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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

Vol. 49, No. 3—June 2021

THE ANNALS OF STATISTICS Vol. 49, No. 3, pp. 1239–1809 June 2021

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

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

Past President: Susan Murphy, Department of Statistics, Harvard University, Cambridge, Massachusetts 02138-2901, USA

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

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

Program Secretary: Ming Yuan, Department of Statistics, Columbia University, New York, NY 10027-5927, USA

IMS EDITORS

The Annals of Statistics. *Editors*: Richard J. Samworth, Statistical Laboratory, Centre for Mathematical Sciences, University of Cambridge, Cambridge, CB3 0WB, UK. Ming Yuan, Department of Statistics, Columbia University, New York, NY 10027, USA

The Annals of Applied Statistics. *Editor-in-Chief*: Karen Kafadar, Department of Statistics, University of Virginia, Heidelberg Institute for Theoretical Studies, Charlottesville, VA 22904-4135, 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*: François Delarue, Laboratoire J. A. Dieudonné, Université de Nice Sophia-Antipolis, France-06108 Nice Cedex 2. Peter Friz, Institut für Mathematik, Technische Universität Berlin, 10623 Berlin, Germany and Weierstrass-Institut für Angewandte Analysis und Stochastik, 10117 Berlin, Germany

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

The IMS Bulletin. *Editor*: Vlada Limic, UMR 7501 de l'Université de Strasbourg et du CNRS, 7 rue René Descartes, 67084 Strasbourg Cedex, France

The Annals of Statistics **[ISSN 0090-5364 (print); ISSN 2168-8966 (online)],** Volume 49, Number 3, June 2021. Published bimonthly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, Ohio 44094, USA. Periodicals postage paid at Cleveland, Ohio, and at additional mailing offices.

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

Copyright © 2021 by the Institute of Mathematical Statistics
Printed in the United States of America

The Annals of Statistics
2021, Vol. 49, No. 3, 1239–1266
https://doi.org/10.1214/20-AOS1980
© Institute of Mathematical Statistics, 2021

A SHRINKAGE PRINCIPLE FOR HEAVY-TAILED DATA: HIGH-DIMENSIONAL ROBUST LOW-RANK MATRIX RECOVERY

BY JIANQING FAN[1,*], WEICHEN WANG[1,†] AND ZIWEI ZHU[2]

[1]*Department of Operations Research and Financial Engineering, Princeton University, [*]jqfan@princeton.edu; [†]nickweichwang@gmail.com*

[2]*Department of Statistics, University of Michigan, Ann Arbor, ziweiz@umich.edu*

This paper introduces a simple principle for robust statistical inference via appropriate shrinkage on the data. This widens the scope of high-dimensional techniques, reducing the distributional conditions from subexponential or sub-Gaussian to more relaxed bounded second or fourth moment. As an illustration of this principle, we focus on robust estimation of the low-rank matrix $\mathbf{\Theta}^*$ from the trace regression model $Y = \mathrm{Tr}(\mathbf{\Theta}^{*\top}\mathbf{X}) + \varepsilon$. It encompasses four popular problems: sparse linear model, compressed sensing, matrix completion and multitask learning. We propose to apply the penalized least-squares approach to the appropriately truncated or shrunk data. Under only bounded $2+\delta$ moment condition on the response, the proposed robust methodology yields an estimator that possesses the same statistical error rates as previous literature with sub-Gaussian errors. For sparse linear model and multitask regression, we further allow the design to have only bounded fourth moment and obtain the same statistical rates. As a byproduct, we give a robust covariance estimator with concentration inequality and optimal rate of convergence in terms of the spectral norm, when the samples only bear bounded fourth moment. This result is of its own interest and importance. We reveal that under high dimensions, the sample covariance matrix is not optimal whereas our proposed robust covariance can achieve optimality. Extensive simulations are carried out to support the theories.

REFERENCES

BELLEC, P. C., DALALYAN, A. S., GRAPPIN, E. and PARIS, Q. (2018). On the prediction loss of the lasso in the partially labeled setting. *Electron. J. Stat.* **12** 3443–3472. MR3864589 https://doi.org/10.1214/18-EJS1457

BICKEL, P. J. and LEVINA, E. (2008). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. MR2485008 https://doi.org/10.1214/08-AOS600

BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. MR2533469 https://doi.org/10.1214/08-AOS620

BROWNLEES, C., JOLY, E. and LUGOSI, G. (2015). Empirical risk minimization for heavy-tailed losses. *Ann. Statist.* **43** 2507–2536. MR3405602 https://doi.org/10.1214/15-AOS1350

CAI, T. and LIU, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *J. Amer. Statist. Assoc.* **106** 672–684. MR2847949 https://doi.org/10.1198/jasa.2011.tm10560

CAI, T. T. and ZHANG, A. (2014). Sparse representation of a polytope and recovery in sparse signals and low-rank matrices. *IEEE Trans. Inf. Theory* **60** 122–132. MR3150915 https://doi.org/10.1109/TIT.2013.2288639

CAI, T. T. and ZHANG, A. (2015). ROP: Matrix recovery via rank-one projections. *Ann. Statist.* **43** 102–138. MR3285602 https://doi.org/10.1214/14-AOS1267

CAI, T. T. and ZHOU, H. H. (2012). Optimal rates of convergence for sparse covariance matrix estimation. *Ann. Statist.* **40** 2389–2420. MR3097607 https://doi.org/10.1214/12-AOS998

CANDES, E. J. (2008). The restricted isometry property and its implications for compressed sensing. *C. R. Math. Acad. Sci. Paris* **346** 589–592. MR2412803 https://doi.org/10.1016/j.crma.2008.03.014

CANDES, E. J. and PLAN, Y. (2010). Matrix completion with noise. *Proc. IEEE* **98** 925–936.

MSC2020 subject classifications. Primary 62F35; secondary 62J05.

Key words and phrases. Robust statistics, shrinkage, heavy-tailed data, trace regression, low-rank matrix recovery, high-dimensional statistics.

CANDES, E. J. and PLAN, Y. (2011). Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Trans. Inf. Theory* **57** 2342–2359. MR2809094 https://doi.org/10.1109/TIT.2011.2111771

CANDES, E. J. and RECHT, B. (2009). Exact matrix completion via convex optimization. *Found. Comput. Math.* **9** 717–772. MR2565240 https://doi.org/10.1007/s10208-009-9045-5

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

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

CATONI, O. (2012). Challenging the empirical mean and empirical variance: A deviation study. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 1148–1185. MR3052407 https://doi.org/10.1214/11-AIHP454

CHEN, S. S., DONOHO, D. L. and SAUNDERS, M. A. (2001). Atomic decomposition by basis pursuit. *SIAM Rev.* **43** 129–159. MR1854649 https://doi.org/10.1137/S003614450037906X

CHEN, Y., FAN, J., MA, C. and YAN, Y. (2019). Inference and uncertainty quantification for noisy matrix completion. *Proc. Natl. Acad. Sci. USA* **116** 22931–22937. MR4036123 https://doi.org/10.1073/pnas.1910053116

CHEN, Y., CHI, Y., FAN, J., MA, C. and YAN, Y. (2020). Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. *SIAM J. Optim.* **30** 3098–3121. MR4167625 https://doi.org/10.1137/19M1290000

DIAKONIKOLAS, I., KAMATH, G., KANE, D. M., LI, J., MOITRA, A. and STEWART, A. (2016). Robust estimators in high dimensions without the computational intractability. In *57th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2016* 655–664. IEEE Computer Soc., Los Alamitos, CA. MR3631028 https://doi.org/10.1109/FOCS.2016.85

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

DONOHO, D. L., JOHNSTONE, I. and MONTANARI, A. (2013). Accurate prediction of phase transitions in compressed sensing via a connection to minimax denoising. *IEEE Trans. Inf. Theory* **59** 3396–3433. MR3061255 https://doi.org/10.1109/TIT.2013.2239356

FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581 https://doi.org/10.1198/016214501753382273

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., 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., LIU, H. and WANG, W. (2018). Large covariance estimation through elliptical factor models. *Ann. Statist.* **46** 1383–1414. MR3819104 https://doi.org/10.1214/17-AOS1588

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

FAN, J., WANG, W. and ZHU, Z. (2021). Supplement to "A shrinkage principle for heavy-tailed data: High-dimensional robust low-rank matrix recovery." https://doi.org/10.1214/20-AOS1980SUPP

FAN, J. and YAO, Q. (2015). *Elements of Financial Econometrics*. Science Press, Beijing.

FIGUEIREDO, M. A. T., NOWAK, R. D. and WRIGHT, S. J. (2007). Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE J. Sel. Top. Signal Process.* **1** 586–597.

GUPTA, S., ELLIS, S. E., ASHAR, F. N., MOES, A., BADER, J. S., ZHAN, J., WEST, A. B. and ARKING, D. E. (2014). Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nat. Commun.* **5** 5748. https://doi.org/10.1038/ncomms6748

HSU, D. and SABATO, S. (2016). Loss minimization and parameter estimation with heavy tails. *J. Mach. Learn. Res.* **17** Paper No. 18, 40. MR3491112

KIM, S. and XING, E. P. (2012). Tree-guided group lasso for multi-response regression with structured sparsity, with an application to EQTL mapping. *Ann. Appl. Stat.* **6** 1095–1117. MR3012522 https://doi.org/10.1214/12-AOAS549

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

KOLTCHINSKII, V., LOUNICI, K. and TSYBAKOV, A. B. (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *Ann. Statist.* **39** 2302–2329. MR2906869 https://doi.org/10.1214/11-AOS894

LAM, C. and FAN, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Statist.* **37** 4254–4278. MR2572459 https://doi.org/10.1214/09-AOS720

LOH, P.-L. (2017). Statistical consistency and asymptotic normality for high-dimensional robust M-estimators. *Ann. Statist.* **45** 866–896. MR3650403 https://doi.org/10.1214/16-AOS1471

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

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

MINSKER, S. (2015). Geometric median and robust estimation in Banach spaces. *Bernoulli* **21** 2308–2335. MR3378468 https://doi.org/10.3150/14-BEJ645

MINSKER, S. (2018). Sub-Gaussian estimators of the mean of a random matrix with heavy-tailed entries. *Ann. Statist.* **46** 2871–2903. MR3851758 https://doi.org/10.1214/17-AOS1642

NEGAHBAN, S. and WAINWRIGHT, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *Ann. Statist.* **39** 1069–1097. MR2816348 https://doi.org/10.1214/10-AOS850

NEGAHBAN, S. and WAINWRIGHT, M. J. (2012). Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *J. Mach. Learn. Res.* **13** 1665–1697. MR2930649

NEGAHBAN, S. N., RAVIKUMAR, P., WAINWRIGHT, M. J. and YU, B. (2012). A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers. *Statist. Sci.* **27** 538–557. MR3025133 https://doi.org/10.1214/12-STS400

NEMIROVSKY, A. S. and YUDIN, D. B. (1982). *Problem Complexity and Method Efficiency in Optimization. A Wiley-Interscience Publication*. Wiley, New York. MR0702836

RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2011). Minimax rates of estimation for high-dimensional linear regression over ℓ_q-balls. *IEEE Trans. Inf. Theory* **57** 6976–6994. MR2882274 https://doi.org/10.1109/TIT.2011.2165799

RECHT, B. (2011). A simpler approach to matrix completion. *J. Mach. Learn. Res.* **12** 3413–3430. MR2877360

RECHT, B., FAZEL, M. and PARRILO, P. A. (2010). Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.* **52** 471–501. MR2680543 https://doi.org/10.1137/070697835

REINSEL, G. C. and VELU, R. P. (2013). *Multivariate Reduced-Rank Regression*: *Theory and Applications*. *Lecture Notes in Statistics* **136**. Springer, New York. MR1719704 https://doi.org/10.1007/978-1-4757-2853-8

ROHDE, A. and TSYBAKOV, A. B. (2011). Estimation of high-dimensional low-rank matrices. *Ann. Statist.* **39** 887–930. MR2816342 https://doi.org/10.1214/10-AOS860

RUDELSON, M. and ZHOU, S. (2013). Reconstruction from anisotropic random measurements. *IEEE Trans. Inf. Theory* **59** 3434–3447. MR3061256 https://doi.org/10.1109/TIT.2013.2243201

STOCK, J. H. and WATSON, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *J. Bus. Econom. Statist.* **20** 147–162. MR1963257 https://doi.org/10.1198/073500102317351921

SUN, Q., ZHOU, W.-X. and FAN, J. (2020). Adaptive Huber regression. *J. Amer. Statist. Assoc.* **115** 254–265. MR4078461 https://doi.org/10.1080/01621459.2018.1543124

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

VERSHYNIN, R. (2010). Introduction to the non-asymptotic analysis of random matrices. Preprint. Available at arXiv:1011.3027.

WANG, L., ZHENG, C., ZHOU, W. and ZHOU, W.-X. (2018). A new principle for tuning-free Huber regression. Preprint.

ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. MR2604701 https://doi.org/10.1214/09-AOS729

ZOU, H. and LI, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.* **36** 1509–1533. MR2435443 https://doi.org/10.1214/009053607000000802

The Annals of Statistics
2021, Vol. 49, No. 3, 1267–1299
https://doi.org/10.1214/20-AOS1992
© Institute of Mathematical Statistics, 2021

STRONG SELECTION CONSISTENCY OF BAYESIAN VECTOR AUTOREGRESSIVE MODELS BASED ON A PSEUDO-LIKELIHOOD APPROACH

BY SATYAJIT GHOSH[1], KSHITIJ KHARE[2] AND GEORGE MICHAILIDIS[3]

[1]*Department of Statistics and Biostatistics, Rutgers University, satyajitghosh90@ufl.edu*
[2]*Department of Statistics, University of Florida, kdkhare@stat.ufl.edu*
[3]*Informatics Institute, University of Florida, gmichail@ufl.edu*

Vector autoregressive (VAR) models aim to capture linear temporal interdependencies among multiple time series. They have been widely used in macroeconomics and financial econometrics and more recently have found novel applications in functional genomics and neuroscience. These applications have also accentuated the need to investigate the behavior of the VAR model in a high-dimensional regime, which will provide novel insights into the role of temporal dependence for regularized estimates of the models parameters. However, hardly anything is known regarding posterior model selection consistency for Bayesian VAR models in such regimes.

In this work, we develop a pseudo-likelihood based Bayesian approach for consistent variable selection in high-dimensional VAR models by considering hierarchical normal priors on the autoregressive coefficients, as well as on the model space. We establish *strong selection consistency* of the proposed method, namely that the posterior probability assigned to the true underlying VAR model converges to one under high-dimensional scaling where the dimension p of the VAR system grows nearly exponentially with the sample size n.

Further, the result is established under mild regularity conditions on the problem parameters. Finally, as a by-product of these results, we also establish strong selection consistency for the sparse high-dimensional linear regression model with serially correlated regressors and errors.

REFERENCES

BAŃBURA, M., GIANNONE, D. and REICHLIN, L. (2010). Large Bayesian vector auto regressions. *J. Appl. Econometrics* **25** 71–92. MR2751790 https://doi.org/10.1002/jae.1137

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

BONDELL, H. D. and REICH, B. J. (2012). Consistent high-dimensional Bayesian variable selection via penalized credible regions. *J. Amer. Statist. Assoc.* **107** 1610–1624. MR3036420 https://doi.org/10.1080/01621459.2012.716344

BROWN, P. J., VANNUCCI, M. and FEARN, T. (2002). Bayes model averaging with selection of regressors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 519–536. MR1924304 https://doi.org/10.1111/1467-9868.00348

CAO, X., KHARE, K. and GHOSH, M. (2019). Posterior graph selection and estimation consistency for high-dimensional Bayesian DAG models. *Ann. Statist.* **47** 319–348. MR3909935 https://doi.org/10.1214/18-AOS1689

CAO, X., KHARE, K. and GHOSH, M. (2020). High-dimensional posterior consistency for hierarchical non-local priors in regression. *Bayesian Anal.* **15** 241–262. MR4050884 https://doi.org/10.1214/19-BA1154

CASELLA, G., GIRÓN, F. J., MARTÍNEZ, M. L. and MORENO, E. (2009). Consistency of Bayesian procedures for variable selection. *Ann. Statist.* **37** 1207–1228. MR2509072 https://doi.org/10.1214/08-AOS606

CASTILLO, I., SCHMIDT-HIEBER, J. and VAN DER VAART, A. (2015). Bayesian linear regression with sparse priors. *Ann. Statist.* **43** 1986–2018. MR3375874 https://doi.org/10.1214/15-AOS1334

MSC2020 subject classifications. 62M10, 62F15, 62J07.
Key words and phrases. Bayesian variable selection, vector autoregression, pseudo-likelihood, strong selection consistency, high-dimensional data.

De Mol, C., Giannone, D. and Reichlin, L. (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *J. Econometrics* **146** 318–328. MR2465176 https://doi.org/10.1016/j.jeconom.2008.08.011

George, E. I. and Foster, D. P. (2000). Calibration and empirical Bayes variable selection. *Biometrika* **87** 731–747. MR1813972 https://doi.org/10.1093/biomet/87.4.731

George, E. I. and McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statist. Sinica* **7** 339–373.

Ghosh, S., Khare, K. and Michailidis, G. (2019a). High-dimensional posterior consistency in Bayesian vector autoregressive models. *J. Amer. Statist. Assoc.* **114** 735–748. MR3963176 https://doi.org/10.1080/01621459.2018.1437043

Ghosh, S., Khare, K. and Michailidis, G. (2019b). Supplemental document for "High dimensional posterior consistency in bayesian vector autoregressive models". *J. Amer. Statist. Assoc.*

Ghosh, S., Khare, K. and Michailidis, G. (2021). Supplement to "Strong selection consistency of Bayesian vector autoregressive models based on a pseudo-likelihood approach." https://doi.org/10.1214/20-AOS1992SUPP

Hans, C., Dobra, A. and West, M. (2007). Shotgun stochastic search for "large p" regression. *J. Amer. Statist. Assoc.* **102** 507–516. MR2370849 https://doi.org/10.1198/016214507000000121

Ishwaran, H. and Rao, J. S. (2005). Spike and slab variable selection: Frequentist and Bayesian strategies. *Ann. Statist.* **33** 730–773. MR2163158 https://doi.org/10.1214/009053604000001147

Johnson, V. E. and Rossell, D. (2012). Bayesian model selection in high-dimensional settings. *J. Amer. Statist. Assoc.* **107** 649–660. MR2980074 https://doi.org/10.1080/01621459.2012.682536

Kinney, S. K. and Dunson, D. B. (2007). Fixed and random effects selection in linear and logistic models. *Biometrics* **63** 690–698. MR2395705 https://doi.org/10.1111/j.1541-0420.2007.00771.x

Liang, F., Paulo, R., Molina, G., Clyde, M. A. and Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection. *J. Amer. Statist. Assoc.* **103** 410–423. MR2420243 https://doi.org/10.1198/016214507000001337

Lin, J. and Michailidis, G. (2017). Regularized estimation and testing for high-dimensional multi-block vector-autoregressive models. *J. Mach. Learn. Res.* **18** Paper No. 117, 49. MR3725456 https://doi.org/10.1631/jzus.a1500279

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer, Berlin. MR2172368 https://doi.org/10.1007/978-3-540-27752-1

Melnyk, I. and Banerjee, A. (2016). Estimating structured vector autoregressive models. In *Proceedings of The 33rd International Conference on Machine Learning* 830–839.

Michailidis, G. and d'Alché-Buc, F. (2013). Autoregressive models for gene regulatory network inference: Sparsity, stability and causality issues. *Math. Biosci.* **246** 326–334. MR3132054 https://doi.org/10.1016/j.mbs.2013.10.003

Narisetty, N. N. and He, X. (2014). Bayesian variable selection with shrinking and diffusing priors. *Ann. Statist.* **42** 789–817. MR3210987 https://doi.org/10.1214/14-AOS1207

Nicholson, W. B., Matteson, D. S. and Bien, J. (2017). Varx-l: Structured regularization for large vector autoregressions with exogenous variables. *Int. J. Forecast.* **33** 627–651.

Robertson, J. C. and Tallman, E. W. (1999). Vector autoregressions: Forecasting and reality. *Econ. Rev.* **84**.

Rudelson, M. and Vershynin, R. (2013). Hanson–Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.* **18** no. 82, 9. MR3125258 https://doi.org/10.1214/ECP.v18-2865

Seth, A. K., Barrett, A. B. and Barnett, L. (2015). Granger causality analysis in neuroscience and neuroimaging. *J. Neurosci.* **35** 3293–3297.

Shin, M., Bhattacharya, A. and Johnson, V. E. (2018). Scalable Bayesian variable selection using nonlocal prior densities in ultrahigh-dimensional settings. *Statist. Sinica* **28** 1053–1078. MR3791100

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica* **48** 1–48.

Sims, C. A. and Zha, T. (1998). Bayesian methods for dynamic multivariate models. *Internat. Econom. Rev.* 949–968.

Song, Q. and Liang, F. (2015). A split-and-merge Bayesian variable selection approach for ultrahigh dimensional regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 947–972. MR3414135 https://doi.org/10.1111/rssb.12095

Stock, J. H. and Watson, M. W. (2005). An empirical comparison of methods for forecasting using many predictors. Princeton Univ. Manuscript.

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

Vershynin, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge Univ. Press, Cambridge. MR2963170

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

The Annals of Statistics
2021, Vol. 49, No. 3, 1300–1317
https://doi.org/10.1214/20-AOS2000
© Institute of Mathematical Statistics, 2021

ON CROSS-VALIDATED LASSO IN HIGH DIMENSIONS

BY DENIS CHETVERIKOV[1,*], ZHIPENG LIAO[1,†] AND VICTOR CHERNOZHUKOV[2]

[1] *Department of Economics, UCLA,* [*] *chetverikov@econ.ucla.edu;* [†] *zhipeng.liao@econ.ucla.edu*
[2] *Department of Economics and Operations Research Center, MIT, vchern@mit.edu*

In this paper, we derive nonasymptotic error bounds for the Lasso estimator when the penalty parameter for the estimator is chosen using K-fold cross-validation. Our bounds imply that the cross-validated Lasso estimator has nearly optimal rates of convergence in the prediction, L^2, and L^1 norms. For example, we show that in the model with the Gaussian noise and under fairly general assumptions on the candidate set of values of the penalty parameter, the estimation error of the cross-validated Lasso estimator converges to zero in the prediction norm with the $\sqrt{s \log p/n} \times \sqrt{\log(pn)}$ rate, where n is the sample size of available data, p is the number of covariates and s is the number of nonzero coefficients in the model. Thus, the cross-validated Lasso estimator achieves the fastest possible rate of convergence in the prediction norm up to a small logarithmic factor $\sqrt{\log(pn)}$, and similar conclusions apply for the convergence rate both in L^2 and in L^1 norms. Importantly, our results cover the case when p is (potentially much) larger than n and also allow for the case of non-Gaussian noise. Our paper therefore serves as a justification for the widely spread practice of using cross-validation as a method to choose the penalty parameter for the Lasso estimator.

REFERENCES

[1] ABADIE, A. and KASY, M. (2018). Choosing among regularized estimators in empirical economics. *Rev. Econ. Stat.* **100** 743–762.
[2] ARLOT, S. and CELISSE, A. (2010). A survey of cross-validation procedures for model selection. *Stat. Surv.* **4** 40–79. MR2602303 https://doi.org/10.1214/09-SS054
[3] BELLEC, P. and ZHANG, C.-H. (2018). Second order Stein: SURE for SURE and other applications in high-dimensional inference. Preprint. Available at arXiv:1811.04121.
[4] BELLEC, P. and ZHANG, C.-H. (2020). De-biasing the Lasso with degrees-of-freedom adjustment. Preprint. Available at arXiv:1902.08885v2.
[5] BELLEC, P. C. (2018). The noise barrier and the large signal bias of the Lasso and other convex estimators. Preprint. Available at arXiv:1804.01230.
[6] BELLONI, A. and CHERNOZHUKOV, V. (2011). High dimensional sparse econometric models: An introduction. In *Inverse Problems and High-Dimensional Estimation*. *Lect. Notes Stat. Proc.* **203** 121–156. Springer, Heidelberg. MR2868201 https://doi.org/10.1007/978-3-642-19989-9_3
[7] BELLONI, A. and CHERNOZHUKOV, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli* **19** 521–547. MR3037163 https://doi.org/10.3150/11-BEJ410
[8] 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
[9] BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. MR2533469 https://doi.org/10.1214/08-AOS620
[10] BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data*: *Methods*, *Theory and Applications*. *Springer Series in Statistics*. Springer, Heidelberg. MR2807761 https://doi.org/10.1007/978-3-642-20192-9
[11] CHATTERJEE, S. and JAFAROV, J. (2015). Prediction error of cross-validated Lasso. Preprint. Available at arXiv:1502.06292.
[12] CHETVERIKOV, D., LIAO, Z. and CHERNOZHUKOV, V. (2021). Supplement to "On cross-validated Lasso in high dimensions." https://doi.org/10.1214/20-AOS2000SUPP

MSC2020 subject classifications. 62J07.
Key words and phrases. Cross-validation, Lasso, high-dimensional models, sparsity, nonasymptotic bounds.

[13] GIRAUD, C. (2015). *Introduction to High-Dimensional Statistics*. *Monographs on Statistics and Applied Probability* **139**. CRC Press, Boca Raton, FL. MR3307991

[14] HASTIE, T., TIBSHIRANI, R. and WAINWRIGHT, M. (2015). *Statistical Learning with Sparsity*: *The Lasso and Generalizations*. *Monographs on Statistics and Applied Probability* **143**. CRC Press, Boca Raton, FL. MR3616141

[15] HOMRIGHAUSEN, D. and MCDONALD, D. (2013). The Lasso, persistence, and cross-validation. In *Proceedings of the* 30*th International Conference on Machine Learning* **28**.

[16] HOMRIGHAUSEN, D. and MCDONALD, D. J. (2014). Leave-one-out cross-validation is risk consistent for Lasso. *Mach*. *Learn*. **97** 65–78. MR3252827 https://doi.org/10.1007/s10994-014-5438-z

[17] HOMRIGHAUSEN, D. and MCDONALD, D. J. (2017). Risk consistency of cross-validation with Lasso-type procedures. *Statist*. *Sinica* **27** 1017–1036. MR3699692

[18] LECUÉ, G. and MENDELSON, S. (2018). Regularization and the small-ball method I: Sparse recovery. *Ann*. *Statist*. **46** 611–641. MR3782379 https://doi.org/10.1214/17-AOS1562

[19] LECUÉ, G. and MITCHELL, C. (2012). Oracle inequalities for cross-validation type procedures. *Electron*. *J*. *Stat*. **6** 1803–1837. MR2988465 https://doi.org/10.1214/12-EJS730

[20] LI, K.-C. (1987). Asymptotic optimality for C_p, C_L, cross-validation and generalized cross-validation: Discrete index set. *Ann*. *Statist*. **15** 958–975. MR0902239 https://doi.org/10.1214/aos/1176350486

[21] MIOLANE, L. and MONTANARI, A. (2018). The distribution of the Lasso: Uniform control over sparse balls and adaptive parameter tuning. Preprint. Available at arXiv:1811.01212.

[22] RIGOLLET, P. and TSYBAKOV, A. (2011). Exponential screening and optimal rates of sparse estimation. *Ann*. *Statist*. **39** 731–771. MR2816337 https://doi.org/10.1214/10-AOS854

[23] SUN, T. and ZHANG, C.-H. (2013). Sparse matrix inversion with scaled Lasso. *J*. *Mach*. *Learn*. *Res*. **14** 3385–3418. MR3144466

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

[25] TIBSHIRANI, R. J. (2013). The lasso problem and uniqueness. *Electron*. *J*. *Stat*. **7** 1456–1490. MR3066375 https://doi.org/10.1214/13-EJS815

[26] 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

[27] VAN DE GEER, S. (2016). *Estimation and Testing Under Sparsity*. *Lecture Notes in Math*. **2159**. Springer, Cham. MR3526202 https://doi.org/10.1007/978-3-319-32774-7

[28] 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

[29] WEGKAMP, M. (2003). Model selection in nonparametric regression. *Ann*. *Statist*. **31** 252–273. MR1962506 https://doi.org/10.1214/aos/1046294464

[30] ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann*. *Statist*. **38** 894–942. MR2604701 https://doi.org/10.1214/09-AOS729

[31] ZHANG, C.-H. and HUANG, J. (2008). The sparsity and bias of the LASSO selection in high-dimensional linear regression. *Ann*. *Statist*. **36** 1567–1594. MR2435448 https://doi.org/10.1214/07-AOS520

[32] ZOU, H., HASTIE, T. and TIBSHIRANI, R. (2007). On the "degrees of freedom" of the Lasso. *Ann*. *Statist*. **35** 2173–2192. MR2363967 https://doi.org/10.1214/009053607000000127

The Annals of Statistics
2021, Vol. 49, No. 3, 1318–1346
https://doi.org/10.1214/20-AOS2001
© Institute of Mathematical Statistics, 2021

FRAME-CONSTRAINED TOTAL VARIATION REGULARIZATION FOR WHITE NOISE REGRESSION

BY MIGUEL DEL ÁLAMO[*], HOUSEN LI[†] AND AXEL MUNK[‡]

Institute for Mathematical Stochastics, University of Göttingen, [*]*miguel.del-alamo@mathematik.uni-goettingen.de;* [†]*housen.li@mathematik.uni-goettingen.de;* [‡]*munk@math.uni-goettingen.de*

Despite the popularity and practical success of total variation (TV) regularization for function estimation, surprisingly little is known about its theoretical performance in a statistical setting. While TV regularization has been known for quite some time to be minimax optimal for denoising one-dimensional signals, for higher dimensions this remains elusive until today. In this paper, we consider frame-constrained TV estimators including many well-known (overcomplete) frames in a white noise regression model, and prove their minimax optimality w.r.t. L^q-risk ($1 \leq q < \infty$) up to a logarithmic factor in any dimension $d \geq 1$. Overcomplete frames are an established tool in mathematical imaging and signal recovery, and their combination with TV regularization has been shown to give excellent results in practice, which our theory now confirms. Our results rely on a novel connection between frame-constraints and certain Besov norms, and on an interpolation inequality to relate them to the risk functional. Additionally, our results explain a phase transition in the minimax risk for *BV* functions.

REFERENCES

[1] ASSOUAD, P. (1983). Deux remarques sur l'estimation. *C. R. Acad. Sci. Paris Sér. I Math.* **296** 1021–1024. MR0777600

[2] BORUP, L. and NIELSEN, M. (2007). Frame decomposition of decomposition spaces. *J. Fourier Anal. Appl.* **13** 39–70. MR2296727 https://doi.org/10.1007/s00041-006-6024-y

[3] BROWN, L. D. and LOW, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24** 2384–2398. MR1425958 https://doi.org/10.1214/aos/1032181159

[4] CAI, T. T. and LOW, M. G. (2005). Nonparametric estimation over shrinking neighborhoods: Superefficiency and adaptation. *Ann. Statist.* **33** 184–213. MR2157801 https://doi.org/10.1214/009053604000000832

[5] CANDÈS, E. J. and DONOHO, D. L. (2000). Curvelets: A surprisingly effective nonadaptive representation for objects with edges. Technical report, Stanford Univ., California, Dept. of Statistics.

[6] CANDÈS, E. J. and DONOHO, D. L. (2004). New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities. *Comm. Pure Appl. Math.* **57** 219–266. MR2012649 https://doi.org/10.1002/cpa.10116

[7] CANDÈS, E. J. and GUO, F. (2002). New multiscale transforms, minimum total variation synthesis: Applications to edge-preserving image reconstruction. *Signal Process.* **82** 1519–1543.

[8] CHAMBOLLE, A. (2004). An algorithm for total variation minimization and applications. *J. Math. Imaging Vision* **20** 89–97. MR2049783 https://doi.org/10.1023/B:JMIV.0000011320.81911.38

[9] CHAMBOLLE, A. and LIONS, P.-L. (1997). Image recovery via total variation minimization and related problems. *Numer. Math.* **76** 167–188. MR1440119 https://doi.org/10.1007/s002110050258

[10] CHAMBOLLE, A. and POCK, T. (2011). A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vision* **40** 120–145. MR2782122 https://doi.org/10.1007/s10851-010-0251-1

[11] CLASON, C., JIN, B. and KUNISCH, K. (2010). A semismooth Newton method for L^1 data fitting with automatic choice of regularization parameters and noise calibration. *SIAM J. Imaging Sci.* **3** 199–231. MR2657627 https://doi.org/10.1137/090758003

MSC2020 subject classifications. Primary 62G05; secondary 62G20, 62M40.
Key words and phrases. Nonparametric regression, minimax estimation, total variation, interpolation inequalities, wavelets, overcomplete dictionaries.

[12] CLASON, C., KRUSE, F. and KUNISCH, K. (2018). Total variation regularization of multi-material topology optimization. *ESAIM Math. Model. Numer. Anal.* **52** 275–303. MR3808161 https://doi.org/10.1051/m2an/2017061

[13] COHEN, A. (2003). *Numerical Analysis of Wavelet Methods. Studies in Mathematics and Its Applications* **32**. North-Holland, Amsterdam. MR1990555

[14] COHEN, A., DAHMEN, W., DAUBECHIES, I. and DEVORE, R. (2003). Harmonic analysis of the space BV. *Rev. Mat. Iberoam.* **19** 235–263. MR1993422 https://doi.org/10.4171/RMI/345

[15] CONDAT, L. (2017). Discrete total variation: New definition and minimization. *SIAM J. Imaging Sci.* **10** 1258–1290. MR3684410 https://doi.org/10.1137/16M1075247

[16] DALALYAN, A. S., HEBIRI, M. and LEDERER, J. (2017). On the prediction performance of the Lasso. *Bernoulli* **23** 552–581. MR3556784 https://doi.org/10.3150/15-BEJ756

[17] DAUBECHIES, I. (1992). *Ten Lectures on Wavelets. CBMS-NSF Regional Conference Series in Applied Mathematics* **61**. SIAM, Philadelphia, PA. MR1162107 https://doi.org/10.1137/1.9781611970104

[18] DAVIES, P. L. and KOVAC, A. (2001). Local extremes, runs, strings and multiresolution. *Ann. Statist.* **29** 1–65. MR1833958 https://doi.org/10.1214/aos/996986501

[19] DEL ÁLAMO, M., LI, H. and MUNK, A. (2021). Supplement to "Frame-constrained total variation regularization for white noise regression." https://doi.org/10.1214/20-AOS2001SUPP

[20] DELYON, B. and JUDITSKY, A. (1996). On minimax wavelet estimators. *Appl. Comput. Harmon. Anal.* **3** 215–228. MR1400080 https://doi.org/10.1006/acha.1996.0017

[21] DENG, H. and ZHANG, C.-H. (2020). Isotonic regression in multi-dimensional spaces and graphs. *Ann. Statist.* To appear.

[22] DONG, Y., HINTERMÜLLER, M. and RINCON-CAMACHO, M. M. (2011). Automated regularization parameter selection in multi-scale total variation models for image restoration. *J. Math. Imaging Vision* **40** 82–104. MR2782120 https://doi.org/10.1007/s10851-010-0248-9

[23] DONOHO, D. L. (1993). Unconditional bases are optimal bases for data compression and for statistical estimation. *Appl. Comput. Harmon. Anal.* **1** 100–115. MR1256530 https://doi.org/10.1006/acha.1993.1008

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

[25] STARCK, J. L., DONOHO, D. L. and CANDÈS, E. J. (2001). Very high quality image restoration by combining wavelets and curvelets. In *Wavelet Applications in Signal and Image Processing IX* **4478** 9–19.

[26] DÜMBGEN, L. and KOVAC, A. (2009). Extensions of smoothing via taut strings. *Electron. J. Stat.* **3** 41–75. MR2471586 https://doi.org/10.1214/08-EJS216

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

[28] DURAND, S. and FROMENT, J. (2001). Artifact free signal denoising with wavelets. In *Acoustics, Speech, and Signal Processing*, 2001. *Proceedings*. (*ICASSP*'01). 2001 *IEEE International Conference on* **6** 3685–3688. IEEE, New York.

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

[30] FRICK, K., MARNITZ, P. and MUNK, A. (2012). Statistical multiresolution Dantzig estimation in imaging: Fundamental concepts and algorithmic framework. *Electron. J. Stat.* **6** 231–268. MR2988407 https://doi.org/10.1214/12-EJS671

[31] FRICK, K., MARNITZ, P. and MUNK, A. (2013). Statistical multiresolution estimation for variational imaging: With an application in Poisson-biophotonics. *J. Math. Imaging Vision* **46** 370–387. MR3068534 https://doi.org/10.1007/s10851-012-0368-5

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

[33] GARNETT, J. B., LE, T. M., MEYER, Y. and VESE, L. A. (2007). Image decompositions using bounded variation and generalized homogeneous Besov spaces. *Appl. Comput. Harmon. Anal.* **23** 25–56. MR2333827 https://doi.org/10.1016/j.acha.2007.01.005

[34] 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

[35] GOBET, E., HOFFMANN, M. and REISS, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.* **32** 2223–2253. MR2102509 https://doi.org/10.1214/009053604000000797

[36] GOLDENSHLUGER, A. and LEPSKI, O. (2014). On adaptive minimax density estimation on R^d. *Probab. Theory Related Fields* **159** 479–543. MR3230001 https://doi.org/10.1007/s00440-013-0512-1

[37] GRASMAIR, M., LI, H. and MUNK, A. (2018). Variational multiscale nonparametric regression: Smooth functions. *Ann. Inst. Henri Poincaré Probab. Stat.* **54** 1058–1097. MR3795077 https://doi.org/10.1214/17-AIHP832

[38] 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

[39] GUO, K., KUTYNIOK, G. and LABATE, D. (2006). Sparse multidimensional representations using anisotropic dilation and shear operators. In *Wavelets and Splines: Athens* 2005. *Mod. Methods Math.* 189–201. Nashboro Press, Brentwood, TN. MR2233452

[40] HADDAD, A. and MEYER, Y. (2007). An improvement of Rudin–Osher–Fatemi model. *Appl. Comput. Harmon. Anal.* **22** 319–334. MR2311857 https://doi.org/10.1016/j.acha.2006.09.001

[41] HALTMEIER, M. and MUNK, A. (2014). Extreme value analysis of empirical frame coefficients and implications for denoising by soft-thresholding. *Appl. Comput. Harmon. Anal.* **36** 434–460. MR3175087 https://doi.org/10.1016/j.acha.2013.07.004

[42] 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

[43] HÄRDLE, W., KERKYACHARIAN, G., PICARD, D. and TSYBAKOV, A. (2012). *Wavelets, Approximation, and Statistical Applications*. *Lecture Notes in Statistics* **129**. Springer, New York. MR1618204 https://doi.org/10.1007/978-1-4612-2222-4

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

[45] JIANG, H. (2014). *Photoacoustic Tomography*. CRC Press, Boca Raton.

[46] LABATE, D., LIM, W. Q., KUTYNIOK, G. and WEISS, G. (2005). Sparse multidimensional representation using shearlets. In *Wavelets XI* **5914** 59140U. International Society for Optics and Photonics.

[47] LABATE, D., MANTOVANI, L. and NEGI, P. (2013). Shearlet smoothness spaces. *J. Fourier Anal. Appl.* **19** 577–611. MR3048591 https://doi.org/10.1007/s00041-013-9261-x

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

[49] LASSAS, M. and SILTANEN, S. (2004). Can one use total variation prior for edge-preserving Bayesian inversion? *Inverse Probl.* **20** 1537–1563. MR2109134 https://doi.org/10.1088/0266-5611/20/5/013

[50] LEDOUX, M. (2003). On improved Sobolev embedding theorems. *Math. Res. Lett.* **10** 659–669. MR2024723 https://doi.org/10.4310/MRL.2003.v10.n5.a9

[51] FANG, B., GUNTUBOYINA, A. and SEN, B. (2020). Multivariate extensions of isotonic regression and total variation denoising via entire monotonicity and Hardy–Krause variation. *Ann. Statist.* To appear.

[52] LEPSKI, O. (2015). Adaptive estimation over anisotropic functional classes via oracle approach. *Ann. Statist.* **43** 1178–1242. MR3346701 https://doi.org/10.1214/14-AOS1306

[53] LEPSKI, O. V., MAMMEN, E. and SPOKOINY, V. G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: An approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.* **25** 929–947. MR1447734 https://doi.org/10.1214/aos/1069362731

[54] LEPSKII, O. (1991). On a problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** 454–466.

[55] LI, H. (2016). Variational estimators in statistical multiscale analysis. Ph.D. thesis, Georg-August-Universität Göttingen.

[56] LI, H., GUO, Q. and MUNK, A. (2019). Multiscale change-point segmentation: Beyond step functions. *Electron. J. Stat.* **13** 3254–3296. MR4010980 https://doi.org/10.1214/19-ejs1608

[57] LI, H., HALTMEIER, M., ZHANG, S., FRAHM, J. and MUNK, A. (2014). Aggregated motion estimation for real-time MRI reconstruction. *Magn. Reson. Med.* **72** 1039–1048.

[58] MALGOUYRES, F. (2001). A unified framework for image restoration. Technical report, Univ. of California, Los Angeles.

[59] MALGOUYRES, F. (2002). Mathematical analysis of a model which combines total variation and wavelet for image restoration. *J. Inf. Process.* **2** 1–10.

[60] 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

[61] MEYER, Y. (2001). *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations: The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures*. *University Lecture Series* **22**. Amer. Math. Soc., Providence, RI. MR1852741 https://doi.org/10.1090/ulect/022

[62] MUNK, A., BISSANTZ, N., WAGNER, T. and FREITAG, G. (2005). On difference-based variance estimation in nonparametric regression when the covariate is high dimensional. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 19–41. MR2136637 https://doi.org/10.1111/j.1467-9868.2005.00486.x

[63] 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

[64] NEMIROVSKIY, A. S. (1985). Nonparametric estimation of smooth regression functions. *Izv. Akad. Nauk SSSR Tekhn. Kibernet*. **3** 50–60. MR0844292

[65] NIINIMÄKI, K., LASSAS, M., HÄMÄLÄINEN, K., KALLONEN, A., KOLEHMAINEN, V., NIEMI, E. and SILTANEN, S. (2016). Multiresolution parameter choice method for total variation regularized tomography. *SIAM J. Imaging Sci*. **9** 938–974. MR3521540 https://doi.org/10.1137/15M1034076

[66] NIRENBERG, L. (1959). On elliptic partial differential equations. *Ann. Sc. Norm. Super. Pisa Cl. Sci*. (3) **13** 115–162. MR0109940

[67] OSHER, S., SOLÉ, A. and VESE, L. (2003). Image decomposition and restoration using total variation minimization and the H^{-1} norm. *Multiscale Model. Simul*. **1** 349–370. MR2030155 https://doi.org/10.1137/S1540345902416247

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

[69] REISS, M. (2008). Asymptotic equivalence for nonparametric regression with multivariate and random design. *Ann. Statist*. **36** 1957–1982. MR2435461 https://doi.org/10.1214/07-AOS525

[70] 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

[71] 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.

[72] SCHERZER, O., GRASMAIR, M., GROSSAUER, H., HALTMEIER, M. and LENZEN, F. (2009). *Variational Methods in Imaging*. *Applied Mathematical Sciences* **167**. Springer, New York. MR2455620

[73] SCHMEISSER, H.-J. and TRIEBEL, H. (1987). *Topics in Fourier Analysis and Function Spaces*. *A Wiley-Interscience Publication*. Wiley, Chichester. MR0891189

[74] SPOKOINY, V. (2002). Variance estimation for high-dimensional regression models. *J. Multivariate Anal*. **82** 111–133. MR1918617 https://doi.org/10.1006/jmva.2001.2023

[75] TRIEBEL, H. (1988). Characterizations of Besov–Hardy–Sobolev spaces: A unified approach. *J. Approx. Theory* **52** 162–203. MR0929302 https://doi.org/10.1016/0021-9045(88)90055-X

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

[77] WAHBA, G. (1977). Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal*. **14** 651–667. MR0471299 https://doi.org/10.1137/0714044

[78] 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

The Annals of Statistics
2021, Vol. 49, No. 3, 1347–1377
https://doi.org/10.1214/20-AOS2002
© Institute of Mathematical Statistics, 2021

ADAPTIVE ROBUST ESTIMATION IN SPARSE VECTOR MODEL

BY L. COMMINGES[1], O. COLLIER[2], M. NDAOUD[3,*] AND A. B. TSYBAKOV[3,†]

[1]*CEREMADE, Université Paris-Dauphine, PSL and CREST, laetitia.comminges@dauphine.fr*
[2]*Modal'X, UPL, Université Paris Nanterre and CREST, olivier.collier@parisnanterre.fr*
[3]*CREST, ENSAE, [*]ndaoudm@gmail.com; [†]alexandre.tsybakov@ensae.fr*

For the sparse vector model, we consider estimation of the target vector, of its ℓ_2-norm and of the noise variance. We construct adaptive estimators and establish the optimal rates of adaptive estimation when adaptation is considered with respect to the triplet "noise level—noise distribution—sparsity." We consider classes of noise distributions with polynomially and exponentially decreasing tails as well as the case of Gaussian noise. The obtained rates turn out to be different from the minimax nonadaptive rates when the triplet is known. A crucial issue is the ignorance of the noise variance. Moreover, knowing or not knowing the noise distribution can also influence the rate. For example, the rates of estimation of the noise variance can differ depending on whether the noise is Gaussian or sub-Gaussian without a precise knowledge of the distribution. Estimation of noise variance in our setting can be viewed as an adaptive variant of robust estimation of scale in the contamination model, where instead of fixing the "nominal" distribution in advance we assume that it belongs to some class of distributions.

REFERENCES

[1] BELLEC, P. C., LECUÉ, G. and TSYBAKOV, A. B. (2018). Slope meets Lasso: Improved oracle bounds and optimality. *Ann. Statist.* **46** 3603–3642. MR3852663 https://doi.org/10.1214/17-AOS1670

[2] BELLONI, A., CHERNOZHUKOV, V. and WANG, L. (2014). Pivotal estimation via square-root Lasso in nonparametric regression. *Ann. Statist.* **42** 757–788. MR3210986 https://doi.org/10.1214/14-AOS1204

[3] BELOMESTNY, D., TRABS, M. and TSYBAKOV, A. B. (2019). Sparse covariance matrix estimation in high-dimensional deconvolution. *Bernoulli* **25** 1901–1938. MR3961235 https://doi.org/10.3150/18-BEJ1040A

[4] BUTUCEA, C. and MATIAS, C. (2005). Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model. *Bernoulli* **11** 309–340. MR2132729 https://doi.org/10.3150/bj/1116340297

[5] CAI, T. T. and JIN, J. (2010). Optimal rates of convergence for estimating the null density and proportion of nonnull effects in large-scale multiple testing. *Ann. Statist.* **38** 100–145. MR2589318 https://doi.org/10.1214/09-AOS696

[6] CARPENTIER, A., COLLIER, O., COMMINGES, L., TSYBAKOV, A. B. and WANG, Y. (2019). Minimax rate of testing in sparse linear regression. *Autom. Remote Control* **80** 1817–1834.

[7] CARPENTIER, A. and VERZELEN, N. (2019). Adaptive estimation of the sparsity in the Gaussian vector model. *Ann. Statist.* **47** 93–126. MR3909928 https://doi.org/10.1214/17-AOS1680

[8] CHEN, M., GAO, C. and REN, Z. (2018). Robust covariance and scatter matrix estimation under Huber's contamination model. *Ann. Statist.* **46** 1932–1960. MR3845006 https://doi.org/10.1214/17-AOS1607

[9] 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

[10] COLLIER, O., COMMINGES, L., TSYBAKOV, A. B. and VERZELEN, N. (2018). Optimal adaptive estimation of linear functionals under sparsity. *Ann. Statist.* **46** 3130–3150. MR3851767 https://doi.org/10.1214/17-AOS1653

[11] COMMINGES, L., COLLIER, O., NDAOUD, M. and TSYBAKOV, A. (2021). Supplement to "Adaptive robust estimation in sparse vector model." https://doi.org/10.1214/20-AOS2002SUPP

MSC2020 subject classifications. 62J05, 62G05.

Key words and phrases. Sparse vector model, variance estimation, functional estimation, robust estimation, adaptive estimation, minimax rate.

[12] 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. MR1157714

[13] GAUTIER, E. and TSYBAKOV, A. B. (2013). Pivotal estimation in high-dimensional regression via linear programming. In *Empirical Inference* 195–204. Springer, Heidelberg. MR3236866 https://doi.org/10.1007/978-3-642-41136-6_17

[14] GOLUBEV, G. K. and KRYMOVA, E. A. (2018). Noise level estimation in high-dimensional linear models. *Probl. Inf. Transm.* **54** 351–371.

[15] GUO, Z., WANG, W., CAI, T. T. and LI, H. (2019). Optimal estimation of genetic relatedness in high-dimensional linear models. *J. Amer. Statist. Assoc.* **114** 358–369. MR3941260 https://doi.org/10.1080/01621459.2017.1407774

[16] HUBER, P. J. (1981). *Robust Statistics*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0606374

[17] IBRAGIMOV, I. A. and HASMINSKII, R. Z. (1981). *Statistical Estimation*: *Asymptotic Theory*. *Applications of Mathematics* **16**. Springer, New York–Berlin. MR0620321

[18] JANSON, L., FOYGEL BARBER, R. and CANDÈS, E. (2017). EigenPrism: Inference for high dimensional signal-to-noise ratios. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1037–1065. MR3689308 https://doi.org/10.1111/rssb.12203

[19] MINSKER, S. and WEI, X. (2017). Estimation of the covariance structure of heavy-tailed distributions. In *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, eds.) **30** 2859–2868. Curran Associates, Red Hook, NY.

[20] PETROV, V. V. (1995). *Limit Theorems of Probability Theory*: *Sequences of Independent Random Variables*. *Oxford Studies in Probability* **4**. Clarendon Press, Oxford. MR1353441

[21] SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York. MR0838963

[22] SUN, T. and ZHANG, C.-H. (2012). Scaled sparse linear regression. *Biometrika* **99** 879–898. MR2999166 https://doi.org/10.1093/biomet/ass043

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

[24] VERZELEN, N. (2012). Minimax risks for sparse regressions: Ultra-high dimensional phenomenons. *Electron. J. Stat.* **6** 38–90. MR2879672 https://doi.org/10.1214/12-EJS666

[25] VERZELEN, N. and GASSIAT, E. (2018). Adaptive estimation of high-dimensional signal-to-noise ratios. *Bernoulli* **24** 3683–3710. MR3788186 https://doi.org/10.3150/17-BEJ975

[26] WASSERMAN, L. (2005). *All of Statistics*: *A Concise Course in Statistical Inference*. *Springer Texts in Statistics*. Springer, New York. MR2055670 https://doi.org/10.1007/978-0-387-21736-9

The Annals of Statistics
2021, Vol. 49, No. 3, 1378–1406
https://doi.org/10.1214/20-AOS2004
© Institute of Mathematical Statistics, 2021

LEARNING MODELS WITH UNIFORM PERFORMANCE VIA DISTRIBUTIONALLY ROBUST OPTIMIZATION

BY JOHN C. DUCHI[1] AND HONGSEOK NAMKOONG[2]

[1]*Departments of Statistics and Electrical Engineering, Stanford University, jduchi@stanford.edu*

[2]*Decision, Risk, and Operations Division, Columbia Business School, namkoong@gsb.columbia.edu*

A common goal in statistics and machine learning is to learn models that can perform well against distributional shifts, such as latent heterogeneous subpopulations, unknown covariate shifts or unmodeled temporal effects. We develop and analyze a distributionally robust stochastic optimization (DRO) framework that learns a model providing good performance against perturbations to the data-generating distribution. We give a convex formulation for the problem, providing several convergence guarantees. We prove finite-sample minimax upper and lower bounds, showing that distributional robustness sometimes comes at a cost in convergence rates. We give limit theorems for the learned parameters, where we fully specify the limiting distribution so that confidence intervals can be computed. On real tasks including generalizing to unknown subpopulations, fine-grained recognition and providing good tail performance, the distributionally robust approach often exhibits improved performance.

REFERENCES

[1] AGARWAL, A., BARTLETT, P. L., RAVIKUMAR, P. and WAINWRIGHT, M. J. (2012). Information-theoretic lower bounds on the oracle complexity of stochastic convex optimization. *IEEE Trans. Inf. Theory* **58** 3235–3249. MR2952543 https://doi.org/10.1109/TIT.2011.2182178

[2] AHMADI-JAVID, A. (2012). Entropic value-at-risk: A new coherent risk measure. *J. Optim. Theory Appl.* **155** 1105–1123. MR3000633 https://doi.org/10.1007/s10957-011-9968-2

[3] AITKIN, M. and RUBIN, D. B. (1985). Estimation and hypothesis testing in finite mixture models. *J. Roy. Statist. Soc. Ser. B* **47** 67–75.

[4] ALI, S. M. and SILVEY, S. D. (1966). A general class of coefficients of divergence of one distribution from another. *J. Roy. Statist. Soc. Ser. B* **28** 131–142. MR0196777

[5] AMODEI, D., ANANTHANARAYANAN, S., ANUBHAI, R., BAI, J., BATTENBERG, E., CASE, C., CASPER, J., CATANZARO, B., CHENG, Q. et al. (2016). Deep speech 2: End-to-end speech recognition in English and Mandarin. In *Proceedings of the* 33*rd International Conference on Machine Learning* 173–182.

[6] ANDERSON, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* **6** 170–176. MR0069229 https://doi.org/10.2307/2032333

[7] ARTZNER, P., DELBAEN, F., EBER, J.-M. and HEATH, D. (1999). Coherent measures of risk. *Math. Finance* **9** 203–228. MR1850791 https://doi.org/10.1111/1467-9965.00068

[8] ASUNCION, A. and NEWMAN, D. J. (2007). UCI Machine Learning Repository.

[9] ATAR, R., CHOWDHARY, K. and DUPUIS, P. (2015). Robust bounds on risk-sensitive functionals via Rényi divergence. *SIAM/ASA J. Uncertain. Quantificat.* **3** 18–33. MR3299141 https://doi.org/10.1137/130939730

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

[11] BEN-DAVID, S., BLITZER, J., CRAMMER, K., KULESZA, A., PEREIRA, F. and VAUGHAN, J. W. (2010). A theory of learning from different domains. *Mach. Learn.* **79** 151–175. MR3108150 https://doi.org/10.1007/s10994-009-5152-4

[12] BEN-DAVID, S., BLITZER, J., CRAMMER, K. and PEREIRA, F. (2007). Analysis of representations for domain adaptation. In *Advances in Neural Information Processing Systems* 20 137–144.

MSC2020 subject classifications. 62F12, 68Q32, 62C20.
Key words and phrases. Robust optimization, minimax optimality, risk-averse learning.

[13] BEN-TAL, A., DEN HERTOG, D., WAEGENAERE, A. D., MELENBERG, B. and RENNEN, G. (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Manage. Sci.* **59** 341–357.

[14] BEN-TAL, A., EL GHAOUI, L. and NEMIROVSKI, A. (2009). *Robust Optimization*. *Princeton Series in Applied Mathematics*. Princeton Univ. Press, Princeton, NJ. MR2546839 https://doi.org/10.1515/9781400831050

[15] BERTSIMAS, D., GUPTA, V. and KALLUS, N. (2018). Data-driven robust optimization. *Math. Program.* **167** 235–292. MR3755733 https://doi.org/10.1007/s10107-017-1125-8

[16] BICKEL, S., BRÜCKNER, M. and SCHEFFER, T. (2007). Discriminative learning for differing training and test distributions. In *Proceedings of the* 24*th International Conference on Machine Learning*.

[17] BLANCHET, J., KANG, Y. and MURTHY, K. (2019). Robust Wasserstein profile inference and applications to machine learning. *J. Appl. Probab.* **56** 830–857. MR4015639 https://doi.org/10.1017/jpr.2019.49

[18] BLANCHET, J. and MURTHY, K. (2019). Quantifying distributional model risk via optimal transport. *Math. Oper. Res.* **44** 565–600. MR3959085 https://doi.org/10.1287/moor.2018.0936

[19] BLITZER, J., MCDONALD, R. and PEREIRA, F. (2006). Domain adaptation with structural correspondence learning. In *Proceedings of the* 2006 *Conference on Empirical Methods in Natural Language Processing* 120–128. Association for Computational Linguistics, Stroudsburg, PA.

[20] BLODGETT, S. L., GREEN, L. and O'CONNOR, B. (2016). Demographic dialectal variation in social media: A case study of African-American English. In *Proceedings of Empirical Methods for Natural Language Processing* 1119–1130.

[21] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities*: *A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford. MR3185193 https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

[22] BOYD, S. and VANDENBERGHE, L. (2004). *Convex Optimization*. Cambridge Univ. Press, Cambridge. MR2061575 https://doi.org/10.1017/CBO9780511804441

[23] BÜHLMANN, P. and MEINSHAUSEN, N. (2016). Magging: Maximin aggregation for inhomogeneous large-scale data. *Proc. IEEE* **104** 126–135.

[24] CAI, Z., FAN, J. and LI, R. (2000). Efficient estimation and inferences for varying-coefficient models. *J. Amer. Statist. Assoc.* **95** 888–902. MR1804446 https://doi.org/10.2307/2669472

[25] CAPPÉ, O., MOULINES, E. and RYDÉN, T. (2005). *Inference in Hidden Markov Models*. *Springer Series in Statistics*. Springer, New York. MR2159833

[26] CARUANA, R. (1998). Multitask learning. In *Learning to Learn* 95–133. Springer, Berlin.

[27] CRESSIE, N. and READ, T. R. C. (1984). Multinomial goodness-of-fit tests. *J. Roy. Statist. Soc. Ser. B* **46** 440–464. MR0790631

[28] CSISZÁR, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.* **2** 299–318. MR0219345

[29] DAUMÉ, H. III and MARCU, D. (2006). Domain adaptation for statistical classifiers. *J. Artificial Intelligence Res.* **26** 101–126. MR2306416 https://doi.org/10.1613/jair.1872

[30] DE CAMPOS, T. E., BABU, B. R. and VARMA, M. (2009). Character recognition in natural images. In *Proceedings of the Fourth International Conference on Computer Vision Theory and Applications*.

[31] DELAGE, E. and YE, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* **58** 595–612. MR2680566 https://doi.org/10.1287/opre.1090.0741

[32] DENKER, J. S., GARDNER, W. R., GRAF, H. P., HENDERSON, D., HOWARD, R. E., HUBBARD, W., JACKEL, L. D., BAIRD, H. S. and GUYON, I. (1988). Neural network recognizer for hand-written zip code digits. In *Advances in Neural Information Processing Systems* 1.

[33] DUCHI, J. C. (2018). Introductory lectures on stochastic optimization. In *The Mathematics of Data*. *IAS/Park City Math. Ser.* **25** 99–185. Amer. Math. Soc., Providence, RI. MR3839168

[34] DUCHI, J. C., GLYNN, P. W. and NAMKOONG, H. (2016). Statistics of robust optimization: A generalized empirical likelihood approach. Available at arXiv:1610.03425.

[35] DUCHI, J. C. and NAMKOONG, H. (2021). Supplement to "Learning models with uniform performance via distributionally robust optimization." https://doi.org/10.1214/20-AOS2004SUPP

[36] DUPUIS, P., KATSOULAKIS, M. A., PANTAZIS, Y. and PLECHÁČ, P. (2016). Path-space information bounds for uncertainty quantification and sensitivity analysis of stochastic dynamics. *SIAM/ASA J. Uncertain. Quantificat.* **4** 80–111. MR3455143 https://doi.org/10.1137/15M1025645

[37] ELDAR, Y. C., BEN-TAL, A. and NEMIROVSKI, A. (2004). Linear minimax regret estimation of deterministic parameters with bounded data uncertainties. *IEEE Trans. Signal Process.* **52** 2177–2188. MR2085579 https://doi.org/10.1109/TSP.2004.831144

[38] FAN, J. and ZHANG, W. (1999). Statistical estimation in varying coefficient models. *Ann. Statist.* **27** 1491–1518. MR1742497 https://doi.org/10.1214/aos/1017939139

[39] FIGUEIREDO, M. A. T. and JAIN, A. K. (2002). Unsupervised learning of finite mixture models. *IEEE Trans. Pattern Anal. Mach. Intell.* **24** 381–396.
[40] GAO, R. and KLEYWEGT, A. J. (2016). Distributionally robust stochastic optimization with Wasserstein distance. Available at arXiv:1604.02199.
[41] GARDNER, R. J. (2002). The Brunn–Minkowski inequality. *Bull. Amer. Math. Soc.* (*N.S.*) **39** 355–405. MR1898210 https://doi.org/10.1090/S0273-0979-02-00941-2
[42] GHOSH, S. and LAM, H. (2019). Robust analysis in stochastic simulation: Computation and performance guarantees. *Oper. Res.* **67** 232–249. MR3919867 https://doi.org/10.1287/opre.2018.1765
[43] GLASSERMAN, P. and XU, X. (2014). Robust risk measurement and model risk. *Quant. Finance* **14** 29–58. MR3175968 https://doi.org/10.1080/14697688.2013.822989
[44] GOTOH, J.-Y., KIM, M. J. and LIM, A. (2015). Robust empirical optimization is almost the same as mean-variance optimization. Available at https://ssrn.com/abstract=2827400.
[45] GROTHER, P. J., QUINN, G. W. and PHILLIPS, P. J. (2010). Report on the evaluation of 2D still-image face recognition algorithms. NIST Interagency/Internal Reports (NISTIR) 7709.
[46] HAND, D. J. (2006). Classifier technology and the illusion of progress. *Statist. Sci.* **21** 1–34. MR2275965 https://doi.org/10.1214/088342306000000060
[47] HANSEN, L. P. and SARGENT, T. J. (2008). *Robustness*. Princeton Univ. Press, Princeton, NJ. MR3617628 https://doi.org/10.1515/9781400829385
[48] HIRIART-URRUTY, J.-B. and LEMARÉCHAL, C. (1993). *Convex Analysis and Minimization Algorithms. I: Fundamentals. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **305**. Springer, Berlin. MR1261420
[49] HOVY, D. and SØGAARD, A. (2015). Tagging performance correlates with author age. In *Proceedings of the* 53*rd Annual Meeting of the Association for Computational Linguistics* (*Short Papers*) **2** 483–488.
[50] HUANG, J., GRETTON, A., BORGWARDT, K. M., SCHÖLKOPF, B. and SMOLA, A. J. (2007). Correcting sample selection bias by unlabeled data. In *Advances in Neural Information Processing Systems* 20 601–608.
[51] HUBER, P. J. (1981). *Robust Statistics. Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0606374
[52] HUBER, P. J. and RONCHETTI, E. M. (2009). *Robust Statistics*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2488795 https://doi.org/10.1002/9780470434697
[53] JIA, Y., SHELHAMER, E., DONAHUE, J., KARAYEV, S., LONG, J., GIRSHICK, R., GUADARRAMA, S. and DARRELL, T. (2014). Caffe: Convolutional architecture for fast feature embedding. Available at arXiv:1408.5093.
[54] JIANG, R. and GUAN, Y. (2016). Data-driven chance constrained stochastic program. *Math. Program.* **158** 291–327. MR3511385 https://doi.org/10.1007/s10107-015-0929-7
[55] KHOSLA, A., JAYADEVAPRAKASH, N., YAO, B. and LI, F.-F. (2011). Novel dataset for fine-grained image categorization. In *First Workshop on Fine-Grained Visual Categorization*, *IEEE Conference on Computer Vision and Pattern Recognition* **2** 1–2.
[56] KING, A. J. and WETS, R. J.-B. (1991). Epi-consistency of convex stochastic programs. *Stoch. Stoch. Rep.* **34** 83–92. MR1104423 https://doi.org/10.1080/17442509108833676
[57] KROKHMAL, P. A. (2007). Higher moment coherent risk measures. *Quant. Finance* **7** 373–387. MR2354775 https://doi.org/10.1080/14697680701458307
[58] KUSUOKA, S. (2001). On law invariant coherent risk measures. In *Advances in Mathematical Economics*, *Vol*. 3. *Adv. Math. Econ.* 83–95. Springer, Tokyo. MR1886557 https://doi.org/10.1007/978-4-431-67891-5_4
[59] LAM, H. (2016). Robust sensitivity analysis for stochastic systems. *Math. Oper. Res.* **41** 1248–1275. MR3544795 https://doi.org/10.1287/moor.2015.0776
[60] LAM, H. (2017). Sensitivity to serial dependency of input processes: A robust approach. *Manage. Sci.* **64** 1311–1327.
[61] LAM, H. and ZHOU, E. (2017). The empirical likelihood approach to quantifying uncertainty in sample average approximation. *Oper. Res. Lett.* **45** 301–307. MR3671280 https://doi.org/10.1016/j.orl.2017.04.003
[62] LE CAM, L. and YANG, G. L. (2000). *Asymptotics in Statistics*: *Some Basic Concepts*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR1784901 https://doi.org/10.1007/978-1-4612-1166-2
[63] LECUN, Y., BOSER, B., DENKER, J. S., HENDERSON, D., HOWARD, R. E., HUBBARD, W. and JACKEL, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1** 541–551.
[64] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces*: *Isoperimetry and Processes. Ergebnisse der Mathematik und Ihrer Grenzgebiete* (3) [*Results in Mathematics and Related Areas* (3)] **23**. Springer, Berlin. MR1102015 https://doi.org/10.1007/978-3-642-20212-4

[65] LEE, J. and RAGINSKY, M. (2017). Minimax statistical learning and domain adaptation with Wasserstein distances. Available at arXiv:1705.07815.

[66] MCLACHLAN, G. and PEEL, D. (2000). *Finite Mixture Models*. *Wiley Series in Probability and Statistics*: *Applied Probability and Statistics*. Wiley Interscience, New York. MR1789474 https://doi.org/10.1002/0471721182

[67] MEINSHAUSEN, N. and BÜHLMANN, P. (2015). Maximin effects in inhomogeneous large-scale data. *Ann. Statist.* **43** 1801–1830. MR3357879 https://doi.org/10.1214/15-AOS1325

[68] MOHAJERIN ESFAHANI, P. and KUHN, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Math. Program.* **171** 115–166. MR3844536 https://doi.org/10.1007/s10107-017-1172-1

[69] NAMKOONG, H. and DUCHI, J. C. (2016). Stochastic gradient methods for distributionally robust optimization with f-divergences. In *Advances in Neural Information Processing Systems* 29.

[70] NAMKOONG, H. and DUCHI, J. C. (2017). Variance regularization with convex objectives. In *Advances in Neural Information Processing Systems* 30.

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

[72] PETERSEN, I. R., JAMES, M. R. and DUPUIS, P. (2000). Minimax optimal control of stochastic uncertain systems with relative entropy constraints. *IEEE Trans. Automat. Control* **45** 398–412. MR1762853 https://doi.org/10.1109/9.847720

[73] PFLUG, G. and WOZABAL, D. (2007). Ambiguity in portfolio selection. *Quant. Finance* **7** 435–442. MR2354780 https://doi.org/10.1080/14697680701455410

[74] RECHT, B., ROELOFS, R., SCHMIDT, L. and SHANKAR, V. (2019). Do ImageNet classifiers generalize to ImageNet? In *Proceedings of the* 36*th International Conference on Machine Learning*.

[75] REDMOND, M. and BAVEJA, A. (2002). A data-driven software tool for enabling cooperative information sharing among police departments. *European J. Oper. Res.* **141** 660–678.

[76] ROCKAFELLAR, R. T. and URYASEV, S. (2000). Optimization of conditional value-at-risk. *J. Risk* **2** 21–42.

[77] ROCKAFELLAR, R. T. and WETS, R. J.-B. (1998). *Variational Analysis*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **317**. Springer, Berlin. MR1491362 https://doi.org/10.1007/978-3-642-02431-3

[78] ROTHENHÄUSLER, D., BÜHLMANN, P., MEINSHAUSEN, N. and PETERS, J. (2018). Anchor regression: Heterogeneous data meets causality. Available at arXiv:1801.06229.

[79] ROTHENHÄUSLER, D., MEINSHAUSEN, N. and BÜHLMANN, P. (2016). Confidence intervals for maximin effects in inhomogeneous large-scale data. In *Statistical Analysis for High-Dimensional Data*. *Abel Symp.* **11** 255–277. Springer, Cham. MR3616272

[80] SAENKO, K., KULIS, B., FRITZ, M. and DARRELL, T. (2010). Adapting visual category models to new domains. In *Proceedings of the European Conference on Computer Vision* 213–226. Springer, Berlin.

[81] SAPIEZYNSKI, P., KASSARNIG, V. and WILSON, C. (2017). Academic performance prediction in a gender-imbalanced environment. In *Proceedings of the Eleventh ACM Conference on Recommender Systems* **1** 48–51.

[82] SHAFIEEZADEH-ABADEH, S., ESFAHANI, P. M. and KUHN, D. (2015). Distributionally robust logistic regression. In *Advances in Neural Information Processing Systems* 28 1576–1584.

[83] SHAKED, M. and SHANTHIKUMAR, J. G. (2007). *Stochastic Orders*. *Springer Series in Statistics*. Springer, New York. MR2265633 https://doi.org/10.1007/978-0-387-34675-5

[84] SHAPIRO, A. (2013). On Kusuoka representation of law invariant risk measures. *Math. Oper. Res.* **38** 142–152. MR3029482 https://doi.org/10.1287/moor.1120.0563

[85] SHAPIRO, A. (2017). Distributionally robust stochastic programming. *SIAM J. Optim.* **27** 2258–2275. MR3715383 https://doi.org/10.1137/16M1058297

[86] SHAPIRO, A., DENTCHEVA, D. and RUSZCZYŃSKI, A. (2009). *Lectures on Stochastic Programming*: *Modeling and Theory*. *MPS/SIAM Series on Optimization* **9**. SIAM, Philadelphia, PA; Mathematical Programming Society (MPS), Philadelphia, PA. MR2562798 https://doi.org/10.1137/1.9780898718751

[87] SHIMODAIRA, H. (2000). Improving predictive inference under covariate shift by weighting the log-likelihood function. *J. Statist. Plann. Inference* **90** 227–244. MR1795598 https://doi.org/10.1016/S0378-3758(00)00115-4

[88] SINHA, A., NAMKOONG, H. and DUCHI, J. C. (2017). Certifiable distributional robustness with principled adversarial training. Available at arXiv:1710.10571.

[89] SUGIYAMA, M., KRAULEDAT, M. and MÜLLER, K.-R. (2007). Covariate shift adaptation by importance weighted cross validation. *J. Mach. Learn. Res.* **8** 985–1005.

[90] SUGIYAMA, M., NAKAJIMA, S., KASHIMA, H., BUENAU, P. V. and KAWANABE, M. (2008). Direct importance estimation with model selection and its application to covariate shift adaptation. In *Advances in Neural Information Processing Systems* 21 1433–1440.

[91] TALAGRAND, M. (1996). A new look at independence. *Ann. Probab.* **24** 1–34. MR1387624 https://doi.org/10.1214/aop/1042644705

[92] TATMAN, R. (2017). Gender and dialect bias in YouTube's automatic captions. In *First Workshop on Ethics in Natural Langauge Processing* **1** 53–59.

[93] TORRALBA, A. and EFROS, A. A. (2011). Unbiased look at dataset bias. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 1521–1528. IEEE, Piscataway, NJ.

[94] TSUBOI, Y., KASHIMA, H., HIDO, S., BICKEL, S. and SUGIYAMA, M. (2009). Direct density ratio estimation for large-scale covariate shift adaptation. *J. Inf. Process.* **17** 138–155.

[95] UDELL, M., MOHAN, K., ZENG, D., HONG, J., DIAMOND, S. and BOYD, S. (2014). Convex optimization in Julia. In *First Workshop on High Performance Technical Computing in Dynamic Languages* 18–28. IEEE, Piscataway, NJ.

[96] USUI, Y. and KONDO, K. (2009). The sift image feature reduction method using the histogram intersection kernel. In *International Symposium on Intelligent Signal Processing and Communication Systems* (*ISPACS*) 517–520. IEEE, Piscataway, NJ.

[97] VAN ERVEN, T. and HARREMOËS, P. (2014). Rényi divergence and Kullback–Leibler divergence. *IEEE Trans. Inf. Theory* **60** 3797–3820. MR3225930 https://doi.org/10.1109/TIT.2014.2320500

[98] 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

[99] 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

[100] WALD, A. (1945). Statistical decision functions which minimize the maximum risk. *Ann. of Math.* (2) **46** 265–280. MR0012402 https://doi.org/10.2307/1969022

[101] WOZABAL, D. (2012). A framework for optimization under ambiguity. *Ann. Oper. Res.* **193** 21–47. MR2874755 https://doi.org/10.1007/s10479-010-0812-0

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

[103] ZOU, H. and HASTIE, T. (2005). Regularization and variable selection via the elastic net. *J. Roy. Statist. Soc. Ser. B* **67** 301–320. MR2137327 https://doi.org/10.1111/j.1467-9868.2005.00503.x

The Annals of Statistics
2021, Vol. 49, No. 3, 1407–1435
https://doi.org/10.1214/20-AOS2006
© Institute of Mathematical Statistics, 2021

BOOTSTRAP LONG MEMORY PROCESSES IN THE FREQUENCY DOMAIN

BY JAVIER HIDALGO

Department of Economics, London School of Economics, f.j.hidalgo@lse.ac.uk

The aim of the paper is to describe a bootstrap, contrary to the sieve bootstrap, valid under either long memory (*LM*) or short memory (*SM*) dependence. One of the reasons of the failure of the sieve bootstrap in our context is that under *LM* dependence, the sieve bootstrap may not be able to capture the true covariance structure of the original data. We also describe and examine the validity of the bootstrap scheme for the least squares estimator of the parameter in a regression model and for model specification. The motivation for the latter example comes from the observation that the asymptotic distribution of the test is intractable.

REFERENCES

ARTECHE, J. (2000). Gaussian semiparametric estimation in seasonal/cyclical long memory time series. *Kybernetika* (*Prague*) **36** 279–310. MR1773505

BELTRÃO, K. I. and BLOOMFIELD, P. (1987). Determining the bandwidth of a kernel spectrum estimate. *J. Time Series Anal.* **8** 21–38. MR0868015 https://doi.org/10.1111/j.1467-9892.1987.tb00418.x

BERK, K. N. (1974). Consistent autoregressive spectral estimates. *Ann. Statist.* **2** 489–502. MR0421010

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217. MR0630103

BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods*. Springer, New York. MR2839251

BÜHLMANN, P. (1997). Sieve bootstrap for time series. *Bernoulli* **3** 123–148. MR1466304 https://doi.org/10.2307/3318584

BÜHLMANN, P. (2002). Bootstraps for time series. *Statist. Sci.* **17** 52–72. MR1910074 https://doi.org/10.1214/ss/1023798998

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

DEHLING, H. and TAQQU, M. S. (1989). The empirical process of some long-range dependent sequences with an application to U-statistics. *Ann. Statist.* **17** 1767–1783. MR1026312 https://doi.org/10.1214/aos/1176347394

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

FRANKE, J. and HÄRDLE, W. (1992). On bootstrapping kernel spectral estimates. *Ann. Statist.* **20** 121–145. MR1150337 https://doi.org/10.1214/aos/1176348515

GIACOMINI, R., POLITIS, D. N. and WHITE, H. (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory* **29** 567–589. MR3064050 https://doi.org/10.1017/S0266466612000655

GIRAITIS, L. and SURGAILIS, D. (1994). A central limit theorem for the empirical process of a long memory linear sequence. Mimeo. Univ. Heidelberg. Belträge zur Statistik, Nr. 24.

GRADSHTEYN, I. S. and RYZHIK, I. M. (2000). *Table of Integrals, Series, and Products*, 6th ed. *Springer Series in Statistics*. Academic Press, San Diego, CA. MR1773820 https://doi.org/10.1016/B978-0-12-294757-5.X5000-4

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

HIDALGO, J. (2003). An alternative bootstrap to moving blocks for time series regression models. *J. Econometrics* **117** 369–399. MR2008775 https://doi.org/10.1016/S0304-4076(03)00154-4

HIDALGO, J. (2005). Semiparametric estimation for stationary processes whose spectra have an unknown pole. *Ann. Statist.* **33** 1843–1889. MR2166564 https://doi.org/10.1214/009053605000000318

HIDALGO, J. (2008). Specification testing for regression models with dependent data. *J. Econometrics* **143** 143–165. MR2384437 https://doi.org/10.1016/j.jeconom.2007.08.013

MSC2020 subject classifications. Primary 62G10, 62F40; secondary 62J20, 62G30.
Key words and phrases. Long memory, bootstrap methods, frequency domain.

HIDALGO, J. (2019). Model diagnostics in time series regression models. Mimeo.

HIDALGO, J. (2021). Supplement to “Bootstrap long memory processes in the frequency domain.” https://doi.org/10.1214/20-AOS2006SUPP

HOROWITZ, J. L. (1997). Bootstrap methods in econometrics. In *Advances in Economics and Econometrics: Theory and Applications*, *Seventh World Congress* (D. M. Kreps and K. R. Wallis, eds.) 188–222. Cambridge Univ. Press, Cambridge.

HURVICH, C. M. and ZEGER, S. J. (1987). Frequency domain bootstrap methods for time series. New York Univ. Working paper.

KIRCH, C. and POLITIS, D. N. (2011). TFT-bootstrap: Resampling time series in the frequency domain to obtain replicates in the time domain. *Ann. Statist.* **39** 1427–1470. MR2850208 https://doi.org/10.1214/10-AOS868

KOUL, H. L., BAILLIE, R. T. and SURGAILIS, D. (2004). Regression model fitting with a long memory covariate process. *Econometric Theory* **20** 485–512. MR2061725 https://doi.org/10.1017/S0266466604203036

KREISS, J. P. (1988). Asymptotical inference for a class of stochastic processes, Univ. Hamburg, Habilitationsschrift.

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

KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217–1241. MR1015147 https://doi.org/10.1214/aos/1176347265

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

LOBATO, I. N. and ROBINSON, P. M. (1998). A nonparametric test for I(0). *Rev. Econ. Stud.* **65** 475–495. MR1635362 https://doi.org/10.1111/1467-937X.00054

MARINUCCI, D. and ROBINSON, P. M. (1999). Alternative forms of fractional Brownian motion. *J. Statist. Plann. Inference* **80** 111–122. MR1713794 https://doi.org/10.1016/S0378-3758(98)00245-6

MARINUCCI, D. and ROBINSON, P. M. (2000). Weak convergence of multivariate fractional processes. *Stochastic Process. Appl.* **86** 103–120. MR1741198 https://doi.org/10.1016/S0304-4149(99)00088-5

POLITIS, D. N. (2003). The impact of bootstrap methods on time series analysis. *Statist. Sci.* **18** 219–230. MR2026081 https://doi.org/10.1214/ss/1063994977

PRESS, H. and TUKEY, J. W. (1956). Power spectral methods of analysis and their application to problems in airplane dynamics. *Bell Tel. System Monograph* **2606** i+41. MR0080430

ROBINSON, P. M. (1991). Automatic frequency domain inference on semiparametric and nonparametric models. *Econometrica* **59** 1329–1363. MR1133037 https://doi.org/10.2307/2938370

ROBINSON, P. M. (1995a). Log-periodogram regression of time series with long range dependence. *Ann. Statist.* **23** 1048–1072. MR1345214 https://doi.org/10.1214/aos/1176324636

ROBINSON, P. M. (1995b). Gaussian semiparametric estimation of long range dependence. *Ann. Statist.* **23** 1630–1661. MR1370301 https://doi.org/10.1214/aos/1176324317

ROBINSON, P. M. (1997). Large sample inference for nonparametric regression with dependent errors. *Ann. Statist.* **25** 2054–2083. MR1474083 https://doi.org/doi.org/10.1214/aos/1069362387

ROBINSON, P. M. and HIDALGO, F. J. (1997). Time series regression with long-range dependence. *Ann. Statist.* **25** 77–104. MR1429918 https://doi.org/10.1214/aos/1034276622

STUTE, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25** 613–641. MR1439316 https://doi.org/10.1214/aos/1031833666

WU, W. B. (2003). Empirical processes of long-memory sequences. *Bernoulli* **9** 809–831. MR2047687 https://doi.org/10.3150/bj/1066418879

YONG, C. H. (1974). *Asymptotic Behaviour of Trigonometric Series*. Chinese Univ. Hong Kong, Hong Kong.

The Annals of Statistics
2021, Vol. 49, No. 3, 1436–1459
https://doi.org/10.1214/20-AOS2007
© Institute of Mathematical Statistics, 2021

TOTAL POSITIVITY IN EXPONENTIAL FAMILIES WITH APPLICATION TO BINARY VARIABLES

BY STEFFEN LAURITZEN[1], CAROLINE UHLER[2] AND PIOTR ZWIERNIK[3]

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

[2]*Laboratory for Information and Decision Systems, and Institute for Data, Systems, and Society Massachusetts Institute of Technology, cuhler@mit.edu*

[3]*Department of Economics and Business, Universitat Pompeu Fabra, piotr.zwiernik@upf.edu*

We study exponential families of distributions that are multivariate totally positive of order 2 (MTP_2), show that these are convex exponential families and derive conditions for existence of the MLE. Quadratic exponential familes of MTP_2 distributions contain attractive Gaussian graphical models and ferromagnetic Ising models as special examples. We show that these are defined by intersecting the space of canonical parameters with a polyhedral cone whose faces correspond to conditional independence relations. Hence MTP_2 serves as an implicit regularizer for quadratic exponential families and leads to sparsity in the estimated graphical model. We prove that the maximum likelihood estimator (MLE) in an MTP_2 binary exponential family exists if and only if both of the sign patterns $(1, -1)$ and $(-1, 1)$ are represented in the sample for every pair of variables; in particular, this implies that the MLE may exist with $n = d$ observations, in stark contrast to unrestricted binary exponential families where 2^d observations are required. Finally, we provide a novel and globally convergent algorithm for computing the MLE for MTP_2 Ising models similar to iterative proportional scaling and apply it to the analysis of data from two psychological disorders.

REFERENCES

[1] AGOSTINI, D. and AMÉNDOLA, C. (2019). Discrete Gaussian distributions via theta functions. *SIAM J. Appl. Algebra Geom.* **3** 1–30. MR3904412 https://doi.org/10.1137/18M1164937

[2] ALEGRIA, M., JACKSON, J. S. J. S., KESSLER, R. C. and TAKEUCHI, D. (2016). Collaborative Psychiatric Epidemiology Surveys (CPES), 2001–2003 [United States].

[3] ALLMAN, E. S., BAÑOS, H., EVANS, R., HOŞTEN, S., KUBJAS, K., LEMKE, D., RHODES, J. A. and ZWIERNIK, P. (2019). Maximum likelihood estimation of the latent class model through model boundary decomposition. *J. Algebr. Stat.* **10** 51–84. MR3947124 https://doi.org/10.18409/jas.v10i1.75

[4] ALLMAN, E. S., RHODES, J. A., STURMFELS, B. and ZWIERNIK, P. (2015). Tensors of nonnegative rank two. *Linear Algebra Appl.* **473** 37–53. MR3338324 https://doi.org/10.1016/j.laa.2013.10.046

[5] ANANDKUMAR, A., TAN, V. Y. F., HUANG, F. and WILLSKY, A. S. (2012). High-dimensional Gaussian graphical model selection: Walk summability and local separation criterion. *J. Mach. Learn. Res.* **13** 2293–2337. MR2973603

[6] BACH, F. (2013). Learning with submodular functions: A convex optimization perspective. *Foundations and Trends® in Machine Learning* **6** 145–373.

[7] BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, Chichester. MR0489333

[8] BARTOLUCCI, F. and BESAG, J. (2002). A recursive algorithm for Markov random fields. *Biometrika* **89** 724–730. MR1929176 https://doi.org/10.1093/biomet/89.3.724

[9] BARTOLUCCI, F. and FORCINA, A. (2000). A likelihood ratio test for MTP_2 within binary variables. *Ann. Statist.* **28** 1206–1218. MR1811325 https://doi.org/10.1214/aos/1015956713

[10] BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B* 192–236.

MSC2020 subject classifications. Primary 60E15, 62H99; secondary 15B48.

Key words and phrases. Graphical models, Ising model, log-supermodular distributions, positive dependence, exponential families.

[11] BHATTACHARYA, B. (2012). Covariance selection and multivariate dependence. *J. Multivariate Anal.* **106** 212–228. MR2887689 https://doi.org/10.1016/j.jmva.2011.11.002

[12] BORSBOOM, D. and CRAMER, A. O. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annu. Rev. Clin. Psychol.* **9** 91–121.

[13] BOYD, S. and VANDENBERGHE, L. (2004). *Convex Optimization*. Cambridge Univ. Press, Cambridge. MR2061575 https://doi.org/10.1017/CBO9780511804441

[14] COLANGELO, A., SCARSINI, M. and SHAKED, M. (2005). Some notions of multivariate positive dependence. *Insurance Math. Econom.* **37** 13–26. MR2156593 https://doi.org/10.1016/j.insmatheco.2004.09.004

[15] COWELL, R. G., DAWID, A. P., LAURITZEN, S. L. and SPIEGELHALTER, D. J. (1999). *Probabilistic Networks and Expert Systems*. *Statistics for Engineering and Information Science*. Springer, New York. MR1697175

[16] DJOLONGA, J. and KRAUSE, A. (2015). Scalable variational inference in log-supermodular models. arXiv:1502.06531.

[17] EGILMEZ, H. E., PAVEZ, E. and ORTEGA, A. (2019). Graph learning from filtered signals: Graph system and diffusion kernel identification. *IEEE Trans. Signal Inf. Process. Netw.* **5** 360–374. MR3949864 https://doi.org/10.1109/TSIPN.2018.2872157

[18] FALLAT, S., LAURITZEN, S., SADEGHI, K., UHLER, C., WERMUTH, N. and ZWIERNIK, P. (2017). Total positivity in Markov structures. *Ann. Statist.* **45** 1152–1184. MR3662451 https://doi.org/10.1214/16-AOS1478

[19] FORTUIN, C. M., KASTELEYN, P. W. and GINIBRE, J. (1971). Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.* **22** 89–103. MR0309498

[20] GEIGER, D., MEEK, C. and STURMFELS, B. (2006). On the toric algebra of graphical models. *Ann. Statist.* **34** 1463–1492. MR2278364 https://doi.org/10.1214/009053606000000263

[21] JIROUŠEK, R. and PŘEUČIL, R. (1995). On the effective implementation of the iterative proportional fitting procedure. *Comput. Statist. Data Anal.* **19** 177–189.

[22] KARLIN, S. and RINOTT, Y. (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *J. Multivariate Anal.* **10** 467–498. MR0599685 https://doi.org/10.1016/0047-259X(80)90065-2

[23] KESSLER, R. C., BERGLUND, P., CHIU, W. T., DEMLER, O., HEERINGA, S., HIRIPI, E., JIN, R., PENNELL, B.-E., WALTERS, E. E. et al. (2004). The US national comorbidity survey replication (NCS-R): Design and field procedures. *Int. J. Methods Psychiatr. Res.* **13** 69–92.

[24] LAURITZEN, S. (2002). Lectures on Contingency Tables, Electronic edition. Earlier editions: 1979, 1982, 1989.

[25] LAURITZEN, S., UHLER, C. and ZWIERNIK, P. (2019). Maximum likelihood estimation in Gaussian models under total positivity. *Ann. Statist.* **47** 1835–1863. MR3953437 https://doi.org/10.1214/17-AOS1668

[26] LAURITZEN, S. L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. The Clarendon Press, Oxford University Press, New York. MR1419991

[27] MALIOUTOV, D. V., JOHNSON, J. K. and WILLSKY, A. S. (2006). Walk-sums and belief propagation in Gaussian graphical models. *J. Mach. Learn. Res.* **7** 2031–2064. MR2274432

[28] MARCHETTI, G. M. and WERMUTH, N. (2016). Palindromic Bernoulli distributions. *Electron. J. Stat.* **10** 2435–2460. MR3545465 https://doi.org/10.1214/16-EJS1175

[29] MOUSSOURIS, J. (1974). Gibbs and Markov random systems with constraints. *J. Stat. Phys.* **10** 11–33. MR0432132 https://doi.org/10.1007/BF01011714

[30] NEWMAN, C. M. (1983). A general central limit theorem for FKG systems. *Comm. Math. Phys.* **91** 75–80. MR0719811

[31] PROPP, J. G. and WILSON, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures Algorithms* **9** 223–252.

[32] ROBEVA, E., STURMFELS, B. and UHLER, C. (2019). Geometry of log-concave density estimation. *Discrete Comput. Geom.* **61** 136–160. MR3925548 https://doi.org/10.1007/s00454-018-0024-y

[33] SLAWSKI, M. and HEIN, M. (2015). Estimation of positive definite M-matrices and structure learning for attractive Gaussian Markov random fields. *Linear Algebra Appl.* **473** 145–179. MR3338330 https://doi.org/10.1016/j.laa.2014.04.020

[34] STUDENÝ, M. (2005). *Probabilistic Conditional Independence Structures*. *Information Science and Statistics*. Springer, London. MR3183760

[35] WASSERMAN, L., RAMDAS, A. and BALAKRISHNAN, S. (2020). Universal inference using the split likelihood ratio test. *Proc. Natl. Acad. Sci. USA* **117** 16880–16890.

The Annals of Statistics
2021, Vol. 49, No. 3, 1460–1488
https://doi.org/10.1214/20-AOS2009
© Institute of Mathematical Statistics, 2021

A CAUSAL BOOTSTRAP

BY GUIDO IMBENS[1] AND KONRAD MENZEL[2]

[1]*Department of Economics and Graduate School of Business, Stanford University, imbens@stanford.edu*
[2]*Department of Economics, New York University, km125@nyu.edu*

The bootstrap, introduced by *The Jackknife, the Bootstrap and Other Resampling Plans* ((1982), SIAM), has become a very popular method for estimating variances and constructing confidence intervals. A key insight is that one can approximate the properties of estimators by using the empirical distribution function of the sample as an approximation for the true distribution function. This approach views the uncertainty in the estimator as coming exclusively from sampling uncertainty. We argue that for causal estimands the uncertainty arises entirely, or partially, from a different source, corresponding to the stochastic nature of the treatment received. We develop a bootstrap procedure for inference regarding the average treatment effect that accounts for this uncertainty, and compare its properties to that of the classical bootstrap. We consider completely randomized and observational designs as well as designs with imperfect compliance.

REFERENCES

[1] ABADIE, A., ATHEY, S., IMBENS, G. and WOOLDRIDGE, J. (2017). Sampling-based vs. Design-based Uncertainty in Regression Analysis. Working paper, MIT, Stanford, Michigan State.

[2] ANGRIST, J., IMBENS, G. and RUBIN, D. (1996). Identification of causal effects using instrumental variables. *J. Amer. Statist. Assoc.* **434** 444–455.

[3] ARONOW, P. M., GREEN, D. P. and LEE, D. K. K. (2014). Sharp bounds on the variance in randomized experiments. *Ann. Statist.* **42** 850–871. MR3210989 https://doi.org/10.1214/13-AOS1200

[4] BECKER, G. (1973). A theory of marriage, part I. *J. Polit. Econ.* **81** 813–846.

[5] BICKEL, P. J. (1968). A distribution free version of the Smirnov two sample test in the p-variate case. *Ann. Math. Stat.* **40** 1–23. MR0256519 https://doi.org/10.1214/aoms/1177697800

[6] BICKEL, P. J. and FREEDMAN, D. A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *Ann. Statist.* **12** 470–482. MR0740906 https://doi.org/10.1214/aos/1176346500

[7] BLOZNELIS, M. and GÖTZE, F. (2001). Orthogonal decomposition of finite population statistics and its applications to distributional asymptotics. *Ann. Statist.* **29** 899–917. MR1865345 https://doi.org/10.1214/aos/1009210694

[8] BOOTH, J. G., BUTLER, R. W. and HALL, P. (1994). Bootstrap methods for finite populations. *J. Amer. Statist. Assoc.* **89** 1282–1289. MR1310222

[9] CAMBANIS, S., SIMONS, G. and STOUT, W. (1976). Inequalities for $Ek(X,Y)$ when the marginals are fixed. *Z. Wahrsch. Verw. Gebiete* **36** 285–294. MR0420778 https://doi.org/10.1007/BF00532695

[10] CHUNG, E. and ROMANO, J. P. (2013). Exact and asymptotically robust permutation tests. *Ann. Statist.* **41** 484–507. MR3099111 https://doi.org/10.1214/13-AOS1090

[11] DING, P. (2017). A paradox from randomization-based causal inference. *Statist. Sci.* **32** 331–345. MR3695995 https://doi.org/10.1214/16-STS571

[12] EFRON, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans. CBMS-NSF Regional Conference Series in Applied Mathematics* **38**. SIAM, Philadelphia, PA. MR0659849

[13] FAN, Y. and PARK, S. S. (2010). Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory* **26** 931–951. MR2646485 https://doi.org/10.1017/S0266466609990168

[14] FAN, Y. and PARK, S. S. (2012). Confidence intervals for the quantile of treatment effects in randomized experiments. *J. Econometrics* **167** 330–344. MR2892079 https://doi.org/10.1016/j.jeconom.2011.09.019

MSC2020 subject classifications. 62D05, 62D20.
Key words and phrases. Causal inference, bootstrap, partial identification.

[15] FAN, Y. and WU, J. (2010). Partial identification of the distribution of treatment effects in switching regime models and its confidence sets. *Rev. Econ. Stud.* **77** 1002–1041. MR2674896 https://doi.org/10.1111/j.1467-937X.2009.00593.x

[16] FANG, Z. and SANTOS, A. (2015). Inference on directionallly differentiable functions. Working paper, Kansas State and UCLA.

[17] FIRPO, S. and RIDDER, G. (2008). Bounds on functionals of the distribution of treatment effects. Working paper, Insper and USC.

[18] FRANK, M. J., NELSEN, R. B. and SCHWEIZER, B. (1987). Best-possible bounds for the distribution of a sum—a problem of Kolmogorov. *Probab. Theory Related Fields* **74** 199–211. MR0871251 https://doi.org/10.1007/BF00569989

[19] FREEDMAN, D. A. (2008). On regression adjustments in experiments with several treatments. *Ann. Appl. Stat.* **2** 176–196. MR2415599 https://doi.org/10.1214/07-AOAS143

[20] HAHN, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica* **66** 315–331. MR1612242 https://doi.org/10.2307/2998560

[21] HECKMAN, J. J. and SMITH, J. (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *Rev. Econ. Stud.* **64** 487–535.

[22] HIRANO, K., IMBENS, G. W. and RIDDER, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71** 1161–1189. MR1995826 https://doi.org/10.1111/1468-0262.00442

[23] HONG, H. and LI, J. (2016). The numerical Delta method and bootstrap. Working paper, Stanford.

[24] HORVITZ, D. G. and THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47** 663–685. MR0053460

[25] IMBENS, G. and MENZEL, K. (2020). Matlab code for “A causal bootstrap”.

[26] IMBENS, G. and MENZEL, K. (2021). Supplement to “A causal bootstrap.” https://doi.org/10.1214/20-AOS2009SUPP.

[27] IMBENS, G. W. and MANSKI, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica* **72** 1845–1857. MR2095534 https://doi.org/10.1111/j.1468-0262.2004.00555.x

[28] IMBENS, G. W. and RUBIN, D. B. (2015). *Causal Inference—for Statistics, Social, and Biomedical Sciences: An introduction*. Cambridge Univ. Press, New York. MR3309951 https://doi.org/10.1017/CBO9781139025751

[29] LI, X. and DING, P. (2017). General forms of finite population central limit theorems with applications to causal inference. *J. Amer. Statist. Assoc.* **112** 1759–1769. MR3750897 https://doi.org/10.1080/01621459.2017.1295865

[30] LIN, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman’s critique. *Ann. Appl. Stat.* **7** 295–318. MR3086420 https://doi.org/10.1214/12-AOAS583

[31] LOH, W. W., RICHARDSON, T. S. and ROBINS, J. M. (2017). An apparent paradox explained [discussion on MR3695995]. *Statist. Sci.* **32** 356–361. MR3695999 https://doi.org/10.1214/17-STS610

[32] LU, J., DING, P. and DASGUPTA, T. (2018). Treatment effects on ordinal outcomes: Causal estimands and sharp bounds. *J. Educ. Behav. Stat.* **20** 1–28.

[33] MAKAROV, G. D. (1981). Estimates for the distribution function of the sum of two random variables with given marginal distributions. *Theory Probab. Appl.* **26** 803–806. MR0636775

[34] MANSKI, C. F. (1997). The mixing problem in programme evaluation. *Rev. Econ. Stud.* **64** 537–553. MR1485827 https://doi.org/10.2307/2971730

[35] NELSEN, R. B. (2006). *An Introduction to Copulas*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2197664 https://doi.org/10.1007/s11229-005-3715-x

[36] ROBINS, J. (1988). Confidence intervals for causal parameters. *Stat. Med.* **7** 773–785.

[37] ROSENBAUM, P. R. (2002). *Observational Studies*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR1899138 https://doi.org/10.1007/978-1-4757-3692-2

[38] RUBIN, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Ann. Statist.* **6** 34–58. MR0472152

[39] SPLAWA-NEYMAN, J. (1923,1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statist. Sci.* **5** 465–472. Translated from the Polish and edited by D. M. D’abrowska and T. P. Speed. MR1092986

[40] STOYE, J. (2009). More on confidence intervals for partially identified parameters. *Econometrica* **77** 1299–1315. MR2547075 https://doi.org/10.3982/ECTA7347

[41] STOYE, J. (2010). Partial identification of spread parameters. *Quant. Econ.* **1** 323–357. MR2945102 https://doi.org/10.3982/QE24

[42] WILLIAMSON, R. C. and DOWNS, T. (1990). Probabilistic arithmetic. I. Numerical methods for calculating convolutions and dependency bounds. *Internat. J. Approx. Reason.* **4** 89–158. MR1042207 https://doi.org/10.1016/0888-613X(90)90022-T

The Annals of Statistics
2021, Vol. 49, No. 3, 1489–1513
https://doi.org/10.1214/20-AOS2010
© Institute of Mathematical Statistics, 2021

PRINCIPAL COMPONENTS IN LINEAR MIXED MODELS WITH GENERAL BULK

BY ZHOU FAN[1], YI SUN[2] AND ZHICHAO WANG[3]

[1]*Department of Statistics and Data Science, Yale University, zhou.fan@yale.edu*

[2]*Department of Statistics, University of Chicago, yisun@statistics.uchicago.edu*

[3]*Department of Mathematics, University of California, San Diego, zhw036@ucsd.edu*

We study the principal components of covariance estimators in multivariate mixed-effects linear models. We show that, in high dimensions, the principal eigenvalues and eigenvectors may exhibit bias and aliasing effects that are not present in low-dimensional settings. We derive the first-order limits of the principal eigenvalue locations and eigenvector projections in a high-dimensional asymptotic framework, allowing for general population spectral distributions for the random effects and extending previous results from a more restrictive spiked model. Our analysis uses free probability techniques, and we develop two general tools of independent interest—strong asymptotic freeness of GOE and deterministic matrices and a free deterministic equivalent approximation for bilinear forms of resolvents.

REFERENCES

[1] BAI, Z. and YAO, J. (2012). On sample eigenvalues in a generalized spiked population model. *J. Multivariate Anal.* **106** 167–177. MR2887686 https://doi.org/10.1016/j.jmva.2011.10.009

[2] BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. MR2165575 https://doi.org/10.1214/009117905000000233

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

[4] BELINSCHI, S., BERCOVICI, H. and CAPITAINE, M. (2017). On the outlying eigenvalues of a polynomial in large independent random matrices. Preprint. Available at arXiv:1703.08102.

[5] BELINSCHI, S. T., BERCOVICI, H., CAPITAINE, M. and FÉVRIER, M. (2017). Outliers in the spectrum of large deformed unitarily invariant models. *Ann. Probab.* **45** 3571–3625. MR3729610 https://doi.org/10.1214/16-AOP1144

[6] BELINSCHI, S. T. and CAPITAINE, M. (2017). Spectral properties of polynomials in independent Wigner and deterministic matrices. *J. Funct. Anal.* **273** 3901–3963. MR3711884 https://doi.org/10.1016/j.jfa.2017.07.010

[7] BENAYCH-GEORGES, F. (2009). Rectangular random matrices, related convolution. *Probab. Theory Related Fields* **144** 471–515. MR2496440 https://doi.org/10.1007/s00440-008-0152-z

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

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

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

[11] BHATTACHARJEE, M. and BOSE, A. (2016). Polynomial generalizations of the sample variance-covariance matrix when $pn^{-1} \to 0$. *Random Matrices Theory Appl.* **5** 1650014, 41. MR3569842 https://doi.org/10.1142/S2010326316500143

MSC2020 subject classifications. 62E20.

Key words and phrases. Random effects models, high-dimensional asymptotics, random matrix theory, free probability.

[12] BHATTACHARJEE, M. and BOSE, A. (2017). Matrix polynomial generalizations of the sample variance-covariance matrix when $pn^{-1} \to y \in (0, \infty)$. *Indian J. Pure Appl. Math.* **48** 575–607. MR3741695 https://doi.org/10.1007/s13226-017-0247-2

[13] BHATTACHARJEE, M. and BOSE, A. (2019). Joint convergence of sample autocovariance matrices when $p/n \to 0$ with application. *Ann. Statist.* **47** 3470–3503. MR4025749 https://doi.org/10.1214/18-AOS1785

[14] BLOWS, M. W. and MCGUIGAN, K. (2015). The distribution of genetic variance across phenotypic space and the response to selection. *Mol. Ecol.* **24** 2056–2072.

[15] BOSE, A. and BHATTACHARJEE, M. (2019). *Large Covariance and Autocovariance Matrices. Monographs on Statistics and Applied Probability* **162**. CRC Press, Boca Raton, FL. MR3821539

[16] BURDA, Z., JURKIEWICZ, J. and WACŁAW, B. (2005). Spectral moments of correlated Wishart matrices. *Phys. Rev. E* (3) **71** 026111, 11. MR2139960 https://doi.org/10.1103/PhysRevE.71.026111

[17] CAPITAINE, M. and CASALIS, M. (2004). Asymptotic freeness by generalized moments for Gaussian and Wishart matrices. Application to beta random matrices. *Indiana Univ. Math. J.* **53** 397–431. MR2060040 https://doi.org/10.1512/iumj.2004.53.2325

[18] CAPITAINE, M. and DONATI-MARTIN, C. (2007). Strong asymptotic freeness for Wigner and Wishart matrices. *Indiana Univ. Math. J.* **56** 767–803. MR2317545 https://doi.org/10.1512/iumj.2007.56.2886

[19] COLLET, J. M., MCGUIGAN, K., ALLEN, S. L., CHENOWETH, S. F. and BLOWS, M. W. (2018). Mutational pleiotropy and the strength of stabilizing selection within and between functional modules of gene expression. *Genetics* **208** 1601–1616. https://doi.org/10.1534/genetics.118.300776

[20] COLLINS, B. (2003). Moments and cumulants of polynomial random variables on unitary groups, the Itzykson–Zuber integral, and free probability. *Int. Math. Res. Not.* **17** 953–982. MR1959915 https://doi.org/10.1155/S107379280320917X

[21] COLLINS, B. and ŚNIADY, P. (2006). Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. *Comm. Math. Phys.* **264** 773–795. MR2217291 https://doi.org/10.1007/s00220-006-1554-3

[22] DOBRIBAN, E. and LIU, S. (2018). A new theory for sketching in linear regression. Preprint. Available at arXiv:1810.06089.

[23] DYKEMA, K. (1993). On certain free product factors via an extended matrix model. *J. Funct. Anal.* **112** 31–60. MR1207936 https://doi.org/10.1006/jfan.1993.1025

[24] FAN, Z. and JOHNSTONE, I. M. (2019). Eigenvalue distributions of variance components estimators in high-dimensional random effects models. *Ann. Statist.* **47** 2855–2886. MR3988775 https://doi.org/10.1214/18-AOS1767

[25] FAN, Z., JOHNSTONE, I. M. and SUN, Y. (2018). Spiked covariances and principal components analysis in high-dimensional random effects models. Preprint. Available at arXiv:1806.09529.

[26] FAN, Z., SUN, Y. and WANG, Z. (2021). Supplement to "Principal components in linear mixed models with general bulk." https://doi.org/10.1214/20-AOS2010SUPP

[27] FISHER, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* **52** 399–433.

[28] HAAGERUP, U. and THORBJØRNSEN, S. (2005). A new application of random matrices: $\mathrm{Ext}(C^*_{\mathrm{red}}(F_2))$ is not a group. *Ann. of Math.* (2) **162** 711–775. MR2183281 https://doi.org/10.4007/annals.2005.162.711

[29] HACHEM, W., LOUBATON, P. and NAJIM, J. (2007). Deterministic equivalents for certain functionals of large random matrices. *Ann. Appl. Probab.* **17** 875–930. MR2326235 https://doi.org/10.1214/105051606000000925

[30] HIAI, F. and PETZ, D. (2000). Asymptotic freeness almost everywhere for random matrices. *Acta Sci. Math.* (*Szeged*) **66** 809–834. MR1804226

[31] HINE, E. and BLOWS, M. W. (2006). Determining the effective dimensionality of the genetic variance–covariance matrix. *Genetics* **173** 1135–1144.

[32] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961 https://doi.org/10.1214/aos/1009210544

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

[34] JOHNSTONE, I. M. and PAUL, D. (2018). PCA in high dimensions: An orientation. *Proc. IEEE* **106** 1277–1292. https://doi.org/10.1109/JPROC.2018.2846730

[35] JOLLIFFE, I. T. (1986). *Principal Component Analysis. Springer Series in Statistics*. Springer, New York. MR0841268 https://doi.org/10.1007/978-1-4757-1904-8

[36] LEEK, J. T. and STOREY, J. D. (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3** e161.

[37] LI, G. and ZHU, H. (2013). Genetic studies: The linear mixed models in genome-wide association studies. *Open Bioinform. J.* **7** 27–33. MR3172942 https://doi.org/10.2174/1875036201307010027

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

[39] LYNCH, M. and WALSH, B. (1998). *Genetics and Analysis of Quantitative Traits* **1**. Sinauer, Sunderland, MA.

[40] MALE, C. (2012). The norm of polynomials in large random and deterministic matrices. *Probab. Theory Related Fields* **154** 477–532. MR3000553 https://doi.org/10.1007/s00440-011-0375-2

[41] MARCENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Sb. Math.* **1** 457–483.

[42] MINGO, J. A. and SPEICHER, R. (2017). *Free Probability and Random Matrices. Fields Institute Monographs* **35**. Springer, New York. MR3585560 https://doi.org/10.1007/978-1-4939-6942-5

[43] NADLER, B. (2008). Finite sample approximation results for principal component analysis: A matrix perturbation approach. *Ann. Statist.* **36** 2791–2817. MR2485013 https://doi.org/10.1214/08-AOS618

[44] NICA, A., SHLYAKHTENKO, D. and SPEICHER, R. (2002). Operator-valued distributions. I. Characterizations of freeness. *Int. Math. Res. Not.* **29** 1509–1538. MR1907203 https://doi.org/10.1155/S1073792802201038

[45] NICA, A. and SPEICHER, R. (2006). *Lectures on the Combinatorics of Free Probability. London Mathematical Society Lecture Note Series* **335**. Cambridge Univ. Press, Cambridge. MR2266879 https://doi.org/10.1017/CBO9780511735127

[46] PATTERSON, N., PRICE, A. L. and REICH, D. (2006). Population structure and eigenanalysis. *PLoS Genet.* **2** e190. https://doi.org/10.1371/journal.pgen.0020190

[47] PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17** 1617–1642. MR2399865

[48] PAUL, D. and AUE, A. (2014). Random matrix theory in statistics: A review. *J. Statist. Plann. Inference* **150** 1–29. MR3206718 https://doi.org/10.1016/j.jspi.2013.09.005

[49] RAO, C. R. (1972). Estimation of variance and covariance components in linear models. *J. Amer. Statist. Assoc.* **67** 112–115. MR0314185

[50] SCHULTZ, H. (2005). Non-commutative polynomials of independent Gaussian random matrices. The real and symplectic cases. *Probab. Theory Related Fields* **131** 261–309. MR2117954 https://doi.org/10.1007/s00440-004-0366-7

[51] SEARLE, S. R., CASELLA, G. and MCCULLOCH, C. E. (2006). *Variance Components. Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2298115

[52] SPEICHER, R. and VARGAS, C. (2012). Free deterministic equivalents, rectangular random matrix models, and operator-valued free probability theory. *Random Matrices Theory Appl.* **1** 1150008, 26. MR2934714 https://doi.org/10.1142/S2010326311500080

[53] STEGLE, O., PARTS, L., PIIPARI, M., WINN, J. and DURBIN, R. (2012). Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat. Protoc.* **7** 500–507. https://doi.org/10.1038/nprot.2011.457

[54] TRACY, C. A. and WIDOM, H. (1996). On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.* **177** 727–754. MR1385083

[55] VOICULESCU, D. (1991). Limit laws for random matrices and free products. *Invent. Math.* **104** 201–220. MR1094052 https://doi.org/10.1007/BF01245072

[56] VOICULESCU, D. (1998). A strengthened asymptotic freeness result for random matrices with applications to free entropy. *Int. Math. Res. Not.* **1** 41–63. MR1601878 https://doi.org/10.1155/S107379289800004X

[57] WANG, L., AUE, A. and PAUL, D. (2017). Spectral analysis of sample autocovariance matrices of a class of linear time series in moderately high dimensions. *Bernoulli* **23** 2181–2209. MR3648029 https://doi.org/10.3150/16-BEJ807

[58] WRIGHT, S. (1935). The analysis of variance and the correlations between relatives with respect to deviations from an optimum. *J. Genet.* **30** 243–256.

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

[60] ZHANG, L. (2006). Spectral analysis of large dimensional random matrices Ph.D. thesis, National Univ. Singapore.

The Annals of Statistics
2021, Vol. 49, No. 3, 1514–1536
https://doi.org/10.1214/20-AOS2011
© Institute of Mathematical Statistics, 2021

STATISTICAL INFERENCE IN SPARSE HIGH-DIMENSIONAL ADDITIVE MODELS

BY KARL GREGORY[1], ENNO MAMMEN[2] AND MARTIN WAHL[3]

[1]*Department of Statistics, University of South Carolina,* gregorkb@stat.sc.edu
[2]*Institute of Applied Mathematics, Heidelberg University,* mammen@math.uni-heidelberg.de
[3]*Institute of Mathematics, Humboldt-Universität zu Berlin,* martin.wahl@math.hu-berlin.de

In this paper, we discuss the estimation of a nonparametric component f_1 of a nonparametric additive model $Y = f_1(X_1) + \cdots + f_q(X_q) + \epsilon$. We allow the number q of additive components to grow to infinity and we make sparsity assumptions about the number of nonzero additive components. We compare this estimation problem with that of estimating f_1 in the oracle model $Z = f_1(X_1) + \epsilon$, for which the additive components $f_2, \dots, f_q$ are known. We construct a two-step presmoothing-and-resmoothing estimator of f_1 and state finite-sample bounds for the difference between our estimator and a corresponding smoothing estimator $\hat{f}_1^{(\text{oracle})}$ in the oracle model. In an asymptotic setting, these bounds can be used to show asymptotic equivalence of our estimator and the oracle estimator; the paper thus shows that, asymptotically, under strong enough sparsity conditions, knowledge of $f_2, \dots, f_q$ has no effect on estimation accuracy. Our first step is to estimate f_1 with an undersmoothed estimator based on near-orthogonal projections with a group Lasso bias correction. In the second step, we construct pseudo responses $\hat{Y}$ by evaluating this undersmoothed estimator of f_1 at the design points and then apply the smoothing method of the oracle estimator $\hat{f}_1^{(\text{oracle})}$ to the nonparametric regression problem with "responses" $\hat{Y}$ and covariates X_1. Our mathematical exposition centers primarily on establishing properties of the presmoothing estimator. We present simulation results demonstrating close-to-oracle performance of our estimator in practical applications.

REFERENCES

[1] AVALOS, M., GRANDVALET, Y. and AMBROISE, C. (2007). Parsimonious additive models. *Comput*. *Statist*. *Data Anal*. **51** 2851–2870. MR2345610 https://doi.org/10.1016/j.csda.2006.10.007
[2] BACH, P., KLAASSEN, S., KUECK, J. and SPINDLER, M. (2020). Uniform inference in high-dimensional generalized additive models Technical Report. Available at arXiv:2004.01623v1.
[3] BARAUD, Y. (2002). Model selection for regression on a random design. *ESAIM Probab*. *Stat*. **6** 127–146. MR1918295 https://doi.org/10.1051/ps:2002007
[4] BAUER, B. and KOHLER, M. (2019). On deep learning as a remedy for the curse of dimensionality in non-parametric regression. *Ann*. *Statist*. **47** 2261–2285. MR3953451 https://doi.org/10.1214/18-AOS1747
[5] BELLONI, A. and CHERNOZHUKOV, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli* **19** 521–547. MR3037163 https://doi.org/10.3150/11-BEJ410
[6] BELLONI, A., CHERNOZHUKOV, V., CHETVERIKOV, D. and WEI, Y. (2018). Uniformly valid post-regularization confidence regions for many functional parameters in z-estimation framework. *Ann*. *Statist*. **46** 3643–3675. MR3852664 https://doi.org/10.1214/17-AOS1671
[7] BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Rev*. *Econ*. *Stud*. **81** 608–650. MR3207983 https://doi.org/10.1093/restud/rdt044
[8] BERTIN, K. and LECUÉ, G. (2008). Selection of variables and dimension reduction in high-dimensional non-parametric regression. *Electron*. *J*. *Stat*. **2** 1224–1241. MR2461900 https://doi.org/10.1214/08-EJS327

MSC2020 subject classifications. Primary 62G08; secondary 62G20.
Key words and phrases. Nonparametric curve estimation, additive models, bias correction, near-orthogonality, Lasso.

[9] BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. MR2533469 https://doi.org/10.1214/08-AOS620

[10] BIRGÉ, L. and MASSART, P. (1997). From model selection to adaptive estimation. In *Festschrift for Lucien Le Cam* 55–87. Springer, New York. MR1462939

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

[12] DEMKO, S. (1977). Inverses of band matrices and local convergence of spline projections. *SIAM J. Numer. Anal.* **14** 616–619. MR0455281 https://doi.org/10.1137/0714041

[13] EFROMOVICH, S. (2013). Nonparametric regression with the scale depending on auxiliary variable. *Ann. Statist.* **41** 1542–1568. MR3113821 https://doi.org/10.1214/13-AOS1126

[14] FAN, J., FENG, Y. and SONG, R. (2011). Nonparametric independence screening in sparse ultra-high-dimensional additive models. *J. Amer. Statist. Assoc.* **106** 544–557. MR2847969 https://doi.org/10.1198/jasa.2011.tm09779

[15] GAÏFFAS, S. (2007). On pointwise adaptive curve estimation based on inhomogeneous data. *ESAIM Probab. Stat.* **11** 344–364. MR2339297 https://doi.org/10.1051/ps:2007023

[16] GREGORY, K., MAMMEN, E. and WAHL, M. (2021). Supplement to "Statistical inference in sparse high-dimensional additive models." https://doi.org/10.1214/20-AOS2011SUPP

[17] GUO, Z. and ZHANG, C.-H. (2019). Local inference in additive models with decorrelated local linear estimator Technical Report. Available at arXiv:1907.12732v1.

[18] HENGARTNER, N. W. and SPERLICH, S. (2005). Rate optimal estimation with the integration method in the presence of many covariates. *J. Multivariate Anal.* **95** 246–272. MR2170397 https://doi.org/10.1016/j.jmva.2004.09.010

[19] HOROWITZ, J., KLEMELÄ, J. and MAMMEN, E. (2006). Optimal estimation in additive regression models. *Bernoulli* **12** 271–298. MR2218556 https://doi.org/10.3150/bj/1145993975

[20] HOROWITZ, J. L. and MAMMEN, E. (2007). Rate-optimal estimation for a general class of nonparametric regression models with unknown link functions. *Ann. Statist.* **35** 2589–2619. MR2382659 https://doi.org/10.1214/009053607000000415

[21] HUANG, J., HOROWITZ, J. L. and WEI, F. (2010). Variable selection in nonparametric additive models. *Ann. Statist.* **38** 2282–2313. MR2676890 https://doi.org/10.1214/09-AOS781

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

[23] KATO, K. (2012). Two-step estimation of high dimensional additive models Technical Report 1207.5313, ArXiv.

[24] KIM, W., LINTON, O. B. and HENGARTNER, N. W. (1999). A computationally efficient oracle estimator for additive nonparametric regression with bootstrap confidence intervals. *J. Comput. Graph. Statist.* **8** 278–297. MR1706361 https://doi.org/10.2307/1390637

[25] KOLTCHINSKII, V. and YUAN, M. (2008). Sparse recovery in large ensembles of kernel machines. In *Colt* (R. A. Servedio and T. Zhang, eds.) 229–238. Omnipress, Madison, WI.

[26] LEE, E. R. and PARK, B. U. (2012). Sparse estimation in functional linear regression. *J. Multivariate Anal.* **105** 1–17. MR2877499 https://doi.org/10.1016/j.jmva.2011.08.005

[27] LEPSKI, O. V., MAMMEN, E. and SPOKOINY, V. G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: An approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.* **25** 929–947. MR1447734 https://doi.org/10.1214/aos/1069362731

[28] LEPSKI, O. V. and SPOKOINY, V. G. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.* **25** 2512–2546. MR1604408 https://doi.org/10.1214/aos/1030741083

[29] LIN, Y. and ZHANG, H. H. (2006). Component selection and smoothing in multivariate nonparametric regression. *Ann. Statist.* **34** 2272–2297. MR2291500 https://doi.org/10.1214/009053606000000722

[30] LU, J., KOLAR, M. and LIU, H. (2020). Kernel meets sieve: Post-regularization confidence bands for high dimensional nonparametric models with local sparsity. *J. Amer. Statist. Assoc.*, *to appear* DOI: 10.1080/01621459.2019.1689984.

[31] MASSART, P. (2007). *Concentration Inequalities and Model Selection. Lecture Notes in Math.* **1896**. Springer, Berlin. MR2319879

[32] MEIER, L., VAN DE GEER, S. and BÜHLMANN, P. (2009). High-dimensional additive modeling. *Ann. Statist.* **37** 3779–3821. MR2572443 https://doi.org/10.1214/09-AOS692

[33] MITRA, R. and ZHANG, C.-H. (2016). The benefit of group sparsity in group inference with de-biased scaled group Lasso. *Electron. J. Stat.* **10** 1829–1873. MR3522662 https://doi.org/10.1214/16-EJS1120

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

[35] NOH, H. S. and PARK, B. U. (2010). Sparse varying coefficient models for longitudinal data. *Statist*. *Sinica* **20** 1183–1202. MR2730179

[36] RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2012). Minimax-optimal rates for sparse additive models over kernel classes via convex programming. *J*. *Mach*. *Learn*. *Res*. **13** 389–427. MR2913704

[37] RAVIKUMAR, P., LAFFERTY, J., LIU, H. and WASSERMAN, L. (2009). Sparse additive models. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **71** 1009–1030. MR2750255 https://doi.org/10.1111/j.1467-9868.2009.00718.x

[38] STONE, C. J. (1985). Additive regression and other nonparametric models. *Ann*. *Statist*. **13** 689–705. MR0790566 https://doi.org/10.1214/aos/1176349548

[39] STUCKY, B. and VAN DE GEER, S. (2018). Asymptotic confidence regions for high-dimensional structured sparsity. *IEEE Trans*. *Signal Process*. **66** 2178–2190. MR3797823 https://doi.org/10.1109/TSP.2018.2807399

[40] SUZUKI, T. and SUGIYAMA, M. (2013). Fast learning rate of multiple kernel learning: Trade-off between sparsity and smoothness. *Ann*. *Statist*. **41** 1381–1405. MR3113815 https://doi.org/10.1214/13-AOS1095

[41] 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

[42] VAN DE GEER, S. and STUCKY, B. (2016). χ^2-confidence sets in high-dimensional regression. In *Statistical Analysis for High-Dimensional Data*. *Abel Symp*. **11** 279–306. Springer, Cham. MR3616273

[43] WAHL, M. (2014). A theory of nonparametric regression in the presence of complex nuisance components. Preprint.

[44] WHITTAKER, E. T. and WATSON, G. N. (1996). *A Course of Modern Analysis*. *Cambridge Mathematical Library*. Cambridge Univ. Press, Cambridge. Reprint of the fourth (1927) edition. MR1424469 https://doi.org/10.1017/CBO9780511608759

[45] YUAN, M. (2007). Nonnegative garrote component selection in functional ANOVA models. In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics* (*AISTATS*-07) 656–662.

[46] 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

[47] ZHOU, S., SHEN, X. and WOLFE, D. A. (1998). Local asymptotics for regression splines and confidence regions. *Ann*. *Statist*. **26** 1760–1782. MR1673277 https://doi.org/10.1214/aos/1024691356

The Annals of Statistics
2021, Vol. 49, No. 3, 1537–1568
https://doi.org/10.1214/20-AOS2012
© Institute of Mathematical Statistics, 2021

A CONVEX OPTIMIZATION APPROACH TO HIGH-DIMENSIONAL SPARSE QUADRATIC DISCRIMINANT ANALYSIS

BY T. TONY CAI[1] AND LINJUN ZHANG[2]

[1]*Department of Statistics, The Wharton School, University of Pennsylvania, tcai@wharton.upenn.edu*
[2]*Department of Statistics, Rutgers University, linjun.zhang@rutgers.edu*

In this paper, we study high-dimensional sparse Quadratic Discriminant Analysis (QDA) and aim to establish the optimal convergence rates for the classification error. Minimax lower bounds are established to demonstrate the necessity of structural assumptions such as sparsity conditions on the discriminating direction and differential graph for the possible construction of consistent high-dimensional QDA rules.

We then propose a classification algorithm called SDAR using constrained convex optimization under the sparsity assumptions. Both minimax upper and lower bounds are obtained and this classification rule is shown to be simultaneously rate optimal over a collection of parameter spaces, up to a logarithmic factor. Simulation studies demonstrate that SDAR performs well numerically. The algorithm is also illustrated through an analysis of prostate cancer data and colon tissue data. The methodology and theory developed for high-dimensional QDA for two groups in the Gaussian setting are also extended to multigroup classification and to classification under the Gaussian copula model.

REFERENCES

[1] ALON, Y., BARKAI, N., NOTTERMAN, DA., GISH, K., YBARRA, S., MACK, D. and LEVINE, AJ. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *PNAS* **96(12)** 6745–6750. https://doi.org/10.1073/pnas.96.12.6745

[2] ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR1990662

[3] AZIZYAN, M., SINGH, A. and WASSERMAN, L. (2013). Minimax theory for high-dimensional Gaussian mixtures with sparse mean separation. In *Advances in Neural Information Processing Systems* 2139–2147.

[4] BANDYOPADHYAY, S., MEHTA, M., KUO, D., SUNG, M.-K., CHUANG, R., JAEHNIG, E. J., BODENMILLER, B., LICON, K., COPELAND, W. et al. (2010). Rewiring of genetic networks in response to DNA damage. *Science* **330** 1385–1389.

[5] BARBER, R. F. and KOLAR, M. (2018). ROCKET: robust confidence intervals via Kendall's tau for transelliptical graphical models. *Ann. Statist.* **46** 3422–3450. MR3852657 https://doi.org/10.1214/17-AOS1663

[6] BICKEL, P. J. and LEVINA, E. (2004). Some theory of Fisher's linear discriminant function, 'naive Bayes', and some alternatives when there are many more variables than observations. *Bernoulli* **10** 989–1010. MR2108040 https://doi.org/10.3150/bj/1106314847

[7] CAI, T. and LIU, W. (2011). A direct estimation approach to sparse linear discriminant analysis. *J. Amer. Statist. Assoc.* **106** 1566–1577. MR2896857 https://doi.org/10.1198/jasa.2011.tm11199

[8] CAI, T. T., LIANG, T. and ZHOU, H. H. (2015). Law of log determinant of sample covariance matrix and optimal estimation of differential entropy for high-dimensional Gaussian distributions. *J. Multivariate Anal.* **137** 161–172. MR3332804 https://doi.org/10.1016/j.jmva.2015.02.003

[9] CAI, T. T., MA, J. and ZHANG, L. (2019). CHIME: Clustering of high-dimensional Gaussian mixtures with EM algorithm and its optimality. *Ann. Statist.* **47** 1234–1267. MR3911111 https://doi.org/10.1214/18-AOS1711

MSC2020 subject classifications. Primary 62H30; secondary 62C20, 62H12.
Key words and phrases. Classification, constrained ℓ_1 minimization, high-dimensional data, minimax lower bound, optimal rate of convergence, quadratic discriminant analysis.

[10] CAI, T. T. and ZHANG, L. (2018). High-dimensional Gaussian copula regression: Adaptive estimation and statistical inference. *Statist. Sinica* **28** 963–993. MR3791096

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

[12] CAI, T. T and ZHANG, L. (2021). Supplement to "A convex optimization approach to high-dimensional sparse quadratic discriminant analysis." https://doi.org/10.1214/20-AOS2012SUPP

[13] CANDES, E. and ROMBERG, J. (2005). l1-magic: Recovery of sparse signals via convex programming.

[14] CARMON, Y., RAGHUNATHAN, A., SCHMIDT, L., DUCHI, J. C. and LIANG, P. S. (2019). Unlabeled data improves adversarial robustness. In *Advances in Neural Information Processing Systems* 11192–11203.

[15] CHURCHILL, G. A. and IACOBUCCI, D. (2006). *Marketing Research: Methodological Foundations*. Dryden Press, New York.

[16] DAN, C., WEI, Y. and RAVIKUMAR, P. (2020). Sharp statistical guarantees for adversarially robust Gaussian classification. Preprint. Available at arXiv:2006.16384.

[17] DENG, Z., KAMMOUN, A. and THRAMPOULIDIS, C. (2019). A model of double descent for high-dimensional binary linear classification. Preprint. Available at arXiv:1911.05822.

[18] DETTLING, M. (2004). BagBoosting for tumor classification with gene expression data. *Bioinformatics* **20(18)** 3853–3593. https://doi.org/10.1093/bioinformatics/bth447

[19] EFRON, B. (2010). *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*, Cambridge University Press, New York.

[20] HAN, F. and LIU, H. (2017). Statistical analysis of latent generalized correlation matrix estimation in transelliptical distribution. *Bernoulli* **23** 23–57. MR3556765 https://doi.org/10.3150/15-BEJ702

[21] HAN, F., ZHAO, T. and LIU, H. (2013). CODA: High dimensional copula discriminant analysis. *J. Mach. Learn. Res.* **14** 629–671. MR3033343

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

[23] HILL, S. M., HEISER, L. M., COKELAER, T., UNGER, M., NESSER, N. K., CARLIN, D. E., ZHANG, Y., SOKOLOV, A., PAULL, E. O. et al. (2016). Inferring causal molecular networks: Empirical assessment through a community-based effort. *Nat. Methods* **13** 310–318.

[24] INAM, F., INAM, A., MIAN, M. A., SHEIKH, A. A. and AWAN, H. M. (2018). Forecasting bankruptcy for organizational sustainability in Pakistan: Using artificial neural networks, logit regression, and discriminant analysis. *J. Econ. Adm. Sci.*

[25] JIANG, B., WANG, X. and LENG, C. (2018). A direct approach for sparse quadratic discriminant analysis. *J. Mach. Learn. Res.* **19** Paper No. 31, 37. MR3862438

[26] JOMBART, T., DEVILLARD, S. and BALLOUX, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genet.* **11** 94.

[27] JU, F., SUN, Y., GAO, J., HU, Y. and YIN, B. (2019). Probabilistic linear discriminant analysis with vectorial representation for tensor data. *IEEE Trans. Neural Netw. Learn. Syst.* **30** 2938–2950. MR4015091

[28] KINI, G. and THRAMPOULIDIS, C. (2020). Analytic study of double descent in binary classification: The impact of loss. Preprint. Available at arXiv:2001.11572.

[29] KOÇHAN, N., TÜTÜNCÜ, G. Y., SMYTH, G. K., GANDOLFO, L. C. and GINER, G. (2019). qtQDA: Quantile transformed quadratic discriminant analysis for high-dimensional RNA-seq data. *BioRxiv* 751370.

[30] LI, H.-D., XU, Q.-S. and LIANG, Y.-Z. (2018). libPLS: An integrated library for partial least squares regression and linear discriminant analysis. *Chemom. Intell. Lab. Syst.* **176** 34–43.

[31] LI, Q. and SHAO, J. (2015). Sparse quadratic discriminant analysis for high dimensional data. *Statist. Sinica* **25** 457–473. MR3379082

[32] LIN, Y. and JEON, Y. (2003). Discriminant analysis through a semiparametric model. *Biometrika* **90** 379–392. MR1986654 https://doi.org/10.1093/biomet/90.2.379

[33] LIU, H., HAN, F., YUAN, M., LAFFERTY, J. and WASSERMAN, L. (2012). High-dimensional semiparametric Gaussian copula graphical models. *Ann. Statist.* **40** 2293–2326. MR3059084 https://doi.org/10.1214/12-AOS1037

[34] MAI, Q. and ZOU, H. (2015). Sparse semiparametric discriminant analysis. *J. Multivariate Anal.* **135** 175–188. MR3306434 https://doi.org/10.1016/j.jmva.2014.12.009

[35] MAI, Q., ZOU, H. and YUAN, M. (2012). A direct approach to sparse discriminant analysis in ultra-high dimensions. *Biometrika* **99** 29–42. MR2899661 https://doi.org/10.1093/biomet/asr066

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

[37] MIGNACCO, F., KRZAKALA, F., LU, Y. M. and ZDEBOROVÁ, L. (2020). The role of regularization in classification of high-dimensional noisy Gaussian mixture. *ICML* 2020.

[38] NEYKOV, M., NING, Y., LIU, J. S. and LIU, H. (2018). A unified theory of confidence regions and testing for high-dimensional estimating equations. *Statist*. *Sci*. **33** 427–443. MR3843384 https://doi.org/10.1214/18-STS661

[39] SCHMIDT, L., SANTURKAR, S., TSIPRAS, D., TALWAR, K. and MADRY, A. (2018). Adversarially robust generalization requires more data. In *Advances in Neural Information Processing Systems* 5014–5026.

[40] SINGH, D., FEBBO, PG., ROSS, K., JACKSON, D., MANOLA, J., LADD, C., TAMAYO, P., RENSHAW, A., D'AMICO, A., RICHIE, JP., LANDER, ES., LODA, M., KANTOFF, P., GOLUB, TR. and SELLERS, WR. (2002). Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* **1(2)** 203–209. https://doi.org/10.1016/s1535-6108(02)00030-2

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

[42] VERSHYNIN, R. (2010). *Introduction to the non-asymptotic analysis of random matrices*, Cambridge University Press, New York.

[43] WRIGHT, J., YANG, A. Y., GANESH, A., SASTRY, S. S. and MA, Y. (2009). Robust face recognition via sparse representation. *IEEE Trans*. *Pattern Anal*. *Mach*. *Intell*. **31** 210–227.

[44] XIA, Y., CAI, T. and CAI, T. T. (2015). Testing differential networks with applications to the detection of gene-gene interactions. *Biometrika* **102** 247–266. MR3371002 https://doi.org/10.1093/biomet/asu074

[45] XU, M., ZHANG, D. and WU, W. B. (2014). L^2 asymptotics for high-dimensional data. Preprint. Available at arXiv:1405.7244.

[46] XUE, L. and ZOU, H. (2012). Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *Ann*. *Statist*. **40** 2541–2571. MR3097612 https://doi.org/10.1214/12-AOS1041

[47] ZHAO, S. D., CAI, T. T. and LI, H. (2014). Direct estimation of differential networks. *Biometrika* **101** 253–268. MR3215346 https://doi.org/10.1093/biomet/asu009

[48] ZHAO, Y. and WEGKAMP, M. (2014). Semiparametric Gaussian copula classification. Preprint. Available at arXiv:1411.2944.

The Annals of Statistics
2021, Vol. 49, No. 3, 1569–1593
https://doi.org/10.1214/20-AOS2013
© Institute of Mathematical Statistics, 2021

CENTRAL LIMIT THEOREM FOR LINEAR SPECTRAL STATISTICS OF LARGE DIMENSIONAL KENDALL'S RANK CORRELATION MATRICES AND ITS APPLICATIONS

BY ZENG LI[1], QINWEN WANG[2] AND RUNZE LI[3]

[1]*Department of Statistics and Data Science, Southern University of Science and Technology, liz9@sustech.edu.cn*

[2]*School of Data Science, Fudan University, wqw@fudan.edu.cn*

[3]*Department of Statistics, Pennsylvania State University, rzli@psu.edu*

This paper is concerned with the limiting spectral behaviors of large dimensional Kendall's rank correlation matrices generated by samples with independent and continuous components. The statistical setting in this paper covers a wide range of highly skewed and heavy-tailed distributions since we do not require the components to be identically distributed, and do not need any moment conditions. We establish the central limit theorem (CLT) for the linear spectral statistics (LSS) of the Kendall's rank correlation matrices under the Marchenko–Pastur asymptotic regime, in which the dimension diverges to infinity proportionally with the sample size. We further propose three nonparametric procedures for high dimensional independent test and their limiting null distributions are derived by implementing this CLT. Our numerical comparisons demonstrate the robustness and superiority of our proposed test statistics under various mixed and heavy-tailed cases.

REFERENCES

ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. 2nd ed. *Wiley Publications in Statistics*. Wiley, New York. MR0091588

BAI, Z. D. and SILVERSTEIN, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.* **26** 316–345. MR1617051 https://doi.org/10.1214/aop/1022855421

BAI, Z. D. and SILVERSTEIN, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** 553–605. MR2040792 https://doi.org/10.1214/aop/1078415845

BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2567175 https://doi.org/10.1007/978-1-4419-0661-8

BAI, Z. D. and YAO, J. (2005). On the convergence of the spectral empirical process of Wigner matrices. *Bernoulli* **11** 1059–1092. MR2189081 https://doi.org/10.3150/bj/1137421640

BANDEIRA, A. S., LODHIA, A. and RIGOLLET, P. (2017). Marčenko–Pastur law for Kendall's tau. *Electron. Commun. Probab.* **22** Paper No. 32, 7. MR3663103 https://doi.org/10.1214/17-ECP59

BAO, Z. (2019a). Tracy–Widom limit for Kendall's tau. *Ann. Statist.* **47** 3504–3532. MR4025750 https://doi.org/10.1214/18-AOS1786

BAO, Z. (2019b). Tracy–Widom limit for spearman's rho. Preprint.

BAO, Z., PAN, G. and ZHOU, W. (2012). Tracy–Widom law for the extreme eigenvalues of sample correlation matrices. *Electron. J. Probab.* **17** no. 88, 32. MR2988403 https://doi.org/10.1214/EJP.v17-1962

BAO, Z., LIN, L.-C., PAN, G. and ZHOU, W. (2015). Spectral statistics of large dimensional Spearman's rank correlation matrix and its application. *Ann. Statist.* **43** 2588–2623. MR3405605 https://doi.org/10.1214/15-AOS1353

BILLINGSLEY, P. (2008). *Probability and Measure*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0534323

CAI, T. T. and JIANG, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39** 1496–1525. MR2850210 https://doi.org/10.1214/11-AOS879

MSC2020 subject classifications. Primary 62H10; secondary 62H15.

Key words and phrases. Kendall's rank correlation matrices, linear spectral statistics, central limit theorem, random matrix theory, high dimensional independent test.

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

GAO, J., HAN, X., PAN, G. and YANG, Y. (2017). High dimensional correlation matrices: The central limit theorem and its applications. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 677–693. MR3641402 https://doi.org/10.1111/rssb.12189

HAN, F., CHEN, S. and LIU, H. (2017). Distribution-free tests of independence in high dimensions. *Biometrika* **104** 813–828. MR3737306 https://doi.org/10.1093/biomet/asx050

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I* 361–379. Univ. California Press, Berkeley, CA. MR0133191

JIANG, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **14** 865–880. MR2052906 https://doi.org/10.1214/105051604000000143

JIANG, D., BAI, Z. and ZHENG, S. (2013). Testing the independence of sets of large-dimensional variables. *Sci. China Math.* **56** 135–147. MR3016588 https://doi.org/10.1007/s11425-012-4501-0

JIANG, T. and QI, Y. (2015). Likelihood ratio tests for high-dimensional normal distributions. *Scand. J. Stat.* **42** 988–1009. MR3426306 https://doi.org/10.1111/sjos.12147

JOHNSTONE, I. M. (2006). High dimensional statistical inference and random matrices. Preprint. Available at arXiv:math/0611589.

LEUNG, D. and DRTON, M. (2018). Testing independence in high dimensions with sums of rank correlations. *Ann. Statist.* **46** 280–307. MR3766953 https://doi.org/10.1214/17-AOS1550

LI, Z., WANG, Q. and LI, R. (2021). Supplement to "Central limit theorem for linear spectral statistics of large dimensional Kendall's rank correlation matrices and its applications." https://doi.org/10.1214/20-AOS2013SUPP

MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb.* (*N.S.*) **72 (114)** 507–536. MR0208649

MUIRHEAD, R. J. (2009). *Aspects of Multivariate Statistical Theory* **97**. Wiley, New York.

PAN, G. M. and ZHOU, W. (2008). Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. *Ann. Appl. Probab.* **18** 1232–1270. MR2418244 https://doi.org/10.1214/07-AAP477

PAUL, D. and AUE, A. (2014). Random matrix theory in statistics: A review. *J. Statist. Plann. Inference* **150** 1–29. MR3206718 https://doi.org/10.1016/j.jspi.2013.09.005

PILLAI, N. S. and YIN, J. (2012). Edge universality of correlation matrices. *Ann. Statist.* **40** 1737–1763. MR3015042 https://doi.org/10.1214/12-AOS1022

TULINO, A. M. and VERDÚ, S. (2004). Random matrix theory and wireless communications. *Found. Trends Commun. Inf. Theory* **1** 1–182.

YANG, Y. and PAN, G. (2015). Independence test for high dimensional data based on regularized canonical correlation coefficients. *Ann. Statist.* **43** 467–500. MR3316187 https://doi.org/10.1214/14-AOS1284

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

ZHENG, S. (2012). Central limit theorems for linear spectral statistics of large dimensional F-matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 444–476. MR2954263 https://doi.org/10.1214/11-AIHP414

ZHENG, S., BAI, Z. and YAO, J. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *Ann. Statist.* **43** 546–591. MR3316190 https://doi.org/10.1214/14-AOS1292

ZHENG, S., CHEN, Z., CUI, H. and LI, R. (2019). Hypothesis testing on linear structures of high-dimensional covariance matrix. *Ann. Statist.* **47** 3300–3334. MR4025743 https://doi.org/10.1214/18-AOS1779

ZHOU, W. (2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Trans. Amer. Math. Soc.* **359** 5345–5363. MR2327033 https://doi.org/10.1090/S0002-9947-07-04192-X

The Annals of Statistics
2021, Vol. 49, No. 3, 1594–1625
https://doi.org/10.1214/20-AOS2015
© Institute of Mathematical Statistics, 2021

APPROXIMATE AND EXACT DESIGNS FOR TOTAL EFFECTS

BY XIANGSHUN KONG[1], MINGAO YUAN[2] AND WEI ZHENG[3]

[1]*Department of Statistics, Beijing Institute of Technology, kongsunday@163.com*
[2]*Department of Statistics, North Dakota State University, mingao.yuan@ndsu.edu*
[3]*Department of Business Analytics and Statistics, University of Tennessee, Knoxville, wzheng9@utk.edu*

This paper considers both approximate and exact designs for estimating the total effects under one crossover and two interference models. They are different from the traditional block designs in the sense that the assigned treatments also affect their neighboring plots, hence a design is understood as a collection of sequences of treatments. A notable result in literature is that the circular neighbor balanced design (CNBD) is optimal among designs, which do not allow treatments to be neighbors of themselves. However, we find it necessary to allow self-neighboring, and further show that it is the best to allocate each treatment in a subblock of adjacent plots with equal or almost equal numbers of replications. This explains why the efficiency of CNBD drops down to 50% as the sequence length, say k, increases. Unlike CNBD or the designs for direct effects, our proposed designs do not try to put as many treatments in a sequence as possible. The optimal number of distinct treatments in a sequence is around $\sqrt{2k}$ for crossover designs and $\sqrt{k/0.96}$ for interference models, whenever they are smaller than the total number of treatments under consideration. We systematically study necessary and sufficient conditions for any design to be universally optimal under the approximate design framework, based on which algorithms for deriving optimal or efficient exact designs are proposed. This hybrid nature of cohesively combining theories with algorithms makes our method more flexible than existing ones in the following aspects. (i) Not only symmetric designs are studied, general procedures for producing asymmetric designs are also provided. (ii) Our method applies to any form of within-block covariance matrix instead of specific forms. (iii) We cover all configurations of the numbers of treatments and sequence lengths, especially for large values of them when purely computational methods are not applicable. (iv) On top of the latter, we cover a continuous spectrum of the number of sequences instead of special numbers decided by combinatorial constraints.

REFERENCES

[1] AFSARINEJAD, K. and HEDAYAT, A. S. (2002). Repeated measurements designs for a model with self and simple mixed carryover effects. *J. Statist. Plann. Inference* **106** 449–459. MR1927725 https://doi.org/10.1016/S0378-3758(02)00227-6

[2] AI, M., YU, Y. and HE, S. (2009). Optimality of circular neighbor-balanced designs for total effects with autoregressive correlated observations. *J. Statist. Plann. Inference* **139** 2293–2304. MR2507991 https://doi.org/10.1016/j.jspi.2008.10.025

[3] ALDRED, R. E. L., BAILEY, R. A., MCKAY, B. D. and WANLESS, I. M. (2014). Circular designs balanced for neighbours at distances one and two. *Biometrika* **101** 943–956. MR3286927 https://doi.org/10.1093/biomet/asu036

[4] ANDREWS, G. E. (1998). *The Theory of Partitions*. *Cambridge Mathematical Library*. Cambridge Univ. Press, Cambridge. MR1634067

[5] AZAIS, J. M., BAILEY, R. A. and MONOD, H. (1993). A catalogue of efficient neighbour-designs with border plots. *Biometrics* **49** 1252–1261.

MSC2020 subject classifications. Primary 62K05; secondary 62J05.
Key words and phrases. Approximate design theory, interference model, linear equations system, symmetric designs, universally optimal designs.

[6] AZAIS, J. M., ONILLON, O. and LEFORT BUSON, M. (1986). Une methode d'etude de phenomenes de competition entre genotypes. Application au colza (*Brassica napus* L.). *Agronomie* **6** 601–614.

[7] BAILEY, R. A. and DRUILHET, P. (2004). Optimality of neighbor-balanced designs for total effects. *Ann. Statist.* **32** 1650–1661. MR2089136 https://doi.org/10.1214/009053604000000481

[8] BAILEY, R. A. and DRUILHET, P. (2014). Optimal cross-over designs for full interaction models. *Ann. Statist.* **42** 2282–2300. MR3269980 https://doi.org/10.1214/14-AOS1247

[9] BAILEY, R. A. and KUNERT, J. (2006). On optimal crossover designs when carryover effects are proportional to direct effects. *Biometrika* **93** 613–625. MR2261446 https://doi.org/10.1093/biomet/93.3.613

[10] BAILEY, R. A. and PAYNE, R. W. (1990). Experimental design: Statistical research and its application. In *Institute of Arable Crops Research Report for* 1989 (J. Abbott, ed.) 107–112. Agriculture and Food Research Council, Institute of Arable Crops Research, Harpenden.

[11] BESAG, J. and KEMPTON, R. A. (1986). Statistical analysis of field experiments using neighbouring plots. *Biometrics* **42** 231–251.

[12] BHALLI, M. A., DAY, A. D., TUCKER, H., THOMSON, R. K. and MASSEY, G. D. (1964). End-border effects in irrigated barley yield trial. *Agron. J.* **1956** 346–348.

[13] BOSE, M. and DEY, A. (2013). Developments in crossover designs.

[14] BRADSHAW, J. E. (1986). Competition between cultivars of fodder kale in yield trials with single-row plots. *Euphytica* **35** 433–439.

[15] BRANDT, A. E. (1938). Tests of significance in reversal or switchback trials. Research Bulletin No. 234, Iowa Agricultural Experimental Station.

[16] CANNELL, M. G. R., NJUGUAN, C. K., FORD, E. D., SMITH, R. and ROSS-PARKER, H. M. (1977). Variation in yield among competing individuals within mixed genotype stands of tea: A selection problem. *J. Appl. Ecol.* **14** 969–985.

[17] CHAN, A. W. and ALTMAN, D. G. (2005). Epidemiology and reporting of randomized trials published in PubMed journals. *Lancet* **365** 1159–1162.

[18] CHÊNG, C. S. and WU, C.-F. (1980). Balanced repeated measurements designs. *Ann. Statist.* **8** 1272–1283. MR0594644

[19] CLEOPHAS, T. J. (1990). A simple method for the estimation of interaction bias in crossover studies. *J. Clin. Pharmacol.* **30** 1036–1040.

[20] COCHRAN, W. G., AUTREY, K. M. and CANNON, C. Y. (1941). A double changeover design for dairy cattle feeding experiments. *J. Dairy Sci.* **24** 937–951.

[21] COTTON, J. W. (1998). *Analyzing Within—Subjects Experiments*. Lawrence Erlbaum Associates, Mahwah, NJ.

[22] COX, D. R. (1958). *Planning of Experiments*. *A Wiley Publication in Applied Statistics*. Wiley, New York. MR0095561

[23] DETTE, H. and MELAS, V. B. (2011). A note on the de la Garza phenomenon for locally optimal designs. *Ann. Statist.* **39** 1266–1281. MR2816354 https://doi.org/10.1214/11-AOS875

[24] DETTE, H. and SCHORNING, K. (2013). Complete classes of designs for nonlinear regression models and principal representations of moment spaces. *Ann. Statist.* **41** 1260–1267. MR3113810 https://doi.org/10.1214/13-AOS1108

[25] DRUILHET, P. (1999). Optimality of neighbour balanced designs. *J. Statist. Plann. Inference* **81** 141–152. MR1718460 https://doi.org/10.1016/S0378-3758(99)00004-X

[26] DRUILHET, P. and TINSSON, W. (2012). Efficient circular neighbour designs for spatial interference model. *J. Statist. Plann. Inference* **142** 1161–1169. MR2879760 https://doi.org/10.1016/j.jspi.2011.11.017

[27] DRUILHET, P. and TINSSON, W. (2014). Optimal cross-over designs for total effects under a model with self and mixed carryover effects. *J. Statist. Plann. Inference* **154** 54–61. MR3258405 https://doi.org/10.1016/j.jspi.2013.07.006

[28] DYKE, G. V. and SHELLEY, C. F. (1976). Serial designs balanced for effects of neighbours on both sides. *J. Agric. Sci.* **87** 303–305.

[29] ELBORNE, D. R., ALTMAN, D. G., HIGGINS, J. P. T., CURTIN, F., WORTHINGTON, H. V. and VAIL, A. (2002). Meta-analyses involving cross-over trials: Methodological issues. *Int. J. Epidemiol.* **31** 140–149.

[30] FILIPIAK, K. (2012). Universally optimal designs under an interference model with equal left- and right-neighbor effects. *Statist. Probab. Lett.* **82** 592–598. MR2887476 https://doi.org/10.1016/j.spl.2011.11.013

[31] FILIPIAK, K. and MARKIEWICZ, A. (2005). Optimality and efficiency of circular neighbor balanced designs for correlated observations. *Metrika* **61** 17–27. MR2126434 https://doi.org/10.1007/s001840400321

[32] FINNEY, D. J. (1956). Cross-over designs in bioassay. *Proc. R. Soc. Lond., B Biol. Sci.* **145** 42–60.

[33] GOMEZ, K. A. (1972). Border effects in rice experimental plots. II. Varietal competition. *Exp. Agric.* **8** 295–298.

[34] HAINES, W. B. and BENZIAN, B. (1956). Some manuring experiments on oil palm in Africa. *Emp. J. Exp. Agric.* **24** 137–160.

[35] HEDAYAT, A. and AFSARINEJAD, K. (1978). Repeated measurements designs. II. *Ann. Statist.* **6** 619–628. MR0488527

[36] HEDAYAT, A. S. and YANG, M. (2003). Universal optimality of balanced uniform crossover designs. *Ann. Statist.* **31** 978–983. MR1994737 https://doi.org/10.1214/aos/1056562469

[37] HEDAYAT, A. S. and YANG, M. (2004). Universal optimality for selected crossover designs. *J. Amer. Statist. Assoc.* **99** 461–466. MR2062831 https://doi.org/10.1198/016214504000000331

[38] HEDAYAT, A. S. and ZHENG, W. (2010). Optimal and efficient crossover designs for test-control study when subject effects are random. *J. Amer. Statist. Assoc.* **105** 1581–1592. MR2796573 https://doi.org/10.1198/jasa.2010.tm10134

[39] HEDAYAT, A. S. and ZHENG, W. (2020). Some notes on types of symmetry for crossover designs. Manuscript.

[40] HIDE, G. A. and READ, P. J. (1990). Effect of neighbouring plants on the yield of potatoes from seed tubers affected with gangrene (*Phoma foveata*) or from plants affected with stem canker (*Rhizoctonia solani*). *Ann. Appl. Biol.* **116** 233–243.

[41] JEEVITHA, M. and SANTHARAM, C. (2013). Optimality and efficiency of circular neighbor balanced design. *Int. J. Stat. Mat.* **7** 19–23.

[42] JENKYN, J. F. and DYKE, G. V. (1985). Interference between plots in experiments with plant pathogens. *Asp. Appl. Biol.* **10** 75–85.

[43] JENSEN, N. F. and FEDERER, W. T. (1964). Adjacent row competition in wheat. *Crop Sci.* **4** 641–645.

[44] JONES, B. and DEPPE, C. (2001). Recent developments in the design of cross-over trials: A brief review and bibliography. In *Recent Advances in Experimental Designs and Related Topics* (*Philadelphia, PA*, 1999) 153–173. Nova Sci. Publ., Huntington, NY. MR2040519 https://doi.org/10.1016/s0378-3758(99)00181-0

[45] JONES, B. and KENWARD, M. G. (2015). *Design and Analysis of Cross-over Trials. Monographs on Statistics and Applied Probability* **34**. CRC Press, London. MR1014893

[46] KAWANO, K., AMAYA, A., DAZA, P. and RIOS, M. (1978). Factors affecting efficiency of hybridization and seletion in cassava. *Crop Sci.* **17** 373–376.

[47] KEMPTON, R. A. (1982). Adjustment for competition between varieties in plant breeding trials. *J. Agric. Sci.* **98** 599–611.

[48] KEMPTON, R. A. (1992). Interference in agricultural experiments. In *Proceedings of the* 2*nd Meeting of the Biometric Society, East, Central and Southern African Network, Harare, Zimbabwe* 1–19.

[49] KEMPTON, R. A., GREGORY, R. S., HUGHES, W. G. and STOEER, P. J. (1986). The effect of interplot competition on yield assessment in triticale trial. *Euphytica* **35** 257–265.

[50] KEMPTON, R. A. and LOCKWOOD, G. (1984). Interplot competition in variety trials of field beans (*Vicia faba* L.). *J. Agric. Sci.* **103** 293–302.

[51] KIEFER, J. (1975). Construction and optimality of generalized Youden designs. In *A Survey of Statistical Design and Linear Models* (*Proc. Internat. Sympos., Colorado State Univ., Ft. Collins, Colo.*, 1973) (J. N. Srivastava, ed.) 333–353. North-Holland, Amsterdam. MR0395079

[52] KUNERT, J. (1984). Designs balanced for circular residual effects. *Comm. Statist. Theory Methods* **13** 2665–2671. MR0759242 https://doi.org/10.1080/03610928408828850

[53] KUNERT, J. (1984). Optimality of balanced uniform repeated measurements designs. *Ann. Statist.* **12** 1006–1017. MR0751288 https://doi.org/10.1214/aos/1176346717

[54] KUNERT, J. and MARTIN, R. J. (2000). On the determination of optimal designs for an interference model. *Ann. Statist.* **28** 1728–1742. MR1835039 https://doi.org/10.1214/aos/1015957478

[55] KUNERT, J. and MERSMANN, S. (2011). Optimal designs for an interference model. *J. Statist. Plann. Inference* **141** 1623–1632. MR2747931 https://doi.org/10.1016/j.jspi.2010.11.020

[56] KUNERT, J. and STUFKEN, J. (2002). Optimal crossover designs in a model with self and mixed carryover effects. *J. Amer. Statist. Assoc.* **97** 898–906. MR1941418 https://doi.org/10.1198/016214502388618681

[57] KUSHNER, H. B. (1997). Optimal repeated measurements designs: The linear optimality equations. *Ann. Statist.* **25** 2328–2344. MR1604457 https://doi.org/10.1214/aos/1030741075

[58] KUSHNER, H. B. (1998). Optimal and efficient repeated-measurements designs for uncorrelated observations. *J. Amer. Statist. Assoc.* **93** 1176–1187. MR1649211 https://doi.org/10.2307/2669860

[59] LANGTON, S. (1990). Avoiding edge effects in agroforestry experiments; the use of neighbour-balanced designs and guard areas. *Agrofor. Syst.* **12** 173–185.

[60] LI, K., ZHENG, W. and AI, M. (2015). Optimal designs for the proportional interference model. *Ann. Statist.* **43** 1596–1616. MR3357872 https://doi.org/10.1214/15-AOS1317

[61] MACLURE, M. (1991). The case-crossover design: A method for studying transient effects on the risk of acute events. *Am. J. Epidemiol.* **133** 144–153.

[62] MAGDA, C. G. (1980). Circular balanced repeated measurements designs. *Comm. Statist. Theory Methods* **9** 1901–1918. MR0603164 https://doi.org/10.1080/03610928008828004

[63] MATTHEWS, J. N. S. (1988). Recent developments in crossover designs. *Int. Stat. Rev.* **56** 117–127. MR0963525 https://doi.org/10.2307/1403636

[64] MCDONALD, G. E. and PECK, N. H. (1976). Border effect in a long-term fertility experiment. *Agron. J.* **68** 530–532.

[65] MURUGESAN, M. and AROKIA RAJ, A. (1978). Inter-plot competition in manurial experiments among different genotypes of sorghum. *Madras Agric. J.* **65** 46–51.

[66] PARKES, K. R. (1982). Occupational stress among student nurses. *J. Appl. Psychol.* **67** 784–796.

[67] PATTERSON, S. and JONES, B. (2006). *Bioequivalence and Statistics in Clinical Pharmacology. Interdisciplinary Statistics*. CRC Press/CRC, Boca Raton, FL. MR2193758

[68] PEARCE, S. C. (1957). Experimenting with organisms as blocks. *Biometrika* **44** 141–149. MR0090207 https://doi.org/10.1093/biomet/44.1-2.141

[69] RAGHAVARAO, D. (1989). Crossover designs in industry. In *Design and Analysis of Experiments, with Applications to Engineering and Physical Sciences* (S. Gosh, ed.). Dekker, New York.

[70] RAGHAVARAO, D. and PADGETT, L. (2014). *Repeated Measurements and Cross-over Designs*. Wiley, New York.

[71] RATKOWSKY, D. A., EVANS, M. A. and ALLDREDGE, J. R. (1992). *Cross-over Experiments: Design, Analysis, and Application*. Dekker, New York.

[72] SENN, S. J. (2002). *Cross-over Trials in Clinical Research*, 2nd ed. Chichester, Wiley.

[73] SIMPSON, T. W. (1938). Experimental methods and human nutrition (with discussion). *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **5** 46–69.

[74] SPECKEL, D., VINCOURT, P., AZAIS, J. M. and KOBILINSKY, A. (1987). Etude de la competition interparcellaire chez le tournesol. *Biom. Praxim.* **27** 21–43.

[75] STADLER (1921). University of Missouri Agricultural Experimental Station Research Bulletin 49.

[76] STUFKEN, J. (1996). Optimal crossover designs. In *Design and Analysis of Experiments. Handbook of Statist.* **13** 63–90. North-Holland, Amsterdam. MR1492565 https://doi.org/10.1016/S0169-7161(96)13005-4

[77] WILLIAMS, E. J. (1949). Experimental designs balanced for the estimation of residual effects of treatments. *Aust. J. Sci. Res., Ser. A* **2** 149–168. MR0033508

[78] BROWN, B. W. JR. (1980). The crossover experiment for clinical trials. *Biometrics* **36** 69–79.

[79] YANG, M. (2010). On the de la Garza phenomenon. *Ann. Statist.* **38** 2499–2524. MR2676896 https://doi.org/10.1214/09-AOS787

[80] YANG, M. and STUFKEN, J. (2009). Support points of locally optimal designs for nonlinear models with two parameters. *Ann. Statist.* **37** 518–541. MR2488361 https://doi.org/10.1214/07-AOS560

[81] YANG, M. and STUFKEN, J. (2012). Identifying locally optimal designs for nonlinear models: A simple extension with profound consequences. *Ann. Statist.* **40** 1665–1681. MR3015039 https://doi.org/10.1214/12-AOS992

[82] ZHENG, W. (2013). Universally optimal crossover designs under subject dropout. *Ann. Statist.* **41** 63–90. MR3059410 https://doi.org/10.1214/12-AOS1074

[83] ZHENG, W. (2013). Optimal crossover designs for the proportional model. *Ann. Statist.* **41** 2218–2235. MR3127864 https://doi.org/10.1214/13-AOS1148

[84] ZHENG, W. (2015). Universally optimal designs for two interference models. *Ann. Statist.* **43** 501–518. MR3316188 https://doi.org/10.1214/14-AOS1287

[85] ZHENG, W., AI, M. and LI, K. (2017). Identification of universally optimal circular designs for the interference model. *Ann. Statist.* **45** 1462–1487. MR3670185 https://doi.org/10.1214/16-AOS1496

The Annals of Statistics
2021, Vol. 49, No. 3, 1626–1655
https://doi.org/10.1214/20-AOS2016
© Institute of Mathematical Statistics, 2021

ON STATISTICAL LEARNING OF SIMPLICES: UNMIXING PROBLEM REVISITED

BY AMIR NAJAFI[1,*], SAEED ILCHI[1,†], AMIR HOSSEIN SABERI[2,§], SEYED ABOLFAZL MOTAHARI[1,‡], BABAK H. KHALAJ[2,¶] AND HAMID R. RABIEE[3]

[1]*Data Analytics Lab (DAL), Computer Engineering Department, Sharif University of Technology, [*]najafy@ce.sharif.edu; [†]silchi@ce.sharif.edu; [‡]motahari@sharif.edu*

[2]*Electrical Engineering Department, Sharif University of Technology, [§]sah.saberi@ee.sharif.edu; [¶]khalaj@sharif.edu*

[3]*Data science and Machine learning Lab (DML), Computer Engineering Department, Sharif University of Technology, rabiee@sharif.edu*

We study the sample complexity of learning a high-dimensional simplex from a set of points uniformly sampled from its interior. Learning of simplices is a long studied problem in computer science and has applications in computational biology and remote sensing, mostly under the name of "spectral unmixing." We theoretically show that a sufficient sample complexity for reliable learning of a K-dimensional simplex up to a total-variation error of ϵ is $O(\frac{K^2}{\epsilon}\log\frac{K}{\epsilon})$, which yields a substantial improvement over existing bounds. Based on our new theoretical framework, we also propose a heuristic approach for the inference of simplices. Experimental results on synthetic and real-world datasets demonstrate a comparable performance for our method on noiseless samples, while we outperform the state-of-the-art in noisy cases.

REFERENCES

[1] AGATHOS, A., LI, J., BIOUCAS-DIAS, J. M. and PLAZA, A. (2014). Robust minimum volume simplex analysis for hyperspectral unmixing. In *Signal Processing Conference* (*EUSIPCO*), 2014 *Proceedings of the* 22nd *European* 1582–1586. IEEE, Piscataway, NJ.

[2] AMBIKAPATHI, A., CHAN, T.-H., MA, W.-K. and CHI, C.-Y. (2011). Chance-constrained robust minimum-volume enclosing simplex algorithm for hyperspectral unmixing. *IEEE Trans*. *Geosci*. *Remote Sens*. **49** 4194–4209.

[3] ANDERSON, J., GOYAL, N. and RADEMACHER, L. (2013). Efficient learning of simplices. In *Conference on Learning Theory* 1020–1045.

[4] ANG, M. and GILLIS, N. (2019). Algorithms and comparisons of non-negative matrix factorization with volume regularization for hyperspectral unmixing. arXiv preprint, arXiv:1903.04362.

[5] BÁRÁNY, I. and BUCHTA, C. (1993). Random polytopes in a convex polytope, independence of shape, and concentration of vertices. *Math*. *Ann*. **297** 467–497. MR1245400 https://doi.org/10.1007/BF01459511

[6] BÁRÁNY, I. and LARMAN, D. G. (1988). Convex bodies, economic cap coverings, random polytopes. *Mathematika* **35** 274–291. MR0986636 https://doi.org/10.1112/S0025579300015266

[7] BIOUCAS-DIAS, J. M. (2009). A variable splitting augmented Lagrangian approach to linear spectral unmixing. In *Hyperspectral Image and Signal Processing*: *Evolution in Remote Sensing*, 2009. *WHISPERS*'09. *First Workshop on* 1–4. IEEE.

[8] BIOUCAS-DIAS, J. M., PLAZA, A., DOBIGEON, N., PARENTE, M., DU, Q., GADER, P. and CHANUSSOT, J. (2012). Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE J*. *Sel*. *Top*. *Appl*. *Earth Obs*. *Remote Sens*. **5** 354–379.

[9] BLUMER, A., EHRENFEUCHT, A., HAUSSLER, D. and WARMUTH, M. K. (1989). Learnability and the Vapnik–Chervonenkis dimension. *J*. *Assoc*. *Comput*. *Mach*. **36** 929–965. MR1072253 https://doi.org/10.1145/76359.76371

[10] BRUNEL, V.-E. (2013). Adaptive estimation of convex polytopes and convex sets from noisy data. *Electron*. *J*. *Stat*. **7** 1301–1327. MR3063609 https://doi.org/10.1214/13-EJS804

MSC2020 subject classifications. 62F99.

Key words and phrases. Inference of simplices, sample complexity, statistical learning theory, high-dimensional geometry, computer simulations.

[11] BRUNEL, V.-E. (2016). Adaptive estimation of convex and polytopal density support. *Probab. Theory Related Fields* **164** 1–16. MR3449384 https://doi.org/10.1007/s00440-014-0605-5

[12] BRUNEL, V.-E. (2018). Methods for estimation of convex sets. *Statist. Sci.* **33** 615–632. MR3881211 https://doi.org/10.1214/18-STS669

[13] CHAN, T.-H., CHI, C.-Y., HUANG, Y.-M. and MA, W.-K. (2009). A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing. *IEEE Trans. Signal Process.* **57** 4418–4432. MR2683183 https://doi.org/10.1109/TSP.2009.2025802

[14] CHAZELLE, B. (1993). An optimal convex hull algorithm in any fixed dimension. *Discrete Comput. Geom.* **10** 377–409. MR1243335 https://doi.org/10.1007/BF02573985

[15] FRIEZE, A., JERRUM, M. and KANNAN, R. (1996). Learning linear transformations. In 37*th Annual Symposium on Foundations of Computer Science* (*Burlington*, *VT*, 1996) 359–368. IEEE Comput. Soc. Press, Los Alamitos, CA. MR1450634 https://doi.org/10.1109/SFCS.1996.548495

[16] FU, X., HUANG, K. and SIDIROPOULOS, N. D. (2018). On identifiability of nonnegative matrix factorization. *IEEE Signal Process. Lett.* **25** 328–332.

[17] FU, X., HUANG, K., SIDIROPOULOS, N. D. and MA, W.-K. (2019). Nonnegative matrix factorization for signal and data analytics: Identifiability, algorithms, and applications. *IEEE Signal Process. Mag.* **36** 59–80.

[18] GRIBONVAL, R., JENATTON, R., BACH, F., KLEINSTEUBER, M. and SEIBERT, M. (2015). Sample complexity of dictionary learning and other matrix factorizations. *IEEE Trans. Inf. Theory* **61** 3469–3486. MR3352510 https://doi.org/10.1109/TIT.2015.2424238

[19] GROEMER, H. (1974). On the mean value of the volume of a random polytope in a convex set. *Arch. Math.* (*Basel*) **25** 86–90. MR0341286 https://doi.org/10.1007/BF01238645

[20] HANNEKE, S. (2016). The optimal sample complexity of PAC learning. *J. Mach. Learn. Res.* **17** Paper No. 38, 15. MR3491132

[21] HYVÄRINEN, A., KARHUNEN, J. and OJA, E. (2004). *Independent Component Analysis* **46**. Wiley.

[22] LI, J., AGATHOS, A., ZAHARIE, D., BIOUCAS-DIAS, J. M., PLAZA, A. and LI, X. (2015). Minimum volume simplex analysis: A fast algorithm for linear hyperspectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **53** 5067–5082.

[23] LI, T. and DING, C.-C. (2018). Nonnegative matrix factorizations for clustering: A survey. In *Data Clustering* 149–176. CRC Press/CRC.

[24] LIN, C.-H., AMBIKAPATHI, A., LI, W.-C. and CHI, C.-Y. (2013). On the endmember identifiability of craig's criterion for hyperspectral unmixing: A statistical analysis for three-source case. In *Acoustics, Speech and Signal Processing* (*ICASSP*), 2013 *IEEE International Conference on* 2139–2143. IEEE.

[25] MOHRI, M., ROSTAMIZADEH, A. and TALWALKAR, A. (2012). *Foundations of Machine Learning*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3057769

[26] NASCIMENTO, J. M. and DIAS, J. M. (2005). Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **43** 898–910.

[27] NGUYEN, P. Q. and REGEV, O. (2009). Learning a parallelepiped: Cryptanalysis of GGH and NTRU signatures. *J. Cryptology* **22** 139–160. MR2496387 https://doi.org/10.1007/s00145-008-9031-0

[28] OSSERMAN, R. (1978). The isoperimetric inequality. *Bull. Amer. Math. Soc.* **84** 1182–1238. MR0500557 https://doi.org/10.1090/S0002-9904-1978-14553-4

[29] PACKER, A. (2002). NP-hardness of largest contained and smallest containing simplices for V- and H-polytopes. *Discrete Comput. Geom.* **28** 349–377. MR1923957 https://doi.org/10.1007/s00454-002-0745-8

[30] SATAS, G. and RAPHAEL, B. J. (2017). Tumor phylogeny inference using tree-constrained importance sampling. *Bioinformatics* **33** i152–i160. https://doi.org/10.1093/bioinformatics/btx270

[31] SCHWARTZ, R. and SHACKNEY, S. E. (2010). Applying unmixing to gene expression data for tumor phylogeny inference. *BMC Bioinform.* **11** 42.

[32] SHEN-ORR, S. S., TIBSHIRANI, R., KHATRI, P., BODIAN, D. L., STAEDTLER, F., PERRY, N. M., HASTIE, T., SARWAL, M. M., DAVIS, M. M. et al. (2010). Cell type-specific gene expression differences in complex tissues. *Nat. Methods* **7** 287–289. https://doi.org/10.1038/nmeth.1439

[33] TOLLIVER, D., TSOURAKAKIS, C., SUBRAMANIAN, A., SHACKNEY, S. and SCHWARTZ, R. (2010). Robust unmixing of tumor states in array comparative genomic hybridization data. *Bioinformatics* **26** i106–i114.

[34] VAPNIK, V. (1982). *Estimation of Dependencies Based on Empirical Data*. *Information and Control*. Springer, New York. MR2240249

[35] WU, R., LI, Q. and MA, W.-K. (2019). Stochastic ml simplex-structured matrix factorization under the Dirichlet mixture model. In *ICASSP* 2019-2019 *IEEE International Conference on Acoustics*, *Speech and Signal Processing* (*ICASSP*) 5561–5565. IEEE, Piscataway, NJ.

[36] ZHANG, S., AGATHOS, A. and LI, J. (2017). Robust minimum volume simplex analysis for hyperspectral unmixing. *IEEE Trans*. *Geosci*. *Remote Sens*. **55** 6431–6439.

[37] ZHU, F., WANG, Y., FAN, B., XIANG, S., MENG, G. and PAN, C. (2014). Spectral unmixing via data-guided sparsity. *IEEE Trans*. *Image Process*. **23** 5412–5427. MR3298248 https://doi.org/10.1109/TIP.2014.2363423

[38] ZUCKERMAN, N. S., NOAM, Y., GOLDSMITH, A. J. and LEE, P. P. (2013). A self-directed method for cell-type identification and separation of gene expression microarrays. *PLoS Comput*. *Biol*. **9** e1003189.

The Annals of Statistics
2021, Vol. 49, No. 3, 1656–1678
https://doi.org/10.1214/20-AOS2017
© Institute of Mathematical Statistics, 2021

FACTOR-DRIVEN TWO-REGIME REGRESSION

BY SOKBAE LEE[1], YUAN LIAO[2], MYUNG HWAN SEO[3] AND YOUNGKI SHIN[4]

[1]*Department of Economics, Columbia University, sl3841@columbia.edu*

[2]*Department of Economics, Rutgers University, yuan.liao@rutgers.edu*

[3]*Department of Economics, Seoul National University, myunghseo@snu.ac.kr*

[4]*Department of Economics, McMaster University, shiny11@mcmaster.ca*

We propose a novel two-regime regression model where regime switching is driven by a vector of possibly unobservable factors. When the factors are latent, we estimate them by the principal component analysis of a panel data set. We show that the optimization problem can be reformulated as mixed integer optimization, and we present two alternative computational algorithms. We derive the asymptotic distribution of the resulting estimator under the scheme that the threshold effect shrinks to zero. In particular, we establish a phase transition that describes the effect of first-stage factor estimation as the cross-sectional dimension of panel data increases relative to the time-series dimension. Moreover, we develop bootstrap inference and illustrate our methods via numerical studies.

REFERENCES

[1] ANDREWS, D. W. K. (2002). Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. *Econometrica* **70** 119–162. MR1926257 https://doi.org/10.1111/1468-0262.00271

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

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

[4] 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

[5] BAI, J. and NG, S. (2008). Extremum estimation when the predictors are estimated from large panels. *Ann. Econ. Financ.* **9** 201–222.

[6] BAI, J. and NG, S. (2009). Boosting diffusion indices. *J. Appl. Econometrics* **24** 607–629. MR2675197 https://doi.org/10.1002/jae.1063

[7] BAI, J. and PERRON, P. (2003). Computation and analysis of multiple structural change models. *J. Appl. Econometrics* **18** 1–22.

[8] BERTSIMAS, D., KING, A. and MAZUMDER, R. (2016). Best subset selection via a modern optimization lens. *Ann. Statist.* **44** 813–852. MR3476618 https://doi.org/10.1214/15-AOS1388

[9] CHAN, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Ann. Statist.* **21** 520–533. MR1212191 https://doi.org/10.1214/aos/1176349040

[10] CHENG, X. and HANSEN, B. E. (2015). Forecasting with factor-augmented regression: A frequentist model averaging approach. *J. Econometrics* **186** 280–293. MR3343787 https://doi.org/10.1016/j.jeconom.2015.02.010

[11] 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

[12] GONÇALVES, S. and PERRON, B. (2014). Bootstrapping factor-augmented regression models. *J. Econometrics* **182** 156–173. MR3212768 https://doi.org/10.1016/j.jeconom.2014.04.015

[13] GONÇALVES, S. and PERRON, B. (2020). Bootstrapping factor models with cross sectional dependence. *J. Econometrics* **218** 476–495. MR4149236 https://doi.org/10.1016/j.jeconom.2020.04.026

MSC2020 subject classifications. Primary 62H25; secondary 62F12.

Key words and phrases. Threshold regression, principal component analysis, mixed integer optimization, phase transition, oracle properties.

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

[15] HANSEN, B. E. (1997). Inference in TAR models. *Stud. Nonlinear Dyn. Econom.* **2** 1–14. MR1467458 https://doi.org/10.2202/1558-3708.1024

[16] HANSEN, B. E. (2000). Sample splitting and threshold estimation. *Econometrica* **68** 575–603. MR1769379 https://doi.org/10.1111/1468-0262.00124

[17] HAWKINS, D. L. (1986). A simple least squares method for estimating a change in mean. *Comm. Statist. B. Simulation Comput.* **15** 655–679. MR0859644 https://doi.org/10.1080/03610918608812531

[18] HORVÁTH, L. and KOKOSZKA, P. (1997). The effect of long-range dependence on change-point estimators. *J. Statist. Plann. Inference* **64** 57–81. MR1492361 https://doi.org/10.1016/S0378-3758(96)00208-X

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

[20] LEE, S., LIAO, Y., SEO, M. H. and SHIN, Y. (2021). Supplement to "Factor-driven two-regime regression." https://doi.org/10.1214/20-AOS2017SUPP

[21] LING, S. (1999). On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model. *J. Appl. Probab.* **36** 688–705. MR1737046 https://doi.org/10.1017/s0021900200017502

[22] LUDVIGSON, S. C. and NG, S. (2009). Macro factors in bond risk premia. *Rev. Financ. Stud.* **22** 5027–5067.

[23] MCKEAGUE, I. W. and SEN, B. (2010). Fractals with point impact in functional linear regression. *Ann. Statist.* **38** 2559–2586. MR2676898 https://doi.org/10.1214/10-AOS791

[24] MERLEVÈDE, F., PELIGRAD, M. and RIO, E. (2011). A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probab. Theory Related Fields* **151** 435–474. MR2851689 https://doi.org/10.1007/s00440-010-0304-9

[25] QU, Z. and TKACHENKO, D. (2017). Global identification in DSGE models allowing for indeterminacy. *Rev. Econ. Stud.* **84** 1306–1345. MR3744244 https://doi.org/10.1093/restud/rdx007

[26] SEIJO, E. and SEN, B. (2011). Change-point in stochastic design regression and the bootstrap. *Ann. Statist.* **39** 1580–1607. MR2850213 https://doi.org/10.1214/11-AOS874

[27] SEO, M. H. and LINTON, O. (2007). A smoothed least squares estimator for threshold regression models. *J. Econometrics* **141** 704–735. MR2413485 https://doi.org/10.1016/j.jeconom.2006.11.002

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

[29] TONG, H. (1990). *Nonlinear Time Series*: *A Dynamical System Approach*. *Oxford Statistical Science Series* **6**. Oxford University Press, New York. MR1079320

[30] 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

The Annals of Statistics
2021, Vol. 49, No. 3, 1679–1701
https://doi.org/10.1214/20-AOS2018
© Institute of Mathematical Statistics, 2021

ROBUST BREGMAN CLUSTERING

BY CLAIRE BRÉCHETEAU[1], AURÉLIE FISCHER[2,*] AND CLÉMENT LEVRARD[2,†]

[1] *Université Rennes 2, claire.brecheteau@univ-rennes2.fr*
[2] *Laboratoire de Probabilités, Statistiques et Modélisation, Université de Paris,* [*] *aurelie.fischer@lpsm.paris;* [†] *clement.levrard@lpsm.paris*

Clustering with Bregman divergences encompasses a wide family of clustering procedures that are well suited to mixtures of distributions from exponential families (*J. Mach. Learn. Res.* **6** (2005) 1705–1749). However, these techniques are highly sensitive to noise. To address the issue of clustering data with possibly adversarial noise, we introduce a robustified version of Bregman clustering based on a trimming approach. We investigate its theoretical properties, showing for instance that our estimator converges at a sub-Gaussian rate $1/\sqrt{n}$, where n denotes the sample size, under mild tail assumptions. We also show that it is robust to a certain amount of noise, stated in terms of breakdown point. We also derive a Lloyd-type algorithm with a trimming parameter, along with a heuristic to select this parameter and the number of clusters from sample. Some numerical experiments assess the performance of our method on simulated and real datasets.

REFERENCES

[1] ARTHUR, D. and VASSILVITSKII, S. (2007). `k-means++`: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms* 1027–1035. ACM, New York. MR2485254

[2] BANERJEE, A., GUO, X. and WANG, H. (2005). On the optimality of conditional expectation as a Bregman predictor. *IEEE Trans. Inf. Theory* **51** 2664–2669. MR2246384 https://doi.org/10.1109/TIT.2005.850145

[3] BANERJEE, A., MERUGU, S., DHILLON, I. S. and GHOSH, J. (2005). Clustering with Bregman divergences. *J. Mach. Learn. Res.* **6** 1705–1749. MR2249870

[4] BIAU, G., DEVROYE, L. and LUGOSI, G. (2008). On the performance of clustering in Hilbert spaces. *IEEE Trans. Inf. Theory* **54** 781–790. MR2444554 https://doi.org/10.1109/TIT.2007.913516

[5] BRÉCHETEAU, C., FISCHER, A. and LEVRARD, C. (2021). Supplement to "Robust Bregman clustering." https://doi.org/10.1214/20-AOS2018SUPP

[6] BRÈGMAN, L. M. (1967). A relaxation method of finding a common point of convex sets and its application to the solution of problems in convex programming. *Zh. Vychisl. Mat. Mat. Fiz.* **7** 620–631. MR0215617

[7] BROWNLEES, C., JOLY, E. and LUGOSI, G. (2015). Empirical risk minimization for heavy-tailed losses. *Ann. Statist.* **43** 2507–2536. MR3405602 https://doi.org/10.1214/15-AOS1350

[8] CARDOT, H., CÉNAC, P. and ZITT, P.-A. (2013). Efficient and fast estimation of the geometric median in Hilbert spaces with an averaged stochastic gradient algorithm. *Bernoulli* **19** 18–43. MR3019484 https://doi.org/10.3150/11-BEJ390

[9] CATONI, O. and GIULINI, I. (2018). Dimension-free PAC-Bayesian bounds for the estimation of the mean of a random vector. ArXiv E-prints.

[10] CESA-BIANCHI, N. and LUGOSI, G. (2006). *Prediction, Learning, and Games*. Cambridge Univ. Press, Cambridge. MR2409394 https://doi.org/10.1017/CBO9780511546921

[11] CHAZAL, F., COHEN-STEINER, D. and MÉRIGOT, Q. (2011). Geometric inference for probability measures. *Found. Comput. Math.* **11** 733–751. MR2859954 https://doi.org/10.1007/s10208-011-9098-0

[12] CHAZAL, F., COHEN-STEINER, D. and MÉRIGOT, Q. (2011). Geometric inference for measures based on distance functions. *Found. Comput. Math.* **11** 733–751. MR2859954 https://doi.org/10.1007/s10208-011-9098-0

MSC2020 subject classifications. Primary 62H30; secondary 62G35.
Key words and phrases. Clustering, Bregman divergence, robustness.

[13] Chazal, F., Guibas, L. J., Oudot, S. Y. and Skraba, P. (2013). Persistence-based clustering in Riemannian manifolds. *J. ACM* **60** Art. 41, 38. MR3144911 https://doi.org/10.1145/2535927

[14] Coe, R. and Stern, R. D. (1982). Fitting models to daily rainfall data. *J. Appl. Meteorol.* **21** 1024–1031. https://doi.org/10.1175/1520-0450(1982)021<1024:FMTDRD>2.0.CO;2

[15] Cuesta-Albertos, J. A., Gordaliza, A. and Matrán, C. (1997). Trimmed k-means: An attempt to robustify quantizers. *Ann. Statist.* **25** 553–576. MR1439314 https://doi.org/10.1214/aos/1031833664

[16] Donoho, D. and Huber, P. J. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann. Wadsworth Statist./Probab. Ser.* 157–184. Wadsworth, Belmont, CA. MR0689745

[17] Duda, R. O., Hart, P. E. and Stork, D. G. (2001). *Pattern Classification*, 2nd ed. Wiley Interscience, New York. MR1802993

[18] Evert, S. (2004). A simple LNRE model for random character sequences. In *In Proceedings of the 7èmes Journées Internationales d'Analyse Statistique des Données Textuelles* (*Louvain-la-Neuve* 411–422.

[19] Fischer, A. (2010). Quantization and clustering with Bregman divergences. *J. Multivariate Anal.* **101** 2207–2221. MR2671211 https://doi.org/10.1016/j.jmva.2010.05.008

[20] Fritz, H., Garcia-Escudero, L. A. and Mayo-Iscar, A. (2012). Tclust: An R package for a trimming approach to cluster analysis. *J. Stat. Softw.* **47** 1–26.

[21] García-Escudero, L. A., Gordaliza, A., Matrán, C. and Mayo-Iscar, A. (2008). A general trimming approach to robust cluster analysis. *Ann. Statist.* **36** 1324–1345. MR2418659 https://doi.org/10.1214/07-AOS515

[22] Gordaliza, A. (1991). Best approximations to random variables based on trimming procedures. *J. Approx. Theory* **64** 162–180. MR1091467 https://doi.org/10.1016/0021-9045(91)90072-I

[23] Gray, R. M., Buzo, A., Gray, A. H. and Matsuyama, Y. (1980). Distortion measures for speech processing. *IEEE Trans. Acoust. Speech Signal Process.* **28** 367–376.

[24] Hahsler, M., Piekenbrock, M. and Doran, D. (2019). Dbscan: Fast density-based clustering with R. *J. Stat. Softw.* **91** 1–30. https://doi.org/10.18637/jss.v091.i01

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

[26] Lecué, G. and Lerasle, M. (2020). Robust machine learning by median-of-means: Theory and practice. *Ann. Statist.* **48** 906–931. MR4102681 https://doi.org/10.1214/19-AOS1828

[27] Levrard, C. (2015). Nonasymptotic bounds for vector quantization in Hilbert spaces. *Ann. Statist.* **43** 592–619. MR3316191 https://doi.org/10.1214/14-AOS1293

[28] Levrard, C. (2018). Quantization/clustering: When and why does k-means work? *J. SFdS* **159** 1–26. MR3803122

[29] Linder, T. (2002). Learning-theoretic methods in vector quantization. In *Principles of Nonparametric Learning* (*Udine*, 2001). *CISM Courses and Lect.* **434** 163–210. Springer, Vienna. MR1987659

[30] Lloyd, S. P. (1982). Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **28** 129–137. MR0651807 https://doi.org/10.1109/TIT.1982.1056489

[31] Maronna, R. A., Martin, R. D. and Yohai, V. J. (2006). *Robust Statistics*: *Theory and Methods. Wiley Series in Probability and Statistics*. Wiley, Chichester. MR2238141 https://doi.org/10.1002/0470010940

[32] Nielsen, F., Boissonnat, J. D. and Nock, R. (2007). Bregman Voronoi diagrams: Properties, algorithms and applications Technical Report No. 6154 INRIA.

[33] Rockafellar, R. T. (1970). *Convex Analysis. Princeton Mathematical Series, No.* 28. Princeton Univ. Press, Princeton, N.J. MR0274683

[34] Strehl, A. and Ghosh, J. (2003). Cluster ensembles—a knowledge reuse framework for combining multiple partitions *J. Mach. Learn. Res.* **3** 583–617. MR1991087 https://doi.org/10.1162/153244303321897735

[35] Tang, C. and Monteleoni, C. (2016). On Lloyd's algorithm: New theoretical insights for clustering in practice. In *Proceedings of the* 19*th International Conference on Artificial Intelligence and Statistics. PMLR* **51** 1280–1289.

[36] Taylor Arnold, L. T. (2015). *Humanities Data in R*: *Exploring Networks, Geospatial Data, Images, and Text*, 1st ed. *Quantitative Methods in the Humanities and Social Sciences*. Springer, Berlin.

[37] Vandev, D. (1993). A note on the breakdown point of the least median of squares and least trimmed squares estimators. *Statist. Probab. Lett.* **16** 117–119. MR1212043 https://doi.org/10.1016/0167-7152(93)90155-C

The Annals of Statistics
2021, Vol. 49, No. 3, 1702–1735
https://doi.org/10.1214/20-AOS2019
© Institute of Mathematical Statistics, 2021

LASSO-DRIVEN INFERENCE IN TIME AND SPACE

BY VICTOR CHERNOZHUKOV[1], WOLFGANG KARL HÄRDLE[2], CHEN HUANG[3] AND WEINING WANG[4]

[1]*Department of Economics and Operations Research Center, Massachusetts Institute of Technology, vchern@mit.edu*
[2]*IRTG1792, Humboldt–Universität zu Berlin, haerdle@hu-berlin.de*
[3]*Department of Economics and Business Economics and CREATES, Aarhus University, chen.huang@econ.au.dk*
[4]*Department of Economics and Related Studies, University of York, weining.wang@york.ac.uk*

We consider the estimation and inference in a system of high-dimensional regression equations allowing for temporal and cross-sectional dependency in covariates and error processes, covering rather general forms of weak temporal dependence. A sequence of regressions with many regressors using LASSO (Least Absolute Shrinkage and Selection Operator) is applied for variable selection purpose, and an overall penalty level is carefully chosen by a block multiplier bootstrap procedure to account for multiplicity of the equations and dependencies in the data. Correspondingly, oracle properties with a jointly selected tuning parameter are derived. We further provide high-quality de-biased simultaneous inference on the many target parameters of the system. We provide bootstrap consistency results of the test procedure, which are based on a general Bahadur representation for the Z-estimators with dependent data. Simulations demonstrate good performance of the proposed inference procedure. Finally, we apply the method to quantify spillover effects of textual sentiment indices in a financial market and to test the connectedness among sectors.

REFERENCES

ANDREWS, D. W. K. (1984). Nonstrong mixing autoregressive processes. *J. Appl. Probab.* **21** 930–934. MR0766830 https://doi.org/10.2307/3213710

ANTWEILER, W. and FRANK, M. Z. (2004). Is all that talk just noise? The information content of Internet stock message boards. *J. Finance* **59** 1259–1294.

AUDRINO, F. and TETEREVA, A. (2019). Sentiment spillover effects for us and European companies. *J. Bank. Financ.* **106** 542–567.

BAKER, M. and WURGLER, J. (2006). Investor sentiment and the cross-section of stock returns. *J. Finance* **61** 1645–1680.

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

BELLONI, A., CHEN, M. and CHERNOZHUKOV, V. (2016). Quantile graphical models: Prediction and conditional independence with applications to financial risk management. Preprint. Available at arXiv:1607.00286.

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

BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2011). Inference for high-dimensional sparse econometric models. Preprint. Available at arXiv:1201.0220.

BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Rev. Econ. Stud.* **81** 608–650. MR3207983 https://doi.org/10.1093/restud/rdt044

BELLONI, A., CHERNOZHUKOV, V. and KATO, K. (2015a). Supplement material for "Uniform post selection inference for least absolute deviation regression and other Z-estimation problems." Available at *Biometrika* online.

MSC2020 subject classifications. Primary 62M10, 62J99; secondary 62F40.
Key words and phrases. LASSO, time series, simultaneous inference, system of equations, Z-estimation, Bahadur representation, martingale decomposition.

BELLONI, A., CHERNOZHUKOV, V. and KATO, K. (2015b). Uniform post-selection inference for least absolute deviation regression and other Z-estimation problems. *Biometrika* **102** 77–94. MR3335097 https://doi.org/10.1093/biomet/asu056

BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. MR2533469 https://doi.org/10.1214/08-AOS620

CHEN, C. Y.-H., HÄRDLE, W. K. and OKHRIN, Y. (2019). Tail event driven networks of SIFIs. *J. Econometrics* **208** 282–298. MR3906972 https://doi.org/10.1016/j.jeconom.2018.09.016

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. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564–1597. MR3262461 https://doi.org/10.1214/14-AOS1230

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. and HANSEN, C. (2008). Instrumental variable quantile regression: A robust inference approach. *J. Econometrics* **142** 379–398. MR2408741 https://doi.org/10.1016/j.jeconom.2007.06.005

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

CHERNOZHUKOV, V., KARL HÄRDLE, W., HUANG, C. and WANG, W. (2021). Supplement to “LASSO-driven inference in time and space.” https://doi.org/10.1214/20-AOS2019SUPP.

DEZEURE, R., BÜHLMANN, P. and ZHANG, C.-H. (2017). High-dimensional simultaneous inference with the bootstrap. *TEST* **26** 685–719. MR3713586 https://doi.org/10.1007/s11749-017-0554-2

DIMITRAKOPOULOU, K., TSIMPOURIS, C., PAPADOPOULOS, G., POMMERENKE, C., WILK, E., SGARBAS, K. N., SCHUGHART, K. and BEZERIANOS, A. (2011). Dynamic gene network reconstruction from gene expression data in mice after influenza A (H1N1) infection. *J. Clin. Bioinformat.* **1** 27. https://doi.org/10.1186/2043-9113-1-27

EPSKAMP, S., WALDORP, L. J., MÕTTUS, R. and BORSBOOM, D. (2018). The Gaussian graphical model in cross-sectional and time-series data. *Multivar. Behav. Res.* **53** 453–480. https://doi.org/10.1080/00273171.2018.1454823

GARMAN, M. B. and KLASS, M. J. (1980). On the estimation of security price volatilities from historical data. *J. Bus.* **53** 67–78.

HÄRDLE, W. K., WANG, W. and YU, L. (2016). TENET: Tail-Event driven NETwork risk. *J. Econometrics* **192** 499–513. MR3488092 https://doi.org/10.1016/j.jeconom.2016.02.013

HÄRDLE, W. K., CHEN, S., LIANG, C. and SCHIENLE, M. (2018). Time-varying limit order book networks. IRTG 1792 Discussion Paper 2018-016, IRTG 1792, Humboldt Universität zu Berlin, Germany.

HAUTSCH, N., SCHAUMBURG, J. and SCHIENLE, M. (2015). Financial network systemic risk contributions. *Review of Finance* **19** 685–738.

HU, M. and LIU, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 168–177.

HUANG, D., YIN, J., SHI, T. and WANG, H. (2016). A statistical model for social network labeling. *J. Bus. Econom. Statist.* **34** 368–374. MR3523781 https://doi.org/10.1080/07350015.2015.1039014

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

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

KOLACZYK, E. D. and CSÁRDI, G. (2014). *Statistical Analysis of Network Data with R*. *Use R!* Springer, New York. MR3288852 https://doi.org/10.1007/978-1-4939-0983-4

KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. *Springer Series in Statistics*. Springer, New York. MR2724368 https://doi.org/10.1007/978-0-387-74978-5

KRAMPE, J. KREISS, J.-P. and PAPARODITIS, E. (2018). Bootstrap based inference for sparse high-dimensional time series models. Preprint. Available at arXiv:1806.11083.

LAHIRI, S. N. (1999). Theoretical comparisons of block bootstrap methods. *Ann. Statist.* **27** 386–404. MR1701117 https://doi.org/10.1214/aos/1018031117

LIN, J. and MICHAILIDIS, G. (2017). Regularized estimation and testing for high-dimensional multi-block vector-autoregressive models. *J. Mach. Learn. Res.* **18** Paper No. 117, 49. MR3725456 https://doi.org/10.1631/jzus.a1500279

LOUGHRAN, T. and MCDONALD, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *J. Finance* **66** 35–65.

LÜTKEPOHL, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer, Berlin. MR2172368 https://doi.org/10.1007/978-3-540-27752-1

MANRESA, E. (2013). Estimating the structure of social interactions using panel data. CEMFI, Madrid. Unpublished manuscript.

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

NEYKOV, M., NING, Y., LIU, J. S. and LIU, H. (2018). A unified theory of confidence regions and testing for high-dimensional estimating equations. *Statist. Sci.* **33** 427–443. MR3843384 https://doi.org/10.1214/18-STS661

OPGEN-RHEIN, R. and STRIMMER, K. (2007). From correlation to causation networks: A simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC Syst. Biol.* **1** 37.

PESARAN, M. H. and YAMAGATA, T. (2017). Testing for alpha in linear factor pricing models with a large number of securities. USC-INET Research Paper No. 17-13, USC Dornsife Institute for New Economic Thinking.

RAMIREZ, R. N., EL-ALI, N. C., MAGER, M. A., WYMAN, D., CONESA, A. and MORTAZAVI, A. (2017). Dynamic gene regulatory networks of human myeloid differentiation. *Cell Systems* **4** 416–429.

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

TETLOCK, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *J. Finance* **62** 1139–1168.

VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. and DEZEURE, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** 1166–1202. MR3224285 https://doi.org/10.1214/14-AOS1221

VAN DER VAART, A. W. 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

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

YUAN, M. and LIN, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94** 19–35. MR2367824 https://doi.org/10.1093/biomet/asm018

ZHANG, X. and CHENG, G. (2017). Simultaneous inference for high-dimensional linear models. *J. Amer. Statist. Assoc.* **112** 757–768. MR3671768 https://doi.org/10.1080/01621459.2016.1166114

ZHANG, D. and WU, W. B. (2017a). Gaussian approximation for high dimensional time series. *Ann. Statist.* **45** 1895–1919. MR3718156 https://doi.org/10.1214/16-AOS1512

ZHANG, D. and WU, W. B. (2017b). Supplement material for "Gaussian approximation for high dimensional time series. Available at *Ann. Statist.* online. https://doi.org/10.1214/16-AOS1512SUPP

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

ZHANG, J. L., HÄRDLE, W. K., CHEN, C. Y. and BOMMES, E. (2016). Distillation of news flow into analysis of stock reactions. *J. Bus. Econom. Statist.* **34** 547–563. MR3547995 https://doi.org/10.1080/07350015.2015.1110525

ZHU, Y. and BRADIC, J. (2018). Linear hypothesis testing in dense high-dimensional linear models. *J. Amer. Statist. Assoc.* **113** 1583–1600. MR3902231 https://doi.org/10.1080/01621459.2017.1356319

ZHU, X., PAN, R., LI, G., LIU, Y. and WANG, H. (2017). Network vector autoregression. *Ann. Statist.* **45** 1096–1123. MR3662449 https://doi.org/10.1214/16-AOS1476

ZHU, X., WANG, W., WANG, H. and HÄRDLE, W. K. (2019). Network quantile autoregression. *J. Econometrics* **212** 345–358. MR3994021 https://doi.org/10.1016/j.jeconom.2019.04.034

The Annals of Statistics
2021, Vol. 49, No. 3, 1736–1754
https://doi.org/10.1214/20-AOS2020
© Institute of Mathematical Statistics, 2021

E-VALUES: CALIBRATION, COMBINATION AND APPLICATIONS

BY VLADIMIR VOVK[1] AND RUODU WANG[2]

[1]*Department of Computer Science, Royal Holloway, University of London,* v.vovk@rhul.ac.uk

[2]*Department of Statistics and Actuarial Science, University of Waterloo,* wang@uwaterloo.ca

Multiple testing of a single hypothesis and testing multiple hypotheses are usually done in terms of p-values. In this paper, we replace p-values with their natural competitor, e-values, which are closely related to betting, Bayes factors and likelihood ratios. We demonstrate that e-values are often mathematically more tractable; in particular, in multiple testing of a single hypothesis, e-values can be merged simply by averaging them. This allows us to develop efficient procedures using e-values for testing multiple hypotheses.

REFERENCES

BENJAMIN, D. J. and BERGER, J. O. (2019). Three recommendations for improving the use of p-values. *Amer. Statist.* **73** 186–191. MR3925724 https://doi.org/10.1080/00031305.2018.1543135

DOBRIBAN, E. (2020). Fast closed testing for exchangeable local tests. *Biometrika* **107** 761–768. MR4138990 https://doi.org/10.1093/biomet/asz082

DUAN, B., RAMDAS, A., BALAKRISHNAN, S. and WASSERMAN, L. (2019). Interactive martingale tests for the global null. Technical report. Available at arXiv:1909.07339 [stat.ME].

EMBRECHTS, P., WANG, B. and WANG, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance Stoch.* **19** 763–790. MR3413935 https://doi.org/10.1007/s00780-015-0273-z

FISHER, R. A. (1932). *Statistical Methods for Research Workers*, 4th ed. Oliver and Boyd, Edinburgh. Section 21.1 on combining independent p-values first appears in this edition and is present in all subsequent editions.

GOEMAN, J. J. and SOLARI, A. (2011). Multiple testing for exploratory research. *Statist. Sci.* **26** 584–597. MR2951390 https://doi.org/10.1214/11-STS356

GRÜNWALD, P., DE HEIDE, R. and KOOLEN, W. M. (2020). Safe testing. Technical report. Available at arXiv:1906.07801 [math.ST].

HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. (1952). *Inequalities*, 2nd ed. Cambridge Univ. Press, Cambridge. MR0046395

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

HOWARD, S. R., RAMDAS, A., MCAULIFFE, J. and SEKHON, J. (2020). Time-uniform, nonparametric, nonasymptotic confidence sequences. Technical report. Available at arXiv:1810.08240 [math.ST]. *Ann. Statist.* To appear.

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

MARCUS, R., PERITZ, E. and GABRIEL, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* **63** 655–660. MR0468056 https://doi.org/10.1093/biomet/63.3.655

RÜGER, B. (1978). Das maximale Signifikanzniveau des Tests: "Lehne H_0 ab, wenn k unter n gegebenen Tests zur Ablehnung führen". *Metrika* **25** 171–178. MR0526476 https://doi.org/10.1007/BF02204362

SHAFER, G. (2019). The language of betting as a strategy for statistical and scientific communication. Technical report. Available at arXiv:1903.06991 [math.ST]. *J. R. Stat. Soc., A*. To appear.

SHAFER, G. and VOVK, V. (2019). *Game-Theoretic Foundations for Probability and Finance*. Wiley, Hoboken, NJ.

SHAFER, G., SHEN, A., VERESHCHAGIN, N. and VOVK, V. (2011). Test martingales, Bayes factors and p-values. *Statist. Sci.* **26** 84–101. MR2849911 https://doi.org/10.1214/10-STS347

SHAFFER, J. P. (1986). Modified sequentially rejective multiple test procedures. *J. Amer. Statist. Assoc.* **81** 826–831.

SIMES, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73** 751–754. MR0897872 https://doi.org/10.1093/biomet/73.3.751

MSC2020 subject classifications. Primary 62G10, 62F03; secondary 62C07, 62C15.

Key words and phrases. Hypothesis testing, multiple hypothesis testing, global null, Bayes factor, test martingale, admissible decisions.

VOVK, V. G. (1993). A logic of probability, with application to the foundations of statistics. *J. Roy. Statist. Soc. Ser. B* **55** 317–351. With discussion and a reply by the author. MR1224399

VOVK, V. and WANG, R. (2020). Combining p-values via averaging. Technical report. Available at arXiv:1212.4966 [math.ST]. *Biometrika*. To appear. https://doi.org/10.1093/biomet/asaa027.

VOVK, V. and WANG, R. (2021). Supplement to "E-values: Calibration, combination, and applications." https://doi.org/10.1214/20-AOS2020SUPP

WALD, A. (1950). *Statistical Decision Functions*. Wiley, New York, NY. MR0036976

WRIGHT, S. P. (1992). Adjusted p-values for simultaneous inference. *Biometrics* **48** 1005–1013.

The Annals of Statistics
2021, Vol. 49, No. 3, 1755–1778
https://doi.org/10.1214/20-AOS2021
© Institute of Mathematical Statistics, 2021

CAUSAL DISCOVERY IN HEAVY-TAILED MODELS

BY NICOLA GNECCO[1,*], NICOLAI MEINSHAUSEN[2], JONAS PETERS[3] AND SEBASTIAN ENGELKE[1,†]

[1]*Research Center for Statistics, University of Geneva,* [*]*nicola.gnecco@unige.ch;* [†]*sebastian.engelke@unige.ch*

[2]*Department of Mathematics, ETH Zurich, meinshausen@stat.math.ethz.ch*

[3]*Department of Mathematical Sciences, University of Copenhagen, jonas.peters@math.ku.dk*

Causal questions are omnipresent in many scientific problems. While much progress has been made in the analysis of causal relationships between random variables, these methods are not well suited if the causal mechanisms only manifest themselves in extremes. This work aims to connect the two fields of causal inference and extreme value theory. We define the causal tail coefficient that captures asymmetries in the extremal dependence of two random variables. In the population case, the causal tail coefficient is shown to reveal the causal structure if the distribution follows a linear structural causal model. This holds even in the presence of latent common causes that have the same tail index as the observed variables. Based on a consistent estimator of the causal tail coefficient, we propose a computationally highly efficient algorithm that estimates the causal structure. We prove that our method consistently recovers the causal order and we compare it to other well-established and nonextremal approaches in causal discovery on synthetic and real data. The code is available as an open-access `R` package.

REFERENCES

ASADI, P., DAVISON, A. C. and ENGELKE, S. (2015). Extremes on river networks. *Ann. Appl. Stat.* **9** 2023–2050. MR3456363 https://doi.org/10.1214/15-AOAS863

BASRAK, B. and SEGERS, J. (2009). Regularly varying multivariate time series. *Stochastic Process. Appl.* **119** 1055–1080. MR2508565 https://doi.org/10.1016/j.spa.2008.05.004

BOLLEN, K. A. (1989). *Structural Equations with Latent Variables. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Wiley, New York. MR0996025 https://doi.org/10.1002/9781118619179

BÜHLMANN, P., PETERS, J. and ERNEST, J. (2014). CAM: Causal additive models, high-dimensional order search and penalized regression. *Ann. Statist.* **42** 2526–2556. MR3277670 https://doi.org/10.1214/14-AOS1260

CAI, J.-J., EINMAHL, J. H. J., DE HAAN, L. and ZHOU, C. (2015). Estimation of the marginal expected shortfall: The mean when a related variable is extreme. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 417–442. MR3310533 https://doi.org/10.1111/rssb.12069

CLAASSEN, T., MOOIJ, J. and HESKES, T. (2013). Learning sparse causal models is not NP-hard. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence* (*UAI*) 172–181. ACM Press, New York.

COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values. Springer Series in Statistics*. Springer London, Ltd., London. MR1932132 https://doi.org/10.1007/978-1-4471-3675-0

COMON, P. (1994). Independent component analysis, a new concept? *Signal Process.* **36** 287–314.

COX, D. R. and WERMUTH, N. (1996). *Multivariate Dependencies: Models, Analysis and Interpretation. Monographs on Statistics and Applied Probability* **67**. CRC Press, London. MR1456990

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

EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (1997). *Modelling Extremal Events: For Insurance and Finance. Applications of Mathematics* (*New York*) **33**. Springer, Berlin. MR1458613 https://doi.org/10.1007/978-3-642-33483-2

MSC2020 subject classifications. Primary 62Hxx; secondary 62G08, 62G32.
Key words and phrases. Causality, extreme value theory, heavy-tailed distributions, nonparametric estimation.

ENGELKE, S. and HITZ, A. S. (2020). Graphical models for extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 871–932. MR4136498

ENGELKE, S. and IVANOVS, J. (2021). Sparse structures for multivariate extremes. *Annu. Rev. Stat. Appl.* **8** 241–270. https://doi.org/10.1146/annurev-statistics-040620-041554

ENTNER, D. and HOYER, P. O. (2010). Discovering unconfounded causal relationships using linear non-Gaussian models. In *Proceedings of the* 2010 *International Conference on New Frontiers in Artificial Intelligence* 181–195. Springer, Berlin.

FAWCETT, L. and WALSHAW, D. (2007). Improved estimation for temporally clustered extremes. *Environmetrics* **18** 173–188. MR2345653 https://doi.org/10.1002/env.810

FORBES, K. J. and RIGOBON, R. (2002). No contagion, only interdependence: Measuring stock market comovements. *J. Finance* **57** 2223–2261.

GISSIBL, N. and KLÜPPELBERG, C. (2018). Max-linear models on directed acyclic graphs. *Bernoulli* **24** 2693–2720. MR3779699 https://doi.org/10.3150/17-BEJ941

GISSIBL, N., KLÜPPELBERG, C. and LAURITZEN, S. (2020). Identifiability and estimation of recursive max-linear models. *Scand. J. Stat.* To appear. https://doi.org/0.1111/sjos.12446

GNECCO, N., MEINSHAUSEN, N., PETERS, J. and ENGELKE, S. (2021). Supplement to "Causal discovery in heavy-tailed models." https://doi.org/10.1214/20-AOS2021SUPP

GRANGER, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37** 424–438.

HARRIS, N. and DRTON, M. (2013). PC algorithm for nonparanormal graphical models. *J. Mach. Learn. Res.* **14** 3365–3383. MR3144465

HILL, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3** 1163–1174. MR0378204

HOYER, P. O., SHIMIZU, S., KERMINEN, A. J. and PALVIAINEN, M. (2008). Estimation of causal effects using linear non-Gaussian causal models with hidden variables. *Internat. J. Approx. Reason.* **49** 362–378. MR2460274 https://doi.org/10.1016/j.ijar.2008.02.006

HYVÄRINEN, A. and SMITH, S. M. (2013). Pairwise likelihood ratios for estimation of non-Gaussian structural equation models. *J. Mach. Learn. Res.* **14** 111–152. MR3033327

IMBENS, G. W. and RUBIN, D. B. (2015). *Causal Inference—for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge Univ. Press, New York. MR3309951 https://doi.org/10.1017/CBO9781139025751

KALISCH, M., MÄCHLER, M., COLOMBO, D., MAATHUIS, M. H. and BÜHLMANN, P. (2012). Causal inference using graphical models with the R package pcalg. *J. Stat. Softw.* **47** 1–26.

KOLMOGOROV, A. N. (1968). Three approaches to the quantitative definition of information. *Int. J. Comput. Math.* **2** 157–168. MR0243922 https://doi.org/10.1080/00207166808803030

LAURITZEN, S. L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. The Clarendon Press, New York. MR1419991

LAURITZEN, S. L., DAWID, A. P., LARSEN, B. N. and LEIMER, H.-G. (1990). Independence properties of directed Markov fields. *Networks* **20** 491–505. MR1064735 https://doi.org/10.1002/net.3230200503

MHALLA, L., CHAVEZ-DEMOULIN, V. and DUPUIS, D. J. (2020). Causal mechanism of extreme river discharges in the upper Danube basin network. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **69** 741–764. MR4133145 https://doi.org/doi.org/10.1111/rssc.12415

MISRA, N. and KURUOGLU, E. E. (2016). Stable graphical models. *J. Mach. Learn. Res.* **17** Paper No. 168, 1–36. MR3567436

NAVEAU, P., RIBES, A., ZWIERS, F., HANNART, A., TUEL, A. and YIOU, P. (2018). Revising return periods for record events in a climate event attribution context. *J. Climate* **31** 3411–3422.

NESTLÉ (2019). Financial Statements 2019. Online release 89.

NOVARTIS (2019). Annual Report 2019. Online release F-25–26.

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

PETERS, J. and BÜHLMANN, P. (2014). Identifiability of Gaussian structural equation models with equal error variances. *Biometrika* **101** 219–228. MR3180667 https://doi.org/10.1093/biomet/ast043

PETERS, J. and BÜHLMANN, P. (2015). Structural intervention distance for evaluating causal graphs. *Neural Comput.* **27** 771–799. MR3865898 https://doi.org/10.1162/neco_a_00708

PETERS, J., JANZING, D. and SCHÖLKOPF, B. (2017). *Elements of Causal Inference: Foundations and Learning Algorithms*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3822088

PETERS, J., JANZING, D., GRETTON, A. and SCHÖLKOPF, B. (2009). Detecting the direction of causal time series. In *Proceedings of the* 26*th International Conference on Machine Learning* (*ICML*) 801–808. ACM Press, New York.

PETERS, J., MOOIJ, J. M., JANZING, D. and SCHÖLKOPF, B. (2014). Causal discovery with continuous additive noise models. *J. Mach. Learn. Res.* **15** 2009–2053. MR3231600

RESNICK, S. I. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer, New York, NY.
ROCHE (2019). Finance Report 2019. Online release 160.
RODRIGUEZ, J. C. (2007). Measuring financial contagion: A copula approach. *J. Empir. Finance* **14** 401–423.
SENEVIRATNE, S. I., CORTI, T., DAVIN, E. L., HIRSCHI, M., JAEGER, E. B., LEHNER, I., ORLOWSKY, B. and TEULING, A. J. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Sci. Rev.* **99** 125–161.
SHIMIZU, S., HOYER, P. O., HYVÄRINEN, A. and KERMINEN, A. (2006). A linear non-Gaussian acyclic model for causal discovery. *J. Mach. Learn. Res.* **7** 2003–2030. MR2274431
SHIMIZU, S., INAZUMI, T., SOGAWA, Y., HYVÄRINEN, A., KAWAHARA, Y., WASHIO, T., HOYER, P. O. and BOLLEN, K. (2011). DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model. *J. Mach. Learn. Res.* **12** 1225–1248. MR2804599
SMITH, R. L. (1990). Regional estimation from spatially dependent data. Unpublished.
SPIRTES, P., GLYMOUR, C. and SCHEINES, R. (2000). *Causation, Prediction, and Search*, 2nd ed. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR1815675
TASHIRO, T., SHIMIZU, S., HYVÄRINEN, A. and WASHIO, T. (2014). ParceLiNGAM: A causal ordering method robust against latent confounders. *Neural Comput.* **26** 57–83. MR3155580 https://doi.org/10.1162/NECO_a_00533
WANG, Y. S. and DRTON, M. (2020). Causal discovery with unobserved confounding and non-gaussian data. Preprint. Available at arXiv:2007.11131.
ZHANG, J. (2008). Causal reasoning with ancestral graphs. *J. Mach. Learn. Res.* **9** 1437–1474. MR2426048
ZOU, N., VOLGUSHEV, S. and BÜCHER, A. (2021). Multiple block sizes and overlapping blocks for multivariate time series extremes. *Ann. Statist.* To appear.

The Annals of Statistics
2021, Vol. 49, No. 3, 1779–1809
https://doi.org/10.1214/20-AOS2022
© Institute of Mathematical Statistics, 2021

SUPERMIX: SPARSE REGULARIZATION FOR MIXTURES

BY Y. DE CASTRO[1], S. GADAT[2], C. MARTEAU[3] AND C. MAUGIS-RABUSSEAU[4]

[1]*Université Lyon, École Centrale de Lyon, Institut Camille Jordan, yohann.de-castro@ec-lyon.fr*

[2]*Toulouse School of Economics, Université Toulouse 1 Capitole, sebastien.gadat@tse-fr.eu*

[3]*Université Lyon, Université Claude Bernard Lyon 1, Institut Camille Jordan, marteau@math.univ-lyon1.fr*

[4]*Institut de Mathématiques de Toulouse, Université de Toulouse, maugis@insa-toulouse.fr*

This paper investigates the statistical estimation of a discrete mixing measure μ^0 involved in a kernel mixture model. Using some recent advances in ℓ_1-regularization over the space of measures, we introduce a "data fitting and regularization" convex program for estimating μ^0 in a grid-less manner from a sample of mixture law, this method is referred to as Beurling-LASSO.

Our contribution is two-fold: we derive a lower bound on the bandwidth of our data fitting term depending only on the support of μ^0 and its so-called "minimum separation" to ensure quantitative support localization error bounds; and under a so-called "nondegenerate source condition" we derive a nonasymptotic support stability property. This latter shows that for a sufficiently large sample size n, our estimator has exactly as many weighted Dirac masses as the target μ^0, converging in amplitude and localization towards the true ones. Finally, we also introduce some tractable algorithms for solving this convex program based on "Sliding Frank–Wolfe" or "Conic Particle Gradient Descent".

Statistical performances of this estimator are investigated designing a so-called "dual certificate", which is appropriate to our setting. Some classical situations as, for example, mixtures of super-smooth distributions (see, e.g., Gaussian distributions) or ordinary-smooth distributions (see, e.g., Laplace distributions), are discussed at the end of the paper.

REFERENCES

[1] AZAÏS, J.-M., DE CASTRO, Y. and GAMBOA, F. (2015). Spike detection from inaccurate samplings. *Appl. Comput. Harmon. Anal.* **38** 177–195.

[2] BALAKRISHNAN, S., WAINWRIGHT, M. J. and YU, B. (2017). Statistical guarantees for the EM algorithm: From population to sample-based analysis. *Ann. Statist.* **45** 77–120.

[3] BERLINET, A. and THOMAS-AGNAN, C. (2004). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer, New York.

[4] BHASKAR, B. N., TANG, G. and RECHT, B. (2012). Atomic norm denoising with applications to line spectral estimation Technical report, University of Wisconsin-Madison.

[5] BORDES, L., MOTTELET, S. and VANDEKERKHOVE, P. (2006). Semiparametric estimation of a two-component mixture model. *Ann. Statist.* **34** 1204–1232.

[6] BUNEA, F., TSYBAKOV, A. B., WEGKAMP, M. H. and BARBU, A. (2010). Spades and mixture models. *Ann. Statist.* **38** 2525–2558.

[7] BUTUCEA, C. and VANDEKERKHOVE, P. (2014). Semiparametric mixtures of symmetric distributions. *Scand. J. Stat.* **41** 227–239.

[8] CANDÈS, E. J. and FERNANDEZ-GRANDA, C. (2014). Towards a mathematical theory of super-resolution. *Comm. Pure Appl. Math.* **67** 906–956.

[9] CHIZAT, L. (2019). Sparse optimization on measures with over-parameterized gradient descent. arXiv preprint. Available at arXiv:1907.10300.

[10] DE CASTRO, Y., GADAT, S., MARTEAU, C. and MAUGIS-RABUSSEAU, C. (2021). Supplement to "SuperMix: Sparse Regularization for Mixtures." https://doi.org/10.1214/20-AOS2022SUPP

[11] DE CASTRO, Y. and GAMBOA, F. (2012). Exact reconstruction using Beurling minimal extrapolation. *J. Math. Anal. Appl.* **395** 336–354.

MSC2020 subject classifications. Primary 62G05, 90C25; secondary 49M29.

Key words and phrases. Beurling Lasso, mixture recovery, dual certificate, kernel approach, super-resolution.

[12] De Castro, Y., Gamboa, F., Henrion, D. and Lasserre, J.-B. (2017). Exact solutions to super resolution on semi-algebraic domains in higher dimensions. *IEEE Trans. Inf. Theory* **63** 621–630.

[13] Denoyelle, Q., Duval, V., Peyré, G. and Soubies, E. (2019). The sliding Frank–Wolfe algorithm and its application to super-resolution microscopy. *Inverse Probl.*.

[14] Duval, V. and Peyré, G. (2015). Exact support recovery for sparse spikes deconvolution. *Found. Comput. Math.* 1–41.

[15] Dwivedi, R., Ho, N., Khamaru, K., Jordan, M. I., Wainwright, M. J. and Yu, B. (2018). Singularity, Misspecification, and the Convergence Rate of EM. arXiv preprint. Available at arXiv:1810.00828.

[16] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272.

[17] Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer, Berlin.

[18] Gadat, S., Kahn, J., Marteau, C. and Maugis-Rabusseau, C. (2020). Parameter recovery in two-component contamination mixtures: The L2 strategy. *Ann. Inst. Henri Poincaré* **56** 1391–1418.

[19] Heinrich, P. and Kahn, J. (2018). Strong identifiability and optimal minimax rates for finite mixture estimation. *Ann. Statist.* **46** 2844–2870.

[20] Le Cam, L. (1990). Maximum likelihood: An introduction. *Int. Stat. Rev.* **58** 153–171.

[21] Maugis-Rabusseau, C. and Michel, B. (2011). A non asymptotic penalized criterion for Gaussian mixture model selection. *ESAIM Probab. Stat.* **15** 41–68.

[22] McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. *Wiley Series in Probability and Statistics*.

[23] Meister, A. (2009). *Deconvolution Problems in Nonparametric Statistics*. *Lecture Notes in Statistics* **193**. Springer, Berlin.

[24] Nguyen, X. (2013). Convergence of latent mixing measures in finite and infinite mixture models. *Ann. Statist.* **41** 370–400.

[25] Poon, C., Keriven, N. and Peyré, G. (2019). Support localization and the Fisher metric for off-the-grid sparse regularization. In *The 22nd International Conference on Artificial Intelligence and Statistics* 1341–1350.

[26] Rudin, W. (1991). *Functional Analysis*. *International Series in Pure and Applied Mathematics*. McGraw-Hill, Inc., New York.

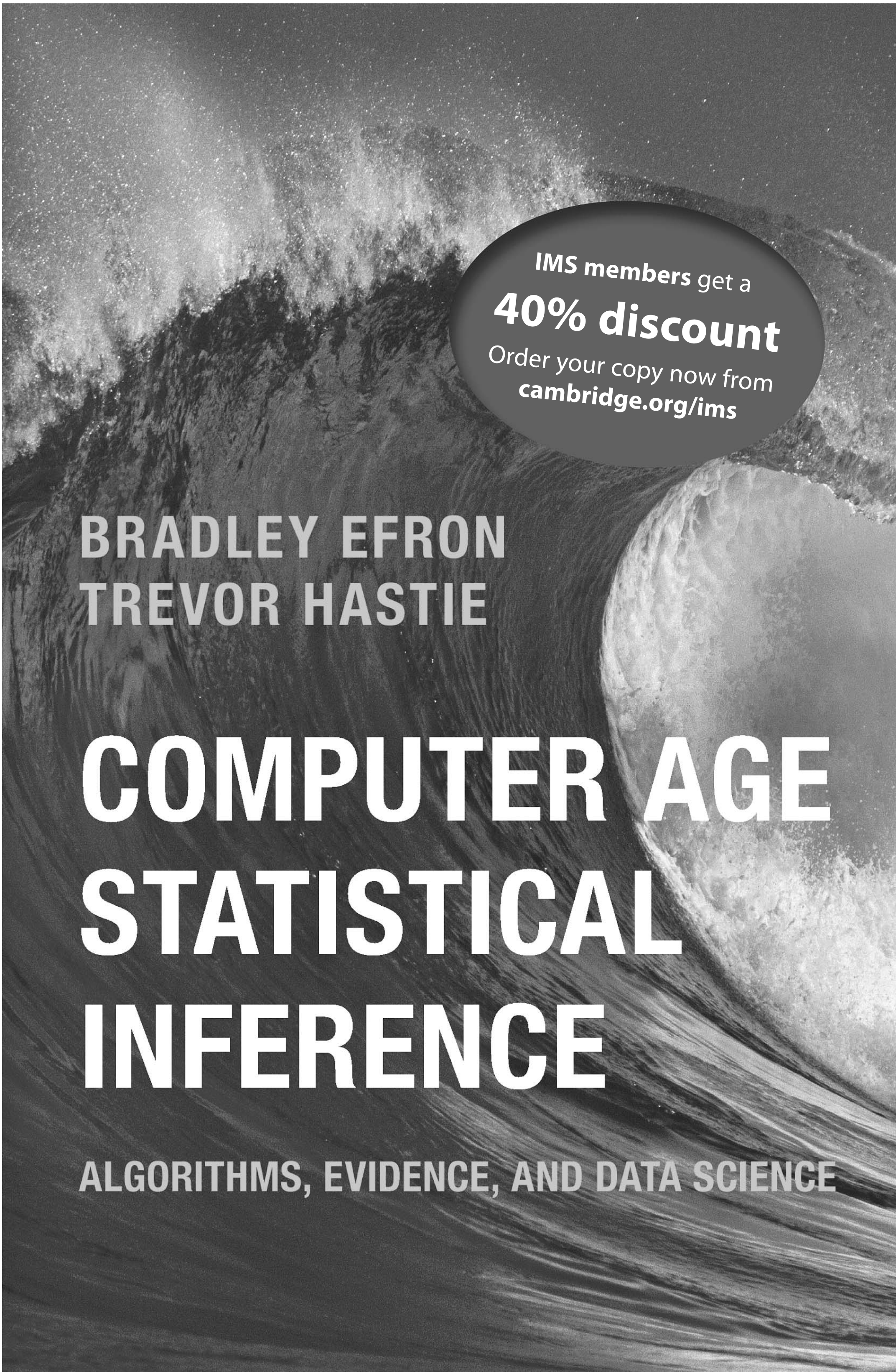
[27] Sriperumbudur, B. K., Fukumizu, K. and Lanckriet, G. R. (2011). Universality, characteristic kernels and RKHS embedding of measures. *J. Mach. Learn. Res.* **12** 2389–2410.

[28] van der Vaart, A. W. (1996). Efficient maximum likelihood estimation in semiparametric mixture models. *Ann. Statist.* **24** 862–878.

[29] Wahba, G. (1990). *Spline Models for Observational Data*. SIAM publisher: SIAM, Philadelphia, PA.

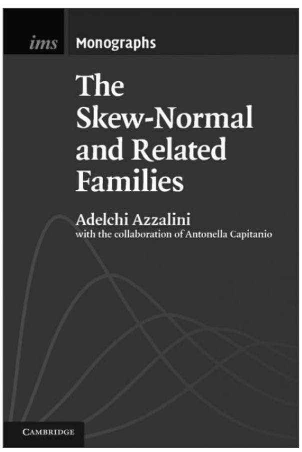
[30] Wu, C. J. (1983). On the convergence properties of the em algorithm. *Ann. Statist.* **11** 95–103.

[31] Xu, J., Hsu, D. J. and Maleki, A. (2016). Global analysis of expectation maximization for mixtures of two Gaussians. In *Advances in Neural Information Processing Systems* 2676–2684.


IMS members get a
40% discount
Order your copy now from
cambridge.org/ims
BRADLEY EFRON
TREVOR HASTIE
COMPUTER AGE
STATISTICAL
INFERENCE
ALGORITHMS, EVIDENCE, AND DATA SCIENCE

The Institute of Mathematical Statistics presents

IMS MONOGRAPHS



The Skew-Normal and Related Families

Adelchi Azzalini
in collaboration with Antonella Capitanio

Interest in the skew-normal and related families of distributions has grown enormously over recent years, as theory has advanced, challenges of data have grown, and computational tools have made substantial progress. This comprehensive treatment, blending theory and practice, will be the standard resource for statisticians and applied researchers. Assuming only basic knowledge of (non-measure-theoretic) probability and statistical inference, the book is accessible to the wide range of researchers who use statistical modelling techniques. Guiding readers through the main concepts and results, it covers both the probability and the statistics sides of the subject, in the univariate and multivariate settings. The theoretical development is complemented by numerous illustrations and applications to a range of fields including quantitative finance, medical statistics, environmental risk studies, and industrial and business efficiency.

The author's freely available R package sn, available from CRAN, equips readers to put the methods into action with their own data.

IMS member? Claim your 40% discount:
www.cambridge.org/ims

Hardback price US$48.00
(non-member price $80.00)

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.