Poisson spaces: Mehler's formula, second order Poincaré inequalities and stabilization. *Probab. Theory Related Fields* **165** 667–723. MR3520016 https://doi.org/10.1007/s00440-015-0643-7

[31] OWADA, T. (2018). Limit theorems for Betti numbers of extreme sample clouds with application to persistence barcodes. *Ann. Appl. Probab.* **28** 2814–2854. MR3847974 https://doi.org/10.1214/17-AAP1375

[32] OWADA, T. and ADLER, R. J. (2017). Limit theorems for point processes under geometric constraints (and topological crackle). *Ann. Probab.* **45** 2004–2055. MR3650420 https://doi.org/10.1214/16-AOP1106

[33] PENROSE, M. D. and YUKICH, J. E. (2001). Central limit theorems for some graphs in computational geometry. *Ann. Appl. Probab.* **11** 1005–1041. MR1878288 https://doi.org/10.1214/aoap/1015345393

[34] PENROSE, M. D. and YUKICH, J. E. (2003). Weak laws of large numbers in geometric probability. *Ann. Appl. Probab.* **13** 277–303. MR1952000 https://doi.org/10.1214/aoap/1042765669

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

[36] PRANAV, P., ADLER, R. J., BUCHERT, T., EDELSBRUNNER, H., JONES, B. J. T., SCHWARTZMAN, A., WAGNER, H. and VAN DE WEYGAERT, R. (2019). Unexpected topology of the temperature fluctuations in the cosmic microwave background. *Astron. Astrophys.* **627** A163.

[37] PRANAV, P., EDELSBRUNNER, H., VAN DE WEYGAERT, R., VEGTER, G., KERBER, M., JONES, B. J. T. and WINTRAECKEN, M. (2016). The topology of the cosmic web in terms of persistent Betti numbers. *Mon. Not. R. Astron. Soc.* **465** 4281–4310.

[38] PRANAV, P., VAN DE WEYGAERT, R., VEGTER, G., JONES, B. J. T., ADLER, R. J., FELDBRUGGE, J., PARK, C., BUCHERT, T. and KERBER, M. (2019). Topology and geometry of Gaussian random fields I: On Betti numbers, Euler characteristic, and Minkowski functionals. *Mon. Not. R. Astron. Soc.* **485** 4167–4208.

[39] ROYCRAFT, B. (2021). github.com/btroycraft/stabilizing_statistics_bootstrap.

[40] ROYCRAFT, B., KREBS, J. and POLONIK, W. (2023). Supplement to "Bootstrapping persistent Betti numbers and other stabilizing statistics." https://doi.org/10.1214/23-AOS2277SUPP

[41] SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. *Monographs on Statistics and Applied Probability*. CRC Press, London. MR0848134 https://doi.org/10.1007/978-1-4899-3324-9

[42] TRINH, K. D. (2019). On central limit theorems in stochastic geometry for add-one cost stabilizing functionals. *Electron. Commun. Probab.* **24** 76. MR4049088 https://doi.org/10.1214/19-ecp279

[43] TURNER, K., MUKHERJEE, S. and BOYER, D. M. (2014). Persistent homology transform for modeling shapes and surfaces. *Inf. Inference* **3** 310–344. MR3311455 https://doi.org/10.1093/imaiai/iau011

[44] ULMER, M., ZIEGELMEIER, L. and TOPAZ, C. M. (2019). A topological approach to selecting models of biological experiments. *PLoS ONE* **14** 1–18.

[45] WASSERMAN, L. (2018). Topological data analysis. *Annu. Rev. Stat. Appl.* **5** 501–535. MR3774757 https://doi.org/10.1146/annurev-statistics-031017-100045

[46] XIA, K., FENG, X., TONG, Y. and WEI, G. W. (2014). Persistent homology for the quantitative prediction of fullerene stability. *J. Comput. Chem.* **36** 408–422.

[47] YOGESHWARAN, D. and ADLER, R. J. (2015). On the topology of random complexes built over stationary point processes. *Ann. Appl. Probab.* **25** 3338–3380. MR3404638 https://doi.org/10.1214/14-AAP1075

[48] YOGESHWARAN, D., SUBAG, E. and ADLER, R. J. (2017). Random geometric complexes in the thermodynamic regime. *Probab. Theory Related Fields* **167** 107–142. MR3602843 https://doi.org/10.1007/s00440-015-0678-9

[49] ZOMORODIAN, A. and CARLSSON, G. (2005). Computing persistent homology. *Discrete Comput. Geom.* **33** 249–274. MR2121296 https://doi.org/10.1007/s00454-004-1146-y

The Annals of Statistics
2023, Vol. 51, No. 4, 1510–1533
https://doi.org/10.1214/23-AOS2279
© Institute of Mathematical Statistics, 2023

ON LOWER BOUNDS FOR THE BIAS-VARIANCE TRADE-OFF

BY ALEXIS DERUMIGNY[1,a] AND JOHANNES SCHMIDT-HIEBER[2,b]

[1]*Department of Applied Mathematics, Delft University of Technology,* [a]*a.f.f.derumigny@tudelft.nl*
[2]*Department of Applied Mathematics, University of Twente,* [b]*a.j.schmidt-hieber@utwente.nl*

It is a common phenomenon that for high-dimensional and nonparametric statistical models, rate-optimal estimators balance squared bias and variance. Although this balancing is widely observed, little is known whether methods exist that could avoid the trade-off between bias and variance. We propose a general strategy to obtain lower bounds on the variance of any estimator with bias smaller than a prespecified bound. This shows to which extent the bias-variance trade-off is unavoidable and allows to quantify the loss of performance for methods that do not obey it. The approach is based on a number of abstract lower bounds for the variance involving the change of expectation with respect to different probability measures as well as information measures such as the Kullback–Leibler or χ^2-divergence. Some of these inequalities rely on a new concept of information matrices. In a second part of the article, the abstract lower bounds are applied to several statistical models including the Gaussian white noise model, a boundary estimation problem, the Gaussian sequence model and the high-dimensional linear regression model. For these specific statistical applications, different types of bias-variance trade-offs occur that vary considerably in their strength. For the trade-off between integrated squared bias and integrated variance in the Gaussian white noise model, we propose to combine the general strategy for lower bounds with a reduction technique. This allows us to reduce the original problem to a lower bound on the bias-variance trade-off for estimators with additional symmetry properties in a simpler statistical model. In the Gaussian sequence model, different phase transitions of the bias-variance trade-off occur. Although there is a non-trivial interplay between bias and variance, the rate of the squared bias and the variance do not have to be balanced in order to achieve the minimax estimation rate.

REFERENCES

[1] BELKIN, M., HSU, D., MA, S. and MANDAL, S. (2019). Reconciling modern machine-learning practice and the classical bias-variance trade-off. *Proc. Natl. Acad. Sci. USA* **116** 15849–15854. MR3997901 https://doi.org/10.1073/pnas.1903070116

[2] BERTHET, Q. and RIGOLLET, P. (2013). Complexity theoretic lower bounds for sparse principal component detection. In *Conference on Learning Theory*, *PMLR* 1046–1066.

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

[4] BROWN, L. D. and LOW, M. G. (1996). A constrained risk inequality with applications to nonparametric functional estimation. *Ann. Statist.* **24** 2524–2535. MR1425965 https://doi.org/10.1214/aos/1032181166

[5] 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

[6] BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data*. *Springer Series in Statistics*. Springer, Heidelberg. MR2807761 https://doi.org/10.1007/978-3-642-20192-9

[7] CAI, T. T. and GUO, Z. (2017). Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *Ann. Statist.* **45** 615–646. MR3650395 https://doi.org/10.1214/16-AOS1461

MSC2020 subject classifications. 62G05, 62C05, 62C20.
Key words and phrases. Bias-variance decomposition, Cramér–Rao inequality, high-dimensional statistics, minimax estimation, nonparametric estimation.

[8] CHAFAÏ, D. (2004). Entropies, convexity, and functional inequalities: On Φ-entropies and Φ-Sobolev inequalities. *J. Math. Kyoto Univ.* **44** 325–363. MR2081075 https://doi.org/10.1215/kjm/1250283556

[9] CHEN, J. (2004). Notes on the bias-variance trade-off phenomenon. In *A Festschrift for Herman Rubin. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **45** 207–217. IMS, Beachwood, OH. MR2126898 https://doi.org/10.1214/lnms/1196285391

[10] CHEN, X., GUNTUBOYINA, A. and ZHANG, Y. (2016). On Bayes risk lower bounds. *J. Mach. Learn. Res.* **17** 219. MR3595153

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

[12] DERUMIGNY, A. and SCHMIDT-HIEBER, J. (2023). Codivergences and information matrices. ArXiv E-prints. Available at arXiv:2303.08122.

[13] DERUMIGNY, A. and SCHMIDT-HIEBER, J. (2023). Supplement to "On lower bounds for the bias-variance trade-off." https://doi.org/10.1214/23-AOS2279SUPP

[14] DONOHO, D. L., ELAD, M. and TEMLYAKOV, V. N. (2006). Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inf. Theory* **52** 6–18. MR2237332 https://doi.org/10.1109/TIT.2005.860430

[15] DONOHO, D. L., JOHNSTONE, I. M., HOCH, J. C. and STERN, A. S. (1992). Maximum entropy and the nearly black object. *J. Roy. Statist. Soc. Ser. B* **54** 41–81. With discussion and a reply by the authors. MR1157714

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

[17] FELDMAN, V., LIGETT, K. and SABATO, S., eds. (2021). *Algorithmic Learning Theory. Proceedings of Machine Learning Research (PMLR)* **132**.

[18] GEMAN, S., BIENENSTOCK, E. and DOURSAT, R. (1992). Neural networks and the bias/variance dilemma. *Neural Comput.* **1** 1–58.

[19] GERCHINOVITZ, S., MÉNARD, P. and STOLTZ, G. (2020). Fano's inequality for random variables. *Statist. Sci.* **35** 178–201. MR4106600 https://doi.org/10.1214/19-STS716

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

[21] JOHNSTONE, I. M. (2019). Gaussian estimation: Sequence and wavelet models. Available at http://statweb.stanford.edu/~imj/GE_09_16_19.pdf.

[22] LEHMANN, E. L. and CASELLA, G. (2006). *Theory of Point Estimation*. Springer, Berlin.

[23] LIU, R. C. and BROWN, L. D. (1993). Nonexistence of informative unbiased estimators in singular problems. *Ann. Statist.* **21** 1–13. MR1212163 https://doi.org/10.1214/aos/1176349012

[24] LOW, M. G. (1995). Bias-variance tradeoffs in functional estimation problems. *Ann. Statist.* **23** 824–835. MR1345202 https://doi.org/10.1214/aos/1176324624

[25] MEISTER, A. and REISS, M. (2013). Asymptotic equivalence for nonparametric regression with non-regular errors. *Probab. Theory Related Fields* **155** 201–229. MR3010397 https://doi.org/10.1007/s00440-011-0396-x

[26] NEAL, B., MITTAL, S., BARATIN, A., TANTIA, V., SCICLUNA, M., LACOSTE-JULIEN, S. and MITLIAGKAS, I. (2018). A modern take on the bias-variance tradeoff in neural networks. ArXiv E-prints. Available at arXiv:1810.08591.

[27] NISHIYAMA, T. (2019). A new lower bound for Kullback–Leibler divergence based on Hammersley–Chapman–Robbins bound. arXiv e-prints. Available at arXiv:1907.00288.

[28] NISHIYAMA, T. (2020). A tight lower bound for the Hellinger distance with given means and variances. arXiv e-prints. Available at arXiv:2010.13548.

[29] PFANZAGL, J. (2001). A nonparametric asymptotic version of the Cramér–Rao bound. In *State of the Art in Probability and Statistics* (*Leiden*, 1999). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **36** 499–517. IMS, Beachwood, OH. MR1836577 https://doi.org/10.1214/lnms/1215090085

[30] POLYANSKIY, Y. (2021). Lecture 2: Dpi and statistics. information theoretic methods in statistics and computer science. Available at http://people.lids.mit.edu/yp/homepage/sdpi_course.html.

[31] RAGINSKY, M. (2016). Strong data processing inequalities and Φ-Sobolev inequalities for discrete channels. *IEEE Trans. Inf. Theory* **62** 3355–3389. MR3506739 https://doi.org/10.1109/TIT.2016.2549542

[32] REISS, M. and SCHMIDT-HIEBER, J. (2020). Posterior contraction rates for support boundary recovery. *Stochastic Process. Appl.* **130** 6638–6656. MR4158798 https://doi.org/10.1016/j.spa.2020.06.005

[33] REISS, M. and SELK, L. (2017). Efficient estimation of functionals in nonparametric boundary models. *Bernoulli* **23** 1022–1055. MR3606758 https://doi.org/10.3150/15-BEJ768

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

[35] STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, *Vol. I* 197–206. Univ. California Press, Berkeley-Los Angeles, CA. MR0084922

[36] SZABÓ, B. and VAN ZANTEN, H. (2020). Adaptive distributed methods under communication constraints. *Ann. Statist.* **48** 2347–2380. MR4134798 https://doi.org/10.1214/19-AOS1890

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

[38] 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

[39] WAHL, M. (2022). Van Trees inequality, group equivariance, and estimation of principal subspaces. *Ann. Inst. Henri Poincaré Probab. Stat.* **58** 1565–1589. MR4452643 https://doi.org/10.1214/21-AIHP1193

[40] 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

[41] ZHANG, Y., DUCHI, J., JORDAN, M. I. and WAINWRIGHT, M. J. (2013). Information-theoretic lower bounds for distributed statistical estimation with communication constraints. In *Advances in Neural Information Processing Systems* 26 2328–2336. Curran Associates, Red Hook.

The Annals of Statistics
2023, Vol. 51, No. 4, 1534–1560
https://doi.org/10.1214/23-AOS2283
© Institute of Mathematical Statistics, 2023

A CROSS-VALIDATION FRAMEWORK FOR SIGNAL DENOISING WITH APPLICATIONS TO TREND FILTERING, DYADIC CART AND BEYOND

BY ANAMITRA CHAUDHURI[a] AND SABYASACHI CHATTERJEE[b]

Department of Statistics, University of Illinois at Urbana Champaign, [a]*ac34@illinois.edu,* [b]*sc1706@illinois.edu*

This paper formulates a general cross-validation framework for signal denoising. The general framework is then applied to nonparametric regression methods such as trend filtering and dyadic CART. The resulting cross-validated versions are then shown to attain nearly the same rates of convergence as are known for the optimally tuned analogues. There did not exist any previous theoretical analyses of cross-validated versions of trend filtering or dyadic CART. To illustrate the generality of the framework, we also propose and study cross-validated versions of two fundamental estimators; lasso for high-dimensional linear regression and singular value thresholding for matrix estimation. Our general framework is inspired by the ideas in Chatterjee and Jafarov (2015) and is potentially applicable to a wide range of estimation methods which use tuning parameters.

REFERENCES

[1] ARNOLD, T. B., TIBSHIRANI, R. J., ARNOLD, M. T. and BYTECOMPILE, T. (2022). Package genlasso. *Statistics* **39** 1335–1371.

[2] BLANCHARD, G., SCHÄFER, C., ROZENHOLC, Y. and MÜLLER, K.-R. (2007). Optimal dyadic decision trees. *Mach*. *Learn*. **66** 209–241.

[3] CAI, J.-F., CANDÈS, E. J. and SHEN, Z. (2010). A singular value thresholding algorithm for matrix completion. *SIAM J*. *Optim*. **20** 1956–1982. MR2600248 https://doi.org/10.1137/080738970

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

[5] CHATTERJEE, S. and GOSWAMI, S. (2021). Adaptive estimation of multivariate piecewise polynomials and bounded variation functions by optimal decision trees. *Ann*. *Statist*. **49** 2531–2551. MR4338374 https://doi.org/10.1214/20-aos2045

[6] CHATTERJEE, S. and GOSWAMI, S. (2021). New risk bounds for 2D total variation denoising. *IEEE Trans*. *Inf*. *Theory* **67** 4060–4091. MR4289366 https://doi.org/10.1109/TIT.2021.3059657

[7] CHATTERJEE, S. and JAFAROV, J. (2015). Prediction error of cross-validated lasso. arXiv preprint arXiv:1502.06291.

[8] CHATTERJEE, S. and MUKHERJEE, S. (2019). Estimation in tournaments and graphs under monotonicity constraints. *IEEE Trans*. *Inf*. *Theory* **65** 3525–3539. MR3959003 https://doi.org/10.1109/TIT.2019.2893911

[9] CHAUDHURI, A. and CHATTERJEE, S. (2023). Supplement to "A cross-validation framework for signal denoising with applications to trend filtering, dyadic CART and beyond." https://doi.org/10.1214/23-AOS2283SUPP

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

[11] DONOHO, D. L. (1997). CART and best-ortho-basis: A connection. *Ann*. *Statist*. **25** 1870–1911. MR1474073 https://doi.org/10.1214/aos/1069362377

[12] 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

[13] GOLUB, G. H., HEATH, M. and WAHBA, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* **21** 215–223. MR0533250 https://doi.org/10.2307/1268518

MSC2020 subject classifications. 62G05, 62G08.
Key words and phrases. Cross-validation, trend filtering, dyadic CART, singular value thresholding, lasso, adaptive risk bounds.

[14] 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

[15] HOEFLING, H. (2010). A path algorithm for the fused lasso signal approximator. *J. Comput. Graph. Statist.* **19** 984–1006. Supplementary material available online. MR2791265 https://doi.org/10.1198/jcgs.2010.09208

[16] HÜTTER, J.-C. and RIGOLLET, P. (2016). Optimal rates for total variation denoising. In *Conference on Learning Theory* 1115–1146.

[17] KI, D., FANG, B. and GUNTUBOYINA, A. (2021). MARS via LASSO. arXiv preprint arXiv:2111.11694.

[18] 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

[19] LI, Q. and RACINE, J. (2004). Cross-validated local linear nonparametric regression. *Statist. Sinica* **14** 485–512. MR2059293

[20] MADRID PADILLA, O. H. and CHATTERJEE, S. (2022). Risk bounds for quantile trend filtering. *Biometrika* **109** 751–768. MR4472846 https://doi.org/10.1093/biomet/asab045

[21] 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

[22] NEMIROVSKI, A. (2000). Topics in non-parametric statistics. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1998). *Lecture Notes in Math.* **1738** 85–277. Springer, Berlin. MR1775640

[23] NOWAK, R., MITRA, U. and WILLETT, R. (2004). Estimating inhomogeneous fields using wireless sensor networks. *IEEE J. Sel. Areas Commun.* **22** 999–1006.

[24] ORTELLI, F. and VAN DE GEER, S. (2020). Adaptive rates for total variation image denoising. *J. Mach. Learn. Res.* **21** Paper No. 247, 38. MR4213428

[25] ORTELLI, F. and VAN DE GEER, S. (2021). Prediction bounds for higher order total variation regularized least squares. *Ann. Statist.* **49** 2755–2773. MR4338382 https://doi.org/10.1214/21-aos2054

[26] ORTELLI, F. and VAN DE GEER, S. (2021). Oracle inequalities for square root analysis estimators with application to total variation penalties. *Inf. Inference* **10** 483–514. MR4270756 https://doi.org/10.1093/imaiai/iaaa002

[27] PADILLA, O. H. M. and CHATTERJEE, S. (2021). Quantile regression by dyadic CART. arXiv preprint arXiv:2110.08665.

[28] POLITSCH, C. A., CISEWSKI-KEHE, J., CROFT, R. A. and WASSERMAN, L. (2020). Trend filtering–I. A modern statistical tool for time-domain astronomy and astronomical spectroscopy. *Mon. Not. R. Astron. Soc.* **492** 4005–4018.

[29] RUDIN, L. I., OSHER, S. and FATEMI, E. (1992). Nonlinear total variation based noise removal algorithms. *Phys. D, Nonlinear Phenom.* **60** 259–268. MR3363401 https://doi.org/10.1016/0167-2789(92)90242-F

[30] SADHANALA, V., WANG, Y.-X. and TIBSHIRANI, R. J. (2016). Total variation classes beyond 1d: Minimax rates, and the limitations of linear smoothers. In *Advances in Neural Information Processing Systems* 3513–3521.

[31] SCOTT, C. and NOWAK, R. D. (2006). Minimax-optimal classification with dyadic decision trees. *IEEE Trans. Inf. Theory* **52** 1335–1353. MR2241192 https://doi.org/10.1109/TIT.2006.871056

[32] SHAH, N., BALAKRISHNAN, S., GUNTUBOYINA, A. and WAINWRIGHT, M. (2016). Stochastically transitive models for pairwise comparisons: Statistical and computational issues. In *International Conference on Machine Learning* 11–20.

[33] SHAO, J. (1993). Linear model selection by cross-validation. *J. Amer. Statist. Assoc.* **88** 486–494. MR1224373

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

[35] TIBSHIRANI, R. J. (2020). Divided differences, falling factorials, and discrete splines: another look at trend filtering and related problems. arXiv preprint arXiv:2003.03886.

[36] TIBSHIRANI, R. J. and TAYLOR, J. (2011). The solution path of the generalized lasso. *Ann. Statist.* **39** 1335–1371. MR2850205 https://doi.org/10.1214/11-AOS878

[37] TIBSHIRANI, R. J. and TAYLOR, J. (2012). Degrees of freedom in lasso problems. *Ann. Statist.* **40** 1198–1232. MR2985948 https://doi.org/10.1214/12-AOS1003

[38] VERSHYNIN, R. (2018). *High-Dimensional Probability*: *An Introduction with Applications in Data Science*. *Cambridge Series in Statistical and Probabilistic Mathematics* **47**. Cambridge Univ. Press, Cambridge. With a foreword by Sara van de Geer. MR3837109 https://doi.org/10.1017/9781108231596

[39] WANG, Y.-X., SHARPNACK, J., SMOLA, A. J. and TIBSHIRANI, R. J. (2016). Trend filtering on graphs. *J. Mach. Learn. Res.* **17** Paper No. 105, 41. MR3543511

[40] WANG, Y.-X., SMOLA, A. J. and TIBSHIRANI, R. J. (2014). The falling factorial basis and its statistical applications. In *ICML* 730–738.

[41] WILLETT, R. M. and NOWAK, R. D. (2007). Multiscale Poisson intensity and density estimation. *IEEE Trans*. *Inf*. *Theory* **53** 3171–3187. MR2417680 https://doi.org/10.1109/TIT.2007.903139

[42] WONG, W. H. (1983). On the consistency of cross-validation in kernel nonparametric regression. *Ann*. *Statist*. **11** 1136–1141. MR0720259 https://doi.org/10.1214/aos/1176346327

The Annals of Statistics
2023, Vol. 51, No. 4, 1561–1585
https://doi.org/10.1214/23-AOS2287
© Institute of Mathematical Statistics, 2023

OFF-POLICY EVALUATION IN PARTIALLY OBSERVED MARKOV DECISION PROCESSES UNDER SEQUENTIAL IGNORABILITY

BY YUCHEN HU[1,a] AND STEFAN WAGER[2,b]

[1]*Management Science and Engineering, Stanford University,* [a]*yuchenhu@stanford.edu*
[2]*Graduate School of Business, Stanford University,* [b]*swager@stanford.edu*

We consider off-policy evaluation of dynamic treatment rules under sequential ignorability, given an assumption that the underlying system can be modeled as a partially observed Markov decision process (POMDP). We propose an estimator, partial history importance weighting, and show that it can consistently estimate the stationary mean rewards of a target policy, given long enough draws from the behavior policy. We provide an upper bound on its error that decays polynomially in the number of observations (i.e., the number of trajectories times their length) with an exponent that depends on the overlap of the target and behavior policies as well as the mixing time of the underlying system. Furthermore, we show that this rate of convergence is minimax, given only our assumptions on mixing and overlap. Our results establish that off-policy evaluation in POMDPs is strictly harder than off-policy evaluation in (fully observed) Markov decision processes but strictly easier than model-free off-policy evaluation.

REFERENCES

ANTOS, A., SZEPESVÁRI, C. and MUNOS, R. (2008). Learning near-optimal policies with Bellman-residual minimization based fitted policy iteration and a single sample path. *Mach. Learn.* **71** 89–129.

ATHEY, S. and WAGER, S. (2021). Policy learning with observational data. *Econometrica* **89** 133–161. MR4220385 https://doi.org/10.3982/ecta15732

AZUMA, K. (1967). Weighted sums of certain dependent random variables. *Tohoku Math. J.* (2) **19** 357–367. MR0221571 https://doi.org/10.2748/tmj/1178243286

BATTALIO, S. L., CONROY, D. E., DEMPSEY, W., LIAO, P., MENICTAS, M., MURPHY, S., NAHUM-SHANI, I., QIAN, T., KUMAR, S. et al. (2021). Sense2Stop: A micro-randomized trial using wearable sensors to optimize a just-in-time-adaptive stress management intervention for smoking relapse prevention. *Contemp. Clin. Trials* 106534.

BENNETT, A. and KALLUS, N. (2021). Proximal reinforcement learning: Efficient off-policy evaluation in partially observed markov decision processes. Preprint. Available at arXiv:2110.15332.

BERK ATAMAN, M., VAN HEERDE, H. J. and MELA, C. F. (2010). The long-term effect of marketing strategy on brand sales. *J. Mark. Res.* **47** 866–882.

BIRGÉ, L. (2001). An alternative point of view on Lepski's method. In *State of the Art in Probability and Statistics* (*Leiden*, 1999). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **36** 113–133. IMS, Beachwood, OH. MR1836557 https://doi.org/10.1214/lnms/1215090065

CLOUSE, J. A. and UTGOFF, P. E. (1992). A teaching method for reinforcement learning. In *Machine Learning Proceedings* 1992 92–101. Elsevier, Amsterdam.

CONNAULT, B. (2016). Hidden rust models. Unpublished manuscript. Available at https://www.sas.upenn.edu/~connault/hidden-rust-models.pdf.

EVEN-DAR, E., KAKADE, S. M. and MANSOUR, Y. (2005). Experts in a Markov decision process. *Adv. Neural Inf. Process. Syst.* **17** 401–408.

FARIAS, V. F., LI, A. A., PENG, T. and ZHENG, A. T. (2022). Markovian interference in experiments. Preprint. Available at arXiv:2206.02371.

FARIAS, V. F., MOALLEMI, C. C., VAN ROY, B. and WEISSMAN, T. (2010). Universal reinforcement learning. *IEEE Trans. Inf. Theory* **56** 2441–2454. MR2729794 https://doi.org/10.1109/TIT.2010.2043762

MSC2020 subject classifications. 62M09, 62D20.
Key words and phrases. Importance weighting, causal inference, mixing time, Lepski's method, sequential ignorability.

GIMELFARB, M., SANNER, S. and LEE, C.-G. (2018). Reinforcement learning with multiple experts: A Bayesian model combination approach. *Adv. Neural Inf. Process. Syst.* **31** 9528–9538.

GOLDENSHLUGER, A. and NEMIROVSKI, A. (1997). On spatially adaptive estimation of nonparametric regression. *Math. Methods Statist.* **6** 135–170. MR1466625

GRIMMETT, G. and STIRZAKER, D. (2020). *Probability and Random Processes*. Oxford University Press, New York.

GUO, Z., THOMAS, P. S. and BRUNSKILL, E. (2017). Using options and covariance testing for long horizon off-policy policy evaluation. *Adv. Neural Inf. Process. Syst.* **30**.

GUSTAFSON, D. H., MCTAVISH, F. M., CHIH, M.-Y., ATWOOD, A. K., JOHNSON, R. A., BOYLE, M. G., LEVY, M. S., DRISCOLL, H., CHISHOLM, S. M. et al. (2014). A smartphone application to support recovery from alcoholism: A randomized clinical trial. *JAMA Psychiatr.* **71** 566–572.

HERNÁN, M. A., BRUMBACK, B. and ROBINS, J. M. (2001). Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *J. Amer. Statist. Assoc.* **96** 440–448. MR1939347 https://doi.org/10.1198/016214501753168154

HERNÁN, M. A. and ROBINS, J. M. (2020). *Causal Inference: What If*. CRC Press/CRC, Boca Raton, FL.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. MR0144363

HU, Y. and WAGER, S. (2023). Supplement to "Off-policy evaluation in partially observed Markov decision processes under sequential ignorability." https://doi.org/10.1214/23-AOS2287SUPP

JIANG, N. and LI, L. (2016). Doubly robust off-policy value evaluation for reinforcement learning. In *Proceedings of the* 33*rd International Conference on International Conference on Machine Learning* **48** 652–661. JMLR.org.

KAELBLING, L. P., LITTMAN, M. L. and CASSANDRA, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence* **101** 99–134. MR1641530 https://doi.org/10.1016/S0004-3702(98)00023-X

KALLUS, N. and UEHARA, M. (2020b). Double reinforcement learning for efficient off-policy evaluation in Markov decision processes. *J. Mach. Learn. Res.* **21** Paper No. 167, 63. MR4209453

KALLUS, N. and UEHARA, M. (2022). Efficiently breaking the curse of horizon in off-policy evaluation with double reinforcement learning. *Oper. Res.* **70** 3282–3302. MR4538517

KATKOVNIK, V. (1999). A new method for varying adaptive bandwidth selection. *IEEE Trans. Signal Process.* **47** 2567–2571.

KATKOVNIK, V., FOI, A., EGIAZARIAN, K. and ASTOLA, J. (2010). From local kernel to nonlocal multiple-model image denoising. *Int. J. Comput. Vis.* **86** 1–32. MR2683762 https://doi.org/10.1007/s11263-009-0272-7

KATKOVNIK, V. and SPOKOINY, V. (2008). Spatially adaptive estimation via fitted local likelihood techniques. *IEEE Trans. Signal Process.* **56** 873–886. MR2518663 https://doi.org/10.1109/TSP.2007.907873

KITAGAWA, T. and TETENOV, A. (2018). Who should be treated? Empirical welfare maximization methods for treatment choice. *Econometrica* **86** 591–616. MR3783340 https://doi.org/10.3982/ECTA13288

KLASNJA, P., HEKLER, E. B., SHIFFMAN, S., BORUVKA, A., ALMIRALL, D., TEWARI, A. and MURPHY, S. A. (2015). Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. *Health Psychology* **34** 1220.

KLASNJA, P., SMITH, S., SEEWALD, N. J., LEE, A., HALL, K., LUERS, B., HEKLER, E. B. and MURPHY, S. A. (2019). Efficacy of contextually tailored suggestions for physical activity: A micro-randomized optimization trial of HeartSteps. *Ann. Behav. Med.* **53** 573–582.

KOEDINGER, K. R., BRUNSKILL, E., SJD BAKER, R., MCLAUGHLIN, E. A. and STAMPER, J. (2013). New potentials for data-driven intelligent tutoring system development and optimization. *AI Mag.* **34** 27–41.

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

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

LIAO, P., KLASNJA, P. and MURPHY, S. (2021). Off-policy estimation of long-term average outcomes with applications to mobile health. *J. Amer. Statist. Assoc.* **116** 382–391. MR4227701 https://doi.org/10.1080/01621459.2020.1807993

LIAO, P., QI, Z., WAN, R., KLASNJA, P. and MURPHY, S. A. (2022). Batch policy learning in average reward Markov decision processes. *Ann. Statist.* **50** 3364–3387. MR4524500 https://doi.org/10.1214/22-aos2231

LINDEBERG, J. W. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Z.* **15** 211–225. MR1544569 https://doi.org/10.1007/BF01494395

LUCKETT, D. J., LABER, E. B., KAHKOSKA, A. R., MAAHS, D. M., MAYER-DAVIS, E. and KOSOROK, M. R. (2020). Estimating dynamic treatment regimes in mobile health using V-learning. *J. Amer. Statist. Assoc.* **115** 692–706. MR4107673 https://doi.org/10.1080/01621459.2018.1537919

MAAHS, D. M., MAYER-DAVIS, E., BISHOP, F. K., WANG, L., MANGAN, M. and MCMURRAY, R. G. (2012). Outpatient assessment of determinants of glucose excursions in adolescents with type 1 diabetes: Proof of concept. *Diabetes Technol. Ther.* **14** 658–664.

MANDEL, T., LIU, Y.-E., LEVINE, S., BRUNSKILL, E. and POPOVIC, Z. (2014). Offline policy evaluation across representations with applications to educational games. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems*.

MIAO, W., GENG, Z. and TCHETGEN TCHETGEN, E. J. (2018). Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika* **105** 987–993. MR3877879 https://doi.org/10.1093/biomet/asy038

MONAHAN, G. E. (1982). A survey of partially observable Markov decision processes: Theory, models, and algorithms. *Manage. Sci.* **28** 1–16. MR0646904 https://doi.org/10.1287/mnsc.28.1.1

MURPHY, S. A. (2003). Optimal dynamic treatment regimes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 331–366. MR1983752 https://doi.org/10.1111/1467-9868.00389

MURPHY, S. A. (2005). A generalization error for Q-learning. *J. Mach. Learn. Res.* **6** 1073–1097. MR2249849

NAHUM-SHANI, I., SMITH, S. N., SPRING, B. J., COLLINS, L. M., WITKIEWITZ, K., TEWARI, A. and MURPHY, S. A. (2018). Just-in-time adaptive interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. *Ann. Behav. Med.* **52** 446–462.

NAIR, Y. and JIANG, N. (2021). A spectral approach to off-policy evaluation for pomdps. Preprint. Available at arXiv:2109.10502.

NIE, X., BRUNSKILL, E. and WAGER, S. (2021). Learning when-to-treat policies. *J. Amer. Statist. Assoc.* **116** 392–409. MR4227702 https://doi.org/10.1080/01621459.2020.1831925

PRECUP, D., SUTTON, R. S. and SINGH, S. P. (2000). Eligibility traces for off-policy policy evaluation. In *Proceedings of the Seventeenth International Conference on Machine Learning*.

QIAN, M. and MURPHY, S. A. (2011). Performance guarantees for individualized treatment rules. *Ann. Statist.* **39** 1180–1210. MR2816351 https://doi.org/10.1214/10-AOS864

ROBINS, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math. Model.* **7** 1393–1512.

ROBINS, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium in Biostatistics. Lect. Notes Stat.* **179** 189–326. Springer, New York. MR2129402 https://doi.org/10.1007/978-1-4419-9076-1_11

ROSENBLATT, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA* **42** 43–47. MR0074711 https://doi.org/10.1073/pnas.42.1.43

RUSSELL, S. and NORVIG, P. (2010). *Artificial Intelligence: A Modern Approach*, 3rd ed. Prentice Hall, New York.

SAGHAFIAN, S. (2018). Ambiguous partially observable Markov decision processes: Structural results and applications. *J. Econom. Theory* **178** 1–35. MR3869284 https://doi.org/10.1016/j.jet.2018.08.006

SMALLWOOD, R. D. and SONDIK, E. J. (1973). The optimal control of partially observable Markov processes over a finite horizon. *Oper. Res.* **21** 1071–1088.

SONDIK, E. J. (1978). The optimal control of partially observable Markov processes over the infinite horizon: Discounted costs. *Oper. Res.* **26** 282–304. MR0484438 https://doi.org/10.1287/opre.26.2.282

SU, Y., SRINATH, P. and KRISHNAMURTHY, A. (2020). Adaptive estimator selection for off-policy evaluation. In *International Conference on Machine Learning* 9196–9205. PMLR.

SWAMINATHAN, A. and JOACHIMS, T. (2015). Batch learning from logged bandit feedback through counterfactual risk minimization. *J. Mach. Learn. Res.* **16** 1731–1755. MR3417797

TENNENHOLTZ, G., SHALIT, U. and MANNOR, S. (2020). Off-policy evaluation in partially observable environments. In *Proceedings of the AAAI Conference on Artificial Intelligence* **34** 10276–10283.

TENORIO-GONZALEZ, A. C., MORALES, E. F. and VILLASENOR-PINEDA, L. (2010). Dynamic reward shaping: Training a robot by voice. In *Ibero-American Conference on Artificial Intelligence* 483–492. Springer, Berlin.

THEOCHAROUS, G., THOMAS, P. S. and GHAVAMZADEH, M. (2015). Personalized ad recommendation systems for life-time value optimization with guarantees. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*.

THOMAS, P. and BRUNSKILL, E. (2016). Data-efficient off-policy policy evaluation for reinforcement learning. In *International Conference on Machine Learning* 2139–2148.

UEHARA, M., KIYOHARA, H., BENNETT, A., CHERNOZHUKOV, V., JIANG, N., KALLUS, N., SHI, C. and SUN, W. (2022). Future-dependent value-based off-policy evaluation in pomdps. Preprint. Available at arXiv:2207.13081.

VAN DANTZIG, S., GELEIJNSE, G. and VAN HALTEREN, A. T. (2013). Toward a persuasive mobile application to reduce sedentary behavior. *Pers. Ubiquitous Comput.* **17** 1237–1246.

VAN ROY, B. (1998). *Learning and Value Function Approximation in Complex Decision Processes*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)–Massachusetts Institute of Technology. MR2716807

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

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

WANG, P., ROWE, J., MOTT, B. and LESTER, J. (2016). Decomposing drama management in educational interactive narrative: A modular reinforcement learning approach. In *International Conference on Interactive Digital Storytelling* 270–282. Springer, Berlin.

WANG, Y.-X., AGARWAL, A. and DUDIK, M. (2017). Optimal and adaptive off-policy evaluation in contextual bandits. In *International Conference on Machine Learning* 3589–3597. PMLR.

YU, B. (1997). Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam* 423–435. Springer, New York. MR1462963

ZHANG, B., TSIATIS, A. A., LABER, E. B. and DAVIDIAN, M. (2013). Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika* **100** 681–694. MR3094445 https://doi.org/10.1093/biomet/ast014

ZHANG, Z. G., CHAN, S.-C., HO, K.-L. and HO, K. C. (2008). On bandwidth selection in local polynomial regression analysis and its application to multi-resolution analysis of non-uniform data. *J. Signal Process*. *Syst*. **52** 263–280.

ZHAO, Y., ZENG, D., RUSH, A. J. and KOSOROK, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *J. Amer. Statist. Assoc*. **107** 1106–1118. MR3010898 https://doi.org/10.1080/01621459.2012.695674

The Annals of Statistics
2023, Vol. 51, No. 4, 1586–1610
https://doi.org/10.1214/23-AOS2297
© Institute of Mathematical Statistics, 2023

OPTIMAL CHANGE-POINT DETECTION AND LOCALIZATION

BY NICOLAS VERZELEN[1,a], MAGALIE FROMONT[2,b], MATTHIEU LERASLE[3,c] AND PATRICIA REYNAUD-BOURET[4,d]

[1]*MISTEA, INRAE,* [a]*nicolas.verzelen@inrae.fr*
[2]*IRMAR, Univ. Rennes,* [b]*magalie.fromont@univ-rennes2.fr*
[3]*CREST, Institut Polytechnique de Paris,* [c]*matthieu.lerasle@ensae.fr*
[4]*LJAD, Université Côte d'Azur,* [d]*Patricia.Reynaud-Bouret@univ-cotedazur.fr*

Given a times series $\mathbf{Y}$ in $\mathbb{R}^n$, with a piecewise constant mean and independent components, the twin problems of change-point detection and change-point localization, respectively amount to detecting the existence of times where the mean varies and estimating the positions of those change-points. In this work, we tightly characterize optimal rates for both problems and uncover the phase transition phenomenon from a global testing problem to a local estimation problem. Introducing a suitable definition of the energy of a change-point, we first establish in the single change-point setting that the optimal detection threshold is $\sqrt{2\log\log(n)}$. When the energy is just above the detection threshold, then the problem of localizing the change-point becomes purely parametric: it only depends on the difference in means and not on the position of the change-point anymore. Interestingly, for most change-point positions, including all those away from the endpoints of the time series, it is possible to detect and localize them at a much smaller energy level. In the multiple change-point setting, we establish the energy detection threshold and show similarly that the optimal localization error of a specific change-point becomes purely parametric. Along the way, tight minimax rates for Hausdorff and l_1 estimation losses of the vector of all change-points positions are also established. Two procedures achieving these optimal rates are introduced. The first one is a least-squares estimator with a new multiscale penalty that favours well spread change-points. The second one is a two-step multiscale post-processing procedure whose computational complexity can be as low as $O(n\log(n))$. Notably, these two procedures accommodate with the presence of possibly many low-energy and therefore undetectable change-points and are still able to detect and localize high-energy change-points even with the presence of those nuisance parameters.

REFERENCES

[1] ARIAS-CASTRO, E., CANDÈS, E. J. and DURAND, A. (2011). Detection of an anomalous cluster in a network. *Ann. Statist.* **39** 278–304. MR2797847 https://doi.org/10.1214/10-AOS839

[2] ARIAS-CASTRO, E., DONOHO, D. L. and HUO, X. (2005). Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Inf. Theory* **51** 2402–2425. MR2246369 https://doi.org/10.1109/TIT.2005.850056

[3] ARLOT, S. (2019). Minimal penalties and the slope heuristics: A survey. *J. SFdS* **160** 1–106. MR4021408

[4] ARLOT, S., CELISSE, A. and HARCHAOUI, Z. (2019). A kernel multiple change-point algorithm via model selection. *J. Mach. Learn. Res.* **20** Paper No. 162, 56. MR4048973

[5] BAI, J. and PERRON, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66** 47–78. MR1616121 https://doi.org/10.2307/2998540

[6] BARANOWSKI, R., CHEN, Y. and FRYZLEWICZ, P. (2019). Narrowest-over-threshold detection of multiple change points and change-point-like features. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 649–672. MR3961502 https://doi.org/10.1111/rssb.12322

MSC2020 subject classifications. 62C20.
Key words and phrases. Change-point analyses, CUSUM, minimax, detection, localization.

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

[8] BASSEVILLE, M. and NIKIFOROV, I. V. (1993). *Detection of Abrupt Changes*: *Theory and Application*. *Prentice Hall Information and System Sciences Series*. Prentice Hall, Inc., Englewood Cliffs, NJ. MR1210954

[9] BELLMAN, R. (1961). On the approximation of curves by line segments using dynamic programming. *Commun*. *ACM* **4** 284.

[10] BIRGÉ, L. and MASSART, P. (2001). Gaussian model selection. *J. Eur. Math. Soc*. (*JEMS*) **3** 203–268. MR1848946 https://doi.org/10.1007/s100970100031

[11] BOYSEN, L., KEMPE, A., LIEBSCHER, V., MUNK, A. and WITTICH, O. (2009). Consistencies and rates of convergence of jump-penalized least squares estimators. *Ann. Statist*. **37** 157–183. MR2488348 https://doi.org/10.1214/07-AOS558

[12] CARLSTEIN, E., MÜLLER, H.-G. and SIEGMUND, D., eds. (1994). *Change-Point Problems*. *Papers from the AMS-IMS-SIAM Summer Research Conference Held at Mt*. *Holyoke College*, *South Hadley*, *MA*, *July* 11–16, 1992. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **23**. IMS, Hayward, CA. MR1477909

[13] CHAN, H.-P. and CHEN, H. (2017). Multi-sequence segmentation via score and higher-criticism tests. arXiv preprint arXiv:1706.07586.

[14] CHAN, H. P. and WALTHER, G. (2013). Detection with the scan and the average likelihood ratio. *Statist*. *Sinica* **23** 409–428. MR3076173

[15] CHO, H. and KIRCH, C. (2019). Localised pruning for data segmentation based on multiscale change point procedures. arXiv preprint arXiv:1910.12486.

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

[17] DÜMBGEN, L. (1991). The asymptotic behavior of some nonparametric change-point estimators. *Ann*. *Statist*. **19** 1471–1495. MR1126333 https://doi.org/10.1214/aos/1176348257

[18] EICHINGER, B. and KIRCH, C. (2018). A MOSUM procedure for the estimation of multiple random change points. *Bernoulli* **24** 526–564. MR3706768 https://doi.org/10.3150/16-BEJ887

[19] FAN, Z. and GUAN, L. (2018). Approximate ℓ_0-penalized estimation of piecewise-constant signals on graphs. *Ann. Statist*. **46** 3217–3245. MR3852650 https://doi.org/10.1214/17-AOS1656

[20] FISHER, W. D. (1958). On grouping for maximum homogeneity. *J. Amer. Statist. Assoc*. **53** 789–798. MR0101577

[21] FRICK, K., MUNK, A. and SIELING, H. (2014). Multiscale change point inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **76** 495–580. With 32 discussions by 47 authors and a rejoinder by the authors. MR3210728 https://doi.org/10.1111/rssb.12047

[22] FRYZLEWICZ, P. (2014). Wild binary segmentation for multiple change-point detection. *Ann. Statist*. **42** 2243–2281. MR3269979 https://doi.org/10.1214/14-AOS1245

[23] FRYZLEWICZ, P. (2018). Tail-greedy bottom-up data decompositions and fast multiple change-point detection. *Ann. Statist*. **46** 3390–3421. MR3852656 https://doi.org/10.1214/17-AOS1662

[24] FRYZLEWICZ, P. (2020). Detecting possibly frequent change-points: Wild Binary Segmentation 2 and steepest-drop model selection. *J. Korean Statist. Soc*. **49** 1027–1070. MR4183039 https://doi.org/10.1007/s42952-020-00060-x

[25] FRYZLEWICZ, P. (2020). Narrowest significance pursuit: Inference for multiple change-points in linear models. arXiv preprint arXiv:2009.05431.

[26] GAO, C., HAN, F. and ZHANG, C.-H. (2020). On estimation of isotonic piecewise constant signals. *Ann. Statist*. **48** 629–654. MR4102670 https://doi.org/10.1214/18-AOS1792

[27] GARREAU, D. and ARLOT, S. (2018). Consistent change-point detection with kernels. *Electron. J. Stat*. **12** 4440–4486. MR3892345 https://doi.org/10.1214/18-ejs1513

[28] GIRSHICK, M. A. and RUBIN, H. (1952). A Bayes approach to a quality control model. *Ann. Math. Stat*. **23** 114–125. MR0047291 https://doi.org/10.1214/aoms/1177729489

[29] 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

[30] HARCHAOUI, Z. and LÉVY-LEDUC, C. (2010). Multiple change-point estimation with a total variation penalty. *J. Amer. Statist. Assoc*. **105** 1480–1493. MR2796565 https://doi.org/10.1198/jasa.2010.tm09181

[31] HAWKINS, D. M. (1977). Testing a sequence of observations for a shift in location. *J. Amer. Statist. Assoc*. **72** 180–186. MR0451496

[32] HINKLEY, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika* **57** 1–17. MR0273727 https://doi.org/10.1093/biomet/57.1.1

[33] HINKLEY, D. V. and HINKLEY, E. A. (1970). Inference about the change-point in a sequence of binomial variables. *Biometrika* **57** 477–488. MR0275556 https://doi.org/10.1093/biomet/57.3.477

[34] HOEFLING, H. (2010). A path algorithm for the fused lasso signal approximator. *J. Comput. Graph. Statist.* **19** 984–1006. Supplementary materials available online. MR2791265 https://doi.org/10.1198/jcgs.2010.09208

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

[36] 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

[37] KOVÁCS, S., BÜHLMANN, P., LI, H. and MUNK, A. (2023). Seeded binary segmentation: A general methodology for fast and optimal changepoint detection. *Biometrika* **110** 249–256. MR4565454 https://doi.org/10.1093/biomet/asac052

[38] LAVIELLE, M. and MOULINES, E. (2000). Least-squares estimation of an unknown number of shifts in a time series. *J. Time Series Anal.* **21** 33–59. MR1766173 https://doi.org/10.1111/1467-9892.00172

[39] LEBARBIER, E. (2005). Detecting multiple change-points in the mean of Gaussian process by model selection. *Signal Process.* **85** 717–736.

[40] LIEHRMANN, A. and RIGAILL, G. (2023). Ms. FPOP: An exact and fast segmentation algorithm with a multiscale penalty. arXiv preprint arXiv:2303.08723.

[41] LIN, K., SHARPNACK, J., RINALDO, A. and TIBSHIRANI, R. J. (2016). Approximate recovery in change-point problems, from $\backslash\ell_2$ estimation error rates. arXiv preprint arXiv:1606.06746.

[42] LIU, H., GAO, C. and SAMWORTH, R. J. (2021). Minimax rates in sparse, high-dimensional change point detection. *Ann. Statist.* **49** 1081–1112. MR4255120 https://doi.org/10.1214/20-aos1994

[43] MAIDSTONE, R., HOCKING, T., RIGAILL, G. and FEARNHEAD, P. (2017). On optimal multiple change-point algorithms for large data. *Stat. Comput.* **27** 519–533. MR3599687 https://doi.org/10.1007/s11222-016-9636-3

[44] NIU, Y. S., HAO, N. and ZHANG, H. (2016). Multiple change-point detection: A selective overview. *Statist. Sci.* **31** 611–623. MR3598742 https://doi.org/10.1214/16-STS587

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

[46] PAGE, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika* **42** 523–527. MR0072412 https://doi.org/10.1093/biomet/42.3-4.523

[47] PAGE, E. S. (1957). On problems in which a change in a parameter occurs at an unknown point. *Biometrika* **44** 248–252.

[48] RIGAILL, G. (2010). Pruned dynamic programming for optimal multiple change-point detection. arXiv preprint arXiv:1004.0887.

[49] SCOTT, A. J. and KNOTT, M. (1974). A cluster analysis method for grouping means in the analysis of variance. *Biometrics* 507–512.

[50] 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

[51] TIBSHIRANI, R., SAUNDERS, M., ROSSET, S., ZHU, J. and KNIGHT, K. (2005). Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 91–108. MR2136641 https://doi.org/10.1111/j.1467-9868.2005.00490.x

[52] TRUONG, C., OUDRE, L. and VAYATIS, N. (2020). Selective review of offline change point detection methods. *Signal Process.* **167** 107299.

[53] VERZELEN, N., FROMONT, M., LERASLE, M. and REYNAUD-BOURET, P. (2023). Supplement to "Optimal change-point detection and localization." https://doi.org/10.1214/23-AOS2297SUPP

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

[55] 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

[56] WANG, T. and SAMWORTH, R. J. (2018). High dimensional change point estimation via sparse projection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 57–83. MR3744712 https://doi.org/10.1111/rssb.12243

[57] XU, H., WANG, D., ZHAO, Z. and YU, Y. (2022). Change point inference in high-dimensional regression models under temporal dependence. arXiv preprint arXiv:2207.12453.

[58] 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

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

[60] ZHANG, N. R. and SIEGMUND, D. O. (2007). A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics* **63** 22–32, 309. MR2345571 https://doi.org/10.1111/j.1541-0420.2006.00662.x

The Annals of Statistics
2023, Vol. 51, No. 4, 1611–1638
https://doi.org/10.1214/23-AOS2301
© Institute of Mathematical Statistics, 2023

NOISY LINEAR INVERSE PROBLEMS UNDER CONVEX CONSTRAINTS: EXACT RISK ASYMPTOTICS IN HIGH DIMENSIONS

BY QIYANG HAN[a]

Department of Statistics, Rutgers University, [a]*qh85@stat.rutgers.edu*

In the standard Gaussian linear measurement model $Y = X\mu_0 + \xi \in \mathbb{R}^m$ with a fixed noise level $\sigma > 0$, we consider the problem of estimating the unknown signal μ_0 under a convex constraint $\mu_0 \in K$, where K is a closed convex set in $\mathbb{R}^n$. We show that the risk of the natural convex constrained least squares estimator (LSE) $\widehat{\mu}(\sigma)$ can be characterized exactly in high-dimensional limits, by that of the convex constrained LSE $\widehat{\mu}_K^{\mathsf{seq}}$ in the corresponding Gaussian sequence model at a different noise level. Formally, we show that

$$\|\widehat{\mu}(\sigma) - \mu_0\|^2/(nr_n^2) \to 1 \quad \text{in probability},$$

where $r_n^2 > 0$ solves the fixed-point equation

$$\mathbb{E}\|\widehat{\mu}_K^{\mathsf{seq}}\big(\sqrt{(r_n^2 + \sigma^2)/(m/n)}\big) - \mu_0\|^2 = nr_n^2.$$

This characterization holds (uniformly) for risks r_n^2 in the maximal regime that ranges from constant order all the way down to essentially the parametric rate, as long as certain necessary nondegeneracy condition is satisfied for $\widehat{\mu}(\sigma)$.

The precise risk characterization reveals a fundamental difference between noiseless (or low noise limit) and noisy linear inverse problems in terms of the sample complexity for signal recovery. A concrete example is given by the isotonic regression problem: While exact recovery of a general monotone signal requires $m \gg n^{1/3}$ samples in the noiseless setting, consistent signal recovery in the noisy setting requires as few as $m \gg \log n$ samples. Such a discrepancy occurs when the low and high noise risk behavior of $\widehat{\mu}_K^{\mathsf{seq}}$ differ significantly. In statistical languages, this occurs when $\widehat{\mu}_K^{\mathsf{seq}}$ estimates 0 at a faster "adaptation rate" than the slower "worst-case rate" for general signals. Several other examples, including nonnegative least squares and generalized Lasso (in constrained forms), are also worked out to demonstrate the concrete applicability of the theory in problems of different types.

The proof relies on a collection of new analytic and probabilistic results concerning estimation error, log likelihood ratio test statistics and degree-of-freedom associated with $\widehat{\mu}_K^{\mathsf{seq}}$, regarded as stochastic processes indexed by the noise level. These results are of independent interest in and of themselves.

REFERENCES

[1] AMELUNXEN, D., LOTZ, M., MCCOY, M. B. and TROPP, J. A. (2014). Living on the edge: Phase transitions in convex programs with random data. *Inf. Inference* **3** 224–294. MR3311453 https://doi.org/10.1093/imaiai/iau005

[2] BARBIER, J., KRZAKALA, F., MACRIS, N., MIOLANE, L. and ZDEBOROVÁ, L. (2019). Optimal errors and phase transitions in high-dimensional generalized linear models. *Proc. Natl. Acad. Sci. USA* **116** 5451–5460. MR3939767 https://doi.org/10.1073/pnas.1802705116

[3] BAYATI, M. and MONTANARI, A. (2011). The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inf. Theory* **57** 764–785. MR2810285 https://doi.org/10.1109/TIT.2010.2094817

MSC2020 subject classifications. Primary 62G08, 60F05; secondary 62G10, 62E17.
Key words and phrases. Fixed-point equation, Gaussian sequence model, high-dimensional asymptotics, linear inverse problem.

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

[5] BELLEC, P. C. (2018). Sharp oracle inequalities for least squares estimators in shape restricted regression. *Ann. Statist.* **46** 745–780. MR3782383 https://doi.org/10.1214/17-AOS1566

[6] BELLEC, P. C. and ZHANG, C.-H. (2023). Debiasing convex regularized estimators and interval estimation in linear models. *Ann. Statist.* **51** 391–436. MR4600987 https://doi.org/10.1214/22-aos2243

[7] BERTHIER, R., MONTANARI, A. and NGUYEN, P.-M. (2020). State evolution for approximate message passing with non-separable functions. *Inf. Inference* **9** 33–79. MR4079177 https://doi.org/10.1093/imaiai/iay021

[8] BOBKOV, S. G. and GÖTZE, F. (1999). Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **163** 1–28. MR1682772 https://doi.org/10.1006/jfan.1998.3326

[9] CANDÈS, E. J., ROMBERG, J. and TAO, T. (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory* **52** 489–509. MR2236170 https://doi.org/10.1109/TIT.2005.862083

[10] CANDÈS, E. J., ROMBERG, J. K. and TAO, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.* **59** 1207–1223. MR2230846 https://doi.org/10.1002/cpa.20124

[11] CANDES, E. J. and TAO, T. (2006). Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inf. Theory* **52** 5406–5425. MR2300700 https://doi.org/10.1109/TIT.2006.885507

[12] CELENTANO, M. and MONTANARI, A. (2022). Fundamental barriers to high-dimensional regression with convex penalties. *Ann. Statist.* **50** 170–196. MR4382013 https://doi.org/10.1214/21-aos2100

[13] CELENTANO, M., MONTANARI, A. and WEI, Y. (2022). The lasso with general gaussian designs with applications to hypothesis testing. Preprint. Available at arXiv:2007.13716v2.

[14] CHANDRASEKARAN, V., RECHT, B., PARRILO, P. A. and WILLSKY, A. S. (2012). The convex geometry of linear inverse problems. *Found. Comput. Math.* **12** 805–849. MR2989474 https://doi.org/10.1007/s10208-012-9135-7

[15] CHANDRASEKHER, K. A., PANANJADY, A. and THRAMPOULIDIS, C. (2023). Sharp global convergence guarantees for iterative nonconvex optimization with random data. *Ann. Statist.* **51** 179–210. MR4564853 https://doi.org/10.1214/22-aos2246

[16] CHATTERJEE, S. (2014). A new perspective on least squares under convex constraint. *Ann. Statist.* **42** 2340–2381. MR3269982 https://doi.org/10.1214/14-AOS1254

[17] CHATTERJEE, S., GUNTUBOYINA, A. and SEN, B. (2015). On risk bounds in isotonic and other shape restricted regression problems. *Ann. Statist.* **43** 1774–1800. MR3357878 https://doi.org/10.1214/15-AOS1324

[18] CHATTERJEE, S., GUNTUBOYINA, A. and SEN, B. (2018). On matrix estimation under monotonicity constraints. *Bernoulli* **24** 1072–1100. MR3706788 https://doi.org/10.3150/16-BEJ865

[19] CHEN, D. and PLEMMONS, R. J. (2010). Nonnegativity constraints in numerical analysis. In *The Birth of Numerical Analysis* 109–139. World Sci. Publ., Hackensack, NJ. MR2604144

[20] DONOHO, D. and MONTANARI, A. (2016). High dimensional robust M-estimation: Asymptotic variance via approximate message passing. *Probab. Theory Related Fields* **166** 935–969. MR3568043 https://doi.org/10.1007/s00440-015-0675-z

[21] DONOHO, D. L. (2006). Compressed sensing. *IEEE Trans. Inf. Theory* **52** 1289–1306. MR2241189 https://doi.org/10.1109/TIT.2006.871582

[22] DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2011). The noise-sensitivity phase transition in compressed sensing. *IEEE Trans. Inf. Theory* **57** 6920–6941. MR2882271 https://doi.org/10.1109/TIT.2011.2165823

[23] EL KAROUI, N. (2018). On the impact of predictor geometry on the performance on high-dimensional ridge-regularized generalized robust regression estimators. *Probab. Theory Related Fields* **170** 95–175. MR3748322 https://doi.org/10.1007/s00440-016-0754-9

[24] 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

[25] GERBELOT, C., ABBARA, A. and KRZAKALA, F. (2020). Asymptotic errors for high-dimensional convex penalized linear regression beyond Gaussian matrices. In *Conference on Learning Theory*, *PMLR* 1682–1713.

[26] GERBELOT, C., ABBARA, A. and KRZAKALA, F. (2023). Asymptotic errors for teacher-student convex generalized linear models (or: How to prove Kabashima's replica formula). *IEEE Trans. Inf. Theory* **69** 1824–1852. MR4564683 https://doi.org/10.1109/tit.2022.3222913

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

[28] GOLDSTEIN, L., NOURDIN, I. and PECCATI, G. (2017). Gaussian phase transitions and conic intrinsic volumes: Steining the Steiner formula. *Ann. Appl. Probab.* **27** 1–47. MR3619780 https://doi.org/10.1214/16-AAP1195

[29] GORDON, Y. (1985). Some inequalities for Gaussian processes and applications. *Israel J. Math.* **50** 265–289. MR0800188 https://doi.org/10.1007/BF02759761

[30] GORDON, Y. (1988). On Milman's inequality and random subspaces which escape through a mesh in $\mathbf{R}^n$. In *Geometric Aspects of Functional Analysis* (1986/87). *Lecture Notes in Math.* **1317** 84–106. Springer, Berlin. MR0950977 https://doi.org/10.1007/BFb0081737

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

[32] GUNTUBOYINA, A. and SEN, B. (2018). Nonparametric shape-restricted regression. *Statist. Sci.* **33** 568–594. MR3881209 https://doi.org/10.1214/18-STS665

[33] HAN, Q. (2023). Supplement to "Noisy linear inverse problems under convex constraints: Exact risk asymptotics in high dimensions." https://doi.org/10.1214/23-AOS2301SUPP

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

[35] HAN, Q. and REN, H. (2022). Gaussian random projections of convex cones: Approximate kinematic formulae and applications. Preprint. Available at arXiv:2212.05545.

[36] HAN, Q., SEN, B. and SHEN, Y. (2022). High-dimensional asymptotics of likelihood ratio tests in the Gaussian sequence model under convex constraints. *Ann. Statist.* **50** 376–406. MR4382021 https://doi.org/10.1214/21-aos2111

[37] HAN, Q., WANG, T., CHATTERJEE, S. and SAMWORTH, R. J. (2019). Isotonic regression in general dimensions. *Ann. Statist.* **47** 2440–2471. MR3988762 https://doi.org/10.1214/18-AOS1753

[38] HAN, Q. and WELLNER, J. A. (2016). Multivariate convex regression: Global risk bounds and adaptation. Preprint. Available at arXiv:1601.06844.

[39] HAN, Q. and XU, X. (2023). The distribution of ridgeless least squares interpolators. Preprint. Available at arXiv:2307.02044.

[40] HAN, Q. and ZHANG, C.-H. (2020). Limit distribution theory for block estimators in multiple isotonic regression. *Ann. Statist.* **48** 3251–3282. MR4185808 https://doi.org/10.1214/19-AOS1928

[41] HU, H. (2021). *Regularized Regression in High Dimensions*: *Asymptotics*, *Optimality and Universality*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)–Harvard University. MR4360796

[42] JAVANMARD, A. and MONTANARI, A. (2013). State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Inf. Inference* **2** 115–144. MR3311445 https://doi.org/10.1093/imaiai/iat004

[43] KABASHIMA, Y. (2008). Inference from correlated patterns: A unified theory for perceptron learning and linear vector channels. In *Journal of Physics*: *Conference Series* **95** 012001. IOP Publishing, Bristol.

[44] KABASHIMA, Y., WADAYAMA, T. and TANAKA, T. (2009). A typical reconstruction limit for compressed sensing based on lp-norm minimization. *J. Stat. Mech. Theory Exp.* **2009** L09003.

[45] KAROUI, N. E., BEAN, D., BICKEL, P. J., LIM, C. and YU, B. (2013). On robust regression with high-dimensional predictors. *Proc. Natl. Acad. Sci. USA* **110** 14557–14562. https://doi.org/10.1073/pnas.1307842110

[46] KATO, K. (2009). On the degrees of freedom in shrinkage estimation. *J. Multivariate Anal.* **100** 1338–1352. MR2514133 https://doi.org/10.1016/j.jmva.2008.12.002

[47] KIM, H. and PARK, H. (2008). Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method. *SIAM J. Matrix Anal. Appl.* **30** 713–730. MR2421467 https://doi.org/10.1137/07069239X

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

[49] KUDÔ, A. (1963). A multivariate analogue of the one-sided test. *Biometrika* **50** 403–418. MR0163386 https://doi.org/10.1093/biomet/50.3-4.403

[50] KUR, G., GAO, F., GUNTUBOYINA, A. and SEN, B. (2020). Convex regression in multidimensions: Suboptimality of least squares estimators. Preprint. Available at arXiv:2006.02044.

[51] LAWSON, C. L. and HANSON, R. J. (1995). *Solving Least Squares Problems*. *Classics in Applied Mathematics* **15**. SIAM, Philadelphia, PA. Revised reprint of the 1974 original. MR1349828 https://doi.org/10.1137/1.9781611971217

[52] LELARGE, M. and MIOLANE, L. (2019). Fundamental limits of symmetric low-rank matrix estimation. *Probab. Theory Related Fields* **173** 859–929. MR3936148 https://doi.org/10.1007/s00440-018-0845-x

[53] LIANG, T. and SUR, P. (2022). A precise high-dimensional asymptotic theory for boosting and minimum-$\ell 1$-norm interpolated classifiers. *Ann. Statist.* **50** 1669–1695. MR4441136 https://doi.org/10.1214/22-aos2170

[54] 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

[55] METZLER, C. A., MALEKI, A. and BARANIUK, R. G. (2016). From denoising to compressed sensing. *IEEE Trans. Inf. Theory* **62** 5117–5144. MR3549470 https://doi.org/10.1109/TIT.2016.2556683

[56] MEYER, M. and WOODROOFE, M. (2000). On the degrees of freedom in shape-restricted regression. *Ann. Statist.* **28** 1083–1104. MR1810920 https://doi.org/10.1214/aos/1015956708

[57] MIOLANE, L. and MONTANARI, A. (2021). The distribution of the Lasso: Uniform control over sparse balls and adaptive parameter tuning. *Ann. Statist.* **49** 2313–2335. MR4319252 https://doi.org/10.1214/20-aos2038

[58] MONTANARI, A. (2018). Mean field asymptotics in high-dimensional statistics: From exact results to efficient algorithms. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro* 2018. *Vol. IV. Invited Lectures* 2973–2994. World Sci. Publ., Hackensack, NJ. MR3966519

[59] MONTANARI, A., RUAN, F., SOHN, Y. and YAN, J. (2023). The generalization error of max-margin linear classifiers: Benign overfitting and high-dimensional asymptotics in the overparametrized regime. Preprint. Available at arXiv:1911.01544v3.

[60] OYMAK, S. and HASSIBI, B. (2016). Sharp MSE bounds for proximal denoising. *Found. Comput. Math.* **16** 965–1029. MR3529131 https://doi.org/10.1007/s10208-015-9278-4

[61] OYMAK, S., RECHT, B. and SOLTANOLKOTABI, M. (2018). Sharp time-data tradeoffs for linear inverse problems. *IEEE Trans. Inf. Theory* **64** 4129–4158. MR3809731 https://doi.org/10.1109/TIT.2017.2773497

[62] OYMAK, S., THRAMPOULIDIS, C. and HASSIBI, B. (2013). The squared-error of generalized lasso: A precise analysis. In 2013 51*st Annual Allerton Conference on Communication*, *Control*, *and Computing* (*Allerton*) 1002–1009. IEEE.

[63] OYMAK, S. and TROPP, J. A. (2018). Universality laws for randomized dimension reduction, with applications. *Inf. Inference* **7** 337–446. MR3858331 https://doi.org/10.1093/imaiai/iax011

[64] RAUBERTAS, R. F., LEE, C.-I. C. and NORDHEIM, E. V. (1986). Hypothesis tests for normal means constrained by linear inequalities. *Comm. Statist. Theory Methods* **15** 2809–2833. MR0855765 https://doi.org/10.1080/03610928608829280

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

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

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

[68] THRAMPOULIDIS, C., OYMAK, S. and HASSIBI, B. (2014). Simple error bounds for regularized noisy linear inverse problems. In 2014 *IEEE International Symposium on Information Theory* 3007–3011. IEEE.

[69] THRAMPOULIDIS, C., OYMAK, S. and HASSIBI, B. (2015). Recovering structured signals in noise: Least-squares meets compressed sensing. In *Compressed Sensing and Its Applications*. *Appl. Numer. Harmon. Anal.* 97–141. Birkhäuser/Springer, Cham. MR3382104

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

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

[72] TROPP, J. A. (2015). Convex recovery of a structured signal from independent random linear measurements. In *Sampling Theory*, *a Renaissance*. *Appl. Numer. Harmon. Anal.* 67–101. Birkhäuser/Springer, Cham. MR3467419

[73] VAN DE GEER, S. and WAINWRIGHT, M. J. (2017). On concentration for (regularized) empirical risk minimization. *Sankhya A* **79** 159–200. MR3707417 https://doi.org/10.1007/s13171-017-0111-9

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

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

[76] WAKEFIELD, J. (2013). *Bayesian and Frequentist Regression Methods*. *Springer Series in Statistics*. Springer, New York. MR3025225 https://doi.org/10.1007/978-1-4419-0925-1

[77] WEI, Y., FANG, B. and WAINWRIGHT, M. J. (2020). From Gauss to Kolmogorov: Localized measures of complexity for ellipses. *Electron*. *J*. *Stat*. **14** 2988–3031. MR4135323 https://doi.org/10.1214/20-EJS1739

[78] ZDEBOROVÁ, L. and KRZAKALA, F. (2016). Statistical physics of inference: Thresholds and algorithms. *Adv. Phys*. **65** 453–552.

[79] ZHANG, C.-H. (2002). Risk bounds in isotonic regression. *Ann*. *Statist*. **30** 528–555. MR1902898 https://doi.org/10.1214/aos/1021379864

The Annals of Statistics
2023, Vol. 51, No. 4, 1639–1665
https://doi.org/10.1214/23-AOS2302
© Institute of Mathematical Statistics, 2023

PROJECTED STATE-ACTION BALANCING WEIGHTS FOR OFFLINE REINFORCEMENT LEARNING

BY JIAYI WANG[1,a], ZHENGLING QI[2,b] AND RAYMOND K. W. WONG[3,c]

[1]*Department of Mathematical Sciences, University of Texas at Dallas,* [a]*jiayi.wang2@utdallas.edu*
[2]*Department of Decision Sciences, George Washington University,* [b]*qizhengling@gwu.edu*
[3]*Department of Statistics, Texas A&M University,* [c]*raywong@tamu.edu*

Off-policy evaluation is considered a fundamental and challenging problem in reinforcement learning (RL). This paper focuses on value estimation of a target policy based on pre-collected data generated from a possibly different policy, under the framework of infinite-horizon Markov decision processes. Motivated by the recently developed marginal importance sampling method in RL and the covariate balancing idea in causal inference, we propose a novel estimator with approximately projected state-action balancing weights for the policy value estimation. We obtain the convergence rate of these weights, and show that the proposed value estimator is asymptotically normal under technical conditions. In terms of asymptotics, our results scale with both the number of trajectories and the number of decision points at each trajectory. As such, consistency can still be achieved with a limited number of subjects when the number of decision points diverges. In addition, we develop a necessary and sufficient condition for establishing the well-posedness of the operator that relates to the nonparametric Q-function estimation in the off-policy setting, which characterizes the difficulty of Q-function estimation and may be of independent interest. Numerical experiments demonstrate the promising performance of our proposed estimator.

REFERENCES

[1] ATHEY, S., IMBENS, G. W. and WAGER, S. (2018). Approximate residual balancing: Debiased inference of average treatment effects in high dimensions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 597–623. MR3849336 https://doi.org/10.1111/rssb.12268

[2] BELLONI, A., CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Some new asymptotic theory for least squares series: Pointwise and uniform results. *J. Econometrics* **186** 345–366. MR3343791 https://doi.org/10.1016/j.jeconom.2015.02.014

[3] BROCKMAN, G., CHEUNG, V., PETTERSSON, L., SCHNEIDER, J., SCHULMAN, J., TANG, J. and ZAREMBA, W. (2016). Openai gym. ArXiv preprint. Available at arXiv:1606.01540.

[4] CHAN, K. C. G., YAM, S. C. P. and ZHANG, Z. (2016). Globally efficient non-parametric inference of average treatment effects by empirical balancing calibration weighting. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 673–700. MR3506798 https://doi.org/10.1111/rssb.12129

[5] CHAN, K. S. (1989). A note on the geometric ergodicity of a Markov chain. *Adv. in Appl. Probab.* **21** 702–704. MR1013658 https://doi.org/10.2307/1427643

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

[7] CHEN, X. and CHRISTENSEN, T. M. (2018). Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression. *Quant. Econ.* **9** 39–84. MR3789729 https://doi.org/10.3982/QE722

[8] CHEN, X. and QI, Z. (2022). On well-posedness and minimax optimal rates of nonparametric Q-function estimation in off-policy evaluation. ArXiv preprint. Available at arXiv:2201.06169.

[9] CHUA, K., CALANDRA, R., MCALLISTER, R. and LEVINE, S. (2018). Deep reinforcement learning in a handful of trials using probabilistic dynamics models. *Adv. Neural Inf. Process. Syst.* **31**.

MSC2020 subject classifications. 62G05, 62M05.
Key words and phrases. Infinite horizons, Markov decision process, Policy evaluation, Reinforcement learning.

[10] DE BOOR, C. (1976). Splines as linear combinations of B-splines. A survey. Technical Report, Wisconsin Univ. Madison Mathematics Research Center.

[11] FAN, J., IMAI, K., LIU, H., NING, Y. and YANG, X. (2016). Improving covariate balancing propensity score: A doubly robust and efficient approach Technical Report Technical report, Princeton Univ.

[12] FARAHMAND, A., GHAVAMZADEH, M., SZEPESVÁRI, C. and MANNOR, S. (2016). Regularized policy iteration with nonparametric function spaces. *J. Mach. Learn. Res.* **17** Paper No. 139. MR3555030

[13] FARAHMAND, A. and SZEPESVÁRI, C. (2012). Regularized least-squares regression: Learning from a β-mixing sequence. *J. Statist. Plann. Inference* **142** 493–505. MR2843052 https://doi.org/10.1016/j.jspi.2011.08.007

[14] GEER, S. A. and VAN DE GEER, S. (2000). *Empirical Processes in M-Estimation* **6**. Cambridge university press.

[15] 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

[16] HANSEN, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50** 1029–1054. MR0666123 https://doi.org/10.2307/1912775

[17] HEARST, M. A., DUMAIS, S. T., OSUNA, E., PLATT, J. and SCHOLKOPF, B. (1998). Support vector machines. *IEEE Intell. Syst. Appl.* **13** 18–28.

[18] HOROWITZ, J. L. and MAMMEN, E. (2004). Nonparametric estimation of an additive model with a link function. *Ann. Statist.* **32** 2412–2443. MR2153990 https://doi.org/10.1214/009053604000000814

[19] HU, X., QIAN, M., CHENG, B. and CHEUNG, Y. K. (2021). Personalized policy learning using longitudinal mobile health data. *J. Amer. Statist. Assoc.* **116** 410–420. MR4227703 https://doi.org/10.1080/01621459.2020.1785476

[20] IMAI, K. and RATKOVIC, M. (2014). Covariate balancing propensity score. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 243–263. MR3153941 https://doi.org/10.1111/rssb.12027

[21] JANNER, M., FU, J., ZHANG, M. and LEVINE, S. (2019). When to trust your model: Model-based policy optimization. *Adv. Neural Inf. Process. Syst.* **32**.

[22] KALLUS, N. (2020). Generalized optimal matching methods for causal inference. *J. Mach. Learn. Res.* **21** Paper No. 62. MR4095341

[23] KALLUS, N. and UEHARA, M. (2020). Double reinforcement learning for efficient off-policy evaluation in Markov decision processes. *J. Mach. Learn. Res.* **21** Paper No. 167. MR4209453

[24] KALLUS, N. and UEHARA, M. (2022). Efficiently breaking the curse of horizon in off-policy evaluation with double reinforcement learning. *Oper. Res.* **70** 3282–3302. MR4538517

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

[26] KOSOROK, M. R. and LABER, E. B. (2019). Precision medicine. *Annu. Rev. Stat. Appl.* **6** 263–286. MR3939521 https://doi.org/10.1146/annurev-statistics-030718-105251

[27] LABER, E. B., LIZOTTE, D. J., QIAN, M., PELHAM, W. E. and MURPHY, S. A. (2014). Dynamic treatment regimes: Technical challenges and applications. *Electron. J. Stat.* **8** 1225–1272. MR3263118 https://doi.org/10.1214/14-EJS920

[28] LE, H., VOLOSHIN, C. and YUE, Y. (2019). Batch policy learning under constraints. In *International Conference on Machine Learning* 3703–3712.

[29] 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.

[30] LIAO, P., DEMPSEY, W., SARKER, H., HOSSAIN, S. M., AL'ABSI, M., KLASNJA, P. and MURPHY, S. (2018). Just-in-time but not too much: Determining treatment timing in mobile health. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2** 179.

[31] LIAO, P., KLASNJA, P. and MURPHY, S. (2021). Off-policy estimation of long-term average outcomes with applications to mobile health. *J. Amer. Statist. Assoc.* **116** 382–391. MR4227701 https://doi.org/10.1080/01621459.2020.1807993

[32] LIAO, P., QI, Z., WAN, R., KLASNJA, P. and MURPHY, S. A. (2022). Batch policy learning in average reward Markov decision processes. *Ann. Statist.* **50** 3364–3387. MR4524500 https://doi.org/10.1214/22-aos2231

[33] LIU, Q., LI, L., TANG, Z. and ZHOU, D. (2018). Breaking the curse of horizon: Infinite-horizon off-policy estimation. In *Advances in Neural Information Processing Systems* 5356–5366.

[34] LUCKETT, D. J., LABER, E. B., KAHKOSKA, A. R., MAAHS, D. M., MAYER-DAVIS, E. and KOSOROK, M. R. (2020). Estimating dynamic treatment regimes in mobile health using V-learning. *J. Amer. Statist. Assoc.* **115** 692–706. MR4107673 https://doi.org/10.1080/01621459.2018.1537919

[35] MARCOLINO, M. S., OLIVEIRA, J. A. Q., D'AGOSTINO, M., RIBEIRO, A. L., ALKMIM, M. B. M. and NOVILLO-ORTIZ, D. (2018). The impact of mHealth interventions: Systematic review of systematic reviews. *JMIR mHealth uHealth* **6** e23. https://doi.org/10.2196/mhealth.8873

[36] MARLING, C. and BUNESCU, R. (2020). The OhioT1DM dataset for blood glucose level prediction: Update 2020. In *CEUR Workshop Proceedings* **2675** 71.

[37] MEYN, S., TWEEDIE, R. and HIBEY, J. (1995). Markov chains and stochastic stability. *IEEE Trans. Automat. Control* **40** 979.

[38] MURPHY, S. A. (2003). Optimal dynamic treatment regimes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 331–366. MR1983752 https://doi.org/10.1111/1467-9868.00389

[39] MURPHY, S. A., VAN DER LAAN, M. J. and ROBINS, J. M. (2001). Marginal mean models for dynamic regimes. *J. Amer. Statist. Assoc.* **96** 1410–1423. MR1946586 https://doi.org/10.1198/016214501753382327

[40] NACHUM, O., CHOW, Y., DAI, B. and LI, L. (2019). Dualdice: Behavior-agnostic estimation of discounted stationary distribution corrections. In *Advances in Neural Information Processing Systems* 2315–2325.

[41] NAHUM-SHANI, I., SMITH, S. N., SPRING, B. J., COLLINS, L. M., WITKIEWITZ, K., TEWARI, A. and MURPHY, S. A. (2016). Just-in-Time Adaptive Interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. *Ann. Behav. Med.* 1–17.

[42] NEWEY, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *J. Econometrics* **79** 147–168. MR1457700 https://doi.org/10.1016/S0304-4076(97)00011-0

[43] NEWEY, W. K. and POWELL, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica* **71** 1565–1578. MR2000257 https://doi.org/10.1111/1468-0262.00459

[44] PADURARU, C. (2007). Planning with approximate and learned models of Markov decision processes. These de Maıtre, University of Alberta.

[45] PRECUP, D. (2000). Eligibility traces for off-policy policy evaluation. *Computer Science Department Faculty Publication Series* 80.

[46] PUTERMAN, M. L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Wiley, New York. MR1270015

[47] ROBINS, J. M., HERNAN, M. A. and BRUMBACK, B. (2000). Marginal structural models and causal inference in epidemiology.

[48] ROBINS, J. M., ROTNITZKY, A. and ZHAO, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89** 846–866. MR1294730

[49] RODBARD, D. (2009). Interpretation of continuous glucose monitoring data: Glycemic variability and quality of glycemic control. *Diabetes Technology & Therapeutics* **11** S–55.

[50] SHI, C., FAN, A., SONG, R. and LU, W. (2018). High-dimensional A-learning for optimal dynamic treatment regimes. *Ann. Statist.* **46** 925–957. MR3797992 https://doi.org/10.1214/17-AOS1570

[51] SHI, C., WAN, R., CHERNOZHUKOV, V. and SONG, R. (2021). Deeply-debiased off-policy interval estimation. ArXiv preprint. Available at arXiv:2105.04646.

[52] SHI, C., WAN, R., SONG, R., LU, W. and LENG, L. (2020). Does the Markov decision process fit the data: Testing for the Markov property in sequential decision making. ArXiv preprint. Available at arXiv:2002.01751.

[53] SHI, C., ZHANG, S., LU, W. and SONG, R. (2022). Statistical inference of the value function for reinforcement learning in infinite-horizon settings. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 765–793. MR4460575

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

[55] TANG, Z., FENG, Y., LI, L., ZHOU, D. and LIU, Q. (2019). Doubly robust bias reduction in infinite horizon off-policy estimation. ArXiv preprint. Available at arXiv:1910.07186.

[56] UEHARA, M., HUANG, J. and JIANG, N. (2020). Minimax weight and q-function learning for off-policy evaluation. In *International Conference on Machine Learning* 9659–9668. PMLR.

[57] UEHARA, M., IMAIZUMI, M., JIANG, N., KALLUS, N., SUN, W. and XIE, T. (2021). Finite sample analysis of minimax offline reinforcement learning: Completeness, fast rates and first-order efficiency. ArXiv preprint. Available at arXiv:2102.02981.

[58] WAHBA, G. (1990). *Spline Models for Observational Data. CBMS-NSF Regional Conference Series in Applied Mathematics* **59**. SIAM, Philadelphia, PA. MR1045442 https://doi.org/10.1137/1.9781611970128

[59] WANG, J., QI, Z. and WONG, R. K. W. (2023). Supplement to "Projected state-action balancing weights for offline reinforcement learning." https://doi.org/10.1214/23-AOS2302SUPP

[60] WANG, L., ZHOU, Y., SONG, R. and SHERWOOD, B. (2018). Quantile-optimal treatment regimes. *J. Amer. Statist. Assoc.* **113** 1243–1254. MR3862354 https://doi.org/10.1080/01621459.2017.1330204

[61] WANG, Y. and ZUBIZARRETA, J. R. (2020). Minimal dispersion approximately balancing weights: Asymptotic properties and practical considerations. *Biometrika* **107** 93–105. MR4064142 https://doi.org/10.1093/biomet/asz050

[62] WONG, R. K. W. and CHAN, K. C. G. (2018). Kernel-based covariate functional balancing for observational studies. *Biometrika* **105** 199–213. MR3768874 https://doi.org/10.1093/biomet/asx069

[63] ZHANG, R., DAI, B., LI, L. and SCHUURMANS, D. (2020). Gen{DICE}: Generalized offline estimation of stationary values. In *International Conference on Learning Representations*.

[64] ZHAO, Y.-Q., ZENG, D., LABER, E. B. and KOSOROK, M. R. (2015). New statistical learning methods for estimating optimal dynamic treatment regimes. *J. Amer. Statist. Assoc.* **110** 583–598. MR3367249 https://doi.org/10.1080/01621459.2014.937488

The Annals of Statistics
2023, Vol. 51, No. 4, 1666–1691
https://doi.org/10.1214/23-AOS2303
© Institute of Mathematical Statistics, 2023

POST-SELECTION INFERENCE VIA ALGORITHMIC STABILITY

BY TIJANA ZRNIC[a] AND MICHAEL I. JORDAN[b]

Department of Electrical Engineering and Computer Sciences, University of California, Berkeley,
[a]*tijana.zrnic@berkeley.edu,* [b]*jordan@cs.berkeley.edu*

When the target of statistical inference is chosen in a data-driven manner, the guarantees provided by classical theories vanish. We propose a solution to the problem of inference after selection by building on the framework of *algorithmic stability*, in particular its branch with origins in the field of differential privacy. Stability is achieved via *randomization* of selection and it serves as a quantitative measure that is sufficient to obtain nontrivial post-selection corrections for classical confidence intervals. Importantly, the underpinnings of algorithmic stability translate directly into computational efficiency—our method computes simple corrections for selective inference without recourse to Markov chain Monte Carlo sampling.

REFERENCES

[1] ANDREWS, I., KITAGAWA, T. and MCCLOSKEY, A. (2019). Inference on winners. National Bureau of Economic Research.

[2] BACHOC, F., PREINERSTORFER, D. and STEINBERGER, L. (2020). Uniformly valid confidence intervals post-model-selection. *Ann. Statist.* **48** 440–463. MR4065169 https://doi.org/10.1214/19-AOS1815

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

[4] BASSILY, R. and FREUND, Y. (2016). Typical stability. Preprint. Available at arXiv:1604.03336.

[5] BASSILY, R., NISSIM, K., SMITH, A., STEINKE, T., STEMMER, U. and ULLMAN, J. (2016). Algorithmic stability for adaptive data analysis. In *STOC'16—Proceedings of the* 48*th Annual ACM SIGACT Symposium on Theory of Computing* 1046–1059. ACM, New York. MR3536635

[6] BEIMEL, A., KASIVISWANATHAN, S. P. and NISSIM, K. (2010). Bounds on the sample complexity for private learning and private data release. In *Theory of Cryptography*. *Lecture Notes in Computer Science* **5978** 437–454. Springer, Berlin. MR2673387 https://doi.org/10.1007/978-3-642-11799-2_26

[7] BENJAMINI, Y. (2010). Simultaneous and selective inference: Current successes and future challenges. *Biom. J.* **52** 708–721. MR2758547 https://doi.org/10.1002/bimj.200900299

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

[9] 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

[10] BI, N., MARKOVIC, J., XIA, L. and TAYLOR, J. (2020). Inferactive data analysis. *Scand. J. Stat.* **47** 212–249. MR4075236 https://doi.org/10.1111/sjos.12425

[11] BOUSQUET, O. and ELISSEEFF, A. (2002). Stability and generalization. *J. Mach. Learn. Res.* **2** 499–526. MR1929416 https://doi.org/10.1162/153244302760200704

[12] BUJA, A., BROWN, L., BERK, R., GEORGE, E., PITKIN, E., TRASKIN, M., ZHANG, K. and ZHAO, L. (2019). Models as approximations I: Consequences illustrated with linear regression. *Statist. Sci.* **34** 523–544. MR4048582 https://doi.org/10.1214/18-STS693

[13] CUMMINGS, R., LIGETT, K., NISSIM, K., ROTH, A. and WU, Z. S. (2016). Adaptive learning with robust generalization guarantees. In *Conference on Learning Theory* (*COLT*) 772–814.

[14] DURFEE, D. and ROGERS, R. M. (2019). Practical differentially private top-k selection with pay-what-you-get composition. In *Advances in Neural Information Processing Systems* (*NeurIPS*) 3532–3542.

[15] DWORK, C., FELDMAN, V., HARDT, M., PITASSI, T. and REINGOLD, O. (2015). Generalization in adaptive data analysis and holdout reuse. In *Advances in Neural Information Processing Systems* (*NIPS*) 2350–2358.

MSC2020 subject classifications. Primary 62J15; secondary 62J05, 62F07.
Key words and phrases. Post-selection inference, selective inference, stability, differential privacy, model selection, linear regression.

[16] DWORK, C., FELDMAN, V., HARDT, M., PITASSI, T., REINGOLD, O. and ROTH, A. (2015). Preserving statistical validity in adaptive data analysis [extended abstract]. In *STOC'15—Proceedings of the* 2015 *ACM Symposium on Theory of Computing* 117–126. ACM, New York. MR3388189
[17] DWORK, C., MCSHERRY, F., NISSIM, K. and SMITH, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography*. *Lecture Notes in Computer Science* **3876** 265–284. Springer, Berlin. MR2241676 https://doi.org/10.1007/11681878_14
[18] DWORK, C. and ROTH, A. (2013). The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.* **9** 211–487. MR3254020 https://doi.org/10.1561/0400000042
[19] DWORK, C., SU, W. and ZHANG, L. (2015). Private false discovery rate control. Available at arXiv:1511.03803.
[20] FAN, J. and LV, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 849–911. MR2530322 https://doi.org/10.1111/j.1467-9868.2008.00674.x
[21] FITHIAN, W., SUN, D. and TAYLOR, J. (2014). Optimal inference after model selection. Available at arXiv:1410.2597.
[22] FRANK, M. and WOLFE, P. (1956). An algorithm for quadratic programming. *Nav. Res. Logist. Q.* **3** 95–110. MR0089102 https://doi.org/10.1002/nav.3800030109
[23] GUYON, I. and ELISSEEFF, A. (2003). An introduction to variable and feature selection. *J. Mach. Learn. Res.* **3** 1157–1182.
[24] JAGGI, M. (2013). Revisiting Frank-Wolfe: Projection-free sparse convex optimization. In *Proceedings of the* 30*th International Conference on Machine Learning* 427–435.
[25] KASIVISWANATHAN, S. P., LEE, H. K., NISSIM, K., RASKHODNIKOVA, S. and SMITH, A. (2011). What can we learn privately? *SIAM J. Comput.* **40** 793–826. MR2823508 https://doi.org/10.1137/090756090
[26] KIVARANOVIC, D. and LEEB, H. (2020). A (tight) upper bound for the length of confidence intervals with conditional coverage. Available at arXiv:2007.12448.
[27] 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
[28] KUCHIBHOTLA, A. K., BROWN, L. D., BUJA, A. and CAI, J. (2019). All of linear regression. Preprint. Available at arXiv:1910.06386.
[29] KUCHIBHOTLA, A. K., BROWN, L. D., BUJA, A., CAI, J., GEORGE, E. I. and ZHAO, L. H. (2020). Valid post-selection inference in model-free linear regression. *Ann. Statist.* **48** 2953–2981. MR4152630 https://doi.org/10.1214/19-AOS1917
[30] 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
[31] LEE, J. D. and TAYLOR, J. E. (2014). Exact post model selection inference for marginal screening. In *NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems*. **1** 136–144. https://doi.org/10.5555/2968826.2968842
[32] LEI, J., CHAREST, A.-S., SLAVKOVIC, A., SMITH, A. and FIENBERG, S. (2018). Differentially private model selection with penalized and constrained likelihood. *J. Roy. Statist. Soc. Ser. A* **181** 609–633. MR3807500 https://doi.org/10.1111/rssa.12324
[33] LIU, K., MARKOVIC, J. and TIBSHIRANI, R. (2018). More powerful post-selection inference, with application to the Lasso. Preprint. Available at arXiv:1801.09037.
[34] MARKOVIC, J. and TAYLOR, J. (2016). Bootstrap inference after using multiple queries for model selection. Preprint. Available at arXiv:1612.07811.
[35] KUCHIBHOTLA, A. K., RINALDO, A. and WASSERMAN, L. (2020). Berry-Esseen bounds for projection parameters and partial correlations with increasing dimension. Preprint. Available at arXiv:2007.09751.
[36] PANIGRAHI, S., MARKOVIC, J. and TAYLOR, J. (2017). An MCMC-free approach to post-selective inference. Available at arXiv:1703.06154.
[37] PANIGRAHI, S. and TAYLOR, J. (2022). Approximate selective inference via maximum likelihood. *J. Amer. Statist. Assoc.* 1–11.
[38] RASINES, D. G. and YOUNG, G. A. (2023). Splitting strategies for post-selection inference. *Biometrika* **110** 597–614. MR4627773 https://doi.org/10.1093/biomet/asac070
[39] RINALDO, A., WASSERMAN, L. and G'SELL, M. (2019). Bootstrapping and sample splitting for high-dimensional, assumption-lean inference. *Ann. Statist.* **47** 3438–3469. MR4025748 https://doi.org/10.1214/18-AOS1784
[40] ROGERS, R., ROTH, A., SMITH, A. and THAKKAR, O. (2016). Max-information, differential privacy, and post-selection hypothesis testing. In 57*th Annual IEEE Symposium on Foundations of*

Computer Science—FOCS 2016 487–494. IEEE Computer Soc., Los Alamitos, CA. MR3631011 https://doi.org/10.1109/FOCS.2016.59

[41] ROSENTHAL, R. (1979). The file drawer problem and tolerance for null results. *Psychol*. *Bull*. **86** 638.

[42] RUSSO, D. and ZOU, J. (2016). Controlling bias in adaptive data analysis using information theory. In *Artificial Intelligence and Statistics* 1232–1240.

[43] SCHEFFÉ, H. (1999). *The Analysis of Variance*. *Wiley Classics Library*. Wiley, New York. MR1673563

[44] STEINBERGER, L. and LEEB, H. (2023). Conditional predictive inference for stable algorithms. *Ann*. *Statist*. **51** 290–311. MR4564857 https://doi.org/10.1214/22-aos2250

[45] STEINKE, T. and ULLMAN, J. (2017). Tight lower bounds for differentially private selection. In 58*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 552–563. IEEE Computer Soc., Los Alamitos, CA. MR3734260 https://doi.org/10.1109/FOCS.2017.57

[46] TALWAR, K., THAKURTA, A. G. and ZHANG, L. (2015). Nearly optimal private LASSO. In *Advances in Neural Information Processing Systems* (*NIPS*) 3025–3033.

[47] TAYLOR, J. and TIBSHIRANI, R. J. (2015). Statistical learning and selective inference. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **112** 7629–7634. MR3371123 https://doi.org/10.1073/pnas.1507583112

[48] THAKURTA, A. G. and SMITH, A. (2013). Differentially private feature selection via stability arguments, and the robustness of the lasso. In *Conference on Learning Theory* (*COLT*) 819–850.

[49] TIAN HARRIS, X., PANIGRAHI, S., MARKOVIC, J., BI, N. and TAYLOR, J. (2016). Selective sampling after solving a convex problem. Preprint. Available at arXiv:1609.05609.

[50] TIAN, X., BI, N. and TAYLOR, J. (2017). MAGIC: a general, powerful and tractable method for selective inference. Preprint. Available at arXiv:1607.02630.

[51] 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

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

[53] 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

[54] ZRNIC, T. and JORDAN, M. I (2023). Supplement to "Post-selection inference via algorithmic stability." https://doi.org/10.1214/23-AOS2303SUPP

The Annals of Statistics
2023, Vol. 51, No. 4, 1692–1717
https://doi.org/10.1214/23-AOS2304
© Institute of Mathematical Statistics, 2023

BRIDGING FACTOR AND SPARSE MODELS

BY JIANQING FAN[1,a] RICARDO P. MASINI[2,b] AND MARCELO C. MEDEIROS[3,c]

[1]*Department of Operations Research and Financial Engineering, Princeton University,* [a]*jqfan@princeton.edu*
[2]*Department of Statistics, University of California, Davis,* [b]*rmasini@ucdavis.edu*
[3]*Department of Economics, University of Illinois at Urbana-Champaign,* [c]*marcelom@illinois.edu*

Factor and sparse models are widely used to impose a low-dimensional structure in high-dimensions. However, they are seemingly mutually exclusive. We propose a lifting method that combines the merits of these two models in a supervised learning methodology that allows for efficiently exploring all the information in high-dimensional datasets. The method is based on a flexible model for high-dimensional panel data with observable and/or latent common factors and idiosyncratic components. The model is called the factor-augmented regression model. It includes principal components and sparse regression as specific models, significantly weakens the cross-sectional dependence, and facilitates model selection and interpretability. The method consists of several steps and a novel test for (partial) covariance structure in high dimensions to infer the remaining cross-section dependence at each step. We develop the theory for the model and demonstrate the validity of the multiplier bootstrap for testing a high-dimensional (partial) covariance structure. A simulation study and applications support the theory.

REFERENCES

ABADIE, A. and GARDEAZABAL, J. (2003). The economic costs of conflict: A case study of the Basque country. *Amer. Econ. Rev.* **93** 113–132.

AGARWAL, A., NEGAHBAN, S. and WAINWRIGHT, M. J. (2012). Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions. *Ann. Statist.* **40** 1171–1197. MR2985947 https://doi.org/10.1214/12-AOS1000

ANDREOU, E. and GHYSELS, E. (2021). Predicting the VIX and the volatility risk premium: The role of short-run funding spreads volatility factors. *J. Econometrics* **220** 366–398. MR4201495 https://doi.org/10.1016/j.jeconom.2020.04.006

ANDREWS, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59** 817–858. MR1106513 https://doi.org/10.2307/2938229

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

BAI, J. (2009). Panel data models with interactive fixed effects. *Econometrica* **77** 1229–1279. MR2547073 https://doi.org/10.3982/ECTA6135

BAI, J. and LIAO, Y. (2017). Inferences in panel data with interactive effects using large covariance matrices. *J. Econometrics* **200** 59–78. MR3683662 https://doi.org/10.1016/j.jeconom.2017.05.014

BAI, J. and NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70** 191–221. MR1926259 https://doi.org/10.1111/1468-0262.00273

BAI, J. and NG, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica* **74** 1133–1150. MR2238213 https://doi.org/10.1111/j.1468-0262.2006.00696.x

BARIGOZZI, M. and BROWNLEES, C. (2019). NETS: Network estimation for time series. *J. Appl. Econometrics* **34** 347–364. MR3948470 https://doi.org/10.1002/jae.2676

BARIGOZZI, M., CHO, H. and OWENS, D. (2022). FNETS: Factor-adjusted network estimation and forecasting for high-dimensional time series Technical Report No. 2201.06110.

BARIGOZZI, M. and HALLIN, M. (2017). A network analysis of the volatility of high dimensional financial series. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **66** 581–605. MR3632343 https://doi.org/10.1111/rssc.12177

MSC2020 subject classifications. 62H25, 62F03.
Key words and phrases. Factor models, penalized least-squares, high-dimensional inference, (partial) covariance structure, prediction, machine learning, supervised learning.

BERNANKE, B. S., BOIVIN, J. and ELIASZ, P. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *Q. J. Econ.* **120** 387–422.

BROWNLEES, C., GUDMUNDSSON, G. S. and LUGOSI, G. (2020). Community detection in partial correlation network models. *J. Bus. Econom. Statist.* forthcoming.

CAI, T. T. (2017). Global testing and large-scale multiple testing for high-dimensional covariance structures. *Annu. Rev. Stat. Appl.* **4** 4.1–4.24.

CARVALHO, C., MASINI, R. and MEDEIROS, M. C. (2018). ArCo: An artificial counterfactual approach for high-dimensional panel time-series data. *J. Econometrics* **207** 352–380. MR3868187 https://doi.org/10.1016/j.jeconom.2018.07.005

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. MR3161448 https://doi.org/10.1214/13-AOS1161

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2019). Inference on causal and structural parameters using many moment inequalities. *Rev. Econ. Stud.* **86** 1867–1900. MR4009488 https://doi.org/10.1093/restud/rdy065

CHERNOZHUKOV, V., CHETVERIKOV, D. and KOIKE, Y. (2020). Nearly optimal central limit theorem and bootstrap approximations in high dimensions.

CHERNOZHUOKOV, V., CHETVERIKOV, D., KATO, K. and KOIKE, Y. (2022). Improved central limit theorem and bootstrap approximations in high dimensions. *Ann. Statist.* **50** 2562–2586. MR4500619 https://doi.org/10.1214/22-aos2193

DIEBOLD, F. X. and YILMAZ, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *J. Econometrics* **182** 119–134. MR3212765 https://doi.org/10.1016/j.jeconom.2014.04.012

FAMA, E. F. and FRENCH, K. R. (1993). Common risk factors in the returns on stocks and bonds. *J. Financ. Econ.* **33** 3–56.

FAMA, E. F. and FRENCH, K. R. (2015). A five-factor asset pricing model. *J. Financ. Econ.* **116** 1–22.

FAN, J., KE, Y. and WANG, K. (2020). Factor-adjusted regularized model selection. *J. Econometrics* **216** 71–85. MR4077382 https://doi.org/10.1016/j.jeconom.2020.01.006

FAN, J., LI, Q. and WANG, Y. (2017). Estimation of high dimensional mean regression in the absence of symmetry and light tail assumptions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 247–265. MR3597972 https://doi.org/10.1111/rssb.12166

FAN, J., LI, R., ZHANG, C. H. and ZOU, H. (2020). *Statistical Foundations of Data Science*. CRC Press, Boca Raton.

FAN, J., LIAO, Y. and MINCHEVA, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 603–680. MR3091653 https://doi.org/10.1111/rssb.12016

FAN, J., MASINI, R. and MEDEIROS, M. C. (2022). Do we exploit all information for counterfactual analysis? Benefits of factor models and idiosyncratic correction. *J. Amer. Statist. Assoc.* **117** 574–590. MR4436297 https://doi.org/10.1080/01621459.2021.2004895

FAN, J., MASINI, R. P. and MEDEIROS, M. C. (2023). Supplement to "Bridging factor and sparse models." https://doi.org/10.1214/23-AOS2304SUPP

FENG, G., GIGLIO, S. and XIU, D. (2020). Taming the factor zoo: A test of new factors. *J. Finance* **75** 1327–1370.

FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Rev. Econ. Stat.* **82** 540–554.

FORNI, M., HALLIN, M., LIPPI, M. and ZAFFARONI, P. (2017). Dynamic factor models with infinite-dimensional factor space: Asymptotic analysis. *J. Econometrics* **199** 74–92. MR3652406 https://doi.org/10.1016/j.jeconom.2017.04.002

GAGLIARDINI, P., OSSOLA, E. and SCAILLET, O. (2019). A diagnostic criterion for approximate factor structure. *J. Econometrics* **212** 503–521. MR4001672 https://doi.org/10.1016/j.jeconom.2019.06.001

GAGLIARDINI, P., OSSOLA, E. and SCAILLET, P. (2020). Estimation of large dimensional conditional factor models in finance. In *Handbook of Econometrics* (S. Durlauf, L. Hansen, J. Heckman and R. Matzkin, eds.) **7A** 219–282.

GIANNONE, D., LENZA, M. and PRIMICERI, G. (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica* **89** 2409–2437.

GIESSING, A. and FAN, J. (2020). Bootstrapping ℓ_p-statistics in high dimensions.

GOBILLON, L. and MAGNAC, T. (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. *Rev. Econ. Stat.* **98** 535–551.

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

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

LEDOIT, O. and WOLF, M. (2021). The power of (non)linear shrinking: A review and guide to covariance matrix estimation. *J. Financ. Econom.*. forthcoming.

LOPES, M. E. (2022). Central limit theorem and bootstrap approximation in high dimensions: Near $1/\sqrt{n}$ rates via implicit smoothing. *Ann. Statist.* **50** 2492–2513. MR4505371 https://doi.org/10.1214/22-aos2184

MASINI, R. P., MEDEIROS, M. C. and MENDES, E. F. (2022). Regularized estimation of high-dimensional vector autoregressions with weakly dependent innovations. *J. Time Series Anal.* **43** 532–557. MR4446423 https://doi.org/10.1111/jtsa.12627

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

MOON, H. R. and WEIDNER, M. (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica* **83** 1543–1579. MR3384227 https://doi.org/10.3982/ECTA9382

MOSKOWITZ, T. J. and GRINBLATT, M. (1999). Do industries explain momentum? *J. Finance* **54** 1249–1290.

PENG, J., ZHOU, N. and ZHU, J. (2009). Partial correlation estimation by joint sparse regression models. *J. Amer. Statist. Assoc.* **104** 735–746. MR2541591 https://doi.org/10.1198/jasa.2009.0126

PESARAN, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* **74** 967–1012. MR2238209 https://doi.org/10.1111/j.1468-0262.2006.00692.x

STOCK, J. H. and WATSON, M. W. (2002). Forecasting using principal components from a large number of predictors. *J. Amer. Statist. Assoc.* **97** 1167–1179. MR1951271 https://doi.org/10.1198/016214502388618960

STOCK, J. H. and WATSON, M. W. (2016). Handbook of macroeconomics. In *Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics* (J. B. Taylor and H. Uhlig, eds.) **2** 415–525. Elsevier, Amsterdam.

TIKHOMIROV, A. N. (1981). On the convergence rate in the central limit theorem for weakly dependent random variables. *Theory Probab. Appl.* **25** 790–809.

VAN DE GEER, S. A. and BÜHLMANN, P. (2009). On the conditions used to prove oracle results for the Lasso. *Electron. J. Stat.* **3** 1360–1392. MR2576316 https://doi.org/10.1214/09-EJS506

WANG, H., LI, B. and LENG, C. (2009). Shrinkage tuning parameter selection with a diverging number of parameters. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 671–683. MR2749913 https://doi.org/10.1111/j.1467-9868.2008.00693.x

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, X. and CHENG, G. (2014). Bootstrapping high dimensional time series.

The Annals of Statistics
2023, Vol. 51, No. 4, 1718–1743
https://doi.org/10.1214/23-AOS2305
© Institute of Mathematical Statistics, 2023

MATCHING RECOVERY THRESHOLD FOR CORRELATED RANDOM GRAPHS

BY JIAN DING[a] AND HANG DU[b]

School of Mathematical Science, Peking University, [a]*dingjian@math.pku.edu.cn,* [b]*duhang@pku.edu.cn*

For two correlated graphs which are independently sub-sampled from a common Erdős–Rényi graph $\mathbf{G}(n, p)$, we wish to recover their *latent* vertex matching from the observation of these two graphs *without labels*. When $p = n^{-\alpha+o(1)}$ for $\alpha \in (0, 1]$, we establish a sharp information-theoretic threshold for whether it is possible to correctly match a positive fraction of vertices. Our result sharpens a constant factor in a recent work by Wu, Xu and Yu.

REFERENCES

[1] ALDOUS, D. and STEELE, J. M. (2004). The objective method: Probabilistic combinatorial optimization and local weak convergence. In *Probability on Discrete Structures*. *Encyclopaedia Math*. *Sci*. **110** 1–72. Springer, Berlin. MR2023650 https://doi.org/10.1007/978-3-662-09444-0_1

[2] ANANTHARAM, V. and SALEZ, J. (2016). The densest subgraph problem in sparse random graphs. *Ann*. *Appl*. *Probab*. **26** 305–327. MR3449319 https://doi.org/10.1214/14-AAP1091

[3] ARRATIA, R. and TAVARÉ, S. (1992). The cycle structure of random permutations. *Ann*. *Probab*. **20** 1567–1591. MR1175278

[4] BARAK, B., CHOU, C.-N., LEI, Z., SCHRAMM, T. and SHENG, Y. (2019). (Nearly) efficient algorithms for the graph matching problem on correlated random graphs. In *Advances in Neural Information Processing Systems* **32**. Curran Associates, Inc., Red Hook.

[5] BERG, A., BERG, T. and MALIK, J. (2005). Shape matching and object recognition using low distortion correspondences. In 2005 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (*CVPR'*05) **1** 26–33.

[6] BOZORG, M., SALEHKALEYBAR, S. and HASHEMI, M. (2019). Seedless graph matching via tail of degree distribution for correlated Erdős–Rényi graphs. Preprint. Available at arXiv:1907.06334.

[7] CAIN, J. A., SANDERS, P. and WORMALD, N. (2007). The random graph threshold for k-orientability and a fast algorithm for optimal multiple-choice allocation. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms* 469–476. ACM, New York, NY, USA. MR2482873

[8] CHEN, S., JIANG, S., MA, Z., NOLAN, G. P. and ZHU, (2022). One-way matching of datasets with low rank signals. Preprint. Available at arXiv:2204.13858.

[9] COUR, T., SRINIVASAN, P. and SHI, J. (2006). Balanced graph matching. In *Proceedings of the* 19*th International Conference on Neural Information Processing Systems* (B. Schölkopf, J. Platt and T. Hoffman, eds.). *Advances in Neural Information Processing Systems* **19**. MIT Press, Cambridge.

[10] CULLINA, D. and KIYAVASH, N. (2016). Improved achievability and converse bounds for Erdős–Rényi graph matching. In *Proceedings of the* 2016 *ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science*, *SIGMETRICS'*16 63–72. Association for Computing Machinery, New York, NY, USA.

[11] CULLINA, D. and KIYAVASH, N. (2017). Exact alignment recovery for correlated Erdős–Rényi graphs. Preprint. Available at arXiv:1711.06783.

[12] CULLINA, D., KIYAVASH, N., MITTAL, P. and POOR, H. V. (2020). Partial recovery of Erdős–Rényi graph alignment via k-core alignment. In *SIGMETRICS'*20 99–100. Association for Computing Machinery, New York, NY, USA.

[13] DAI, O. E., CULLINA, D., KIYAVASH, N. and GROSSGLAUSER, M. (2019). Analysis of a canonical labeling algorithm for the alignment of correlated Erdős–Rényi graphs. *Proc*. *ACM Meas*. *Anal*. *Comput*. *Syst*. **3**.

[14] DING, J. and DU, H. (2023). Supplement to "Matching recovery threshold for correlated random graphs." https://doi.org/10.1214/23-AOS2305SUPP

MSC2020 subject classifications. Primary 05C80; secondary 68Q87.
Key words and phrases. Correlated Erdős–Rényi random graph, matching recovery, information threshold.

[15] DING, J. and DU, H. (2023). Detection threshold for correlated Erdős–Rényi graphs via densest subgraph. *IEEE Trans. Inf. Theory* **69** 5289–5298. MR4624732

[16] DING, J. and LI, Z. (2022). A polynomial time iterative algorithm for matching Gaussian matrices with non-vanishing correlation. Preprint. Available at arXiv:2212.13677.

[17] DING, J. and LI, Z. (2023). A polynomial-time iterative algorithm for random graph matching with non-vanishing correlation. Preprint. Available at arXiv:2306.00266.

[18] DING, J., MA, Z., WU, Y. and XU, J. (2021). Efficient random graph matching via degree profiles. *Probab. Theory Related Fields* **179** 29–115. MR4221654 https://doi.org/10.1007/s00440-020-00997-4

[19] DING, J., WU, Y., XU, J. and YANG, D. (2023). The planted matching problem: Sharp threshold and infinite-order phase transition. *Probab. Theory Related Fields* **187** 1–71. MR4634336 https://doi.org/10.1007/s00440-023-01208-6

[20] FAN, Z., MAO, C., WU, Y. and XU, J. (2019). Spectral graph matching and regularized quadratic relaxations II: Erdős–Rényi graphs and universality. Preprint. Available at arXiv:1907.08883.

[21] FAN, Z., MAO, C., WU, Y. and XU, J. (2020). Spectral graph matching and regularized quadratic relaxations: Algorithm and theory. In *Proceedings of the* 37*th International Conference on Machine Learning*. *Proceedings of Machine Learning Research* **119** 2985–2995. PMLR, Vienna, Austria.

[22] FEIZI, S., QUON, G., MEDARD, M., KELLIS, M. and JADBABAIE, A. (2016). Spectral alignment of networks. Preprint. Available at arXiv:1602.04181.

[23] FERNHOLZ, D. and RAMACHANDRAN, V. (2007). The k-orientability thresholds for $G_{n,p}$. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms* 459–468. ACM, New York, NY, USA. MR2482872

[24] FOUNTOULAKIS, N., KHOSLA, M. and PANAGIOTOU, K. (2016). The multiple-orientability thresholds for random hypergraphs. *Combin. Probab. Comput.* **25** 870–908. MR3568951 https://doi.org/10.1017/S0963548315000334

[25] FRIEZE, A. and KAROŃSKI, M. (2022). *Introduction to Random Graphs*. With electronic preprint available at https://www.math.cmu.edu/~af1p/BOOK.pdf.

[26] GANASSALI, L. and MASSOULIÉ, L. (2020). From tree matching to sparse graph alignment. In *Proceedings of Thirty Third Conference on Learning Theory* (J. Abernethy and S. Agarwal, eds.). *Proceedings of Machine Learning Research* **125** 1633–1665. PMLR, Graz, Austria.

[27] GANASSALI, L., MASSOULIÉ, L. and LELARGE, M. Impossibility of partial recovery in the graph alignment problem. In *Proceedings of Thirty Fourth Conference on Learning Theory* (M. M. Belkin and S. S. Kpotufe, eds.). *Proceedings of Machine Learning Research* **134** 2080–2102. PMLR, Boulder, CO, USA. 15–19 August 2021.

[28] GANASSALI, L., MASSOULIÉ, L. and SEMERJIAN, G. Statistical limits of correlation detection in trees. Available at arXiv:2209.13723.

[29] GAO, P. and WORMALD, N. C. (2010). Load balancing and orientability thresholds for random hypergraphs [extended abstract]. In *STOC'*10*—Proceedings of the* 2010 *ACM International Symposium on Theory of Computing* 97–103. ACM, New York, NY, USA. MR2743258

[30] HAGHIGHI, A., NG, A. and MANNING, C. (2005). Robust textual inference via graph matching. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing* 387–394. Association for Computational Linguistics, Vancouver, BC, Canada.

[31] HAJEK, B. (1990). Performance of global load balancing by local adjustment. *IEEE Trans. Inf. Theory* **36** 1398–1414. MR1080823 https://doi.org/10.1109/18.59935

[32] HALL, G. and MASSOULIÉ, L. (2023). Partial recovery in the graph alignment problem. *Oper. Res.* **71** 259–272. MR4585566

[33] KAZEMI, E., HASSANI, S. H. and GROSSGLAUSER, M. (2015). Growing a graph matching from a handful of seeds. *Proc. VLDB Endow.* **8** 1010–1021.

[34] LYZINSKI, V., FISHKIND, D. E. and PRIEBE, C. E. (2014). Seeded graph matching for correlated Erdős–Rényi graphs. *J. Mach. Learn. Res.* **15** 3513–3540. MR3291408

[35] MAO, C., RUDELSON, M. and TIKHOMIROV, K. (2023). Exact matching of random graphs with constant correlation. *Probab. Theory Related Fields* **186** 327–389. MR4586222 https://doi.org/10.1007/s00440-022-01184-3

[36] MAO, C., WU, Y., XU, J. and YU, S. H. (2021). Testing network correlation efficiently via counting trees. Preprint. Available at arXiv:2110.11816.

[37] MAO, C., WU, Y., XU, J. and YU, S. H. (2023). Random graph matching at Otter's threshold via counting chandeliers. In *STOC'*23*—Proceedings of the* 55*th Annual ACM Symposium on Theory of Computing* 1345–1356. ACM, New York, NY, USA. MR4617472 https://doi.org/10.1145/3564246.3585156

[38] MOSSEL, E. and XU, J. (2020). Seeded graph matching via large neighborhood statistics. *Random Structures Algorithms* **57** 570–611. MR4144076 https://doi.org/10.1002/rsa.20934

[39] NARAYANAN, A. and SHMATIKOV, V. (2008). Robust de-anonymization of large sparse datasets. In 2008 *IEEE Symposium on Security and Privacy* (*sp* 2008) 111–125.

[40] NARAYANAN, A. and SHMATIKOV, V. (2009). De-anonymizing social networks. In 2009 30*th IEEE Symposium on Security and Privacy* 173–187.

[41] PEDARSANI, P. and GROSSGLAUSER, M. (2011). On the privacy of anonymized networks. In *Proceedings of the* 17*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *KDD'*11 1235–1243. Association for Computing Machinery, New York, NY, USA.

[42] RACZ, M. Z. and SRIDHAR, A. (2021). Correlated stochastic block models: Exact graph matching with applications to recovering communities. In *Advances in Neural Information Processing Systems*.

[43] RÁCZ, M. Z. and SRIDHAR, A. (2022). Correlated randomly growing graphs. *Ann. Appl. Probab.* **32** 1058–1111. MR4414701 https://doi.org/10.1214/21-aap1703

[44] SHIRANI, F., GARG, S. and ERKIP, E. (2017). Seeded graph matching: Efficient algorithms and theoretical guarantees. In 2017 51*st Asilomar Conference on Signals*, *Systems*, *and Computers* 253–257.

[45] SINGH, R., XU, J. and BERGER, B. (2008). Global alignment of multiple protein interaction networks with application to functional orthology detection. *Proc. Natl. Acad. Sci. USA* **105** 12763–12768. https://doi.org/10.1073/pnas.0806627105

[46] VOGELSTEIN, J. T., CONROY, J. M., LYZINSKI, V., PODRAZIK, L. J., KRATZER, S. G., HARLEY, E. T., FISHKIND, D. E., VOGELSTEIN, R. J. and PRIEBE, C. E. (2015). Fast approximate quadratic programming for graph matching. *PLoS ONE* **10** 1–17.

[47] WANG, H., WU, Y., XU, J. and YOLOU, I. (2021). Random graph matching in geometric models: The case of complete graphs. Preprint. Available at arXiv:2202.10662.

[48] WU, Y., XU, J. and YU, S. H. (2022). Settling the sharp reconstruction thresholds of random graph matching. *IEEE Trans. Inf. Theory* **68** 5391–5417. MR4476403

[49] WU, Y., XU, J. and YU, S. H. (2023). Testing correlation of unlabeled random graphs. *Ann. Appl. Probab.* **33** 2519–2558. MR4612649 https://doi.org/10.1214/22-aap1786

[50] YARTSEVA, L. and GROSSGLAUSER, M. (2013). On the performance of percolation graph matching. In *Proceedings of the First ACM Conference on Online Social Networks*, *COSN'*13 119–130. Association for Computing Machinery, New York, NY, USA.

The Annals of Statistics
2023, Vol. 51, No. 4, 1744–1769
https://doi.org/10.1214/23-AOS2306
© Institute of Mathematical Statistics, 2023

LEARNING LOW-DIMENSIONAL NONLINEAR STRUCTURES FROM HIGH-DIMENSIONAL NOISY DATA: AN INTEGRAL OPERATOR APPROACH

BY XIUCAI DING[1,a] AND RONG MA[2,b]

[1]*Department of Statistics, University of California, Davis,* [a]*xcading@ucdavis.edu*
[2]*Department of Statistics, Stanford University,* [b]*rongm@stanford.edu*

We propose a kernel-spectral embedding algorithm for learning low-dimensional nonlinear structures from noisy and high-dimensional observations, where the data sets are assumed to be sampled from a nonlinear manifold model and corrupted by high-dimensional noise. The algorithm employs an adaptive bandwidth selection procedure which does not rely on prior knowledge of the underlying manifold. The obtained low-dimensional embeddings can be further utilized for downstream purposes such as data visualization, clustering and prediction. Our method is theoretically justified and practically interpretable. Specifically, for a general class of kernel functions, we establish the convergence of the final embeddings to their noiseless counterparts when the dimension grows polynomially with the size, and characterize the effect of the signal-to-noise ratio on the rate of convergence and phase transition. We also prove the convergence of the embeddings to the eigenfunctions of an integral operator defined by the kernel map of some reproducing kernel Hilbert space capturing the underlying nonlinear structures. Our results hold even when the dimension of the manifold grows with the sample size. Numerical simulations and analysis of real data sets show the superior empirical performance of the proposed method, compared to many existing methods, on learning various nonlinear manifolds in diverse applications.

REFERENCES

[1] ABBE, E., FAN, J. and WANG, K. (2022). An ℓ_p theory of PCA and spectral clustering. *Ann. Statist.* **50** 2359–2385.
[2] ADELI, E., WU, G., SAGHAFI, B., AN, L., SHI, F. and SHEN, D. (2017). Kernel-based joint feature selection and max-margin classification for early diagnosis of Parkinson's disease. *Sci. Rep.* **7** 1–14.
[3] AMINI, A. A. and RAZAEE, Z. S. (2021). Concentration of kernel matrices with application to kernel spectral clustering. *Ann. Statist.* **49** 531–556.
[4] ARIAS-CASTRO, E. and PELLETIER, B. (2013). On the convergence of maximum variance unfolding. *J. Mach. Learn. Res.* **14**.
[5] ARORA, S., HU, W. and KOTHARI, P. K. (2018). An analysis of the t-sne algorithm for data visualization. In *Conference on Learning Theory* 1455–1462. PMLR, Stockholm.
[6] BAO, Z., DING, X., WANG, J. and WANG, K. (2022). Statistical inference for principal components of spiked covariance matrices. *Ann. Statist.* **50** 1144–1169.
[7] BAO, Z., DING, X. and WANG, K. (2021). Singular vector and singular subspace distribution for the matrix denoising model. *Ann. Statist.* **49** 370–392.
[8] BARP, A., OATES, C., PORCU, E. and GIROLAMI, M. (2018). A Riemann-Stein kernel method. ArXiv preprint. Available at arXiv:1810.04946.
[9] BAYATI, M. and MONTANARI, A. (2011). The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inf. Theory* **57** 764–785.
[10] BELKIN, M. and NIYOGI, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **15** 1373–1396.

MSC2020 subject classifications. Primary 62R07, 62R30; secondary 47G10.
Key words and phrases. High-dimensional data, kernel method, manifold learning, nonlinear dimension reduction, spectral method.

[11] BELKIN, M. and NIYOGI, P. (2007). Convergence of Laplacian eigenmaps. *Adv. Neural Inf. Process. Syst.* **19** 129.

[12] BERLINET, A. and THOMAS-AGNAN, C. (2011). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer, Berlin.

[13] BILLINGSLEY, P. (2008). *Probability and Measure*. Wiley, New York.

[14] BISHOP, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer, Berlin.

[15] BLANCHARD, G., BOUSQUET, O. and ZWALD, L. (2007). Statistical properties of kernel principal component analysis. *Mach. Learn.* **66** 259–294.

[16] BORDENAVE, C. (2013). On Euclidean random matrices in high dimension. *Electron. Commun. Probab.* **18** 1–8.

[17] BORG, I. and GROENEN, P. J. (2005). *Modern Multidimensional Scaling: Theory and Applications*. Springer, Berlin.

[18] BRAUN, M. L. (2006). Accurate error bounds for the eigenvalues of the kernel matrix. *J. Mach. Learn. Res.* **7** 2303–2328.

[19] CAI, T. T. and MA, R. (2021). Theoretical foundations of t-SNE for visualizing high-dimensional clustered data. ArXiv preprint. Available at arXiv:2105.07536.

[20] CALDER, J., TRILLOS, N. G. and LEWICKA, M. (2020). Lipschitz regularity of graph Laplacians on random data clouds. ArXiv preprint. Available at arXiv:2007.06679.

[21] CANNOODT, R., SAELENS, W., SICHIEN, D., TAVERNIER, S., JANSSENS, S., GUILLIAMS, M., LAMBRECHT, B., DE PRETER, K. and SAEYS, Y. (2016). SCORPIUS improves trajectory inference and identifies novel modules in dendritic cell development. *BioRxiv* 079509.

[22] CHEN, Y., CHI, Y., FAN, J. and MA, C. (2021). Spectral methods for data science: A statistical perspective. *Found. Trends Mach. Learn.* **14** 566–806.

[23] CHENG, M.-Y. and WU, H.-T. (2013). Local linear regression on manifolds and its geometric interpretation. *J. Amer. Statist. Assoc.* **108** 1421–1434.

[24] CHENG, X. and SINGER, A. (2013). The spectrum of random inner-product kernel matrices. *Random Matrices Theory Appl.* **2** 1350010.

[25] COIFMAN, R. R. and LAFON, S. (2006). Diffusion maps. *Appl. Comput. Harmon. Anal.* **21** 5–30.

[26] CUTURI, M. (2009). Positive definite kernels in machine learning. ArXiv preprint. Available at arXiv:0911.5367.

[27] DING, X. and MA, R. (2023). Supplement to "Learning low-dimensional nonlinear structures from high-dimensional noisy data: An integral operator approach." https://doi.org/10.1214/23-AOS2306SUPP

[28] DING, X. and WU, H.-T. (2021). On the spectral property of kernel-based sensor fusion algorithms of high dimensional data. *IEEE Trans. Inf. Theory* **67** 640–670.

[29] DING, X. and WU, H.-T. (2021). How do kernel-based sensor fusion algorithms behave under high dimensional noise? ArXiv preprint. Available at arXiv:2111.10940.

[30] DING, X. and WU, H.-T. (2023). Impact of signal-to-noise ratio and bandwidth on graph Laplacian spectrum from high-dimensional noisy point cloud. *IEEE Trans. Inf. Theory* **69** 1899–1931.

[31] DO, Y. and VU, V. (2013). The spectrum of random kernel matrices: Universality results for rough and varying kernels. *Random Matrices Theory Appl.* **2** 1350005.

[32] DONOHO, D. L. and GRIMES, C. (2003). Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proc. Natl. Acad. Sci. USA* **100** 5591–5596.

[33] DUNSON, D. B., WU, H.-T. and WU, N. (2020). Graph based Gaussian processes on restricted domains. ArXiv preprint. Available at arXiv:2010.07242.

[34] DUNSON, D. B., WU, H.-T. and WU, N. (2021). Spectral convergence of graph Laplacian and heat kernel reconstruction in L_∞ from random samples. *Appl. Comput. Harmon. Anal.* **55** 282–336.

[35] EL KAROUI, N. (2010). The spectrum of kernel random matrices. *Ann. Statist.* **38** 1–50.

[36] EL KAROUI, N. (2010). On information plus noise kernel random matrices. *Ann. Statist.* **38** 3191–3216.

[37] EL KAROUI, N. and WU, H. T. (2016). Connection graph Laplacian methods can be made robust to noise. *Ann. Statist.* **44** 346–372.

[38] FAN, J., WANG, W. and ZHONG, Y. (2018). An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. *J. Mach. Learn. Res.* **18** 1–42.

[39] FAN, Z. and MONTANARI, A. (2019). The spectral norm of random inner-product kernel matrices. *Probab. Theory Related Fields* **173** 27–85.

[40] FOGEL, F., D'ASPREMONT, A. and VOJNOVIC, M. (2014). Serialrank: Spectral ranking using seriation. *Adv. Neural Inf. Process. Syst.* **27** 900–908.

[41] GARCÍA TRILLOS, N., GERLACH, M., HEIN, M. and SLEPCEV, D. (2020). Error estimates for spectral convergence of the graph Laplacian on random geometric graphs toward the Laplace–Beltrami operator. *Found. Comput. Math.* **20** 827–887.

[42] GINÉ, E. and KOLTCHINSKII, V. (2006). Empirical graph Laplacian approximation of Laplace–Beltrami operators: Large sample results. In *High Dimensional Probability* 238–259. IMS, Beachwood, OH.

[43] GUO, F., YAN, L., GUO, H., LI, L., HU, B., ZHAO, Y., YONG, J., HU, Y., WANG, X. et al. (2015). The transcriptome and DNA methylome landscapes of human primordial germ cells. *Cell* **161** 1437–1452.

[44] HOFMANN, T., SCHÖLKOPF, B. and SMOLA, A. J. (2008). Kernel methods in machine learning. *Ann. Statist.* **36** 1171–1220.

[45] HOFMEYR, D. P. (2019). Improving spectral clustering using the asymptotic value of the normalized cut. *J. Comput. Graph. Statist.* **28** 980–992.

[46] THE SMITHSONIAN INSTITUTE (2020). Mammuthus primigenius (blumbach). Available at https://3d.si.edu/object/3d/mammuthus-primigenius-blumbach:341c96cd-f967-4540-8ed1-d3fc56d31f12.

[47] JI, Z. and JI, H. (2016). TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. *Nucleic Acids Res.* **44** e117. https://doi.org/10.1093/nar/gkw430

[48] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327.

[49] JOLLIFFE, I. T. (2002). *Principal Component Analysis. Springer*, Berlin.

[50] JÖRGENS, K. (1982). *Linear Integral Operators* **13**. Pitman, London.

[51] KOBAK, D. and BERENS, P. (2019). The art of using t-SNE for single-cell transcriptomics. *Nat. Commun.* **10** 1–14.

[52] KOLTCHINSKII, V. and GINÉ, E. (2000). Random matrix approximation of spectra of integral operators. *Bernoulli* **6** 113–167.

[53] KUNG, S. Y. (2014). *Kernel Methods and Machine Learning*. Cambridge Univ. Press, Cambridge.

[54] LEE, J. A. and VERLEYSEN, M. (2007). *Nonlinear Dimensionality Reduction*. Springer, Berlin.

[55] LINDERMAN, G. C. and STEINERBERGER, S. (2019). Clustering with t-SNE, provably. *SIAM J. Math. Data Sci.* **1** 313–332. https://doi.org/10.1137/18m1216134

[56] LÖFFLER, M., ZHANG, A. Y. and ZHOU, H. H. (2021). Optimality of spectral clustering in the Gaussian mixture model. *Ann. Statist.* **49** 2506–2530.

[57] MANTON, J. H. and AMBLARD, P.-O. (2015). A primer on reproducing kernel Hilbert spaces. *Found. Trends Signal Process.* **8** 1–126.

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

[59] MCINNES, L., HEALY, J. and MELVILLE, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. ArXiv preprint. Available at arXiv:1802.03426.

[60] MOON, K. R., VAN DIJK, D., WANG, Z., GIGANTE, S., BURKHARDT, D. B., CHEN, W. S., YIM, K., VAN DEN ELZEN, A., HIRN, M. J. et al. (2019). Visualizing structure and transitions in high-dimensional biological data. *Nat. Biotechnol.* **37** 1482–1492.

[61] NASH, J. (1956). The imbedding problem for Riemannian manifolds. *Ann. Math.* **63** 20–63.

[62] ROSASCO, L., BELKIN, M. and VITO, E. D. (2010). On learning with integral operators. *J. Mach. Learn. Res.* **11** 905–934.

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

[64] SCHÖLKOPF, B., SMOLA, A. and MÜLLER, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* **10** 1299–1319.

[65] SCHÖLKOPF, B. and SMOLA, A. J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. Adaptive Computation and Machine Learning*. MIT Press, Cambridge.

[66] SHAWE-TAYLOR, J. and CRISTIANINI, N. (2004). *Kernel Methods for Pattern Analysis*. Cambridge Univ. Press, Cambridge.

[67] SHEN, C., LIN, Y.-T. and WU, H.-T. (2022). Robust and scalable manifold learning via landmark diffusion for long-term medical signal processing. *J. Mach. Learn. Res.* **23** 1–30.

[68] SHEN, C. and WU, H.-T. (2020). Scalability and robustness of spectral embedding: Landmark diffusion is all you need. ArXiv preprint. Available at arXiv:2001.00801.

[69] SHI, T., BELKIN, M. and YU, B. (2009). Data spectroscopy: Eigenspaces of convolution operators and clustering. *Ann. Statist.* **37** 3960–3984.

[70] SHIOKAWA, Y., DATE, Y. and KIKUCHI, J. (2018). Application of kernel principal component analysis and computational machine learning to exploration of metabolites strongly associated with diet. *Sci. Rep.* **8** 1–8.

[71] SINGER, A. (2006). From graph to manifold Laplacian: The convergence rate. *Appl. Comput. Harmon. Anal.* **21** 128–134.

[72] SINGER, A. and WU, H.-T. (2012). Vector diffusion maps and the connection Laplacian. *Comm. Pure Appl. Math.* **65** 1067–1144.

[73] SINGER, A. and WU, H.-T. (2017). Spectral convergence of the connection Laplacian from random samples. *Inf. Inference* **6** 58–123.

[74] SMALE, S. and ZHOU, D.-X. (2007). Learning theory estimates via integral operators and their approximations. *Constr. Approx.* **26** 153–172.

[75] SMALE, S. and ZHOU, D.-X. (2009). Geometry on probability spaces. *Constr. Approx.* **30** 311–323.

[76] SMITH, A., ZHA, H. and WU, X.-M. (2008). Convergence and rate of convergence of a manifold-based dimension reduction algorithm. In *Advances in Neural Information Processing Systems* **21**.

[77] TENENBAUM, J. B., SILVA, V. D. and LANGFORD, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science* **290** 2319–2323.

[78] VAN DER MAATEN, L. and HINTON, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**.

[79] VAN DER MAATEN, L., POSTMA, E. and VAN DEN HERIK, J. (2009). Dimensionality reduction: A comparative review. *J. Mach. Learn. Res.* **10** 13.

[80] VERSHYNIN, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science* **47**. Cambridge Univ. Press, Cambridge.

[81] VON LUXBURG, U. (2007). A tutorial on spectral clustering. *Stat. Comput.* **17** 395–416.

[82] VON LUXBURG, U., BELKIN, M. and BOUSQUET, O. (2008). Consistency of spectral clustering. *Ann. Statist.* **36** 555–586.

[83] WANG, B., ZHU, J., PIERSON, E., RAMAZZOTTI, D. and BATZOGLOU, S. (2017). Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. *Nat. Methods* **14** 414–416.

[84] WANG, Y., HUANG, H., RUDIN, C. and SHAPOSHNIK, Y. (2021). Understanding how dimension reduction tools work: An empirical approach to deciphering t-SNE, UMAP, TriMap, and PaCMAP for data visualization. *J. Mach. Learn. Res.* **22** 1–73.

[85] WEINBERGER, K. Q. and SAUL, L. K. (2006). An introduction to nonlinear dimensionality reduction by maximum variance unfolding. In *AAAI* **6** 1683–1686.

[86] WILLIAMS, C. K. and RASMUSSEN, C. E. (2006). *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA.

[87] WORMELL, C. L. and REICH, S. (2021). Spectral convergence of diffusion maps: Improved error bounds and an alternative normalization. *SIAM J. Numer. Anal.* **59** 1687–1734.

[88] WU, H.-T. and WU, N. (2018). Think globally, fit locally under the manifold setup: Asymptotic analysis of locally linear embedding. *Ann. Statist.* **46** 3805–3837.

[89] ZHANG, Z. and ZHA, H. (2004). Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM J. Sci. Comput.* **26** 313–338.

[90] ZHOU, Z. and AMINI, A. A. (2019). Analysis of spectral clustering algorithms for community detection: The general bipartite setting. *J. Mach. Learn. Res.* **20** 1–47.

The Annals of Statistics
2023, Vol. 51, No. 4, 1770–1798
https://doi.org/10.1214/23-AOS2307
© Institute of Mathematical Statistics, 2023

SINGLE INDEX FRÉCHET REGRESSION

BY SATARUPA BHATTACHARJEE[a] AND HANS-GEORG MÜLLER

Department of Statistics, University of California, Davis, [a]*sbhattacharjee@ucdavis.edu*

Single index models provide an effective dimension reduction tool in regression, especially for high-dimensional data, by projecting a general multivariate predictor onto a direction vector. We propose a novel single-index model for regression models where metric space-valued random object responses are coupled with multivariate Euclidean predictors. The responses in this regression model include complex, non-Euclidean data, including covariance matrices, graph Laplacians of networks and univariate probability distribution functions, among other complex objects that lie in abstract metric spaces. While Fréchet regression has proved useful for modeling the conditional mean of such random objects given multivariate Euclidean vectors, it does not provide for regression parameters such as slopes or intercepts, since the metric space-valued responses are not amenable to linear operations. As a consequence, distributional results for Fréchet regression have been elusive. We show here that for the case of multivariate Euclidean predictors, the parameters that define a single index and projection vector can be used to substitute for the inherent absence of parameters in Fréchet regression. Specifically, we derive the asymptotic distribution of suitable estimates of these parameters, which then can be utilized to test linear hypotheses for the parameters, subject to an identifiability condition. Consistent estimation of the link function of the single index Fréchet regression model is obtained through local linear Fréchet regression. We demonstrate the finite sample performance of estimation and inference for the proposed single index Fréchet regression model through simulation studies, including the special cases where responses are probability distributions and graph adjacency matrices. The method is illustrated for resting-state functional Magnetic Resonance Imaging (fMRI) data from the ADNI study.

REFERENCES

[1] AFSARI, B. (2011). Riemannian L^p center of mass: Existence, uniqueness, and convexity. *Proc. Amer. Math. Soc.* **139** 655–673. MR2736346 https://doi.org/10.1090/S0002-9939-2010-10541-5

[2] ALLEN, E., DAMARAJU, E., PLIS, S., ERHARDT, E., EICHELE, T. and CALHOUN, V. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cereb. Cortex* **24** 663–676.

[3] ANDREWS-HANNA, J. R., REIDLER, J. S., SEPULCRE, J., POULIN, R. and BUCKNER, R. L. (2010). Functional-anatomic fractionation of the brain's default network. *Neuron* **65** 550–562. https://doi.org/10.1016/j.neuron.2010.02.005

[4] BHATTACHARJEE, S. and MÜLLER, H.-G. (2022). Concurrent object regression. *Electron. J. Stat.* **16** 4031–4089. MR4456783 https://doi.org/10.1214/22-ejs2040

[5] BHATTACHARJEE, S. and MÜLLER, H.-G (2023). Supplement to "Single index Fréchet regression." https://doi.org/10.1214/23-AOS2307SUPP

[6] BHATTACHARYA, R. and PATRANGENARU, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann. Statist.* **31** 1–29. MR1962498 https://doi.org/10.1214/aos/1046294456

[7] BHATTACHARYA, R. and PATRANGENARU, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *Ann. Statist.* **33** 1225–1259. MR2195634 https://doi.org/10.1214/009053605000000093

[8] BUCHINSKY, M. (1995). Estimating the asymptotic covariance matrix for quantile regression models a Monte Carlo study. *J. Econometrics* **68** 303–338.

MSC2020 subject classifications. Primary 62R20, 62G10; secondary 62G05, 62G20.
Key words and phrases. Non-Euclidean data, dimension reduction, random objects, M-estimator, inference.

[9] BURAGO, D., BURAGO, Y. and IVANOV, S. (2001). *A Course in Metric Geometry*. *Graduate Studies in Mathematics* **33**. Amer. Math. Soc., Providence, RI. MR1835418 https://doi.org/10.1090/gsm/033

[10] CARROLL, R. J., FAN, J., GIJBELS, I. and WAND, M. P. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* **92** 477–489. MR1467842 https://doi.org/10.2307/2965697

[11] CHANG, Z., XUE, L. and ZHU, L. (2010). On an asymptotically more efficient estimation of the single-index model. *J. Multivariate Anal.* **101** 1898–1901. MR2651964 https://doi.org/10.1016/j.jmva.2010.02.005

[12] CHEN, D., HALL, P. and MÜLLER, H.-G. (2011). Single and multiple index functional regression models with nonparametric link. *Ann. Statist.* **39** 1720–1747. MR2850218 https://doi.org/10.1214/11-AOS882

[13] CHEN, Y. and MÜLLER, H.-G. (2022). Uniform convergence of local Fréchet regression with applications to locating extrema and time warping for metric space valued trajectories. *Ann. Statist.* **50** 1573–1592. MR4441132 https://doi.org/10.1214/21-aos2163

[14] COOK, R. D. (1994). Using dimension-reduction subspaces to identify important inputs in models of physical systems. In *Proceedings of the Section on Physical and Engineering Sciences* 18–25.

[15] CUI, X., HÄRDLE, W. K. and ZHU, L. (2011). The EFM approach for single-index models. *Ann. Statist.* **39** 1658–1688. MR2850216 https://doi.org/10.1214/10-AOS871

[16] DAMOISEAUX, J. S., PRATER, K. E., MILLER, B. L. and GREICIUS, M. D. (2012). Functional connectivity tracks clinical deterioration in Alzheimer's disease. *Neurobiol. Aging* **33** e19–E30. https://doi.org/10.1016/j.neurobiolaging.2011.06.024

[17] DAVISON, A. C. and HINKLEY, D. V. (1997). *Bootstrap Methods and Their Application*. *Cambridge Series in Statistical and Probabilistic Mathematics* **1**. Cambridge Univ. Press, Cambridge. MR1478673 https://doi.org/10.1017/CBO9780511802843

[18] DELICADO, P. and VIEU, P. (2017). Choosing the most relevant level sets for depicting a sample of densities. *Comput. Statist.* **32** 1083–1113. MR3703579 https://doi.org/10.1007/s00180-017-0746-y

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

[20] FAN, J. and HUANG, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli* **11** 1031–1057. MR2189080 https://doi.org/10.3150/bj/1137421639

[21] FERRATY, F., PARK, J. and VIEU, P. (2011). Estimation of a functional single index model. In *Recent Advances in Functional Data Analysis and Related Topics*. *Contrib. Statist.* 111–116. Physica-Verlag/Springer, Heidelberg. MR2815569 https://doi.org/10.1007/978-3-7908-2736-1_17

[22] FERREIRA, L. R. K. and BUSATTO, G. F. (2013). Resting-state functional connectivity in normal brain aging. *Neurosci. Biobehav. Rev.* **37** 384–400.

[23] FRÉCHET, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. Henri Poincaré* **10** 215–310. MR0027464

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

[25] GAO, J. and LIANG, H. (1997). Statistical inference in single-index and partially nonlinear models. *Ann. Inst. Statist. Math.* **49** 493–517. MR1482369 https://doi.org/10.1023/A:1003118812392

[26] GHOSAL, A., MEIRING, W. and PETERSEN, A. (2023). Fréchet single index models for object response regression. *Electron. J. Stat.* **17** 1074–1112. MR4575027 https://doi.org/10.1214/23-ejs2120

[27] GONÇALVES, S. and WHITE, H. (2005). Bootstrap standard error estimates for linear regression. *J. Amer. Statist. Assoc.* **100** 970–979. MR2201023 https://doi.org/10.1198/016214504000002087

[28] HALL, P. (1989). On projection pursuit regression. *Ann. Statist.* **17** 573–588. MR0994251 https://doi.org/10.1214/aos/1176347126

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

[30] HÄRDLE, W. and STOKER, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.* **84** 986–995. MR1134488

[31] HRISTACHE, M., JUDITSKY, A. and SPOKOINY, V. (2001). Direct estimation of the index coefficient in a single-index model. *Ann. Statist.* **29** 595–623. MR1865333 https://doi.org/10.1214/aos/1009210681

[32] HUH, J. and PARK, B. U. (2002). Likelihood-based local polynomial fitting for single-index models. *J. Multivariate Anal.* **80** 302–321. MR1889778 https://doi.org/10.1006/jmva.2000.1984

[33] ICHIMURA, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J. Econometrics* **58** 71–120. MR1230981 https://doi.org/10.1016/0304-4076(93)90114-K

[34] JIANG, C.-R. and WANG, J.-L. (2011). Functional single index models for longitudinal data. *Ann. Statist.* **39** 362–388. MR2797850 https://doi.org/10.1214/10-AOS845

[35] KATO, K. (2011). A note on moment convergence of bootstrap M-estimators. *Statist. Decisions* **28** 51–61. MR2776694 https://doi.org/10.1524/stnd.2011.1078

[36] KERETA, Ž., KLOCK, T. and NAUMOVA, V. (2021). Nonlinear generalization of the monotone single index model. *Inf. Inference* **10** 987–1029. MR4312085 https://doi.org/10.1093/imaiai/iaaa013

[37] KLOECKNER, B. (2010). A geometric study of Wasserstein spaces: Euclidean spaces. *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **9** 297–323. MR2731158

[38] KUCHIBHOTLA, A. K. and PATRA, R. K. (2020). Efficient estimation in single index models through smoothing splines. *Bernoulli* **26** 1587–1618. MR4058379 https://doi.org/10.3150/19-BEJ1183

[39] KUEPER, J. K., SPEECHLEY, M. and MONTERO-ODASSO, M. (2018). The Alzheimer's disease assessment scale-cognitive subscale (ADAS-Cog): Modifications and responsiveness in pre-dementia populations. A narrative review. *J. Alzheimers Dis.* **63** 423–444. https://doi.org/10.3233/JAD-170991

[40] 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

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

[42] 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

[43] LIANG, H., LIU, X., LI, R. and TSAI, C.-L. (2010). Estimation and testing for partially linear single-index models. *Ann. Statist.* **38** 3811–3836. MR2766869 https://doi.org/10.1214/10-AOS835

[44] LIN, W. and KULASEKERA, K. B. (2007). Identifiability of single-index models and additive-index models. *Biometrika* **94** 496–501. MR2380574 https://doi.org/10.1093/biomet/asm029

[45] LIN, Z. and MÜLLER, H.-G. (2021). Total variation regularized Fréchet regression for metric-space valued data. *Ann. Statist.* **49** 3510–3533. MR4352539 https://doi.org/10.1214/21-aos2095

[46] LOPEZ, O., PATILEA, V. and VAN KEILEGOM, I. (2013). Single index regression models in the presence of censoring depending on the covariates. *Bernoulli* **19** 721–747. MR3079294 https://doi.org/10.3150/12-BEJ464

[47] MARRON, J. S. and ALONSO, A. M. (2014). Overview of object oriented data analysis. *Biom. J.* **56** 732–753. MR3258083 https://doi.org/10.1002/bimj.201300072

[48] MÜLLER, H.-G. (2016). Peter Hall, functional data analysis and random objects. *Ann. Statist.* **44** 1867–1887. MR3546436 https://doi.org/10.1214/16-AOS1492

[49] NISHIYAMA, Y. (2010). Moment convergence of M-estimators. *Stat. Neerl.* **64** 505–507. MR2757870 https://doi.org/10.1111/j.1467-9574.2010.00469.x

[50] NOVO, S., ANEIROS, G. and VIEU, P. (2019). Automatic and location-adaptive estimation in functional single-index regression. *J. Nonparametr. Stat.* **31** 364–392. MR3941218 https://doi.org/10.1080/10485252.2019.1567726

[51] PATRANGENARU, V. and ELLINGSON, L. (2016). *Nonparametric Statistics on Manifolds and Their Applications to Object Data Analysis*. CRC Press, Boca Raton, FL. MR3444169

[52] PENNEC, X. (2018). Barycentric subspace analysis on manifolds. *Ann. Statist.* **46** 2711–2746. MR3851753 https://doi.org/10.1214/17-AOS1636

[53] PETERSEN, A., DEONI, S. and MÜLLER, H.-G. (2019). Fréchet estimation of time-varying covariance matrices from sparse data, with application to the regional co-evolution of myelination in the developing brain. *Ann. Appl. Stat.* **13** 393–419. MR3937434 https://doi.org/10.1214/18-AOAS1195

[54] PETERSEN, A., LIU, X. and DIVANI, A. A. (2021). Wasserstein F-tests and confidence bands for the Fréchet regression of density response curves. *Ann. Statist.* **49** 590–611. MR4206692 https://doi.org/10.1214/20-AOS1971

[55] PETERSEN, A. and MÜLLER, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *Ann. Statist.* **47** 691–719. MR3909947 https://doi.org/10.1214/17-AOS1624

[56] PEYRÉ, G. (2009). Manifold models for signals and images. *Comput. Vis. Image Underst.* **113** 249–260.

[57] SCHÖTZ, C. (2019). Convergence rates for the generalized Fréchet mean via the quadruple inequality. *Electron. J. Stat.* **13** 4280–4345. MR4023955 https://doi.org/10.1214/19-EJS1618

[58] SCHÖTZ, C. (2020). Regression in nonstandard spaces with Fréchet and geodesic approaches. ArXiv preprint. Available at arXiv:2012.13332.

[59] SHAO, J. and TU, D. S. (1995). *The Jackknife and Bootstrap*. *Springer Series in Statistics*. Springer, New York. MR1351010 https://doi.org/10.1007/978-1-4612-0795-5

[60] SILVERMAN, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.* **6** 177–184. MR0471166

[61] SMALL, C. G. (1996). *The Statistical Theory of Shape*. *Springer Series in Statistics*. Springer, New York. MR1418639 https://doi.org/10.1007/978-1-4612-4032-7

[62] STOKER, T. M. (1986). Consistent estimation of scaled coefficients. *Econometrica* **54** 1461–1481. MR0868152 https://doi.org/10.2307/1914309

[63] THOMAS YEO, B., KRIENEN, F. M., SEPULCRE, J., SABUNCU, M. R., LASHKARI, D., HOLLINSHEAD, M., ROFFMAN, J. L., SMOLLER, J. W., ZÖLLEI, L. et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106** 1125–1165.

[64] TSOCHANTARIDIS, I., HOFMANN, T., JOACHIMS, T. and ALTUN, Y. (2004). Support vector machine learning for interdependent and structured output spaces. In *Proceedings of the Twenty-First International Conference on Machine Learning* 104.

[65] VAN DER VAART, A. W. and WELLNER, J. A. (2023). *Weak Convergence and Empirical Processes—with Applications to Statistics*, 2nd ed. *Springer Series in Statistics*. Springer, Cham. MR4628026 https://doi.org/10.1007/978-3-031-29040-4

[66] WANG, J.-L., XUE, L., ZHU, L. and CHONG, Y. S. (2010). Estimation for a partial-linear single-index model. *Ann. Statist.* **38** 246–274. MR2589322 https://doi.org/10.1214/09-AOS712

[67] XIA, Y. (2007). A constructive approach to the estimation of dimension reduction directions. *Ann. Statist.* **35** 2654–2690. MR2382662 https://doi.org/10.1214/009053607000000352

[68] XIA, Y. and HÄRDLE, W. (2006). Semi-parametric estimation of partially linear single-index models. *J. Multivariate Anal.* **97** 1162–1184. MR2276153 https://doi.org/10.1016/j.jmva.2005.11.005

[69] XIA, Y., TONG, H., LI, W. K. and ZHU, L.-X. (2009). An adaptive estimation of dimension reduction space. In *Exploration of a Nonlinear World: An Appreciation of Howell Tong's Contributions to Statistics* 299–346. World Scientific, Singapore.

[70] 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

[71] YU, Y. and RUPPERT, D. (2002). Penalized spline estimation for partially linear single-index models. *J. Amer. Statist. Assoc.* **97** 1042–1054. MR1951258 https://doi.org/10.1198/016214502388618861

[72] ZHANG, H.-Y., WANG, S.-J., LIU, B., MA, Z.-L., YANG, M., ZHANG, Z.-J. and TENG, G.-J. (2010). Resting brain connectivity: Changes during the progress of Alzheimer disease. *Radiology* **256** 598–606.

[73] ZHANG, Q., XUE, L. and LI, B. (2021). Dimension reduction and data visualization for Fréchet regression. ArXiv preprint. Available at arXiv:2110.00467.

[74] ZHAO, W., LIAN, H. and LIANG, H. (2017). Quantile regression for the single-index coefficient model. *Bernoulli* **23** 1997–2027. MR3624884 https://doi.org/10.3150/16-BEJ802

[75] ZHAO, W., ZHANG, F., LI, R. and LIAN, H. (2020). Principal single-index varying-coefficient models for dimension reduction in quantile regression. *J. Stat. Comput. Simul.* **90** 800–818. MR4067347 https://doi.org/10.1080/00949655.2019.1707831

[76] ZHOU, J. and HE, X. (2008). Dimension reduction based on constrained canonical correlation and variable filtering. *Ann. Statist.* **36** 1649–1668. MR2435451 https://doi.org/10.1214/07-AOS529

[77] ZHU, L. and XUE, L. (2006). Empirical likelihood confidence regions in a partially linear single-index model. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 549–570. MR2278341 https://doi.org/10.1111/j.1467-9868.2006.00556.x

[78] ZHU, L.-P. and ZHU, L.-X. (2009). On distribution-weighted partial least squares with diverging number of highly correlated predictors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 525–548. MR2649607 https://doi.org/10.1111/j.1467-9868.2008.00697.x

The Annals of Statistics
2023, Vol. 51, No. 4, 1799–1823
https://doi.org/10.1214/23-AOS2309
© Institute of Mathematical Statistics, 2023

UNIVERSALITY OF REGULARIZED REGRESSION ESTIMATORS IN HIGH DIMENSIONS

BY QIYANG HAN[1,a] AND YANDI SHEN[2,b]

[1]*Department of Statistics, Rutgers University,* [a]*qh85@stat.rutgers.edu*
[2]*Department of Statistics, University of Chicago,* [b]*ydshen@uchicago.edu*

The Convex Gaussian Min–Max Theorem (CGMT) has emerged as a prominent theoretical tool for analyzing the precise stochastic behavior of various statistical estimators in the so-called high-dimensional proportional regime, where the sample size and the signal dimension are of the same order. However, a well-recognized limitation of the existing CGMT machinery rests in its stringent requirement on the exact Gaussianity of the design matrix, therefore rendering the obtained precise high-dimensional asymptotics, largely a specific Gaussian theory in various important statistical models.

This paper provides a structural universality framework for a broad class of regularized regression estimators that is particularly compatible with the CGMT machinery. Here, universality means that if a "structure" is satisfied by the regression estimator $\widehat{\mu}_G$ for a standard Gaussian design G, then it will also be satisfied by $\widehat{\mu}_A$ for a general non-Gaussian design A with independent entries. In particular, we show that with a good enough ℓ_∞ bound for the regression estimator $\widehat{\mu}_A$, any "structural property" that can be detected via the CGMT for $\widehat{\mu}_G$ also holds for $\widehat{\mu}_A$ under a general design A with independent entries.

As a proof of concept, we demonstrate our new universality framework in three key examples of regularized regression estimators: the Ridge, Lasso and regularized robust regression estimators, where new universality properties of risk asymptotics and/or distributions of regression estimators and other related quantities are proved. As a major statistical implication of the Lasso universality results, we validate inference procedures using the degrees-of-freedom adjusted debiased Lasso under general design and error distributions. We also provide a counterexample, showing that universality properties for regularized regression estimators do not extend to general isotropic designs.

The proof of our universality results relies on new comparison inequalities for the optimum of a broad class of cost functions and Gordon's max–min (or min–max) costs, over arbitrary structure sets subject to ℓ_∞ constraints. These results may be of independent interest and broader applicability.

REFERENCES

[1] ABBASI, E., SALEHI, F. and HASSIBI, B. (2019). Universality in learning from linear measurements. *Adv. Neural Inf. Process. Syst.* **32**.

[2] BARBIER, J., KRZAKALA, F., MACRIS, N., MIOLANE, L. and ZDEBOROVÁ, L. (2019). Optimal errors and phase transitions in high-dimensional generalized linear models. *Proc. Natl. Acad. Sci. USA* **116** 5451–5460. MR3939767 https://doi.org/10.1073/pnas.1802705116

[3] BAYATI, M., LELARGE, M. and MONTANARI, A. (2015). Universality in polytope phase transitions and message passing algorithms. *Ann. Appl. Probab.* **25** 753–822. MR3313755 https://doi.org/10.1214/14-AAP1010

[4] BAYATI, M. and MONTANARI, A. (2011). The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inf. Theory* **57** 764–785. MR2810285 https://doi.org/10.1109/TIT.2010.2094817

MSC2020 subject classifications. Primary 60F17; secondary 62E17.
Key words and phrases. Gaussian comparison inequalities, high-dimensional asymptotics, Lasso, Lindeberg's principle, random matrix theory, robust regression, ridge regression, universality.

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

[6] BELLEC, P. C. (2022). Observable adjustments in single-index models for regularized M-estimators. arXiv preprint. Available at arXiv:2204.06990.

[7] BELLEC, P. C. and SHEN, Y. (2022). Derivatives and residual distribution of regularized M-estimators with application to adaptive tuning. In *Conference on Learning Theory* **178** 1912–1947. PMLR.

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

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

[10] BELLEC, P. C. and ZHANG, C.-H. (2023). Debiasing convex regularized estimators and interval estimation in linear models. *Ann. Statist.* **51** 391–436. MR4600987 https://doi.org/10.1214/22-aos2243

[11] CELENTANO, M. and MONTANARI, A. (2022). Fundamental barriers to high-dimensional regression with convex penalties. *Ann. Statist.* **50** 170–196. MR4382013 https://doi.org/10.1214/21-aos2100

[12] CELENTANO, M., MONTANARI, A. and WEI, Y. (2022). The lasso with general gaussian designs with applications to hypothesis testing. arXiv preprint. Available at arXiv:2007.13716v2.

[13] CHATTERJEE, S. (2006). A generalization of the Lindeberg principle. *Ann. Probab.* **34** 2061–2076. MR2294976 https://doi.org/10.1214/009117906000000575

[14] CHEN, W.-K. and LAM, W.-K. (2021). Universality of approximate message passing algorithms. *Electron. J. Probab.* **26** Paper No. 36, 44. MR4235487 https://doi.org/10.1214/21-EJP604

[15] CHERNOZHUKOV, V., CHETVERIKOV, D., KATO, K. and KOIKE, Y. (2023). High-dimensional data bootstrap. *Annu. Rev. Stat. Appl.* **10** 427–449. MR4567800 https://doi.org/10.1146/annurev-statistics-040120-022239

[16] CHERNOZHUKOV, V., CHETVERIKOV, D. and KOIKE, Y. (2023). Nearly optimal central limit theorem and bootstrap approximations in high dimensions. *Ann. Appl. Probab.* **33** 2374–2425. MR4583674 https://doi.org/10.1214/22-aap1870

[17] DENG, Z., KAMMOUN, A. and THRAMPOULIDIS, C. (2022). A model of double descent for high-dimensional binary linear classification. *Inf. Inference* **11** 435–495. MR4474343 https://doi.org/10.1093/imaiai/iaab002

[18] DICKER, L. H. (2016). Ridge regression and asymptotic minimax estimation over spheres of growing dimension. *Bernoulli* **22** 1–37. MR3449775 https://doi.org/10.3150/14-BEJ609

[19] 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

[20] DONOHO, D. and MONTANARI, A. (2016). High dimensional robust M-estimation: Asymptotic variance via approximate message passing. *Probab. Theory Related Fields* **166** 935–969. MR3568043 https://doi.org/10.1007/s00440-015-0675-z

[21] DUDEJA, R., LU, Y. M. and SEN, S. (2022). Universality of approximate message passing with semi-random matrices. arXiv preprint. Available at arXiv:2204.04281.

[22] DUDEJA, R., SEN, S. and LU, Y. M. (2022). Spectral universality of regularized linear regression with nearly deterministic sensing matrices. arXiv preprint. Available at arXiv:2208.02753.

[23] EL KAROUI, N. (2013). Asymptotic behavior of unregularized and ridge-regularized high-dimensional robust regression estimators: Rigorous results. arXiv preprint. Available at arXiv:1311.2445.

[24] EL KAROUI, N. (2018). On the impact of predictor geometry on the performance on high-dimensional ridge-regularized generalized robust regression estimators. *Probab. Theory Related Fields* **170** 95–175. MR3748322 https://doi.org/10.1007/s00440-016-0754-9

[25] FAN, Z. (2022). Approximate message passing algorithms for rotationally invariant matrices. *Ann. Statist.* **50** 197–224. MR4382014 https://doi.org/10.1214/21-aos2101

[26] FANG, X. and KOIKE, Y. (2021). High-dimensional central limit theorems by Stein's method. *Ann. Appl. Probab.* **31** 1660–1686. MR4312842 https://doi.org/10.1214/20-aap1629

[27] GERACE, F., KRZAKALA, F., LOUREIRO, B., STEPHAN, L. and ZDEBOROVÁ, L. (2022). Gaussian universality of linear classifiers with random labels in high-dimension. arXiv preprint. Available at arXiv:2205.13303.

[28] GERACE, F., LOUREIRO, B., KRZAKALA, F., MÉZARD, M. and ZDEBOROVÁ, L. (2020). Generalisation error in learning with random features and the hidden manifold model. In *International Conference on Machine Learning* **119** 3452–3462. PMLR.

[29] GERBELOT, C., ABBARA, A. and KRZAKALA, F. (2020). Asymptotic errors for high-dimensional convex penalized linear regression beyond Gaussian matrices. In *Conference on Learning Theory* **125** 1682–1713 PMLR.

[30] GERBELOT, C., ABBARA, A. and KRZAKALA, F. (2023). Asymptotic errors for teacher–student convex generalized linear models (or: How to prove Kabashima's replica formula). *IEEE Trans. Inf. Theory* **69** 1824–1852. MR4564683 https://doi.org/10.1109/tit.2022.3222913

[31] GOLDT, S., LOUREIRO, B., REEVES, G., KRZAKALA, F., MÉZARD, M. and ZDEBOROVÁ, L. (2022). The Gaussian equivalence of generative models for learning with shallow neural networks. In *Mathematical and Scientific Machine Learning* **145** 426–471. PMLR.

[32] GORDON, Y. (1988). On Milman's inequality and random subspaces which escape through a mesh in $\mathbf{R}^n$. In *Geometric Aspects of Functional Analysis* (1986/87). *Lecture Notes in Math.* **1317** 84–106. Springer, Berlin. MR0950977 https://doi.org/10.1007/BFb0081737

[33] HAN, Q. (2022). Noisy linear inverse problems under convex constraints: Exact risk asymptotics in high dimensions. arXiv preprint. Available at arXiv:2201.08435.

[34] HAN, Q. and SHEN, Y. (2023). Supplement to "Universality of regularized regression estimators in high dimensions." https://doi.org/10.1214/23-AOS2309SUPP

[35] HAN, Q. and XU, X. (2023). The distribution of ridgeless least squares interpolators. arXiv preprint. Available at arXiv:2307.02044.

[36] HASTIE, T., MONTANARI, A., ROSSET, S. and TIBSHIRANI, R. J. (2022). Surprises in high-dimensional ridgeless least squares interpolation. *Ann. Statist.* **50** 949–986. MR4404925 https://doi.org/10.1214/21-aos2133

[37] HOERL, A. E. and KENNARD, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12** 55–67.

[38] HU, H. and LU, Y. M. (2019). Asymptotics and optimal designs of slope for sparse linear regression. In 2019 *IEEE International Symposium on Information Theory* (*ISIT*) 375–379. IEEE, Los Alamitos.

[39] HU, H. and LU, Y. M. (2023). Universality laws for high-dimensional learning with random features. *IEEE Trans. Inf. Theory* **69** 1932–1964. MR4564688

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

[41] HUBER, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821. MR0356373

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

[43] JAVANMARD, A. and MONTANARI, A. (2014). Hypothesis testing in high-dimensional regression under the Gaussian random design model: Asymptotic theory. *IEEE Trans. Inf. Theory* **60** 6522–6554. MR3265038 https://doi.org/10.1109/TIT.2014.2343629

[44] JAVANMARD, A. and MONTANARI, A. (2018). Debiasing the Lasso: Optimal sample size for Gaussian designs. *Ann. Statist.* **46** 2593–2622. MR3851749 https://doi.org/10.1214/17-AOS1630

[45] KAROUI, N. E., BEAN, D., BICKEL, P. J., LIM, C. and YU, B. (2013). On robust regression with high-dimensional predictors. *Proc. Natl. Acad. Sci. USA* **110** 14557–14562. https://doi.org/10.1073/pnas.1307842110

[46] KORADA, S. B. and MONTANARI, A. (2011). Applications of the Lindeberg principle in communications and statistical learning. *IEEE Trans. Inf. Theory* **57** 2440–2450. MR2809100 https://doi.org/10.1109/TIT.2011.2112231

[47] LELARGE, M. and MIOLANE, L. (2019). Fundamental limits of symmetric low-rank matrix estimation. *Probab. Theory Related Fields* **173** 859–929. MR3936148 https://doi.org/10.1007/s00440-018-0845-x

[48] LIANG, T. and SUR, P. (2022). A precise high-dimensional asymptotic theory for boosting and minimum-$\ell 1$-norm interpolated classifiers. *Ann. Statist.* **50** 1669–1695. MR4441136 https://doi.org/10.1214/22-aos2170

[49] LINDEBERG, J. W. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Z.* **15** 211–225. MR1544569 https://doi.org/10.1007/BF01494395

[50] LOPES, M. E. (2022). Central limit theorem and bootstrap approximation in high dimensions: Near $1/\sqrt{n}$ rates via implicit smoothing. *Ann. Statist.* **50** 2492–2513. MR4505371 https://doi.org/10.1214/22-aos2184

[51] LOUREIRO, B., GERBELOT, C., CUI, H., GOLDT, S., KRZAKALA, F., MÉZARD, M. and ZDEBOROVÁ, L. (2021). Learning curves of generic features maps for realistic datasets with a teacher-student model. In *Advances in Neural Information Processing Systems* **34** 18137–18151.

[52] MEI, S. and MONTANARI, A. (2022). The generalization error of random features regression: Precise asymptotics and the double descent curve. *Comm. Pure Appl. Math.* **75** 667–766. MR4400901 https://doi.org/10.1002/cpa.22008

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

[54] MIOLANE, L. and MONTANARI, A. (2021). The distribution of the Lasso: Uniform control over sparse balls and adaptive parameter tuning. *Ann. Statist.* **49** 2313–2335. MR4319252 https://doi.org/10.1214/20-aos2038
[55] MONTANARI, A. (2018). Mean field asymptotics in high-dimensional statistics: From exact results to efficient algorithms. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro* 2018. *Invited Lectures* **4** 2973–2994. World Sci. Publ., Hackensack, NJ. MR3966519
[56] MONTANARI, A. and NGUYEN, P.-M. (2017). Universality of the elastic net error. In 2017 *IEEE International Symposium on Information Theory* (*ISIT*) 2338–2342. IEEE, Los Alamitos.
[57] MONTANARI, A., RUAN, F., SOHN, Y. and YAN, J. (2023). The generalization error of max-margin linear classifiers: Benign overfitting and high-dimensional asymptotics in the overparametrized regime. arXiv preprint. Available at arXiv:1911.01544v3.
[58] MONTANARI, A. and SAEED, B. (2022). Universality of empirical risk minimization. In *Conference on Learning Theory* **178** 4310–4312. PMLR.
[59] OYMAK, S. and TROPP, J. A. (2018). Universality laws for randomized dimension reduction, with applications. *Inf. Inference* **7** 337–446. MR3858331 https://doi.org/10.1093/imaiai/iax011
[60] PANAHI, A. and HASSIBI, B. (2017). A universal analysis of large-scale regularized least squares solutions. *Adv. Neural Inf. Process. Syst.* **30**.
[61] SALEHI, F., ABBASI, E. and HASSIBI, B. (2019). The impact of regularization on high-dimensional logistic regression. *Adv. Neural Inf. Process. Syst.* **32**.
[62] STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. MR0630098
[63] STOJANOVIC, S., DONHAUSER, K. and YANG, F. (2022). Tight bounds for maximum ℓ_1-margin classifiers. arXiv preprint. Available at arXiv:2212.03783.
[64] STOJNIC, M. (2013). A framework to characterize performance of lasso algorithms. arXiv preprint. Available at arXiv:1303.7291.
[65] SUR, P. and CANDÈS, E. J. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. *Proc. Natl. Acad. Sci. USA* **116** 14516–14525. MR3984492 https://doi.org/10.1073/pnas.1810420116
[66] TALAGRAND, M. (2011). *Mean Field Models for Spin Glasses, Vol. I: Basic Examples. Ergebnisse der Mathematik und Ihrer Grenzgebiete.* 3. *Folge. A Series of Modern Surveys in Mathematics* [*Results in Mathematics and Related Areas.* 3*rd Series. A Series of Modern Surveys in Mathematics*] **54**. Springer, Berlin. MR2731561 https://doi.org/10.1007/978-3-642-15202-3
[67] TALAGRAND, M. (2011). *Mean Field Models for Spin Glasses, Vol. II: Advanced Replica-Symmetry and Low Temperature. Ergebnisse der Mathematik und Ihrer Grenzgebiete.* 3. *Folge. A Series of Modern Surveys in Mathematics* [*Results in Mathematics and Related Areas.* 3*rd Series. A Series of Modern Surveys in Mathematics*] **55**. Springer, Heidelberg. MR3024566
[68] THRAMPOULIDIS, C., ABBASI, E. and HASSIBI, B. (2018). Precise error analysis of regularized M-estimators in high dimensions. *IEEE Trans. Inf. Theory* **64** 5592–5628. MR3832326 https://doi.org/10.1109/TIT.2018.2840720
[69] THRAMPOULIDIS, C., OYMAK, S. and HASSIBI, B. (2015). Regularized linear regression: A precise analysis of the estimation error. In *Conference on Learning Theory* **40** 1683–1709. PMLR.
[70] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242
[71] 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
[72] 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
[73] WANG, S., WENG, H. and MALEKI, A. (2022). Does SLOPE outperform bridge regression? *Inf. Inference* **11** 1–54. MR4409197 https://doi.org/10.1093/imaiai/iaab025
[74] WANG, T., ZHONG, X. and FAN, Z. (2022). Universality of approximate message passing algorithms and tensor networks. arXiv preprint. Available at arXiv:2206.13037.
[75] 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
[76] ZHANG, X., ZHOU, H. and YE, H. (2022). A modern theory for high-dimensional Cox regression models. arXiv preprint. Available at arXiv:2204.01161.
[77] ZHOU, L., KOEHLER, F., SUR, P., SUTHERLAND, D. J. and SREBRO, N. (2022). A non-asymptotic moreau envelope theory for high-dimensional generalized linear models. arXiv preprint. Available at arXiv:2210.12082.

The Annals of Statistics
2023, Vol. 51, No. 4, 1824–1849
https://doi.org/10.1214/23-AOS2314
© Institute of Mathematical Statistics, 2023

STATISTICAL INFERENCE ON A CHANGING EXTREME VALUE DEPENDENCE STRUCTURE

BY HOLGER DREES[a]

Department of Mathematics, University of Hamburg, [a]*holger.drees@uni-hamburg.de*

We analyze the extreme value dependence of independent, not necessarily identically distributed multivariate regularly varying random vectors. More specifically, we propose estimators of the spectral measure locally at some time point and of the spectral measures integrated over time. The uniform asymptotic normality of these estimators is proved under suitable nonparametric smoothness and regularity assumptions. We then use the process convergence of the integrated spectral measure to devise consistent tests for the null hypothesis that the spectral measure does not change over time.

REFERENCES

[1] AHMAD, A. A., DEME, E. H., DIOP, A., GIRARD, S. and USSEGLIO-CARLEVE, A. (2020). Estimation of extreme quantiles from heavy-tailed distributions in a location-dispersion regression model. *Electron. J. Stat.* **14** 4421–4456. MR4194267 https://doi.org/10.1214/20-EJS1779

[2] CASTRO-CAMILO, D., DE CARVALHO, M. and WADSWORTH, J. (2018). Time-varying extreme value dependence with application to leading European stock markets. *Ann. Appl. Stat.* **12** 283–309. MR3773394 https://doi.org/10.1214/17-AOAS1089

[3] DAOUIA, A., GARDES, L. and GIRARD, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli* **19** 2557–2589. MR3160564 https://doi.org/10.3150/12-BEJ466

[4] DE CARVALHO, M. and DAVISON, A. C. (2014). Spectral density ratio models for multivariate extremes. *J. Amer. Statist. Assoc.* **109** 764–776. MR3223748 https://doi.org/10.1080/01621459.2013.872651

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

[6] DE HAAN, L. and ZHOU, C. (2021). Trends in extreme value indices. *J. Amer. Statist. Assoc.* **116** 1265–1279. MR4309271 https://doi.org/10.1080/01621459.2019.1705307

[7] DELGADO, M. A. (1993). Testing the equality of nonparametric regression curves. *Statist. Probab. Lett.* **17** 199–204. MR1229937 https://doi.org/10.1016/0167-7152(93)90167-H

[8] DREES, H. (2023). Supplement to "Statistical inference on a changing extreme value dependence structure." https://doi.org/10.1214/23-AOS2314SUPP

[9] DREES, H. and MÜLLER, P. (2008). Fitting and validation of a bivariate model for large claims. *Insurance Math. Econom.* **42** 638–650. MR2404319 https://doi.org/10.1016/j.insmatheco.2007.07.001

[10] EINMAHL, J. H. J., DE HAAN, L. and PITERBARG, V. I. (2001). Nonparametric estimation of the spectral measure of an extreme value distribution. *Ann. Statist.* **29** 1401–1423. MR1873336 https://doi.org/10.1214/aos/1013203459

[11] EINMAHL, J. H. J., DE HAAN, L. and SINHA, A. K. (1997). Estimating the spectral measure of an extreme value distribution. *Stochastic Process. Appl.* **70** 143–171. MR1475660 https://doi.org/10.1016/S0304-4149(97)00065-3

[12] EINMAHL, J. H. J., DE HAAN, L. and ZHOU, C. (2016). Statistics of heteroscedastic extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 31–51. MR3453645 https://doi.org/10.1111/rssb.12099

[13] ENGELKE, S. and IVANOVS, J. (2021). Sparse structures for multivariate extremes. *Annu. Rev. Stat. Appl.* **8** 241–270. MR4243547 https://doi.org/10.1146/annurev-statistics-040620-041554

[14] ESCOBAR-BACH, M., GOEGEBEUR, Y. and GUILLOU, A. (2018). Local robust estimation of the Pickands dependence function. *Ann. Statist.* **46** 2806–2843. MR3851756 https://doi.org/10.1214/17-AOS1640

[15] GARDES, L. and GIRARD, S. (2015). Nonparametric estimation of the conditional tail copula. *J. Multivariate Anal.* **137** 1–16. MR3332795 https://doi.org/10.1016/j.jmva.2015.01.018

MSC2020 subject classifications. Primary 62G32; secondary 62G05, 62G10, 62G20, 60G70.
Key words and phrases. Extreme value dependence, integrated spectral measure, local estimation, multivariate regular variation, test of nonstationarity.

[16] GARDES, L., GIRARD, S. and LEKINA, A. (2010). Functional nonparametric estimation of conditional extreme quantiles. *J. Multivariate Anal.* **101** 419–433. MR2564351 https://doi.org/10.1016/j.jmva.2009.06.007

[17] GOEGEBEUR, Y., GUILLOU, A. and SCHORGEN, A. (2014). Nonparametric regression estimation of conditional tails: The random covariate case. *Statistics* **48** 732–755. MR3234058 https://doi.org/10.1080/02331888.2013.800064

[18] GONZÁLEZ-MANTEIGA, W. and CRUJEIRAS, R. M. (2013). An updated review of goodness-of-fit tests for regression models. *TEST* **22** 361–411. MR3093195 https://doi.org/10.1007/s11749-013-0327-5

[19] HASHORVA, E. (2010). Boundary non-crossings of Brownian pillow. *J. Theoret. Probab.* **23** 193–208. MR2591910 https://doi.org/10.1007/s10959-008-0191-5

[20] HOGA, Y. (2021). Modeling time-varying tail dependence, with application to systemic risk forecasting. *J. Financ. Econom.* 1–31.

[21] KONING, A. J. and PROTASOV, V. (2003). Tail behaviour of Gaussian processes with applications to the Brownian pillow. *J. Multivariate Anal.* **87** 370–397. MR2016944 https://doi.org/10.1016/S0047-259X(03)00059-9

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

[23] WANG, H. J. and LI, D. (2013). Estimation of extreme conditional quantiles through power transformation. *J. Amer. Statist. Assoc.* **108** 1062–1074. MR3174684 https://doi.org/10.1080/01621459.2013.820134

[24] YUEN, R., STOEV, S. and COOLEY, D. (2020). Distributionally robust inference for extreme Value-at-Risk. *Insurance Math. Econom.* **92** 70–89. MR4079575 https://doi.org/10.1016/j.insmatheco.2020.03.003

The Annals of Statistics
2023, Vol. 51, No. 4, 1850–1876
https://doi.org/10.1214/23-AOS2315
© Institute of Mathematical Statistics, 2023

RELAXING THE I.I.D. ASSUMPTION: ADAPTIVELY MINIMAX OPTIMAL REGRET VIA ROOT-ENTROPIC REGULARIZATION

BY BLAIR BILODEAU[1,a], JEFFREY NEGREA[2,c] AND DANIEL M. ROY[1,b]

[1]*Department of Statistical Sciences, University of Toronto,* [a]*blair.bilodeau@mail.utoronto.ca,* [b]*daniel.roy@utoronto.ca*
[2]*Department of Statistics and Actuarial Science, University of Waterloo,* [c]*jnegrea@uwaterloo.ca*

We consider prediction with expert advice when data are generated from distributions varying arbitrarily within an unknown constraint set. This *semi-adversarial* setting includes (at the extremes) the classical i.i.d. setting, when the unknown constraint set is restricted to be a singleton, and the unconstrained adversarial setting, when the constraint set is the set of all distributions. The Hedge algorithm—long known to be minimax (rate) optimal in the adversarial regime—was recently shown to be simultaneously minimax optimal for i.i.d. data. In this work, we propose to relax the i.i.d. assumption by seeking adaptivity at all levels of a natural ordering on constraint sets. We provide matching upper and lower bounds on the minimax regret at all levels, show that Hedge with deterministic learning rates is suboptimal outside of the extremes and prove that one can adaptively obtain minimax regret at all levels. We achieve this optimal adaptivity using the follow-the-regularized-leader (FTRL) framework, with a novel adaptive regularization scheme that implicitly scales as the square root of the entropy of the current predictive distribution, rather than the entropy of the initial predictive distribution. Finally, we provide novel technical tools to study the statistical performance of FTRL along the semi-adversarial spectrum.

REFERENCES

[1] AMIR, I., ATTIAS, I., KOREN, T., LIVNI, R. and MANSOUR, Y. (2020). Prediction with corrupted expert advice. In *Advances in Neural Information Processing Systems* 33.

[2] ANSCOMBE, F. J. (1953). Sequential estimation. *J. Roy. Statist. Soc. Ser. B* **15** 1–29. MR0056895

[3] AUDIBERT, J.-Y. (2009). Fast learning rates in statistical inference through aggregation. *Ann. Statist.* **37** 1591–1646. MR2533466 https://doi.org/10.1214/08-AOS623

[4] AUDIBERT, J.-Y. and BUBECK, S. (2009). Minimax policies for adversarial and stochastic bandits. In *Proceedings of the* 22*nd Conference on Learning Theory*.

[5] AUDIBERT, J.-Y. and BUBECK, S. (2010). Regret bounds and minimax policies under partial monitoring. *J. Mach. Learn. Res.* **11** 2785–2836. MR2738783

[6] AUER, P., CESA-BIANCHI, N. and FISCHER, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Mach. Learn.* **47** 235–256.

[7] AUER, P., CESA-BIANCHI, N., FREUND, Y. and SCHAPIRE, R. E. (2003). The nonstochastic multiarmed bandit problem. *SIAM J. Comput.* **32** 48–77. MR1954855 https://doi.org/10.1137/S0097539701398375

[8] AUER, P. and CHIANG, C.-K. (2016). An algorithm with nearly optimal pseudo-regret for both stochastic and adversarial bandits. In *Proceedings of the* 29*th Conference on Learning Theory*.

[9] BARTLETT, P. L. and MENDELSON, S. (2006). Empirical minimization. *Probab. Theory Related Fields* **135** 311–334. MR2240689 https://doi.org/10.1007/s00440-005-0462-3

[10] BEGG, C. B. and MEHTA, C. R. (1979). Sequential analysis of comparative clinical trials. *Biometrika* **66** 97–103. MR0529153 https://doi.org/10.1093/biomet/66.1.97

[11] BILODEAU, B., NEGREA, J. and ROY, D. M. (2023). Supplement to "Relaxing the i.i.d. assumption: Adaptively minimax optimal regret via root-entropic regularization." https://doi.org/10.1214/23-AOS2315SUPP

MSC2020 subject classifications. Primary 62L10, 62C20; secondary 60G25, 62M20, 68Q32, 68T05.
Key words and phrases. Sequential decision theory, prediction with expert advice, adaptive minimax regret, aggregation, robust prediction.

[12] BUBECK, S. and SLIVKINS, A. (2012). The best of both worlds: Stochastic and adversarial bandits. In *Proceedings of the* 25*th Conference on Learning Theory*.

[13] CAI, T. T. and LOW, M. G. (2004). An adaptation theory for nonparametric confidence intervals. *Ann. Statist.* **32** 1805–1840. MR2102494 https://doi.org/10.1214/009053604000000049

[14] CESA-BIANCHI, N. and LUGOSI, G. (1999). On prediction of individual sequences. *Ann. Statist.* **27** 1865–1895. MR1765620 https://doi.org/10.1214/aos/1017939242

[15] CESA-BIANCHI, N. and LUGOSI, G. (2006). *Prediction, Learning, and Games*. Cambridge Univ. Press, Cambridge. MR2409394 https://doi.org/10.1017/CBO9780511546921

[16] CESA-BIANCHI, N., MANSOUR, Y. and STOLTZ, G. (2007). Improved second-order bounds for prediction with expert advice. *Mach. Learn.* **66** 321–352.

[17] CESA-BIANCHI, N. and SHAMIR, O. (2018). Bandit regret scaling with the effective loss range. In *Algorithmic Learning Theory* 2018. *Proc. Mach. Learn. Res.* (*PMLR*) **83** 24. Proceedings of Machine Learning Research PMLR. MR3857304

[18] CHAMBAZ, A., ZHENG, W. and VAN DER LAAN, M. J. (2017). Targeted sequential design for targeted learning inference of the optimal treatment rule and its mean reward. *Ann. Statist.* **45** 2537–2564. MR3737901 https://doi.org/10.1214/16-AOS1534

[19] CHAUDHURI, K., FREUND, Y. and HSU, D. J. (2009). A parameter-free hedging algorithm. In *Advances in Neural Information Processing Systems* 22.

[20] COVER, T. M. (1967). Behavior of sequential predictors of binary sequences. In *Trans. Fourth Prague Conf. on Information Theory, Statistical Decision Functions, Random Processes* (*Prague*, 1965) 263–272. Academia, Prague. MR0217944

[21] DE ROOIJ, S., VAN ERVEN, T., GRÜNWALD, P. D. and KOOLEN, W. M. (2014). Follow the leader if you can, hedge if you must. *J. Mach. Learn. Res.* **15** 1281–1316. MR3214784

[22] FREUND, Y. and SCHAPIRE, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. System Sci.* **55** 119–139. MR1473055 https://doi.org/10.1006/jcss.1997.1504

[23] GAILLARD, P., STOLTZ, G. and VAN ERVEN, T. (2014). A second-order bound with excess losses. In *Proceedings of the* 27*th Conference on Learning Theory*.

[24] GILLILAND, D. C. (1968). Sequential compound estimation. *Ann. Math. Stat.* **39** 1890–1904. MR0285067 https://doi.org/10.1214/aoms/1177698019

[25] GRÜNWALD, P. D. and MEHTA, N. A. (2020). Fast rates for general unbounded loss functions: From ERM to generalized Bayes. *J. Mach. Learn. Res.* **21** Paper No. 56, 80 pp. MR4095335

[26] GUPTA, A., KOREN, T. and TALWAR, K. (2019). Better algorithms for stochastic bandits with adversarial corruptions. In *Proceedings of the* 32*nd Conference on Learning Theory*.

[27] HANNAN, J. (1957). Approximation to Bayes risk in repeated play. In *Contributions to the Theory of Games, Vol.* 3. *Ann. of Math. Stud.* **39** 97–139. Princeton Univ. Press, Princeton, NJ. MR0109094

[28] HANNEKE, S. (2017). Learning whenever learning is possible: Universal learning under general stochastic processes. Preprint. Available at arXiv:1706.01418.

[29] HAZAN, E. and KALE, S. (2010). Extracting certainty from uncertainty: Regret bounded by variation in costs. *Mach. Learn.* **80** 165–188. MR3108164 https://doi.org/10.1007/s10994-010-5175-x

[30] HAZAN, E. and KALE, S. (2011). Better algorithms for benign bandits. *J. Mach. Learn. Res.* **12** 1287–1311. MR2804601

[31] HUTTER, M. and POLAND, J. (2004). Prediction with expert advice by following the perturbed leader for general weights. In *Algorithmic Learning Theory. Lecture Notes in Computer Science* **3244** 279–293. Springer, Berlin. MR2167508 https://doi.org/10.1007/978-3-540-30215-5_22

[32] JUDITSKY, A. and NEMIROVSKI, A. (2000). Functional aggregation for nonparametric regression. *Ann. Statist.* **28** 681–712. MR1792783 https://doi.org/10.1214/aos/1015951994

[33] KOMIYA, H. (1988). Elementary proof for Sion's minimax theorem. *Kodai Math. J.* **11** 5–7. MR0930413 https://doi.org/10.2996/kmj/1138038812

[34] KOOLEN, W. M., GRÜNWALD, P. D. and VAN ERVEN, T. (2016). Combining adversarial guarantees and stochastic fast rates in online learning. In *Advances in Neural Information Processing Systems* 29.

[35] KOOLEN, W. M. and VAN ERVEN, T. (2015). Second-order quantile methods for experts and combinatorial games. In *Proceedings of the* 28*th Conference on Learning Theory*.

[36] KOTŁOWSKI, W. (2018). On minimaxity of follow the leader strategy in the stochastic setting. *Theoret. Comput. Sci.* **742** 50–65. MR3844274 https://doi.org/10.1016/j.tcs.2017.12.030

[37] LAI, T. L. (1988). Nearly optimal sequential tests of composite hypotheses. *Ann. Statist.* **16** 856–886. MR0947582 https://doi.org/10.1214/aos/1176350840

[38] LATTIMORE, T. and SZEPESVÁRI, C. (2020). *Bandit Algorithms*. Cambridge Univ. Press, Cambridge.

[39] LITTLESTONE, N. and WARMUTH, M. K. (1994). The weighted majority algorithm. *Inform. and Comput.* **108** 212–261. MR1265851 https://doi.org/10.1006/inco.1994.1009

[40] LUO, H. and SCHAPIRE, R. E. (2015). Achieving all with no parameters: Adanormalhedge. In 28*th Annual Conference on Learning Theory*.

[41] LYKOURIS, T., MIRROKNI, V. and PAES LEME, R. (2018). Stochastic bandits robust to adversarial corruptions. In *STOC'*18*—Proceedings of the* 50*th Annual ACM SIGACT Symposium on Theory of Computing* 114–122. ACM, New York. MR3826238 https://doi.org/10.1145/3188745.3188918

[42] MCMAHAN, H. B. (2017). A survey of algorithms and analysis for adaptive online learning. *J. Mach. Learn. Res.* **18** Paper No. 90, 50 pp. MR3714253

[43] MCMAHAN, H. B. and STREETER, M. J. (2010). Adaptive bound optimization for online convex optimization. In *Proceedings of the* 23*rd Conference on Learning Theory*.

[44] MOURTADA, J. and GAÏFFAS, S. (2019). On the optimality of the hedge algorithm in the stochastic regime. *J. Mach. Learn. Res.* **20** Paper No. 83, 28 pp. MR3960937

[45] NEU, G. (2015). First-order regret bounds for combinatorial semi-bandits. In *Proceedings of the* 28*th Conference on Learning Theory*.

[46] RAKHLIN, A., SRIDHARAN, K. and TEWARI, A. (2011). Online learning: Stochastic, constrained, and smoothed adversaries. In *Advances in Neural Information Processing Systems* 25.

[47] RAKHLIN, A., SRIDHARAN, K. and TSYBAKOV, A. B. (2017). Empirical entropy, minimax regret and minimax risk. *Bernoulli* **23** 789–824. MR3606751 https://doi.org/10.3150/14-BEJ679

[48] RASMUSSEN, S. L. (1980). A Bayesian approach to a problem in sequential estimation. *Ann. Statist.* **8** 1229–1243. MR0594640

[49] ROBBINS, H. (1952). Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.* **58** 527–535. MR0050246 https://doi.org/10.1090/S0002-9904-1952-09620-8

[50] SELDIN, Y. and SLIVKINS, A. (2014). One practical algorithm for both stochastic and adversarial bandits. In *Proceedings of the* 31*st International Conference on Machine Learning*.

[51] SION, M. (1958). On general minimax theorems. *Pacific J. Math.* **8** 171–176. MR0097026

[52] THUNE, T. and SELDIN, Y. (2018). Adaptation to easy data in prediction with limited advice. In *Advances in Neural Information Processing Systems* 31.

[53] TSYBAKOV, A. B. (2003). Optimal rates of aggregation. In *Proceedings of the* 16*th Conference on Computational Learning Theory*.

[54] TSYBAKOV, A. B. (2004). Optimal aggregation of classifiers in statistical learning. *Ann. Statist.* **32** 135–166. MR2051002 https://doi.org/10.1214/aos/1079120131

[55] VAN ERVEN, T., GRÜNWALD, P. D., KOOLEN, W. M. and DE ROOIJ, S. (2011). Adaptive hedge. In *Advances in Neural Information Processing Systems* 24.

[56] VAN ERVEN, T. and KOOLEN, W. M. (2016). Metagrad: Multiple learning rates in online learning. In *Advances in Neural Information Processing Systems* 29.

[57] VOVK, V. (1998). A game of prediction with expert advice. *J. Comput. System Sci.* **56** 153–173. MR1629690 https://doi.org/10.1006/jcss.1997.1556

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

[59] WEI, C.-Y. and LUO, H. (2018). More adaptive algorithms for adversarial bandits. In *Proceedings of the* 31*st Annual Conference on Learning Theory*.

[60] WOLFOWITZ, J. (1947). Consistency of sequential binomial estimates. *Ann. Math. Stat.* **18** 131–135. MR0019899 https://doi.org/10.1214/aoms/1177730502

[61] YANG, Y. (2004). Aggregating regression procedures to improve performance. *Bernoulli* **10** 25–47. MR2044592 https://doi.org/10.3150/bj/1077544602

[62] ZIMMERT, J. and SELDIN, Y. (2019). An optimal algorithm for stochastic and adversarial bandits. In *The* 22*nd International Conference on Artificial Intelligence and Statistics*.

The Annals of Statistics
2023, Vol. 51, No. 4, 1877–1894
https://doi.org/10.1214/23-AOS2317
© Institute of Mathematical Statistics, 2023

ORDER-OF-ADDITION ORTHOGONAL ARRAYS TO STUDY THE EFFECT OF TREATMENT ORDERING

BY ERIC D. SCHOEN[1,a] AND ROBERT W. MEE[2,b]

[1]*Faculty of Bioscience Engineering, KU Leuven,* [a]*eric.schoen@kuleuven.be*
[2]*Department of Business Analytics and Statistics, University of Tennessee,* [b]*rmee@utk.edu*

The effect of the order in which a set of m treatments is applied can be modeled by relative-position factors that indicate whether treatment i is carried out before or after treatment j, or by the absolute position for treatment i in the sequence. A design with the same normalized information matrix as the design with all $m!$ sequences is D- and G-optimal for the main-effects model involving the relative-position factors. We prove that such designs are also I-optimal for this model and D-optimal as well as G- and I-optimal for the first-order model in the absolute-position factors. We propose a methodology for a complete or partial enumeration of nonequivalent designs that are optimal for both models.

REFERENCES

[1] ATKINSON, A. C., DONEV, A. N. and TOBIAS, R. D. (2007). *Optimum Experimental Designs*, *with SAS*. *Oxford Statistical Science Series* **34**. Oxford Univ. Press, Oxford. MR2323647

[2] CHANG, J. Y. and HSU, J. C. (1992). Optimal designs for multiple comparisons with the best. *J. Statist. Plann. Inference* **30** 45–62. MR1157749 https://doi.org/10.1016/0378-3758(92)90106-3

[3] CHEN, J., MUKERJEE, R. and LIN, D. K. J. (2020). Construction of optimal fractional order-of-addition designs via block designs. *Statist. Probab. Lett.* **161** 108728, 6. MR4070440 https://doi.org/10.1016/j.spl.2020.108728

[4] FEDOROV, V. V. (1972). *Theory of Optimal Experiments*. *Probability and Mathematical Statistics*. Academic Press, New York. Translated from the Russian and edited by W. J. Studden and E. M. Klimko. MR0403103

[5] HEDAYAT, A. S., SLOANE, N. J. A. and STUFKEN, J. (1999). *Orthogonal Arrays*: *Theory and Applications*. *Springer Series in Statistics*. Springer, New York. MR1693498 https://doi.org/10.1007/978-1-4612-1478-6

[6] MEE, R. W. (2020). Order-of-addition modeling. *Statist. Sinica* **30** 1543–1559. MR4257544 https://doi.org/10.5705/ss.20

[7] MEE, R. W. and LI, H. (2024). Multiple comparison with the best for regression models with discrete inputs. *Technometrics*. To appear. https://doi.org/10.1080/00401706.2023.2252476

[8] NGUYEN, N.-K. and MILLER, A. J. (1992). A review of some exchange algorithms for constructing discrete *D*-optimal designs. *Comput. Statist. Data Anal.* **14** 489–498. MR1192218 https://doi.org/10.1016/0167-9473(92)90064-M

[9] PENG, J., MUKERJEE, R. and LIN, D. K. J. (2019). Design of order-of-addition experiments. *Biometrika* **106** 683–694. MR3992397 https://doi.org/10.1093/biomet/asz025

[10] SCHOEN, E. D., EENDEBAK, P. T. and NGUYEN, M. V. M. (2010). Complete enumeration of pure-level and mixed-level orthogonal arrays. *J. Combin. Des.* **18** 123–140. MR2604638 https://doi.org/10.1002/jcd.20236

[11] SCHOEN, E. D. and MEE, R. W. (2023). Supplement to "Order-of-addition orthogonal arrays to study the effect of treatment ordering." https://doi.org/10.1214/23-AOS2317SUPPA, https://doi.org/10.1214/23-AOS2317SUPPB, https://doi.org/10.1214/23-AOS2317SUPPC

[12] SCHOEN, E. D., VO-THANH, N. and GOOS, P. (2017). Two-level orthogonal screening designs with 24, 28, 32, and 36 runs. *J. Amer. Statist. Assoc.* **112** 1354–1369. MR3735383 https://doi.org/10.1080/01621459.2017.1279547

MSC2020 subject classifications. Primary 62K05; secondary 05B20, 62J05.

Key words and phrases. Average prediction variance, component-position model, D-optimality, equivalence class, I-optimality, pairwise order factors, projection.

[13] STOKES, Z. and XU, H. (2022). A position-based approach for design and analysis of order-of-addition experiments. *Statist*. *Sinica* **32** 1467–1488. MR4449902 https://doi.org/10.5705/ss.202020.0178

[14] TSAI, S.-F. (2022). Generating optimal order-of-addition designs with flexible run sizes. *J*. *Statist*. *Plann*. *Inference* **218** 147–163. MR4343850 https://doi.org/10.1016/j.jspi.2021.11.001

[15] VAN NOSTRAND, R. C. (1995). Design of experiments where order of addition is important. In *ASA Proceedings of the Section on Physical and Engineering Sciences* 155–160. Amer. Statist. Assoc., Alexandria, VA.

[16] VOELKEL, J. G. (2019). The design of order-of-addition experiments. *J*. *Qual*. *Technol*. **51** 230–241.

[17] VOELKEL, J. G. and GALLAGHER, K. P. (2019). The design and analysis of order-of-addition experiments: An introduction and case study. *Qual*. *Eng*. **31** 627–638.

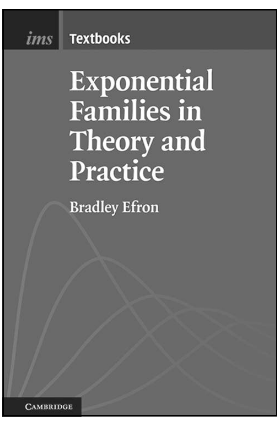
[18] YANG, J.-F., SUN, F. and XU, H. (2021). A component-position model, analysis and design for order-of-addition experiments. *Technometrics* **63** 212–224. MR4251495 https://doi.org/10.1080/00401706.2020.1764394

[19] ZHAO, Y., LIN, D. K. J. and LIU, M.-Q. (2022). Optimal designs for order-of-addition experiments. *Comput*. *Statist*. *Data Anal*. **165** Paper No. 107320, 15. MR4293002 https://doi.org/10.1016/j.csda.2021.107320

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

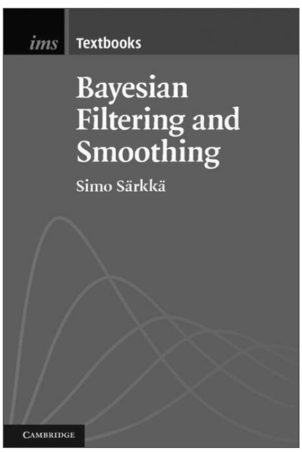
IMS members are entitled to a 40% discount: email ims@imstat.org to request your code

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.

The Institute of Mathematical Statistics presents

IMS TEXTBOOKS



Bayesian Filtering and Smoothing

Simo Särkkä

Filtering and smoothing methods are used to produce an accurate estimate of the state of a time-varying system based on multiple observational inputs (data). Interest in these methods has exploded in recent years, with numerous applications emerging in fields such as navigation, aerospace engineering, telecommunications and medicine. This compact, informal introduction for graduate students and advanced undergraduates presents the current state-of-the-art filtering and smoothing methods in a unified Bayesian framework. Readers learn what non-linear Kalman filters and particle filters are, how they are related, and their relative advantages and disadvantages. They also discover how state-of-the-art Bayesian parameter estimation methods can be combined with state-of-the-art filtering and smoothing algorithms. The book's practical and algorithmic approach assumes only modest mathematical prerequisites. Examples include MATLAB computations, and the numerous end-of-chapter exercises include computational assignments. MATLAB/GNU Octave source code is available for download at www.cambridge.org/sarkka, promoting hands-on work with the methods.

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

Hardback US$59.40
Paperback US$22.19

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