

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

- A new scope of penalized empirical likelihood with high-dimensional estimating equations JINYUAN CHANG, CHENG YONG TANG AND TONG TONG WU 3185
- Approximate ℓ_0 -penalized estimation of piecewise-constant signals on graphs
ZHOU FAN AND LEYING GUAN 3217
- Multiclass classification, information, divergence and surrogate risk
JOHN DUCHI, KHASHAYAR KHOSRAVI AND FENG RUAN 3246
- Halfspace depths for scatter, concentration and shape matrices
DAVY PAINDAVEINE AND GERMAIN VAN BEVER 3276
- Multilayer tensor factorization with applications to recommender systems
XUAN BI, ANNIE QU AND XIAOTONG SHEN 3308
- Principal component analysis for functional data on Riemannian manifolds and spheres
XIONGTAO DAI AND HANS-GEORG MÜLLER 3334
- Assessing robustness of classification using an angular breakdown point
JUNLONG ZHAO, GUAN YU AND YUFENG LIU 3362
- Tail-greedy bottom-up data decompositions and fast multiple change-point detection
PIOTR FRYZLEWICZ 3390
- ROCKET: Robust confidence intervals via Kendall's tau for transelliptical graphical models RINA FOYGEL BARBER AND MLADEN KOLAR 3422
- Adaptive invariant density estimation for ergodic diffusions over anisotropic classes
CLAUDIA STRAUCH 3451
- Robust low-rank matrix estimation ANDREAS ELSENER AND SARA VAN DE GEER 3481
- Sieve bootstrap for functional time series EFSTATHIOS PAPARODITIS 3510
- Restricted strong convexity implies weak submodularity ETHAN R. ELENBERG,
RAJIV KHANNA, ALEXANDROS G. DIMAKIS AND SAHAND NEGAHBAN 3539
- Multiscale scanning in inverse problems
KATHARINA PROKSCH, FRANK WERNER AND AXEL MUNK 3569
- Slope meets Lasso: Improved oracle bounds and optimality
PIERRE C. BELLEC, GUILLAUME LECUÉ AND ALEXANDRE B. TSYBAKOV 3603
- Uniformly valid post-regularization confidence regions for many functional parameters in z-estimation framework ALEXANDRE BELLONI, VICTOR CHERNOZHUKOV,
DENIS CHETVERIKOV AND YING WEI 3643
- Local asymptotic equivalence of pure states ensembles and quantum Gaussian white noise
CRISTINA BUTUCEA, MĂDĂLIN GUȚĂ AND MICHAEL NUSSBAUM 3676
- Extremal quantile treatment effects YICHONG ZHANG 3707
- Optimal maximin L_1 -distance Latin hypercube designs based on good lattice point designs LIN WANG, QIAN XIAO AND HONGQUAN XU 3741
- Rho-estimators revisited: General theory and applications
YANNICK BARAUD AND LUCIEN BIRGÉ 3767
- Think globally, fit locally under the manifold setup: Asymptotic analysis of locally linear embedding HAU-TIENG WU AND NAN WU 3805
- Nonparametric covariate-adjusted response-adaptive design based on a functional urn model GIACOMO ALETTI, ANDREA GHIGLIETTI
AND WILLIAM F. ROSENBERGER 3838

THE ANNALS OF STATISTICS

Vol. 46, No. 6B, pp. 3185–3866 December 2018

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: Xiao-Li Meng, Department of Statistics, Harvard University, Cambridge, Massachusetts 02138-2901, USA

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

Past President: Alison Etheridge, Department of Statistics, University of Oxford, Oxford, OX1 3LB, United Kingdom

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:* Edward I. George, Department of Statistics, University of Pennsylvania, Philadelphia, PA 19104, USA; Tailen Hsing, Department of Statistics, University of Michigan, Ann Arbor, MI 48109-1107 USA

The Annals of Applied Statistics. *Editor-in-Chief:* Tilmann Gneiting, Heidelberg Institute for Theoretical Studies, HITS gGmbH, Schloss-Wolfsbrunnengasse 35, 69118 Heidelberg, Germany

The Annals of Probability. *Editor:* Amir Dembo, Department of Statistics and Department of Mathematics, Stanford University, Stanford, California 94305, USA

The Annals of Applied Probability. *Editor:* Bálint Tóth, School of Mathematics, University of Bristol, University Walk, BS8 1TW, Bristol, UK and Alfréd Rényi, Institute of Mathematics, Hungarian Academy of Sciences, Budapest, Hungary

Statistical Science. *Editor:* Cun-Hui Zhang, Department of Statistics, Rutgers University, Piscataway, NJ 08854, USA

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 46, Number 6B, December 2018. Published bimonthly by the Institute of Mathematical Statistics, 3163 Somerset Drive, Cleveland, Ohio 44122, 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, 9650 Rockville Pike, Suite L 2310, Bethesda, Maryland 20814-3998, USA.

A NEW SCOPE OF PENALIZED EMPIRICAL LIKELIHOOD WITH HIGH-DIMENSIONAL ESTIMATING EQUATIONS

BY JINYUAN CHANG¹, CHENG YONG TANG² AND TONG TONG WU³

*Southwestern University of Finance and Economics, Temple University and
 University of Rochester*

Statistical methods with empirical likelihood (EL) are appealing and effective especially in conjunction with estimating equations for flexibly and adaptively incorporating data information. It is known that EL approaches encounter difficulties when dealing with high-dimensional problems. To overcome the challenges, we begin our study with investigating high-dimensional EL from a new scope targeting at high-dimensional sparse model parameters. We show that the new scope provides an opportunity for relaxing the stringent requirement on the dimensionality of the model parameters. Motivated by the new scope, we then propose a new penalized EL by applying two penalty functions respectively regularizing the model parameters and the associated Lagrange multiplier in the optimizations of EL. By penalizing the Lagrange multiplier to encourage its sparsity, a drastic dimension reduction in the number of estimating equations can be achieved. Most attractively, such a reduction in dimensionality of estimating equations can be viewed as a selection among those high-dimensional estimating equations, resulting in a highly parsimonious and effective device for estimating high-dimensional sparse model parameters. Allowing both the dimensionalities of model parameters and estimating equations growing exponentially with the sample size, our theory demonstrates that our new penalized EL estimator is sparse and consistent with asymptotically normally distributed nonzero components. Numerical simulations and a real data analysis show that the proposed penalized EL works promisingly.

REFERENCES

- BARTOLUCCI, F. (2007). A penalized version of the empirical likelihood ratio for the population mean. *Statist. Probab. Lett.* **77** 104–110. [MR2339024](#)
- CANDES, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n . *Ann. Statist.* **35** 2313–2351. [MR2382644](#)
- CHANG, J., CHEN, S. X. and CHEN, X. (2015). High dimensional generalized empirical likelihood for moment restrictions with dependent data. *J. Econometrics* **185** 283–304. [MR3300347](#)
- CHANG, J., TANG, C. Y. and WU, Y. (2013). Marginal empirical likelihood and sure independence feature screening. *Ann. Statist.* **41** 2123–2148. [MR3127860](#)
- CHANG, J., TANG, C. Y. and WU, Y. (2016). Local independence feature screening for nonparametric and semiparametric models by marginal empirical likelihood. *Ann. Statist.* **44** 515–539. [MR3476608](#)

MSC2010 subject classifications. Primary 62G99; secondary 62F40.

Key words and phrases. Empirical likelihood, estimating equations, high-dimensional statistical methods, moment selection, penalized likelihood.

- CHANG, J., TANG, C. Y. and WU, T. T. (2018). Supplement to “A new scope of penalized empirical likelihood with high-dimensional estimating equations.” DOI:[10.1214/17-AOS1655SUPP](https://doi.org/10.1214/17-AOS1655SUPP).
- CHEN, X. (2007). Large sample sieve estimation of semi-nonparametric models. In *The Handbook of Econometrics*, 6B (J. J. Heckman and E. Leamer, eds.). North- Holland, Amsterdam.
- CHEN, J. and CHEN, Z. (2008). Extended Bayesian information criterion for model selection with large model space. *Biometrika* **95** 759–771.
- CHEN, S. X. and CUI, H. (2006). On Bartlett correction of empirical likelihood in the presence of nuisance parameters. *Biometrika* **93** 215–220.
- CHEN, S. X. and CUI, H. (2007). On the second-order properties of empirical likelihood with moment restrictions. *J. Econometrics* **141** 492–516. [MR2413478](#)
- CHEN, S. X., PENG, L. and QIN, Y. L. (2009). Effects of data dimension on empirical likelihood. *Biometrika* **96** 711–722.
- CHEN, X. and POUZO (2012). Sieve quasi likelihood ratio inference on semi/nonparametric conditional moment models. *Econometrica* **80** 277–321.
- CHENG, X. and LIAO, Z. (2015). Select the valid and relevant moments: An information-based LASSO for GMM with many moments. *J. Econometrics* **186** 443–464. [MR3343796](#)
- FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360.
- FRIEDMAN, J., HASTIE, T., HOEFLING, H. and TIBSHIRANI, R. (2007). Pathwise coordinate optimization. *Ann. Appl. Stat.* **2** 302–332. [MR2415737](#)
- GAUTIER, E. and TSYBAKOV, A. B. (2014). High-dimensional instrumental variables regression and confidence sets. Manuscript, [arXiv:1105.2454v4](https://arxiv.org/abs/1105.2454v4).
- HANSEN, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50** 1029–1054.
- HJORT, N. L., MCKEAGUE, I. W. and VAN KEILEGOM, I. (2009). Extending the scope of empirical likelihood. *Ann. Statist.* **37** 1079–1111. [MR2509068](#)
- LAHIRI, S. N. and MUKHOPADHYAY, S. (2012). A penalized empirical likelihood method in high dimensions. *Ann. Statist.* **40** 2511–2540. [MR3097611](#)
- LENG, C. and TANG, C. Y. (2012). Penalized empirical likelihood and growing dimensional general estimating equations. *Biometrika* **99** 703–716. [MR2966779](#)
- LIANG, K. Y. and ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73** 13–22. [MR0836430](#)
- LV, J. and FAN, Y. (2009). A unified approach to model selection and sparse recovery using regularized least squares. *Ann. Statist.* **37** 3498–3528. [MR2549567](#)
- NEWWEY, W. K. and SMITH, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* **72** 219–255.
- OWEN, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75** 237–249.
- OWEN, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18** 90–120. [MR1041387](#)
- OWEN, A. (2001). *Empirical Likelihood*. Chapman & Hall-CRC, New York.
- PETROV, V. V. (1995). *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*. Oxford Univ. Press, Oxford.
- QIN, J. and LAWLESS, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22** 300–325.
- QU, A., LINDSAY, B. G. and LI, B. (2000). Improving estimating equations using quadratic inference functions. *Biometrika* **87** 823–836.
- RUDIN, W. (1976). *Principles of Mathematical Analysis*. McGraw-Hill, New York.
- SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464. [MR0468014](#)
- SHI, Z. (2016). Econometric estimation with high-dimensional moment equalities. *J. Econometrics* **195** 104–119.

- TANG, C. Y. and LENG, C. (2010). Penalized high dimensional empirical likelihood. *Biometrika* **97** 905–920.
- TANG, C. Y. and WU, T. T. (2014). Nested coordinate descent algorithms for empirical likelihood. *J. Stat. Comput. Simul.* **84** 1917–1930. [MR3215732](#)
- TSAO, M. (2004). Bounds on coverage probabilities of the empirical likelihood ratio confidence regions. *Ann. Statist.* **32** 1215–1221. [MR2065203](#)
- TSAO, M. and WU, F. (2013). Empirical likelihood on the full parameter space. *Ann. Statist.* **41** 2176–2196. [MR3127862](#)
- TSAO, M. and WU, F. (2014). Extended empirical likelihood for estimating equations. *Biometrika* **101** 703–710. [MR3254910](#)
- WANG, H., LI, B. and LENG, C. (2009). Shrinkage tuning parameter selection with a diverging number of parameters. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 671–683.
- WU, T. T. and LANGE, K. (2008). Coordinate descent algorithms for lasso penalized regression. *Ann. Appl. Stat.* **2** 224–244. [MR2415601](#)
- ZHANG, C. H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. [MR2604701](#)
- ZHAO, P. and YU, B. (2006). On model selection consistency of Lasso. *J. Mach. Learn. Res.* **7** 2541–2563. [MR2274449](#)

APPROXIMATE ℓ_0 -PENALIZED ESTIMATION OF PIECEWISE-CONSTANT SIGNALS ON GRAPHS

BY ZHOU FAN¹ AND LEYING GUAN

Stanford University

We study recovery of piecewise-constant signals on graphs by the estimator minimizing an ℓ_0 -edge-penalized objective. Although exact minimization of this objective may be computationally intractable, we show that the same statistical risk guarantees are achieved by the α -expansion algorithm which computes an approximate minimizer in polynomial time. We establish that for graphs with small average vertex degree, these guarantees are minimax rate-optimal over classes of edge-sparse signals. For spatially inhomogeneous graphs, we propose minimization of an edge-weighted objective where each edge is weighted by its effective resistance or another measure of its contribution to the graph's connectivity. We establish minimax optimality of the resulting estimators over corresponding edge-weighted sparsity classes. We show theoretically that these risk guarantees are not always achieved by the estimator minimizing the ℓ_1 /total-variation relaxation, and empirically that the ℓ_0 -based estimates are more accurate in high signal-to-noise settings.

REFERENCES

- [1] ADDARIO-BERRY, L., BROUTIN, N., DEVROYE, L. and LUGOSI, G. (2010). On combinatorial testing problems. *Ann. Statist.* **38** 3063–3092. [MR2722464](#)
- [2] ARIAS-CASTRO, E., CANDÈS, E. J. and DURAND, A. (2011). Detection of an anomalous cluster in a network. *Ann. Statist.* **39** 278–304. [MR2797847](#)
- [3] ARIAS-CASTRO, E., CANDÈS, E. J., HELGASON, H. and ZEITOUNI, O. (2008). Searching for a trail of evidence in a maze. *Ann. Statist.* **36** 1726–1757. [MR2435454](#)
- [4] ARIAS-CASTRO, E., DONOHO, D. L. and HUO, X. (2005). Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Inform. Theory* **51** 2402–2425.
- [5] ARIAS-CASTRO, E. and GRIMMETT, G. R. (2013). Cluster detection in networks using percolation. *Bernoulli* **19** 676–719. [MR3037169](#)
- [6] AUGER, I. E. and LAWRENCE, C. E. (1989). Algorithms for the optimal identification of segment neighborhoods. *Bull. Math. Biol.* **51** 39–54. [MR0978902](#)
- [7] BARRON, A., BIRGÉ, L. and MASSART, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113** 301–413.
- [8] BARRY, D. and HARTIGAN, J. A. (1993). A Bayesian analysis for change point problems. *J. Amer. Statist. Assoc.* **88** 309–319.
- [9] BESAG, J. (1986). On the statistical analysis of dirty pictures. *J. Roy. Statist. Soc. Ser. B* **48** 259–302. [MR0876840](#)
- [10] BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of Lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. [MR2533469](#)

MSC2010 subject classifications. 62G05.

Key words and phrases. Approximation algorithm, graph cut, effective resistance, total-variation denoising.

- [11] BIRGÉ, L. and MASSART, P. (2007). Minimal penalties for Gaussian model selection. *Probab. Theory Related Fields* **138** 33–73.
- [12] BOYKOV, Y. and KOLMOGOROV, V. (2004). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Intell.* **26** 1124–1137.
- [13] BOYKOV, Y., VEKSLER, O. and ZABIH, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* **23** 1222–1239.
- [14] BOYSEN, L., KEMPE, A., LIEBSCHER, V., MUNK, A. and WITTICH, O. (2009). Consistencies and rates of convergence of jump-penalized least squares estimators. *Ann. Statist.* **37** 157–183. [MR2488348](#)
- [15] CHAMBOLLE, A. (2005). Total variation minimization and a class of binary MRF models. In *EMMCVPR 2005* 136–152. Springer, Berlin.
- [16] CHAMBOLLE, A. and LIONS, P.-L. (1997). Image recovery via total variation minimization and related problems. *Numer. Math.* **76** 167–188. [MR1440119](#)
- [17] CHEN, S. S., DONOHO, D. L. and SAUNDERS, M. A. (2001). Atomic decomposition by basis pursuit. *SIAM Rev.* **43** 129–159.
- [18] CHERNOFF, H. and ZACKS, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *Ann. Math. Stat.* **35** 999–1018.
- [19] DALALYAN, A. S., HEBIRI, M. and LEDERER, J. (2017). On the prediction performance of the Lasso. *Bernoulli* **23** 552–581. [MR3556784](#)
- [20] DARBON, J. and SIGELLE, M. (2005). A fast and exact algorithm for total variation minimization. In *Iberian Conference on Pattern Recognition and Image Analysis* 351–359. Springer, Berlin.
- [21] DAVIES, P. L. and KOVAC, A. (2001). Local extremes, runs, strings and multiresolution. *Ann. Statist.* **29** 1–65. [MR1833958](#)
- [22] DONOHO, D. L. (1999). Wedgelets: Nearly minimax estimation of edges. *Ann. Statist.* **27** 859–897. [MR1724034](#)
- [23] DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455.
- [24] EFRON, B. (2004). Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. *J. Amer. Statist. Assoc.* **99** 96–104.
- [25] FAN, Z. and GUAN, L. (2018). Supplement to “Approximate ℓ_0 -penalized estimation of piecewise-constant signals on graphs.” DOI:[10.1214/17-AOS1656SUPP](#).
- [26] GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **6** 721–741.
- [27] GHOSH, A., BOYD, S. and SABERI, A. (2008). Minimizing effective resistance of a graph. *SIAM Rev.* **50** 37–66. [MR2403057](#)
- [28] GOLDSTEIN, T. and OSHER, S. (2009). The split Bregman method for L1-regularized problems. *SIAM J. Imaging Sci.* **2** 323–343.
- [29] GREIG, D. M., PORTEOUS, B. T. and SEHEULT, A. H. (1989). Exact maximum a posteriori estimation for binary images. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **51** 271–279.
- [30] GUNTUBOYINA, A., LIEU, D., CHATTERJEE, S. and SEN, B. (2017). Spatial adaptation in trend filtering. Available at [arXiv:1702.05113](#).
- [31] HARCHAoui, Z. and LÉVY-LEDUC, C. (2010). Multiple change-point estimation with a total variation penalty. *J. Amer. Statist. Assoc.* **105** 1480–1493. [MR2796565](#)
- [32] HARRIS, X. T. (2016). Prediction error after model search. Available at [arXiv:1610.06107](#).
- [33] HOEFLING, H. (2010). A path algorithm for the fused lasso signal approximator. *J. Comput. Graph. Statist.* **19** 984–1006. Supplementary materials available online. [MR2791265](#)
- [34] HÜTTER, J.-C. and RIGOLLET, P. (2016). Optimal rates for total variation denoising. In *Conf. Learning Theory* 1115–1146.

- [35] JACKSON, B., SCARGLE, J. D. et al. (2005). An algorithm for optimal partitioning of data on an interval. *IEEE Signal Process. Lett.* **12** 105–108.
- [36] JOHNSTONE, I. (2015). *Gaussian Estimation: Sequence and Wavelet Models*. Available at statweb.stanford.edu/~imj/GE09-08-15.pdf.
- [37] KARGER, D. R. and STEIN, C. (1996). A new approach to the minimum cut problem. *J. ACM* **43** 601–640.
- [38] KILLICK, R., FEARNHEAD, P. and ECKLEY, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *J. Amer. Statist. Assoc.* **107** 1590–1598. [MR3036418](#)
- [39] KOLMOGOROV, V. and ZABIN, R. (2004). What energy functions can be minimized via graph cuts? *IEEE Trans. Pattern Anal. Mach. Intell.* **26** 147–159.
- [40] KOROSTELEV, A. P. and TSYBAKOV, A. B. (1993). *Minimax Theory of Image Reconstruction. Lecture Notes in Statistics* **82**. Springer, New York.
- [41] KOVAC, A. and SMITH, A. D. (2011). Nonparametric regression on a graph. *J. Comput. Graph. Statist.* **20** 432–447.
- [42] LAND, S. R. and FRIEDMAN, J. H. (1997). Variable fusion: A new adaptive signal regression method. Technical Report 656, Dept. Statistics, Carnegie Mellon Univ., Pittsburgh, PA.
- [43] LEBARBIER, É. (2005). Detecting multiple change-points in the mean of Gaussian process by model selection. *Signal Process.* **85** 717–736.
- [44] LIN, K., SHARPNACK, J., RINALDO, A. and TIBSHIRANI, R. J. (2016). Approximate recovery in changepoint problems, from ℓ_2 estimation error rates. Available at [arXiv:1606.06746](https://arxiv.org/abs/1606.06746).
- [45] LIVNE, O. E. and BRANDT, A. (2012). Lean algebraic multigrid (LAMG): Fast graph Laplacian linear solver. *SIAM J. Sci. Comput.* **34** B499–B522. [MR2970416](#)
- [46] LOVÁSZ, L. (1996). Random walks on graphs: A survey. In *Combinatorics: Paul Erdős Is Eighty, Vol. 2 (Keszthely, 1993)*. *Bolyai Soc. Math. Stud.* **2** 353–397. János Bolyai Math. Soc., Budapest. [MR1395866](#)
- [47] MADRID PADILLA, O. H., SCOTT, J. G., SHARPNACK, J. and TIBSHIRANI, R. J. (2016). The DFS fused lasso: Nearly optimal linear-time denoising over graphs and trees. Available at [arXiv:1608.03384](https://arxiv.org/abs/1608.03384).
- [48] MAMMEN, E. and VAN DE GEER, S. (1997). Locally adaptive regression splines. *Ann. Statist.* **25** 387–413. [MR1429931](#)
- [49] MOORE, C. and NEWMAN, M. E. (2000). Epidemics and percolation in small-world networks. *Phys. Rev. E* **61** 5678–5682.
- [50] MUMFORD, D. and SHAH, J. (1989). Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.* **42** 577–685.
- [51] RINALDO, A. (2009). Properties and refinements of the fused lasso. *Ann. Statist.* **37** 2922–2952. [MR2541451](#)
- [52] RUDIN, L. I., OSHER, S. and FATEMI, E. (1992). Nonlinear total variation based noise removal algorithms. *Phys. D* **60** 259–268. Experimental mathematics: Computational issues in nonlinear science (Los Alamos, NM, 1991). [MR3363401](#)
- [53] SADHANALA, V., WANG, Y.-X. and TIBSHIRANI, R. (2016). Graph sparsification approaches for Laplacian smoothing. In *Int. Conf. Artif. Intell. Statist.* 1250–1259.
- [54] SADHANALA, V., WANG, Y.-X. and TIBSHIRANI, R. J. (2016). Total variation classes beyond $1d$: Minimax rates, and the limitations of linear smoothers. In *Adv. Neural Inform. Process. Syst.* 3513–3521.
- [55] SHARPNACK, J., RINALDO, A. and SINGH, A. (2012). Sparsistency of the edge lasso over graphs. In *Int. Conf. Artif. Intell. Statist.* 1028–1036.
- [56] SHARPNACK, J., SINGH, A. and RINALDO, A. (2013). Detecting activations over graphs using spanning tree wavelet bases. In *Int. Conf. Artif. Intell. Statist.* 545–553.

- [57] SHARPNACK, J. L., KRISHNAMURTHY, A. and SINGH, A. (2013). Near-optimal anomaly detection in graphs using Lovasz extended scan statistic. In *Adv. Neural Inform. Process. Syst.* 1959–1967.
- [58] SPIELMAN, D. A. and SRIVASTAVA, N. (2011). Graph sparsification by effective resistances. *SIAM J. Comput.* **40** 1913–1926.
- [59] SPIELMAN, D. A. and TENG, S.-H. (2004). Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *ACM Symp. Theory Comput.* 81–90. ACM, New York.
- [60] TANSEY, W. and SCOTT, J. G. (2015). A fast and flexible algorithm for the graph-fused lasso. Available at [arXiv:1505.06475](https://arxiv.org/abs/1505.06475).
- [61] TIAN, X. and TAYLOR, J. E. (2015). Selective inference with a randomized response. Available at [arXiv:1507.06739](https://arxiv.org/abs/1507.06739).
- [62] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **58** 267–288.
- [63] TIBSHIRANI, R., SAUNDERS, M., ROSSET, S., ZHU, J. and KNIGHT, K. (2005). Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 91–108. [MR2136641](https://arxiv.org/abs/2136641)
- [64] TIBSHIRANI, R. J. and TAYLOR, J. (2011). The solution path of the generalized lasso. *Ann. Statist.* **39** 1335–1371. [MR2850205](https://arxiv.org/abs/2850205)
- [65] VAN DE GEER, S. A. and BÜHLMANN, P. (2009). On the conditions used to prove oracle results for the Lasso. *Electron. J. Stat.* **3** 1360–1392. [MR2576316](https://arxiv.org/abs/2576316)
- [66] 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. [MR3543511](https://arxiv.org/abs/3543511)
- [67] WINKLER, G. and LIEBSCHER, V. (2002). Smoothers for discontinuous signals. *J. Non-parametr. Stat.* **14** 203–222.
- [68] XIN, B., KAWAHARA, Y., WANG, Y. and GAO, W. (2014). Efficient generalized fused lasso and its application to the diagnosis of Alzheimer’s disease. In *Proc. Assoc. Adv. Artificial. Intell. Conf.* 2163–2169.
- [69] YAO, Y.-C. (1984). Estimation of a noisy discrete-time step function: Bayes and empirical Bayes approaches. *Ann. Statist.* **12** 1434–1447. [MR0760698](https://arxiv.org/abs/0760698)
- [70] YAO, Y.-C. (1988). Estimating the number of change-points via Schwarz’ criterion. *Statist. Probab. Lett.* **6** 181–189. [MR0919373](https://arxiv.org/abs/0919373)
- [71] YAO, Y.-C. and AU, S.-T. (1989). Least-squares estimation of a step function. *Sankhyā Ser. A* **51** 370–381. [MR1175613](https://arxiv.org/abs/1175613)
- [72] ZHANG, Y., WAINWRIGHT, M. J. and JORDAN, M. I. (2014). Lower bounds on the performance of polynomial-time algorithms for sparse linear regression. In *Conf. Learning Theory* **35** 1–28.
- [73] ZHANG, Y., WAINWRIGHT, M. J. and JORDAN, M. I. (2017). Optimal prediction for sparse linear models? Lower bounds for coordinate-separable M-estimators. *Electron. J. Stat.* **11** 752–799.

MULTICLASS CLASSIFICATION, INFORMATION, DIVERGENCE AND SURROGATE RISK

BY JOHN DUCHI¹, KHASHAYAR KHOSRAVI AND FENG RUAN¹

Stanford University

We provide a unifying view of statistical information measures, multiway Bayesian hypothesis testing, loss functions for multiclass classification problems and multidistribution f -divergences, elaborating equivalence results between all of these objects, and extending existing results for binary outcome spaces to more general ones. We consider a generalization of f -divergences to multiple distributions, and we provide a constructive equivalence between divergences, statistical information (in the sense of DeGroot) and losses for multiclass classification. A major application of our results is in multiclass classification problems in which we must both infer a discriminant function γ —for making predictions on a label Y from datum X —and a data representation (or, in the setting of a hypothesis testing problem, an experimental design), represented as a quantizer q from a family of possible quantizers Q . In this setting, we characterize the equivalence between loss functions, meaning that optimizing either of two losses yields an optimal discriminant and quantizer q , complementing and extending earlier results of Nguyen et al. [*Ann. Statist.* **37** (2009) 876–904] to the multiclass case. Our results provide a more substantial basis than standard classification calibration results for comparing different losses: we describe the convex losses that are consistent for jointly choosing a data representation and minimizing the (weighted) probability of error in multiclass classification problems.

REFERENCES

- [1] 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](#)
- [2] BARTLETT, P. L., JORDAN, M. I. and MCAULIFFE, J. D. (2006). Convexity, classification, and risk bounds. *J. Amer. Statist. Assoc.* **101** 138–156. [MR2268032](#)
- [3] BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300.
- [4] BERGER, J. O. (2006). The case for objective Bayesian analysis. *Bayesian Anal.* **1** 385–402.
- [5] BERNARDO, J. M. (2005). Reference analysis. In *Bayesian Thinking, Modeling and Computation* (D. Day and C. R. Rao, eds.). *Handbook of Statistics* **25** 17–90. Elsevier, Amsterdam.
- [6] BLACKWELL, D. (1951). Comparison of experiments. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* 93–102. Univ. California Press, Berkeley, CA. [MR0046002](#)
- [7] COVER, T. M. and THOMAS, J. A. (2006). *Elements of Information Theory*, 2nd ed. Wiley, Hoboken, NJ. [MR2239987](#)

MSC2010 subject classifications. 62G10, 62K05, 62C05, 94A17, 68Q32.

Key words and phrases. f -divergence, risk, surrogate loss function, hypothesis test.

- [8] CSISZÁR, I. (1967). Information-type measures of difference of probability distributions and indirect observation. *Stud. Sci. Math. Hung.* **2** 299–318.
- [9] DEGROOT, M. H. (1962). Uncertainty, information, and sequential experiments. *Ann. Math. Stat.* **33** 404–419.
- [10] DUCHI, J. C., KHOSRAVI, K. and RUAN, F. (201). Supplement to “Multiclass classification, information, divergence and surrogate risk.” DOI:[10.1214/17-AOS1657SUPP](https://doi.org/10.1214/17-AOS1657SUPP).
- [11] GARCÍA-GARCÍA, D. and WILLIAMSON, R. C. (2012). Divergences and risks for multiclass experiments. In *Proceedings of the Twenty Fifth Annual Conference on Computational Learning Theory*.
- [12] GNEITING, T. and RAFTERY, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.* **102** 359–378. [MR2345548](#)
- [13] GRÜNWARD, P. D. and DAWID, A. P. (2004). Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory. *Ann. Statist.* **32** 1367–1433. [MR2089128](#)
- [14] GYÖRFI, L. and NEMETZ, T. (1978). f -dissimilarity: A generalization of the affinity of several distributions. *Ann. Inst. Statist. Math.* **30** 105–113.
- [15] HIRIART-URRUTY, J.-B. and LEMARÉCHAL, C. (1993). *Convex Analysis and Minimization Algorithms. II. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **306**. Springer, Berlin. [MR1295240](#)
- [16] KAILATH, T. (1967). The divergence and Bhattacharyya distance measures in signal selection. *IEEE Trans. Commun. Technol.* **15** 52–60.
- [17] LIESE, F. and VAJDA, I. (2006). On divergences and informations in statistics and information theory. *IEEE Trans. Inform. Theory* **52** 4394–4412.
- [18] LONGO, M., LOOKABAUGH, T. D. and GRAY, R. M. (1990). Quantization for decentralized hypothesis testing under communication constraints. *IEEE Trans. Inform. Theory* **36** 241–255. [MR1052776](#)
- [19] LUGOSI, G. and VAYATIS, N. (2004). On the Bayes-risk consistency of regularized boosting methods. *Ann. Statist.* **32** 30–55. [MR2051000](#)
- [20] NGUYEN, X., WAINWRIGHT, M. J. and JORDAN, M. I. (2009). On surrogate loss functions and f -divergences. *Ann. Statist.* **37** 876–904. [MR2502654](#)
- [21] ÖSTERREICHER, F. and VAJDA, I. (1993). Statistical information and discrimination. *IEEE Trans. Inform. Theory* **39** 1036–1039. [MR1237725](#)
- [22] POOR, H. V. and THOMAS, J. B. (1977). Applications of Ali–Silvey distance measures in the design of generalized quantizers for binary decision systems. *IEEE Trans. Commun.* **25** 893–900.
- [23] PUKELSHEIM, F. (2006). *Optimal Design of Experiments. Classics in Applied Mathematics* **50**. SIAM, Philadelphia, PA. [MR2224698](#)
- [24] REID, M. and WILLIAMSON, R. (2011). Information, divergence, and risk for binary experiments. *J. Mach. Learn. Res.* **12** 731–817.
- [25] ROBBINS, H. (1952). Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.* **55** 527–535.
- [26] SCHAPIRE, R. E. and FREUND, Y. (2012). *Boosting: Foundations and Algorithms*. MIT Press, Cambridge, MA. [MR2920188](#)
- [27] STEINWART, I. (2007). How to compare different loss functions and their risks. *Constr. Approx.* **26** 225–287. [MR2327600](#)
- [28] TEWARI, A. and BARTLETT, P. L. (2007). On the consistency of multiclass classification methods. *J. Mach. Learn. Res.* **8** 1007–1025.
- [29] TISHBY, N., PEREIRA, F. and BIALEK, W. (1999). The information bottleneck method. In *The 37'th Allerton Conference on Communication, Control, and Computing*.
- [30] TSITSIKLIS, J. N. (1993). Decentralized detection. In *Advances in Signal Processing* **2** 297–344. JAI Press, London.

- [31] VAJDA, I. (1972). On the f -divergence and singularity of probability measures. *Period. Math. Hungar.* **2** 223–234.
- [32] WILLIAMSON, R. C., VERNET, E. and REID, M. D. (2016). Composite multiclass losses. *J. Mach. Learn. Res.* **17**. [MR3595157](#)
- [33] ZHANG, T. (2003/04). Statistical analysis of some multi-category large margin classification methods. *J. Mach. Learn. Res.* **5** 1225–1251. [MR2248016](#)

HALFSPACE DEPTHS FOR SCATTER, CONCENTRATION AND SHAPE MATRICES

BY DAVY PAINDAVEINE¹ AND GERMAIN VAN BEVER²

Université libre de Bruxelles

We propose halfspace depth concepts for scatter, concentration and shape matrices. For scatter matrices, our concept is similar to those from Chen, Gao and Ren [Robust covariance and scatter matrix estimation under Huber's contamination model (2018)] and Zhang [*J. Multivariate Anal.* **82** (2002) 134–165]. Rather than focusing, as in these earlier works, on deepest scatter matrices, we thoroughly investigate the properties of the proposed depth and of the corresponding depth regions. We do so under minimal assumptions and, in particular, we do not restrict to elliptical distributions nor to absolutely continuous distributions. Interestingly, fully understanding scatter halfspace depth requires considering different geometries/topologies on the space of scatter matrices. We also discuss, in the spirit of Zuo and Serfling [*Ann. Statist.* **28** (2000) 461–482], the structural properties a scatter depth should satisfy, and investigate whether or not these are met by scatter halfspace depth. Companion concepts of depth for concentration matrices and shape matrices are also proposed and studied. We show the practical relevance of the depth concepts considered in a real-data example from finance.

REFERENCES

- ARCONES, M. A. and GINÉ, E. (1993). Limit theorems for U -processes. *Ann. Probab.* **21** 1494–1542. [MR1235426](#)
- BERGER, M. (2003). *A Panoramic View of Riemannian Geometry*. Springer, Berlin. [MR2002701](#)
- BHATIA, R. (2007). *Positive Definite Matrices*. Princeton Univ. Press, Princeton, NJ.
- BHATIA, R. and HOLBROOK, J. (2006). Riemannian geometry and matrix geometric means. *Linear Algebra Appl.* **413** 594–618. [MR2198952](#)
- CARDOT, H., CÉNAC, P. and GODICHON-BAGGIONI, A. (2017). Online estimation of the geometric median in Hilbert spaces: Nonasymptotic confidence balls. *Ann. Statist.* **45** 591–614. [MR3650394](#)
- CARTAN, E. (1929). Groupes simples clos et ouverts et géométrie riemannienne. *J. Math. Pures Appl.* **8** 1–33.
- CHAKRABORTY, A. and CHAUDHURI, P. (2014). The spatial distribution in infinite dimensional spaces and related quantiles and depths. *Ann. Statist.* **42** 1203–1231.
- CHAUDHURI, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.* **91** 862–872.
- CHEN, M., GAO, C. and REN, Z. (2018). Robust covariance and scatter matrix estimation under Huber's contamination model. *Ann. Statist.* To appear.
- CLAESKENS, G., HUBERT, M., SLAETS, L. and VAKILI, K. (2014). Multivariate functional halfspace depth. *J. Amer. Statist. Assoc.* **109** 411–423. [MR3180573](#)

MSC2010 subject classifications. Primary 62H20; secondary 62G35.

Key words and phrases. Curved parameter spaces, elliptical distributions, robustness, scatter matrices, shape matrices, statistical depth.

- CUEVAS, A., FEBRERO, M. and FRAIMAN, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Comput. Statist.* **22** 481–496.
- DANG, X. and SERFLING, R. J. (2010). Nonparametric depth-based multivariate outlier identifiers, and masking robustness properties. *J. Statist. Plann. Inference* **140** 198–213.
- DONOHU, D. L. and GASKO, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.* **20** 1803–1827. [MR1193313](#)
- DÜMBGEN, L. and TYLER, D. E. (2016). Geodesic convexity and regularized scatter estimators. Available at [arXiv:1607.05455v2](#).
- FAN, Y., JIN, J. and YAO, Z. (2013). Optimal classification in sparse Gaussian graphic model. *Ann. Statist.* **41** 2537–2571.
- FAN, Y. and LV, J. (2016). Innovated scalable efficient estimation in ultra-large Gaussian graphical models. *Ann. Statist.* **44** 2098–2126. [MR3546445](#)
- HALL, P. and JIN, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.* **38** 1686–1732. [MR2662357](#)
- HALLIN, M., PAINDAVEINE, D. and ŠIMAN, M. (2010). Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth. *Ann. Statist.* **38** 635–669. [MR2604670](#)
- HE, Y. and EINMAHL, J. H. J. (2017). Estimation of extreme depth-based quantile regions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 449–461.
- HUBERT, M., ROUSSEUW, P. J. and SEGHAERT, P. (2015). Multivariate functional outlier detection. *Stat. Methods Appl.* **24** 177–202. [MR3376852](#)
- ILMONEN, P. and PAINDAVEINE, D. (2011). Semiparametrically efficient inference based on signed ranks in symmetric independent component models. *Ann. Statist.* **39** 2448–2476. [MR2906874](#)
- LIU, R. Y. (1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18** 405–414.
- LIU, R. Y., PARELIUS, J. M. and SINGH, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *Ann. Statist.* **27** 783–858. [MR1724033](#)
- LÓPEZ-PINTADO, S. and ROMO, J. (2009). On the concept of depth for functional data. *J. Amer. Statist. Assoc.* **104** 718–734.
- MIZERA, I. (2002). On depth and deep points: A calculus. *Ann. Statist.* **30** 1681–1736. [MR1969447](#)
- MIZERA, I. and MÜLLER, C. H. (2004). Location-scale depth. *J. Amer. Statist. Assoc.* **99** 949–989.
- NIETO-REYES, A. and BATTEY, H. (2016). A topologically valid definition of depth for functional data. *Statist. Sci.* **31** 61–79. [MR3458593](#)
- PAINDAVEINE, D. and VAN BEVER, G. (2014). Inference on the shape of elliptical distributions based on the MCD. *J. Multivariate Anal.* **129** 125–144.
- PAINDAVEINE, D. and VAN BEVER, G. (2015). Nonparametrically consistent depth-based classifiers. *Bernoulli* **21** 62–82. [MR3322313](#)
- PAINDAVEINE, D. and VAN BEVER, G. (2018). Supplement to “Halfspace depths for scatter, concentration and shape matrices.” DOI:[10.1214/17-AOS1658SUPP](#).
- ROUSSEUW, P. J. and HUBERT, M. (1999). Regression depth. *J. Amer. Statist. Assoc.* **94** 388–433.
- ROUSSEUW, P. J. and RUTS, I. (1999). The depth function of a population distribution. *Metrika* **49** 213–244.
- ROUSSEUW, P. J. and STRUYF, A. (2004). Characterizing angular symmetry and regression symmetry. *J. Statist. Plann. Inference* **122** 161–173.
- SERFLING, R. J. (2004). Some perspectives on location and scale depth functions. *J. Amer. Statist. Assoc.* **99** 970–973.
- SERFLING, R. (2010). Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardisation. *J. Nonparametr. Stat.* **22** 915–936. [MR2738875](#)
- TUKEY, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974)*, Vol. 2 523–531. Canad. Math. Congress, Montreal.

- VARDI, Y. and ZHANG, C.-H. (2000). The multivariate L_1 -median and associated data depth. *Proc. Natl. Acad. Sci. USA* **97** 1423–1426.
- ZHANG, J. (2002). Some extensions of Tukey’s depth function. *J. Multivariate Anal.* **82** 134–165. [MR1918618](#)
- ZUO, Y. (2003). Projection-based depth functions and associated medians. *Ann. Statist.* **31** 1460–1490. [MR2012822](#)
- ZUO, Y. and SERFLING, R. (2000). General notions of statistical depth function. *Ann. Statist.* **28** 461–482. [MR1790005](#)

MULTILAYER TENSOR FACTORIZATION WITH APPLICATIONS TO RECOMMENDER SYSTEMS¹

BY XUAN BI^{*}, ANNIE QU^{†,‡} AND XIAOTONG SHEN[§]

Yale University,^{} Yunnan University,[†] University of Illinois at Urbana-Champaign[‡] and University of Minnesota[§]*

Recommender systems have been widely adopted by electronic commerce and entertainment industries for individualized prediction and recommendation, which benefit consumers and improve business intelligence. In this article, we propose an innovative method, namely the recommendation engine of multilayers (REM), for tensor recommender systems. The proposed method utilizes the structure of a tensor response to integrate information from multiple modes, and creates an additional layer of nested latent factors to accommodate between-subjects dependency. One major advantage is that the proposed method is able to address the “cold-start” issue in the absence of information from new customers, new products or new contexts. Specifically, it provides more effective recommendations through sub-group information. To achieve scalable computation, we develop a new algorithm for the proposed method, which incorporates a maximum block improvement strategy into the cyclic blockwise-coordinate-descent algorithm. In theory, we investigate algorithmic properties for convergence from an arbitrary initial point and local convergence, along with the asymptotic consistency of estimated parameters. Finally, the proposed method is applied in simulations and IRI marketing data with 116 million observations of product sales. Numerical studies demonstrate that the proposed method outperforms existing competitors in the literature.

REFERENCES

- [1] ADOMAVICIUS, G. and TUZHILIN, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowl. Data Eng.* **17** 734–749.
- [2] ADOMAVICIUS, G. and TUZHILIN, A. (2011). Context-aware recommender systems. In *Recommender Systems Handbook* 217–253. Springer, Berlin.
- [3] ASWANI, A. (2016). Low-rank approximation and completion of positive tensors. *SIAM J. Matrix Anal. Appl.* **37** 1337–1364. [MR3551210](#)
- [4] BHOJANAPALLI, S. and SANGHAVI, S. (2015). A new sampling technique for tensors. Preprint. Available at [arXiv:1502.05023](#).
- [5] BI, X., QU, A. and SHEN, X. (2018). Supplement to “Multilayer tensor factorization with applications to recommender systems.” DOI:[10.1214/17-AOS1659SUPP](#).
- [6] BI, X., QU, A., WANG, J. and SHEN, X. (2017). A group specific recommender system. *J. Amer. Statist. Assoc.* **112** 1344–1353.

MSC2010 subject classifications. Primary 62M20; secondary 90C26, 68T05.

Key words and phrases. Cold-start problem, context-aware recommender system, maximum block improvement, nonconvex optimization, tensor completion.

- [7] BOBADILLA, J., ORTEGA, F., HERNANDO, A. and GUTIÉRREZ, A. (2013). Recommender systems survey. *Knowl.-Based Syst.* **46** 109–132.
- [8] BRONNENBERG, B. J., KRUGER, M. W. and MELA, C. F. (2008). Database paper—The IRI marketing data set. *Mark. Sci.* **27** 745–748.
- [9] CHEN, B., HE, S., LI, Z. and ZHANG, S. (2012). Maximum block improvement and polynomial optimization. *SIAM J. Optim.* **22** 87–107.
- [10] CHI, E. C. and KOLDA, T. G. (2012). On tensors, sparsity, and nonnegative factorizations. *SIAM J. Matrix Anal. Appl.* **33** 1272–1299.
- [11] CLAUSEN, J. (1999). Branch and bound algorithms—principles and examples. Technical Report, Univ. Copenhagen.
- [12] COLOMBO-MENDOZA, L. O., VALENCIA-GARCÍA, R., RODRÍGUEZ-GONZÁLEZ, A., ALOR-HERNÁNDEZ, G. and SAMPER-ZAPATER, J. J. (2015). RecomMetz: A context-aware knowledge-based mobile recommender system for movie showtimes. *Expert Syst. Appl.* **42** 1202–1222.
- [13] DE SILVA, V. and LIM, L.-H. (2008). Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM J. Matrix Anal. Appl.* **30** 1084–1127. [MR2447444](#)
- [14] DEVORE, R. A. and LORENTZ, G. G. (1993). *Constructive Approximation*. Springer, Berlin.
- [15] FANG, Y. and WANG, J. (2012). Selection of the number of clusters via the bootstrap method. *Comput. Statist. Data Anal.* **56** 468–477.
- [16] FEUERVERGER, A., HE, Y. and KHATRI, S. (2012). Statistical significance of the Netflix challenge. *Statist. Sci.* **27** 202–231. [MR2963993](#)
- [17] FORBES, P. and ZHU, M. (2011). Content-boosted matrix factorization for recommender systems: Experiments with recipe recommendation. In *Proceedings of the Fifth ACM Conference on Recommender Systems* 261–264. ACM, New York.
- [18] GOLDBERG, K., ROEDER, T., GUPTA, D. and PERKINS, C. (2001). Eigentaste: A constant time collaborative filtering algorithm. *Inf. Retr.* **4** 133–151.
- [19] KARATZOGLOU, A., AMATRIAIN, X., BALTRUNAS, L. and OLIVER, N. (2010). Multiverse recommendation: N-dimensional tensor factorization for context-aware collaborative filtering. In *Proceedings of the Fourth ACM Conference on Recommender Systems* 79–86. ACM, New York.
- [20] KOLDA, T. G. and BADER, B. W. (2009). Tensor decompositions and applications. *SIAM Rev.* **51** 455–500.
- [21] KOREN, Y. (2010). Collaborative filtering with temporal dynamics. *Commun. ACM* **53** 89–97.
- [22] KRUSKAL, J. B. (1977). Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra Appl.* **18** 95–138. [MR0444690](#)
- [23] LAIRD, N. M. and WARE, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38** 963–974.
- [24] LAND, A. H. and DOIG, A. G. (1960). An automatic method of solving discrete programming problems. *Econometrica* **28** 497–520.
- [25] LI, L. and ZHANG, X. (2017). Parsimonious Tensor Response Regression. *J. Amer. Statist. Assoc.* **112** 1131–1146. [MR3735365](#)
- [26] LI, Z., SUK, H.-I., SHEN, D. and LI, L. (2016). Sparse multi-response tensor regression for Alzheimer’s disease study with multivariate clinical assessments. *IEEE Trans. Med. Imag.* **35** 1927–1936.
- [27] LI, Z., USCHMAJEW, A. and ZHANG, S. (2015). On convergence of the maximum block improvement method. *SIAM J. Optim.* **25** 210–233.
- [28] LOMBARDI, S., ANAND, S. S. and GORGOLIONE, M. (2009). Context and customer behaviour in recommendation. In *Workshop on Context-Aware Recommender Systems*.

- [29] MIRANDA, M., ZHU, H. and IBRAHIM, J. G. (2015). TPRM: Tensor partition regression models with applications in imaging biomarker detection. Preprint. Available at [arXiv:1505.05482](https://arxiv.org/abs/1505.05482).
- [30] NGUYEN, J. and ZHU, M. (2013). Content-boosted matrix factorization techniques for recommender systems. *Stat. Anal. Data Min.* **6** 286–301. [MR3092059](#)
- [31] NGUYEN, T. V., KARATZOGLOU, A. and BALTRUNAS, L. (2014). Gaussian process factorization machines for context-aware recommendations. In *Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval* 63–72. ACM, New York.
- [32] PAATERO, P. (1999). The multilinear engine—a table-driven, least squares program for solving multilinear problems, including the n -way parallel factor analysis model. *J. Comput. Graph. Statist.* **8** 854–888. [MR1748971](#)
- [33] PAATERO, P. (2000). Construction and analysis of degenerate PARAFAC models. *J. Chemom.* **14** 285–299.
- [34] PALMISANO, C., TUZHILIN, A. and GORGOLIONE, M. (2008). Using context to improve predictive modeling of customers in personalization applications. *IEEE Trans. Knowl. Data Eng.* **20** 1535–1549.
- [35] PARK, S.-T., PENNOCK, D., MADANI, O., GOOD, N. and DECOSTE, D. (2006). Naïve filterbots for robust cold-start recommendations. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM 699–705.
- [36] RENDLE, S. (2012). Factorization machines with libFM. *ACM Trans. Intell. Syst. Technol.* **3** 57.
- [37] RENDLE, S., GANTNER, Z., FREUDENTHALER, C. and SCHMIDT-THIEME, L. (2011). Fast context-aware recommendations with factorization machines. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval* 635–644. ACM, New York.
- [38] SALAKHUTDINOV, R., MNIH, A. and HINTON, G. (2007). Restricted Boltzmann machines for collaborative filtering. In *Proceedings of the 24th International Conference on Machine Learning* 791–798. ACM, New York.
- [39] SHEN, X. (1998). On the method of penalization. *Statist. Sinica* **8** 337–357. [MR1624410](#)
- [40] SHEN, X., TSENG, G. C., ZHANG, X. and WONG, W. H. (2003). On ψ -learning. *J. Amer. Statist. Assoc.* **98** 724–734.
- [41] SHI, Y., LARSON, M. and HANJALIC, A. (2014). Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. *ACM Computing Surveys (CSUR)* **47** 3.
- [42] SREBRO, N., ALON, N. and JAAKKOLA, T. S. (2005). Generalization error bounds for collaborative prediction with low-rank matrices. In *Advances in Neural Information Processing Systems* 17 5–27.
- [43] VERBERT, K., MANOUSELIS, N., OCHOA, X., WOLPERS, M., DRACHSLER, H., BOSNIC, I. and DUVAL, E. (2012). Context-aware recommender systems for learning: A survey and future challenges. *IEEE Transactions on Learning Technologies* **5** 318–335.
- [44] WANG, J. (2010). Consistent selection of the number of clusters via crossvalidation. *Biometrika* **97** 893–904.
- [45] WANG, P., TSAI, G. and QU, A. (2012). Conditional inference functions for mixed-effects models with unspecified random-effects distribution. *J. Amer. Statist. Assoc.* **107** 725–736. [MR2980080](#)
- [46] WELLING, M. and WEBER, M. (2001). Positive tensor factorization. *Pattern Recogn. Lett.* **22** 1255–1261.
- [47] XIONG, L., CHEN, X., HUANG, T.-K., SCHNEIDER, J. and CARBONELL, J. G. (2010). Temporal collaborative filtering with Bayesian probabilistic tensor factorization. In *Proceed-*

ings of the 2010 SIAM International Conference on Data Mining SIAM, Philadelphia, PA.

- [48] YUAN, M. and ZHANG, C.-H. (2016). Incoherent tensor norms and their applications in higher order tensor completion. Preprint. Available at [arXiv:1606.03504](https://arxiv.org/abs/1606.03504).
- [49] YUAN, M. and ZHANG, C.-H. (2016). On tensor completion via nuclear norm minimization. *Found. Comput. Math.* **16** 1031–1068.
- [50] ZHOU, H., LI, L. and ZHU, H. (2013). Tensor regression with applications in neuroimaging data analysis. *J. Amer. Statist. Assoc.* **108** 540–552.
- [51] ZHU, Y., SHEN, X. and YE, C. (2016). Personalized prediction and sparsity pursuit in latent factor models. *J. Amer. Statist. Assoc.* **111** 241–252.

PRINCIPAL COMPONENT ANALYSIS FOR FUNCTIONAL DATA ON RIEMANNIAN MANIFOLDS AND SPHERES¹

BY XIONGTAO DAI AND HANS-GEORG MÜLLER

University of California, Davis

Functional data analysis on nonlinear manifolds has drawn recent interest. Sphere-valued functional data, which are encountered, for example, as movement trajectories on the surface of the earth are an important special case. We consider an intrinsic principal component analysis for smooth Riemannian manifold-valued functional data and study its asymptotic properties. Riemannian functional principal component analysis (RFPCA) is carried out by first mapping the manifold-valued data through Riemannian logarithm maps to tangent spaces around the Fréchet mean function, and then performing a classical functional principal component analysis (FPCA) on the linear tangent spaces. Representations of the Riemannian manifold-valued functions and the eigenfunctions on the original manifold are then obtained with exponential maps. The tangent-space approximation yields upper bounds to residual variances if the Riemannian manifold has nonnegative curvature. We derive a central limit theorem for the mean function, as well as root- n uniform convergence rates for other model components. Our applications include a novel framework for the analysis of longitudinal compositional data, achieved by mapping longitudinal compositional data to trajectories on the sphere, illustrated with longitudinal fruit fly behavior patterns. RFPCA is shown to outperform an unrestricted FPCA in terms of trajectory recovery and prediction in applications and simulations.

REFERENCES

- ADRIAENSSENS, N., COENEN, S., VERSPORTEN, A., MULLER, A., MINALU, G., FAES, C., VANKERCKHOVEN, V., AERTS, M., HENS, N. and MOLENBERGHS, G. (2011). European Surveillance of Antimicrobial Consumption (ESAC): Outpatient antibiotic use in Europe (1997–2009). *J. Antimicrob. Chemother.* **66** vi3–vi12.
- AITCHISON, J. (1986). *The Statistical Analysis of Compositional Data*. Chapman & Hall, London. [MR0865647](#)
- ANIRUDH, R., TURAGA, P., SU, J. and SRIVASTAVA, A. (2015). Elastic functional coding of human actions: From vector-fields to latent variables. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 3147–3155.
- ANIRUDH, R., TURAGA, P., SU, J. and SRIVASTAVA, A. (2017). Elastic functional coding of Riemannian trajectories. *IEEE Trans. Pattern Anal. Mach. Intell.* **39** 922–936.
- BHATTACHARYA, R. and LIN, L. (2017). Omnibus CLTs for Fréchet means and nonparametric inference on non-Euclidean spaces. *Proc. Amer. Math. Soc.* **145** 413–428.

MSC2010 subject classifications. Primary 62G05; secondary 62G20, 62G99.

Key words and phrases. Compositional data, dimension reduction, functional data analysis, functional principal component analysis, principal geodesic analysis, Riemannian manifold, trajectory, central limit theorem, uniform convergence.

- BHATTACHARYA, R. and PATRANGENARU, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds.I. *Ann. Statist.* **31** 1–29. [MR1962498](#)
- BHATTACHARYA, R. and PATRANGENARU, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds.II. *Ann. Statist.* **33** 1225–1259. [MR2195634](#)
- BOSQ, D. (2000). *Linear Processes in Function Spaces: Theory and Applications. Lecture Notes in Statistics* **149**. Springer, New York. [MR1783138](#)
- CAREY, J. R., PAPADOPOULOS, N. T., KOULOSSIS, N. A., KATSOYANNOS, B. I., MÜLLER, H.-G., WANG, J.-L. and TSENG, Y.-K. (2006). Age-specific and lifetime behavior patterns in *Drosophila melanogaster* and the Mediterranean fruit fly, *Ceratitis capitata*. *Exp. Gerontol.* **41** 93–97.
- CASTRO, P. E., LAWTON, W. H. and SYLVESTRE, E. A. (1986). Principal modes of variation for processes with continuous sample curves. *Technometrics* **28** 329–337.
- CHAVEL, I. (2006). *Riemannian Geometry*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **98**. Cambridge Univ. Press, Cambridge. [MR2229062](#)
- CHEN, D. and MÜLLER, H.-G. (2012). Nonlinear manifold representations for functional data. *Ann. Statist.* **40** 1–29.
- CHIOU, J.-M., CHEN, Y.-T. and YANG, Y.-F. (2014). Multivariate functional principal component analysis: A normalization approach. *Statist. Sinica* **24** 1571–1596.
- CORNEA, E., ZHU, H., KIM, P. and IBRAHIM, J. G. (2017). Regression models on Riemannian symmetric spaces. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 463–482. [MR3611755](#)
- DAI, X. and MÜLLER, H.-G. (2018). Supplement to “Principal component analysis for functional data on Riemannian manifolds and spheres.” DOI:[10.1214/17-AOS1660SUPP](#).
- FISHER, N. I., LEWIS, T. and EMBLETON, B. J. J. (1987). *Statistical Analysis of Spherical Data*. Cambridge Univ. Press, Cambridge.
- FLETCHER, P. T., LU, C., PIZER, S. M. and JOSHI, S. (2004). Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Trans. Med. Imag.* **23** 995–1005.
- HSING, T. and EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley, Chichester. [MR3379106](#)
- HUCKEMANN, S. F. and ELTZNER, B. (2018). Backward nested descriptors asymptotics with inference on stem cell differentiation. *Ann. Statist.* **46** 1994–2019.
- HUCKEMANN, S., HOTZ, T. and MUNK, A. (2010). Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions. *Statist. Sinica* **20** 1–58. [MR2640651](#)
- JAIN, N. C. and MARCUS, M. B. (1975). Central limit theorems for $C(S)$ -valued random variables. *J. Funct. Anal.* **19** 216–231.
- JUNG, S., DRYDEN, I. L. and MARRON, J. S. (2012). Analysis of principal nested spheres. *Biometrika* **99** 551–568.
- JUPP, P. E. and KENT, J. T. (1987). Fitting smooth paths to spherical data. *J. Roy. Statist. Soc. Ser. C* **36** 34–46.
- KENDALL, D. G., BARDEN, D., CARNE, T. K. and LE, H. (2009). *Shape and Shape Theory*. Wiley, Hoboken.
- KENT, J. T., MARDIA, K. V., MORRIS, R. J. and AYKROYD, R. G. (2001). Functional models of growth for landmark data. In *Proceedings in Functional and Spatial Data Analysis* 109–115.
- KNEIP, A. and UTIKAL, K. J. (2001). Inference for density families using functional principal component analysis. *J. Amer. Statist. Assoc.* **96** 519–542.
- LILA, E., ASTON, J. A. D. and SANGALLI, L. M. (2016). Smooth principal component analysis over two-dimensional manifolds with an application to neuroimaging. *Ann. Appl. Stat.* **10** 1854–1879. [MR3592040](#)
- LIN, Z. and YAO, F. (2017). Functional regression with unknown manifold structures. Available at [arXiv:1704.03005](#).
- LIN, L., THOMAS, B. S., ZHU, H. and DUNSON, D. B. (2017). Extrinsic local regression on manifold-valued data. *J. Amer. Statist. Assoc.* **112** 1261–1273.

- MARDIA, K. V. and JUPP, P. E. (2009). *Directional Statistics*. Wiley, Hoboken.
- NADARAYA, E. A. (1964). On estimating regression. *Theory Probab. Appl.* **9** 141–142.
- PATRANGENARU, V. and ELLINGSON, L. (2015). *Nonparametric Statistics on Manifolds and Their Applications to Object Data Analysis*. CRC Press, Boca Raton, FL.
- PETERSEN, A. and MÜLLER, H.-G. (2016). Functional data analysis for density functions by transformation to a Hilbert space. *Ann. Statist.* **44** 183–218. [MR3449766](#)
- PETERSEN, A. and MÜLLER, H. G. (2018). Fréchet regression for random objects. *Ann. Statist.* To appear. Available at [arXiv:1608.03012](#).
- QIU, Z., SONG, X. K. and TAN, M. (2008). Simplex mixed-effects models for longitudinal proportional data. *Scand. J. Stat.* **35** 577–596.
- RAHMAN, I. U., DRORI, I., STODDEN, V. C., DONOHO, D. L. and SCHRÖDER, P. (2005). Multi-scale representations for manifold-valued data. *Multiscale Model. Simul.* **4** 1201–1232.
- RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. Springer, New York.
- SU, J., KURTEK, S., KLASSEN, E. and SRIVASTAVA, A. (2014). Statistical analysis of trajectories on Riemannian manifolds: Bird migration, hurricane tracking and video surveillance. *Ann. Appl. Stat.* **8** 530–552. [MR3192001](#)
- TELSCHOW, F. J. E., HUCKEMANN, S. F. and PIERRYNOWSKI, M. R. (2016). Functional inference on rotational curves and identification of human gait at the knee joint. Available at [arXiv:1611.03665](#).
- TOURNIER, M., WU, X., COURTY, N., ARNAUD, E. and REVERET, L. (2009). Motion compression using principal geodesics analysis. In *Computer Graphics Forum* **28** 355–364.
- VAN DER VAART, A. and WELLNER, J. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York.
- WANG, J.-L., CHIOU, J.-M. and MÜLLER, H.-G. (2016). Functional data analysis. *Annu. Rev. Stat. Appl.* **3** 257–295.
- WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26** 359–372. [MR0185765](#)
- ZHENG, Y. (2015). Trajectory data mining: An overview. *ACM Trans. Intell. Syst. Technol.* **6** 29:1–29:41.

ASSESSING ROBUSTNESS OF CLASSIFICATION USING AN ANGULAR BREAKDOWN POINT

BY JUNLONG ZHAO¹, GUAN YU² AND YUFENG LIU²

*Beijing Normal University, State University of New York at Buffalo and
University of North Carolina at Chapel Hill*

Robustness is a desirable property for many statistical techniques. As an important measure of robustness, the breakdown point has been widely used for regression problems and many other settings. Despite the existing development, we observe that the standard breakdown point criterion is not directly applicable for many classification problems. In this paper, we propose a new breakdown point criterion, namely angular breakdown point, to better quantify the robustness of different classification methods. Using this new breakdown point criterion, we study the robustness of binary large margin classification techniques, although the idea is applicable to general classification methods. Both bounded and unbounded loss functions with linear and kernel learning are considered. These studies provide useful insights on the robustness of different classification methods. Numerical results further confirm our theoretical findings.

REFERENCES

- BIGGIO, B., NELSON, B. and LASKOV, P. (2012). Poisoning attacks against support vector machines. In *Proceedings of the 29th International Conference on Machine Learning* (J. Langford and J. Pineau, eds.) 1807–1814. ACM, New York.
- CHRISTMANN, A. and STEINWART, I. (2004). On robustness properties of convex risk minimization methods for pattern recognition. *J. Mach. Learn. Res.* **5** 1007–1034.
- DONOHU, D. and HUBER, P. J. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann* 157–184. Wadsworth, Belmont, CA. [MR0689745](#)
- FREUND, Y. and SCHAPIRE, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. System Sci.* **55** 119–139. [MR1473055](#)
- GENTON, M. G. and LUCAS, A. (2003). Comprehensive definitions of breakdown points for independent and dependent observations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 81–94. [MR1959094](#)
- HABLE, R. and CHRISTMANN, A. (2011). On qualitative robustness of support vector machines. *J. Multivariate Anal.* **102** 993–1007. [MR2793871](#)
- HAMPEL, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Stat.* **42** 1887–1896. [MR0301858](#)
- HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393.
- HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, New York. [MR2722294](#)

MSC2010 subject classifications. Primary 62H30; secondary 62G35.

Key words and phrases. Breakdown point, classification, loss function, reproducing kernel Hilbert spaces, robustness.

- HORST, R. and THOAI, N. V. (1999). DC programming: Overview. *J. Optim. Theory Appl.* **103** 1–43.
- HUBERT, M., ROUSSEEUW, P. J. and VAN AELST, S. (2008). High-breakdown robust multivariate methods. *Statist. Sci.* 92–119.
- KANAMORI, T., FUJIWARA, S. and TAKEDA, A. (2014). Breakdown point of robust support vector machine. Preprint. Available at [arXiv:1409.0934](https://arxiv.org/abs/1409.0934).
- KIMELDORF, G. S. and WAHBA, G. (1970). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Stat.* **41** 495–502. [MR0254999](#)
- KRAUSE, N. and SINGER, Y. (2004). Leveraging the margin more carefully. In *Proceedings of the Twenty-First International Conference on Machine Learning* 63. ACM, New York.
- LIN, X., WAHBA, G., XIANG, D., GAO, F., KLEIN, R. and KLEIN, B. (2000). Smoothing spline ANOVA models for large data sets with Bernoulli observations and the randomized GACV. *Ann. Statist.* **28** 1570–1600. [MR1835032](#)
- LIU, Y. and SHEN, X. (2006). Multicategory ψ -learning. *J. Amer. Statist. Assoc.* **101** 500–509.
- MASON, L., BAXTER, J., BARTLETT, P. and FREAN, M. (2000). Boosting algorithms as gradient descent. In *Advances in Neural Information Processing Systems* **12** 512–518. MIT Press, Cambridge.
- RONCHETTI, E. (1997). Robust inference by influence functions. *J. Statist. Plann. Inference* **57** 59–72.
- SAKATA, S. and WHITE, H. (1995). An alternative definition of finite-sample breakdown point with applications to regression model estimators. *J. Amer. Statist. Assoc.* **90** 1099–1106. [MR1354027](#)
- SCHÖLKOPF, B. and SMOLA, A. J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA.
- SCHOLKOPF, B., SMOLA, A. J., WILLIAMSON, R. C. and BARTLETT, P. L. (2000). New support vector algorithms. *Neural Comput.* **12** 1207–1245.
- SHEN, X., TSENG, G. C., ZHANG, X. and WONG, W. H. (2003). On ψ -learning. *J. Amer. Statist. Assoc.* **98** 724–734.
- STROMBERG, A. J. and RUPPERT, D. (1992). Breakdown in nonlinear regression. *J. Amer. Statist. Assoc.* **87** 991–997.
- VAPNIK, V. N. (1998). *Statistical Learning Theory*. Wiley, New York. [MR1641250](#)
- WAHBA, G. (1990). *Spline Models for Observational Data* **59**. SIAM, Philadelphia, PA.
- WU, Y. and LIU, Y. (2007). Robust truncated hinge loss support vector machines. *J. Amer. Statist. Assoc.* **102** 974–983. [MR2411659](#)
- XU, L., CRAMMER, K. and SCHUURMANS, D. (2006). Robust support vector machine training via convex outlier ablation. In *American Association for Artificial Intelligence* **6** 536–542.
- YU, Y., YANG, M., XU, L., WHITE, M. and SCHUURMANS, D. (2010). Relaxed clipping: A global training method for robust regression and classification. In *Advances in Neural Information Processing Systems* 2532–2540.
- ZHAO, J., YU, G. and LIU, Y. (2018). Supplement to “Assessing robustness of classification using an angular breakdown point.” DOI:[10.1214/17-AOS1661SUPP](https://doi.org/10.1214/17-AOS1661SUPP).

TAIL-GREEDY BOTTOM-UP DATA DECOMPOSITIONS AND FAST MULTIPLE CHANGE-POINT DETECTION¹

BY PIOTR FRYZLEWICZ

London School of Economics

This article proposes a “tail-greedy”, bottom-up transform for one-dimensional data, which results in a nonlinear but conditionally orthonormal, multiscale decomposition of the data with respect to an adaptively chosen unbalanced Haar wavelet basis. The “tail-greediness” of the decomposition algorithm, whereby multiple greedy steps are taken in a single pass through the data, both enables fast computation and makes the algorithm applicable in the problem of consistent estimation of the number and locations of multiple change-points in data. The resulting agglomerative change-point detection method avoids the disadvantages of the classical divisive binary segmentation, and offers very good practical performance. It is implemented in the R package `breakfast`, available from CRAN.

REFERENCES

- AGGARWAL, R., INCLAN, C. and LEAL, R. (1999). Volatility in emerging stock markets. *J. Financ. Quant. Anal.* **34** 33–55.
- AUGER, I. E. and LAWRENCE, C. E. (1989). Algorithms for the optimal identification of segment neighborhoods. *Bull. Math. Biol.* **51** 39–54. [MR0978902](#)
- BAI, J. (1997). Estimating multiple breaks one at a time. *Econometric Theory* **13** 315–352. [MR1455175](#)
- BAI, J. and PERRON, P. (2003). Computation and analysis of multiple structural change models. *J. Appl. Econometrics* **18** 1–22.
- BARANOWSKI, R., CHEN, Y. and FRYZLEWICZ, P. (2016). Narrowest-Over-Threshold detection of multiple change-points and change-point-like features. Preprint.
- BIRGE, L. and MASSART, P. (2001). Gaussian model selection. *J. Eur. Math. Soc. (JEMS)* **3** 203–268.
- BOYSEN, L., KEMPE, A., LIEBSCHER, V., MUNK, A. and WITTICH, O. (2009). Consistencies and rates of convergence of jump-penalized least squares estimators. *Ann. Statist.* **37** 157–183. [MR2488348](#)
- BRAUN, J., BRAUN, R. and MUELLER, H.-G. (2000). Multiple changepoint fitting via quasilikelihood, with application to DNA sequence segmentation. *Biometrika* **87** 301–314.
- BRAUN, J. and MUELLER, H.-G. (1998). Statistical methods for DNA sequence segmentation. *Statist. Sci.* **13** 142–162.
- BRODSKY, B. and DARKHOVSKY, B. (1993). *Nonparametric Methods in Change-Point Problems*. Kluwer, Dordrecht.
- CHEN, K.-M., COHEN, A. and SACKROWITZ, H. (2011). Consistent multiple testing for change points. *J. Multivariate Anal.* **102** 1339–1343.

MSC2010 subject classifications. 62G05.

Key words and phrases. Tail-greediness, bottom-up methods, multiscale methods, segmentation, thresholding, sparsity.

- CHO, H. and FRYZLEWICZ, P. (2011). Multiscale interpretation of taut string estimation and its connection to Unbalanced Haar wavelets. *Stat. Comput.* **21** 671–681.
- CHO, H. and FRYZLEWICZ, P. (2012). Multiscale and multilevel technique for consistent segmentation of nonstationary time series. *Statist. Sinica* **22** 207–229.
- CHO, H. and FRYZLEWICZ, P. (2015). Multiple change-point detection for high-dimensional time series via Sparsified Binary Segmentation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 475–507.
- CHOI, F. Y. Y. (2000). Advances in domain independent linear text segmentation. In *NAACL 2000 Proceedings of the 1st North American Chapter of the Association for Computational Linguistics Conference* 26–33.
- CHU, P.-S. and ZHAO, X. (2004). Bayesian change-point analysis of tropical cyclone activity: The central north Pacific case. *J. Climate* **17** 4893–4901.
- CIUPERCA, G. (2011). A general criterion to determine the number of change-points. *Statist. Probab. Lett.* **81** 1267–1275.
- CIUPERCA, G. (2014). Model selection by LASSO methods in a change-point model. *Statist. Papers* **55** 349–374.
- DAVIES, P. L. and KOVAC, A. (2001). Local extremes, runs, strings and multiresolution. *Ann. Statist.* **29** 1–48.
- DAVIS, R., LEE, T. and RODRIGUEZ-YAM, G. (2006). Structural break estimation for nonstationary time series models. *J. Amer. Statist. Assoc.* **101** 223–239.
- DESMOND, R., WEISS, H., ARANI, R., SOONG, S.-J., WOOD, M., FIDDIAN, P., GNANN, J. and WHITLEY, R. (2002). Clinical applications for change-point analysis of herpes zoster pain. *Journal of Pain and Symptom Management* **23** 510–516.
- DONOHU, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455.
- DU, C., KAO, C.-L. and KOU, S. (2016). Stepwise signal extraction via marginal likelihood. *J. Amer. Statist. Assoc.* **111** 314–330.
- EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499. [MR2060166](#)
- EICHINGER, B. and KIRCH, C. (2018). A MOSUM procedure for the estimation of multiple random change points. *Bernoulli* **24** 526–564.
- ERDMAN, C. and EMERSON, J. (2008). A fast Bayesian change point analysis for the segmentation of microarray data. *Bioinformatics* **24** 2143–2148.
- FRICK, K., MUNK, A. and SIELING, H. (2014). Multiscale change-point inference (with discussion). *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 495–580.
- FRYZLEWICZ, P. (2007). Unbalanced Haar technique for nonparametric function estimation. *J. Amer. Statist. Assoc.* **102** 1318–1327.
- FRYZLEWICZ, P. (2014). Wild binary segmentation for multiple change-point detection. *Ann. Statist.* **42** 2243–2281. [MR3269979](#)
- FRYZLEWICZ, P. (2018). Supplement to “Tail-greedy bottom-up data decompositions and fast multiple change-point detection”. DOI:[10.1214/17-AOS1662SUPP](#).
- FRYZLEWICZ, P. and SUBBA RAO, S. (2014). Multiple-change-point detection for auto-regressive conditional heteroscedastic processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 903–924.
- FRYZLEWICZ, P. and TIMMERMAN, C. (2016). SHAH: SHape-Adaptive Haar wavelets for image processing. *J. Comput. Graph. Statist.* **25** 879–898.
- HARCHAoui, Z. and LÉVY-LEDUC, C. (2010). Multiple change-point estimation with a total variation penalty. *J. Amer. Statist. Assoc.* **105** 1480–1493.
- HUSKOVA, M. and SLABY, A. (2001). Permutation tests for multiple changes. *Kybernetika (Prague)* **37** 605–622.
- JACKSON, B., SARGLE, J., BARNES, D., ARABHI, S., ALT, A., GIOUMOUSIS, P., GWIN, E., SANGTRAKULCHAROEN, P., TAN, L. and TSAI, T. T. (2005). An algorithm for optimal partitioning of data on an interval. *IEEE Signal Process. Lett.* **12** 105–108.

- JANSEN, M., NASON, G. and SILVERMAN, B. (2009). Multiscale methods for data on graphs and irregular multidimensional situations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 97–125.
- KILLICK, R., FEARNHEAD, P. and ECKLEY, I. (2012). Optimal detection of changepoints with a linear computational cost. *J. Amer. Statist. Assoc.* **107** 1590–1598.
- KILLICK, R., NAM, C., ASTON, J. and ECKLEY, I. (2012). changepoint.info: The changepoint repository.
- KOPRINSKA, I. and CARRATO, S. (2001). Temporal video segmentation: A survey. *Signal Process. Image Commun.* **16** 477–500.
- LAVIELLE, M. (1999). Detection of multiple changes in a sequence of dependent variables. *Stochastic Process. Appl.* **83** 79–102.
- LAVIELLE, M. (2005). Using penalized contrasts for the change-point problem. *Signal Process.* **85** 1501–1510.
- LAVIELLE, M. and MOULINES, E. (2000). Least-squares estimation of an unknown number of shifts in a time series. *J. Time Series Anal.* **21** 33–59.
- LEBARBIER, E. (2005). Detecting multiple change-points in the mean of Gaussian process by model selection. *Signal Process.* **85** 717–736.
- LEE, C.-B. (1995). Estimating the number of change points in a sequence of independent normal random variables. *Statist. Probab. Lett.* **25** 241–248.
- LI, H., MUNK, A. and SIELING, H. (2016). FDR-control in multiscale change-point segmentation. *Electron. J. Stat.* **10** 918–959.
- LIO, P. and VANUCCI, M. (2000). Wavelet change-point prediction of transmembrane proteins. *Bioinformatics* **16** 376–382.
- LU, L., ZHANG, H.-J. and JIANG, H. (2002). Content analysis for audio classification and segmentation. *IEEE Trans. Speech Audio Process.* **10** 504–516.
- MAHMOUD, M., PARKER, P., WOODALL, W. and HAWKINS, D. (2007). A change point method for linear profile data. *Qual. Reliab. Eng. Int.* **23** 247–268.
- MAIDSTONE, R., HOCKING, T., RIGAILL, G. and FEARNHEAD, P. (2017). On optimal multiple changepoint algorithms for large data. *Stat. Comput.* **27** 519–533.
- MATTESON, D. and JAMES, N. (2014). A nonparametric approach for multiple change point analysis of multivariate data. *J. Amer. Statist. Assoc.* **109** 334–345.
- MININ, V., DORMAN, K., FANG, F. and SUCHARD, M. (2005). Dual multiple change-point model leads to more accurate recombination detection. *Bioinformatics* **21** 3034–3042.
- OLSHEN, A., VENKATRAMAN, E. S., LUCITO, R. and WIGLER, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostat.* **5** 557–572.
- PAN, J. and CHEN, J. (2006). Application of modified information criterion to multiple change point problems. *J. Multivariate Anal.* **97** 2221–2241.
- RIGAILL, G. (2015). A pruned dynamic programming algorithm to recover the best segmentations with 1 to $K_{\max}$ change-points. *J. SFdS* **156** 180–205.
- RINALDO, A. (2009). Properties and refinements of the fused lasso. *Ann. Statist.* **37** 2922–2952. [MR2541451](#)
- ROBBINS, M., GALLAGHER, C., LUND, R. B. and AUE, A. (2011). Mean shift testing in correlated data. *J. Time Series Anal.* **32** 498–511.
- ROJAS, C. and WAHLBERG, B. (2014). On change point detection using the fused lasso method. Preprint.
- SCHROEDER, A. L. and FRYZLEWICZ, P. (2013). Adaptive trend estimation in financial time series via multiscale change-point-induced basis recovery. *Stat. Interface* **6** 449–461.
- SHRIBERGA, E., STOLCKEA, A., HAKKANI-TÜRB, D. and TÜRB, G. (2000). Prosody-based automatic segmentation of speech into sentences and topics. *Speech Commun.* **32** 127–154.
- TARTAKOVSKY, A., ROZOVSKII, B., BLAZEK, R. and KIM, H. (2006). A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods. *IEEE Trans. Signal Process.* **54** 3372–3382.

- TIBSHIRANI, R., SAUNDERS, M., ROSSET, S., ZHU, J. and KNIGHT, K. (2005). Sparsity and smoothness via the fused lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 91–108. [MR2136641](#)
- VENKATRAMAN, E. S. (1992). Consistency results in multiple change-point problems. Technical Report 24, Dept. Statistics, Stanford Univ. Available from <https://statistics.stanford.edu/resources/technical-reports>.
- VENKATRAMAN, E. S. and OLSHEN, A. (2007). A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23** 657–663.
- VOSTRIKOVA, L. (1981). Detecting ‘disorder’ in multidimensional random processes. *Sov. Math., Dokl.* **24** 55–59.
- WANG, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrika* **82** 385–397.
- WANG, H., ZHANG, D. and SHIN, K. (2004). Change-point monitoring for the detection of DoS attacks. *IEEE Trans. Dependable Secure Comput.* **1** 193–208.
- WU, Y. (2008). Simultaneous change point analysis and variable selection in a regression problem. *J. Multivariate Anal.* **99** 2154–2171.
- YAO, Y.-C. (1988). Estimating the number of change-points via Schwarz’ criterion. *Statist. Probab. Lett.* **6** 181–189.
- YAO, Y.-C. and AU, S. T. (1989). Least-squares estimation of a step function. *Sankhyā Ser. A* **51** 370–381. [MR1175613](#)
- ZHANG, N. R. and SIEGMUND, D. O. (2007). A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics* **63** 22–32.

ROCKET: ROBUST CONFIDENCE INTERVALS VIA KENDALL'S TAU FOR TRANSELLIPTICAL GRAPHICAL MODELS¹

BY RINA FOYGEL BARBER AND MLADEN KOLAR

University of Chicago

Understanding complex relationships between random variables is of fundamental importance in high-dimensional statistics, with numerous applications in biological and social sciences. Undirected graphical models are often used to represent dependencies between random variables, where an edge between two random variables is drawn if they are conditionally dependent given all the other measured variables. A large body of literature exists on methods that estimate the structure of an undirected graphical model, however, little is known about the distributional properties of the estimators beyond the Gaussian setting. In this paper, we focus on inference for edge parameters in a high-dimensional transelliptical model, which generalizes Gaussian and nonparanormal graphical models. We propose ROCKET, a novel procedure for estimating parameters in the latent inverse covariance matrix. We establish asymptotic normality of ROCKET in an ultra high-dimensional setting under mild assumptions, without relying on oracle model selection results. ROCKET requires the same number of samples that are known to be necessary for obtaining a $\sqrt{n}$ consistent estimator of an element in the precision matrix under a Gaussian model. Hence, it is an optimal estimator under a much larger family of distributions. The result hinges on a tight control of the sparse spectral norm of the nonparametric Kendall's tau estimator of the correlation matrix, which is of independent interest. Empirically, ROCKET outperforms the nonparanormal and Gaussian models in terms of achieving accurate inference on simulated data. We also compare the three methods on real data (daily stock returns), and find that the ROCKET estimator is the only method whose behavior across subsamples agrees with the distribution predicted by the theory.

REFERENCES

- BARBER, R. F. and KOLAR, M. (2018). Supplement to “ROCKET: Robust confidence intervals via Kendall's tau for transelliptical graphical models.” DOI:[10.1214/17-AOS1663SUPP](https://doi.org/10.1214/17-AOS1663SUPP).
- BELLONI, A. and CHERNOZHUKOV, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli* **19** 521–547. [MR3037163](#)
- 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](#)
- BELLONI, A., CHERNOZHUKOV, V. and KATO, K. (2013a). Uniform post selection inference for LAD regression models. Preprint. Available at [arXiv:1304.0282](#).

MSC2010 subject classifications. Primary 62G10; secondary 62F12, 62G20.

Key words and phrases. Graphical model selection, transelliptical graphical models, covariance selection, uniformly valid inference, post-model selection inference, rank-based estimation.

- BELLONI, A., CHERNOZHUKOV, V. and KATO, K. (2013b). Robust inference in high-dimensional approximately sparse quantile regression models. Preprint. Available at [arXiv:1312.7186](#).
- BELLONI, A., CHERNOZHUKOV, V. and WEI, Y. (2013). Honest confidence regions for logistic regression with a large number of controls. Preprint. Available at [arXiv:1304.3969](#).
- BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer, Heidelberg. [MR2807761](#)
- CAI, T., LIU, W. and LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.* **106** 594–607. [MR2847973](#)
- CANDES, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n . *Ann. Statist.* **35** 2313–2351. [MR2382644](#)
- CHEN, M., REN, Z., ZHAO, H. and ZHOU, H. (2016). Asymptotically normal and efficient estimation of covariate-adjusted Gaussian graphical model. *J. Amer. Statist. Assoc.* **111** 394–406. [MR3494667](#)
- CHENG, J., LI, T., LEVINA, E. and ZHU, J. (2017). High-dimensional mixed graphical models. *J. Comput. Graph. Statist.* **26** 367–378. [MR3640193](#)
- D’ASPREMONT, A., BANERJEE, O. and EL GHAOU, L. (2008). First-order methods for sparse covariance selection. *SIAM J. Matrix Anal. Appl.* **30** 56–66. [MR2399568](#)
- DE LA PEÑA, V. H. and GINÉ, E. (1999). *Decoupling: From Dependence to Independence*. Springer, New York. [MR1666908](#)
- EMBRECHTS, P., LINDSKOG, F. and MCNEIL, A. (2003). Modelling dependence with copulas and applications to risk management. In *Handbook of Heavy Tailed Distributions in Finance* (S. T. Rachev, ed.) 329–384. Elsevier, Amsterdam.
- FAN, J., FENG, Y. and WU, Y. (2009). Network exploration via the adaptive lasso and SCAD penalties. *Ann. Appl. Stat.* **3** 521–541. [MR2750671](#)
- FAN, J., HAN, F. and LIU, H. (2014). PAGE: Robust pattern guided estimation of large covariance matrix. Technical report, Princeton Univ., Princeton, NJ.
- FANG, K. T., KOTZ, S. and NG, K. W. (1990). *Symmetric Multivariate and Related Distributions. Monographs on Statistics and Applied Probability* **36**. Chapman and Hall, Ltd., London. [MR1071174](#)
- FARRELL, M. H. (2015). Robust inference on average treatment effects with possibly more covariates than observations. *J. Econometrics* **189** 1–23. [MR3397349](#)
- FRIEDMAN, J. H., HASTIE, T. J. and TIBSHIRANI, R. J. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441.
- GU, Q., CAO, Y., NING, Y. and LIU, H. (2015). Local and global inference for high dimensional Gaussian copula graphical models. Preprint. Available at [arXiv:1502.02347](#).
- GUO, J., LEVINA, E., MICHAILIDIS, G. and ZHU, J. (2011a). Joint estimation of multiple graphical models. *Biometrika* **98** 1–15.
- GUO, J., LEVINA, E., MICHAILIDIS, G. and ZHU, J. (2011b). Asymptotic properties of the joint neighborhood selection method for estimating categorical Markov networks. Technical report, Univ. Michigan, Ann Arbor, MI.
- HAN, F. and LIU, H. (2013). Optimal rates of convergence for latent generalized correlation matrix estimation in transelliptical distribution. Preprint. Available at [arXiv:1305.6916](#).
- HÖFLING, H. and TIBSHIRANI, R. (2009). Estimation of sparse binary pairwise Markov networks using pseudo-likelihoods. *J. Mach. Learn. Res.* **10** 883–906. [MR2505138](#)
- JAVANMARD, A. and MONTANARI, A. (2013). Nearly optimal sample size in hypothesis testing for high-dimensional regression. Preprint. Available at [arXiv:1311.0274](#).
- JAVANMARD, A. and MONTANARI, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909. [MR3277152](#)
- KLÜPPELBERG, C., KUHN, G. and PENG, L. (2008). Semi-parametric models for the multivariate tail dependence function—The asymptotically dependent case. *Scand. J. Stat.* **35** 701–718. [MR2468871](#)

- LAM, C. and FAN, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Statist.* **37** 4254–4278. [MR2572459](#)
- LAURITZEN, S. L. (1996). *Graphical Models. Oxford Statistical Science Series 17*. The Clarendon Press, Oxford Univ. Press, New York. [MR1419991](#)
- LEE, J. D. and HASTIE, T. J. (2012). Learning mixed graphical models. Preprint. Available at [arXiv:1205.5012](#).
- LEE, J. D., SUN, D. L., SUN, Y. and TAYLOR, J. E. (2016). Exact post-selection inference, with application to the lasso. *Ann. Statist.* **44** 907–927. [MR3485948](#)
- LINDSKOG, F., MCNEIL, A. and SCHMOCK, U. (2003). Kendall’s tau for elliptical distributions. In *Credit Risk* 149–156.
- LIU, W. (2013). Gaussian graphical model estimation with false discovery rate control. *Ann. Statist.* **41** 2948–2978. [MR3161453](#)
- LIU, H., HAN, F. and ZHANG, C.-H. (2012). Transelliptical graphical models. In *Proc. of NIPS* 809–817.
- LIU, H., LAFFERTY, J. and WASSERMAN, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *J. Mach. Learn. Res.* **10** 2295–2328. [MR2563983](#)
- LIU, H. and WANG, L. (2017). TIGER: A tuning-insensitive approach for optimally estimating Gaussian graphical models. *Electron. J. Stat.* **11** 241–294. [MR3606771](#)
- 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](#)
- LOCKHART, R., TAYLOR, J., TIBSHIRANI, R. J. and TIBSHIRANI, R. (2014). A significance test for the lasso. *Ann. Statist.* **42** 413–468. [MR3210970](#)
- LOH, P.-L. and WAINWRIGHT, M. J. (2015). Regularized M -estimators with nonconvexity: Statistical and algorithmic theory for local optima. *J. Mach. Learn. Res.* **16** 559–616. [MR3335800](#)
- MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. [MR2278363](#)
- MITRA, R. and ZHANG, C.-H. (2014). Multivariate analysis of nonparametric estimates of large correlation matrices. Preprint. Available at [arXiv:1403.6195](#).
- RAVIKUMAR, P., WAINWRIGHT, M. J. and LAFFERTY, J. D. (2010). High-dimensional Ising model selection using ℓ_1 -regularized logistic regression. *Ann. Statist.* **38** 1287–1319. [MR2662343](#)
- RAVIKUMAR, P., WAINWRIGHT, M. J., RASKUTTI, G. and YU, B. (2011). High-dimensional covariance estimation by minimizing ℓ_1 -penalized log-determinant divergence. *Electron. J. Stat.* **5** 935–980. [MR2836766](#)
- REN, Z., SUN, T., ZHANG, C.-H. and ZHOU, H. H. (2015). Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *Ann. Statist.* **43** 991–1026. [MR3346695](#)
- ROTHMAN, A. J., BICKEL, P. J., LEVINA, E. and ZHU, J. (2008). Sparse permutation invariant covariance estimation. *Electron. J. Stat.* **2** 494–515. [MR2417391](#)
- SUN, T. and ZHANG, C.-H. (2012a). Comment: “Minimax estimation of large covariance matrices under ℓ_1 -norm”. *Statist. Sinica* **22** 1354–1358.
- SUN, T. and ZHANG, C.-H. (2012b). Sparse matrix inversion with scaled lasso. Preprint. Available at [arXiv:1202.2723](#).
- TIBSHIRANI, R. J., TAYLOR, J., LOCKHART, R. and TIBSHIRANI, R. (2016). Exact post-selection inference for sequential regression procedures. *J. Amer. Statist. Assoc.* **111** 600–620. [MR3538689](#)
- 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](#)
- WEGKAMP, M. and ZHAO, Y. (2016). Adaptive estimation of the copula correlation matrix for semiparametric elliptical copulas. *Bernoulli* **22** 1184–1226. [MR3449812](#)
- XUE, L. and ZOU, H. (2012). Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *Ann. Statist.* **40** 2541–2571. [MR3097612](#)

- XUE, L., ZOU, H. and CAI, T. (2012). Nonconcave penalized composite conditional likelihood estimation of sparse Ising models. *Ann. Statist.* **40** 1403–1429. [MR3015030](#)
- YANG, E., ALLEN, G. I., LIU, Z. and RAVIKUMAR, P. (2012). Graphical models via generalized linear models. In *Advances in Neural Information Processing Systems* 25 1358–1366. Curran Associates, Red Hook, NY.
- YANG, E., BAKER, Y., RAVIKUMAR, P., ALLEN, G. I. and LIU, Z. (2014). Mixed graphical models via exponential families. In *Proc. 17th Int. Conf. Artif. Intel. Stat.* 1042–1050.
- YANG, E., RAVIKUMAR, P., ALLEN, G. I. and LIU, Z. (2015). Graphical models via univariate exponential family distributions. *J. Mach. Learn. Res.* **16** 3813–3847. [MR3450553](#)
- YUAN, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *J. Mach. Learn. Res.* **11** 2261–2286. [MR2719856](#)
- YUAN, M. and LIN, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94** 19–35. [MR2367824](#)
- 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](#)
- ZHAO, T. and LIU, H. (2014). Calibrated precision matrix estimation for high-dimensional elliptical distributions. *IEEE Trans. Inform. Theory* **60** 7874–7887. [MR3285751](#)

ADAPTIVE INVARIANT DENSITY ESTIMATION FOR ERGODIC DIFFUSIONS OVER ANISOTROPIC CLASSES

BY CLAUDIA STRAUCH¹

Universität Mannheim

Consider some multivariate diffusion process $\mathbf{X} = (X_t)_{t \geq 0}$ with unique invariant probability measure and associated invariant density ρ , and assume that a continuous record of observations $X^T = (X_t)_{0 \leq t \leq T}$ of $\mathbf{X}$ is available. Recent results on functional inequalities for symmetric Markov semigroups are used in the statistical analysis of kernel estimators $\hat{\rho}_T = \hat{\rho}_T(X^T)$ of ρ . For the basic problem of estimation with respect to sup-norm risk under anisotropic Hölder smoothness constraints, the proposed approach yields an adaptive estimator which converges at a substantially faster rate than in standard multivariate density estimation from i.i.d. observations.

REFERENCES

- [1] BAKRY, D., BOLLEY, F. and GENTIL, I. (2011). Around Nash inequalities. In *Journées Équations aux Dérivées Partielles* (2010), Biarritz, France.
- [2] BAKRY, D., CATTIAUX, P. and GUILLIN, A. (2008). Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.* **254** 727–759. [MR2381160](#)
- [3] BAKRY, D., GENTIL, I. and LEDOUX, M. (2014). *Analysis and Geometry of Markov Diffusion Operators. Grundlehren der Mathematischen Wissenschaften* **348**. Springer, Cham.
- [4] CARLEN, E. A., KUSUOKA, S. and STROOCK, D. W. (1987). Upper bounds for symmetric Markov transition functions. *Ann. Inst. Henri Poincaré Probab. Stat.* **23** 245–287.
- [5] CATTIAUX, P. and GUILLIN, A. (2008). Deviation bounds for additive functionals of Markov processes. *ESAIM Probab. Stat.* **12** 12–29.
- [6] CHOROWSKI, J. and TRABS, M. (2016). Spectral estimation for diffusions with random sampling times. *Stochastic Process. Appl.* **126** 2976–3008. [MR3542623](#)
- [7] DALALYAN, A. and REISS, M. (2007). Asymptotic statistical equivalence for ergodic diffusions: The multidimensional case. *Probab. Theory Related Fields* **137** 25–47.
- [8] DAVIES, E. B. (1990). *Heat Kernels and Spectral Theory. Cambridge Tracts in Mathematics* **92**. Cambridge Univ. Press, Cambridge.
- [9] GOBET, E., HOFFMANN, M. and REISS, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.* **32** 2223–2253. [MR2102509](#)
- [10] GOLDENSHLUGER, A. and LEPSKI, O. (2008). Universal pointwise selection rule in multivariate function estimation. *Bernoulli* **14** 1150–1190.
- [11] GOLDENSHLUGER, A. and LEPSKI, O. (2009). Structural adaptation via L_p -norm inequalities. *Probab. Theory Related Fields* **143** 41–71.
- [12] GOLDENSHLUGER, A. and LEPSKI, O. (2011). Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39** 1608–1632. [MR2850214](#)

MSC2010 subject classifications. 62M05, 62G07, 62G20.

Key words and phrases. Ergodic diffusion, anisotropic density estimation, adaptation.

- [13] KUTOYANTS, Y. A. (2004). *Statistical Inference for Ergodic Diffusion Processes*. Springer, New York.
- [14] LEDOUX, M. (2001). Logarithmic Sobolev inequalities for unbounded spin systems revisited. In *Séminaire de Probabilités XXXV. Lecture Notes in Math.* **1755** 167–194. Springer, Berlin.
- [15] LEPSKI, O. (2013). Multivariate density estimation under sup-norm loss: Oracle approach, adaptation and independence structure. *Ann. Statist.* **41** 1005–1034. [MR3099129](#)
- [16] LEPSKI, O. (2013). Upper functions for positive random functionals. I. General setting and Gaussian random functions. *Math. Methods Statist.* **22** 1–27. [MR3040410](#)
- [17] LEPSKI, O. (2013). Upper functions for positive random functionals. II. Application to the empirical processes theory, Part 1. *Math. Methods Statist.* **22** 83–99. [MR3071956](#)
- [18] LEPSKI, O. (2013). Upper functions for positive random functionals. II. Application to the empirical processes theory, Part 2. *Math. Methods Statist.* **22** 193–212. [MR3107668](#)
- [19] LEPSKI, O. V. and LEVIT, B. Y. (1999). Adaptive nonparametric estimation of smooth multivariate functions. *Math. Methods Statist.* **8** 344–370.
- [20] LEZAUD, P. (2001). Chernoff and Berry–Esséen inequalities for Markov processes. *ESAIM Probab. Stat.* **5** 183–201.
- [21] NASH, J. (1958). Continuity of solutions of parabolic and elliptic equations. *Amer. J. Math.* **80** 931–954.
- [22] NELSON, E. (1958). The adjoint Markoff process. *Duke Math. J.* **25** 671–690.
- [23] NIKOLSKII, S. M. (1975). *Approximation of Functions of Several Variables and Imbedding Theorems. Die Grundlehren der Mathematischen Wissenschaften* **205**. Springer, New York. Translated from the Russian by J. M. Danskin, Jr. [MR0374877](#)
- [24] PAULIN, D. (2015). Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electron. J. Probab.* **20** 1–32. [MR3383563](#)
- [25] SÖHL, J. and TRABS, M. (2016). Adaptive confidence bands for Markov chains and diffusions: Estimating the invariant measure and the drift. *ESAIM Probab. Stat.* **20** 432–462. [MR3581829](#)
- [26] STRAUCH, C. (2018). Supplement to “Adaptive invariant density estimation for ergodic diffusions over anisotropic classes.” DOI:[10.1214/17-AOS1664SUPP](#).
- [27] VAN DER VAART, A. W. and VAN ZANTEN, H. (2005). Donsker theorems for diffusions: Necessary and sufficient conditions. *Ann. Probab.* **33** 1422–1451. [MR2150194](#)

ROBUST LOW-RANK MATRIX ESTIMATION

BY ANDREAS ELSENER AND SARA VAN DE GEER

ETH Zürich

Many results have been proved for various nuclear norm penalized estimators of the uniform sampling matrix completion problem. However, most of these estimators are not robust: in most of the cases the quadratic loss function and its modifications are used. We consider robust nuclear norm penalized estimators using two well-known robust loss functions: the absolute value loss and the Huber loss. Under several conditions on the sparsity of the problem (i.e., the rank of the parameter matrix) and on the regularity of the risk function sharp and nonsharp oracle inequalities for these estimators are shown to hold with high probability. As a consequence, the asymptotic behavior of the estimators is derived. Similar error bounds are obtained under the assumption of weak sparsity, that is, the case where the matrix is assumed to be only approximately low-rank. In all of our results, we consider a high-dimensional setting. In this case, this means that we assume $n \leq pq$. Finally, various simulations confirm our theoretical results.

REFERENCES

- BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer, Heidelberg. [MR2807761](#)
- CAMBIER, L. and ABSIL, P.-A. (2016). Robust low-rank matrix completion by Riemannian optimization. *SIAM J. Sci. Comput.* **38** S440–S460. [MR3565571](#)
- CANDÈS, E. J. and PLAN, Y. (2010). Matrix completion with noise. *Proc. IEEE* **98** 925–936.
- CANDÈS, E. J., LI, X., MA, Y. and WRIGHT, J. (2011). Robust principal component analysis? *J. ACM* **58** 11.
- CHANDRASEKARAN, V., SANGHAVI, S., PARRILO, P. A. and WILLSKY, A. S. (2011). Rank-sparsity incoherence for matrix decomposition. *SIAM J. Optim.* **21** 572–596. [MR2817479](#)
- CHEN, Y., JALALI, A., SANGHAVI, S. and CARAMANIS, C. (2013). Low-rank matrix recovery from errors and erasures. *IEEE Trans. Inform. Theory* **59** 4324–4337.
- CHERAPANAMJERI, Y., GUPTA, K. and JAIN, P. (2016). Nearly-optimal robust matrix completion. Preprint. Available at [arXiv:1606.07315](#).
- CVX RESEARCH INC. (2012). CVX: Matlab Software for Disciplined Convex Programming, version 2.0. Available at <http://cvxr.com/cvx>.
- ELSENER, A. and VAN DE GEER, S. (2018). Supplement to “Robust low-rank matrix estimation.” DOI:[10.1214/17-AOS1666SUPP](#).
- FOYGEL, R., SHAMIR, O., SREBRO, N. and SALAKHUTDINOV, R. R. (2011). Learning with the weighted trace-norm under arbitrary sampling distributions. *Adv. Neural Inf. Process. Syst.* 2133–2141.
- KLOPP, O. (2014). Noisy low-rank matrix completion with general sampling distribution. *Bernoulli* 282–303.

MSC2010 subject classifications. Primary 62J05, 62F30; secondary 62H12.

Key words and phrases. Matrix completion, robustness, empirical risk minimization, oracle inequality, nuclear norm, sparsity.

- KLOPP, O., LOUNICI, K. and TSYBAKOV, A. B. (2016). Robust matrix completion. *Probab. Theory Related Fields* 1–42.
- 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](#)
- LAFOND, J. (2015). Low rank matrix completion with exponential family noise. *J. Mach. Learn. Res.: Workshop and Conference Proceedings. COLT 2015 Proceedings* **40** 1–20.
- LI, X. (2013). Compressed sensing and matrix completion with constant proportion of corruptions. *Constr. Approx.* **37** 73–99.
- 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](#)
- 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.
- ROHDE, A. and TSYBAKOV, A. B. (2011). Estimation of high-dimensional low-rank matrices. *Ann. Statist.* **39** 887–930. [MR2816342](#)
- SREBRO, N., RENNIE, J. and JAAKKOLA, T. S. (2004). Maximum-margin matrix factorization. In *Proceedings of the NIPS Conference* 1329–1336. Vancouver.
- SREBRO, N. and SHRAIBMAN, A. (2005). Rank, trace-norm and max-norm. In *Learning Theory* 545–560. Springer, Berlin.
- TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* 267–288.
- VAN DE GEER, S. (2001). Least squares estimation with complexity penalties. *Math. Methods Statist.* **10** 355–374. [MR1867165](#)
- VAN DE GEER, S. (2016). *Estimation and Testing Under Sparsity: École d’Été de Probabilités de Saint-Flour XLV-2015*. Springer, Berlin.

SIEVE BOOTSTRAP FOR FUNCTIONAL TIME SERIES

BY EFSTATHIOS PAPARODITIS¹

University of Cyprus

A bootstrap procedure for functional time series is proposed which exploits a general vector autoregressive representation of the time series of Fourier coefficients appearing in the Karhunen–Loève expansion of the functional process. A double sieve-type bootstrap method is developed, which avoids the estimation of process operators and generates functional pseudo-time series that appropriately mimics the dependence structure of the functional time series at hand. The method uses a finite set of functional principal components to capture the essential driving parts of the infinite dimensional process and a finite order vector autoregressive process to imitate the temporal dependence structure of the corresponding vector time series of Fourier coefficients. By allowing the number of functional principal components as well as the autoregressive order used to increase to infinity (at some appropriate rate) as the sample size increases, consistency of the functional sieve bootstrap can be established. We demonstrate this by proving a basic bootstrap central limit theorem for functional finite Fourier transforms and by establishing bootstrap validity in the context of a fully functional testing problem. A novel procedure to select the number of functional principal components is introduced while simulations illustrate the good finite sample performance of the new bootstrap method proposed.

REFERENCES

- ANEIROS-PÉREZ, G., CAO, R. and VILAR-FERNÁNDEZ, J. M. (2011). Functional methods for time series prediction: A nonparametric approach. *J. Forecast.* **30** 377–392. [MR2815433](#)
- AUE, A., NORINHO, D. D. and HÖRMANN, S. (2015). On the prediction of stationary functional time series. *J. Amer. Statist. Assoc.* **110** 378–392. [MR3338510](#)
- BOSQ, D. (2000). *Linear Process in Function Spaces*. Springer, Berlin.
- BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. Springer, New York. [MR1093459](#)
- CEROVECKI, C. and HÖRMANN, S. (2017). On the CLT for discrete Fourier transforms of functional time series. *J. Multivariate Anal.* **154** 282–295. [MR3588570](#)
- CHENG, R. and POURAHMADI, M. (1993). Baxter’s inequality and convergence of finite predictors of multivariate stochastic processes. *Probab. Theory Related Fields* **95** 115–124.
- DEHLING, H., SHARIPOV, O. S. and WENDLER, M. (2015). Bootstrap for dependent Hilbert space valued random variables with application to von Mises statistics. *J. Multivariate Anal.* **133** 200–215.
- FERNÁNDEZ DE CASTRO, B., GUILLAS, S. and GONZÁLEZ MANTEIGA, W. (2005). Functional samples and bootstrap for predicting sulfur dioxide levels. *Technometrics* **47** 212–222.

MSC2010 subject classifications. Primary 62M10, 62M15; secondary 62G09.

Key words and phrases. Bootstrap, Fourier transform, principal components, Karhunen–Loève expansion, spectral density operator.

- FRANKE, J. and NYARIGUE, E. (2016). Residual-based bootstrap for functional autoregressions. Preprint.
- HÖRMANN, S. and KIDZIŃSKI, L. (2015). A note on estimation in Hilbertian linear models. *Scand. J. Stat.* **42** 43–62.
- HÖRMANN, S., KIDZIŃSKI, L. and HALLIN, M. (2015). Dynamic functional principal components. *J. Roy. Statist. Soc. Ser. B* **77** 319–348.
- HÖRMANN, S. and KOKOSZKA, P. (2010). Weakly dependent functional data. *Ann. Statist.* **38** 1845–1884. [MR2662361](#)
- HÖRMANN, S. and KOKOSZKA, P. (2012). Functional time series. In *Time Series Analysis-Methods and Applications* 157–186.
- HORVÁTH, L. and KOKOSZKA, P. (2012). *Inference for Functional Data with Applications*. Springer, New York. [MR2920735](#)
- HORVÁTH, L., KOKOSZKA, P. and REEDER, R. (2013). Estimation of the mean of functional time series and a two sample problem. *J. Roy. Statist. Soc. Ser. B* **75** 103–122.
- HORVÁTH, L., KOKOSZKA, P. and RICE, G. (2014). Testing stationarity of functional time series. *J. Econometrics* **179** 66–82.
- HURVICH, C. M. and TSAI, C.-L. (1993). A corrected Akaike information criterion for vector autoregressive model selection. *J. Time Series Anal.* **14** 271–279.
- HYNDMAN, R. J. and SHANG, H. L. (2009). Forecasting functional time series. *J. Korean Statist. Soc.* **38** 199–211.
- KREISS, J.-P. (1988). Asymptotic statistical inference for a class of stochastic processes. Habilitationsschrift, Univ. Hamburg.
- KREISS, J.-P. and PAPARODITIS, E. (2011). Bootstrap methods for dependent data: A review. *J. Korean Statist. Soc.* **40** 357–378.
- 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.
- LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data*. Springer, Berlin.
- LI, Y., WANG, N. and CAROROL, R. J. (2013). Selecting the number of principal components in functional data. *J. Amer. Statist. Assoc.* **108** 1284–1294.
- MCMURRY, T. and POLITIS, D. N. (2011). Resampling methods for functional data. In *The Oxford Handbook of Functional Data Analysis* (F. Ferraty and Y. Romain, eds.) 189–209. Oxford Univ. Press, London.
- MERLEVÈDE, F., PELIGRAD, M. and UTEV, S. (1997). Sharp conditions for the CLT of linear processes in a Hilbert space. *J. Theoret. Probab.* **10** 681–693. [MR1468399](#)
- MEYER, M., JENTSCH, C. and KREISS, J.-P. (2017). Baxter’s inequality and sieve bootstrap for random fields. *Bernoulli* **23** 2988–3020.
- MEYER, M. and KREISS, J.-P. (2015). On the vector autoregressive sieve bootstrap. *J. Time Series Anal.* **36** 377–397.
- MINGOTTI, N., LILLO, R. E. and ROMO, J. (2015). A random walk test for functional time series. UC3M working papers, Statistics and Econometrics.
- PANARETOS, V. and TAVAKOLI, S. (2013). Fourier analysis of stationary time series in function spaces. *Ann. Statist.* **41** 568–603.
- PAPARODITIS, E. (2018). Supplement to “Sieve bootstrap for functional time series.” DOI:[10.1214/17-AOS1667SUPP](#).
- POLITIS, D. N. and ROMANO, J. (1994). Limit theorems for weakly dependent Hilbert space valued random variables with applications to the stationary bootstrap. *Statist. Sinica* **4** 461–476.
- POURAHMADI, M. (2001). *Foundation of Time Series Analysis and Prediction Theory*. Wiley, New York.
- SHANG, L. H. (2016). Resampling methods for dependent functional data. Preprint.
- SHARIPOV, O., TEWES, J. and WENDLER, M. (2016). Sequential block bootstrap in a Hilbert space with application to change point analysis. *Canad. J. Statist.* **44** 300–322.

- WIENER, N. and MASANI, P. (1958). The prediction theory of multivariate stochastic processes, II. *Acta Math.* **99** 93–137.
- YAO, F., MÜLLER, H. G. and WANG, J. L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590.
- ZHOU, T. and POLITIS, D. N. (2016). Kernel estimation of first-order nonparametric functional autoregression model and its bootstrap approximation. Preprint.

RESTRICTED STRONG CONVEXITY IMPLIES WEAK SUBMODULARITY¹

BY ETHAN R. ELENBERG*, RAJIV KHANNA*, ALEXANDROS G. DIMAKIS*
AND SAHAND NEGAHBAN[†]

University of Texas at Austin and Yale University[†]*

We connect high-dimensional subset selection and submodular maximization. Our results extend the work of Das and Kempe [In *ICML* (2011) 1057–1064] from the setting of linear regression to arbitrary objective functions. For greedy feature selection, this connection allows us to obtain strong multiplicative performance bounds on several methods without statistical modeling assumptions. We also derive recovery guarantees of this form under standard assumptions. Our work shows that greedy algorithms perform within a constant factor from the best possible subset-selection solution for a broad class of general objective functions. Our methods allow a direct control over the number of obtained features as opposed to regularization parameters that only implicitly control sparsity. Our proof technique uses the concept of *weak submodularity* initially defined by Das and Kempe. We draw a connection between convex analysis and submodular set function theory which may be of independent interest for other statistical learning applications that have combinatorial structure.

REFERENCES

- [1] ALTSCHULER, J., BHASKARA, A., FU, G. T., MIRROKNI, V., ROSTAMIZADEH, A. and ZADIMOGHADDAM, M. (2016). Greedy column subset selection: New bounds and distributed algorithms. In *ICML* 2539–2548.
- [2] BACH, F. R. (2013). Learning with submodular functions: A convex optimization perspective. *Found. Trends Mach. Learn.* **6**.
- [3] BAHMANI, S., RAJ, B. and BOUFONOS, P. T. (2013). Greedy sparsity-constrained optimization. *J. Mach. Learn. Res.* **14** 807–841.
- [4] BARBOSA, R. D. P., ENE, A., NGUYEN, H. L. and WARD, J. (2015). The power of randomization: Distributed submodular maximization on massive datasets. In *ICML* 1236–1244.
- [5] BARRON, A. R., COHEN, A., DAHMEN, W. and DEVORE, R. A. (2008). Approximation and learning by greedy algorithms. *Ann. Statist.* **36** 64–94. [MR2387964](#)
- [6] BIAN, A. A., BUHMANN, J. M., KRAUSE, A. and TSCHIATSCHKE, S. (2017). Guarantees for greedy maximization of non-submodular functions with applications. In *ICML* 498–507.
- [7] BIAN, A. A., MIRZASOLEIMAN, B., BUHMANN, J. M. and KRAUSE, A. (2017). Guaranteed non-convex optimization: Submodular maximization over continuous domains. In *AISTATS* 111–120.
- [8] BLUMENSATH, T. and DAVIES, M. E. (2007). On the difference between orthogonal matching pursuit and orthogonal least squares. Available at <https://eprints.soton.ac.uk/142469/>.

MSC2010 subject classifications. Primary 68W25; secondary 62J02, 68W40.

Key words and phrases. Submodular functions, greedy algorithms, restricted strong convexity, subset selection.

- [9] BROWN, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory. Lecture Notes—Monograph Series 9*. IMS, Hayward, CA. [MR882001](#)
- [10] BUCHBINDER, N. and FELDMAN, M. (2016). Constrained submodular maximization via a non-symmetric technique. Available at [arXiv:1611.03253](#).
- [11] CHEN, L., FELDMAN, M. and KARBASI, A. (2017). Weakly submodular maximization beyond cardinality constraints: does randomization help greedy? Available at [arXiv:1707.04347](#).
- [12] CONFORTI, M. and CORNUÉJOLS, G. (1984). Submodular set functions, matroids and the greedy algorithm: Tight worst-case bounds and some generalizations of the Rado–Edmonds theorem. *Discrete Appl. Math.* **7** 251–274.
- [13] DAS, A. and KEMPE, D. (2011). Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection. In *ICML* 1057–1064.
- [14] DEY, D. K., GHOSH, S. K. and MALLICK, B. K., eds. (2000). *Generalized Linear Models: A Bayesian Perspective*, 1st ed. CRC Press, Boca Raton, FL.
- [15] EGHBALI, R. and FAZEL, M. (2016). Designing smoothing functions for improved worst-case competitive ratio in online optimization. In *NIPS* 3279–3287.
- [16] ELENBERG, E. R., DIMAKIS, A. G., FELDMAN, M. and KARBASI, A. (2017). Streaming weak submodularity: Interpreting neural networks on the fly. In *NIPS* 4047–4057.
- [17] FUJISHIGE, S. (2005). *Submodular Functions and Optimization*. Elsevier Science, Amsterdam.
- [18] HASSANI, H., SOLTANOLKOTABI, M. and KARBASI, A. (2017). Gradient methods for submodular maximization. In *NIPS* 5843–5853.
- [19] HAZAN, E. and KALE, S. (2012). Online submodular minimization. *J. Mach. Learn. Res.* **13** 2903–2922.
- [20] HOI, S. C. H., JIN, R., ZHU, J. and LYU, M. R. (2006). Batch mode active learning and its application to medical image classification. In *ICML* 417–424.
- [21] HOREL, T. and SINGER, Y. (2016). Maximization of approximately submodular functions. In *NIPS* 3045–3053.
- [22] IYER, R. K. and BILMES, J. (2015). Polyhedral aspects of submodularity, convexity and concavity. Available at [arXiv:1506.07329](#).
- [23] JAIN, P., TEWARI, A. and KAR, P. (2014). On iterative hard thresholding methods for high-dimensional M-estimation. In *NIPS* 685–693.
- [24] JALALI, A., JOHNSON, C. and RAVIKUMAR, P. (2011). On learning discrete graphical models using greedy methods. In *NIPS* 1935–1943.
- [25] KAKADE, S. M., SHAMIR, O., SRIDHARAN, K. and TEWARI, A. (2010). Learning exponential families in high-dimensions: Strong convexity and sparsity. In *AISTATS* 381–388.
- [26] KHANNA, R., ELENBERG, E. R., DIMAKIS, A. G., GHOSH, J. and NEGAHBAN, S. (2017). On approximation guarantees for greedy low rank optimization. In *ICML* 1837–1846.
- [27] KHANNA, R., ELENBERG, E. R., DIMAKIS, A. G., NEGAHBAN, S. and GHOSH, J. (2017). Scalable greedy feature selection via weak submodularity. In *AISTATS* 1560–1568.
- [28] KRAUSE, A. and CEVHER, V. (2010). Submodular dictionary selection for sparse representation. In *ICML* 567–574.
- [29] KRAUSE, A. and GOLOVIN, D. (2014). Submodular function maximization. In *Tractability* 71–104. Cambridge Univ. Press, Cambridge. [MR3220900](#)
- [30] LEWIS, D. D., YANG, Y., ROSE, T. G. and LI, F. (2004). RCV1: A new benchmark collection for text categorization research. *J. Mach. Learn. Res.* **5** 361–397.
- [31] LIU, J., YE, J. and FUJIMAKI, R. (2014). Forward-backward greedy algorithms for general convex smooth functions over a cardinality constraint. In *ICML* 503–511.
- [32] LOH, P.-L. and WAINWRIGHT, M. J. (2015). Regularized M-estimators with nonconvexity: Statistical and algorithmic theory for local optima. *J. Mach. Learn. Res.* **16** 559–616.

- [33] LOVÁSZ, L. (1983). Submodular functions and convexity. In *Mathematical Programming the State of the Art: Bonn 1982* 235–257. Springer, Berlin.
- [34] LOZANO, A. C., ŚWIRSZCZ, G. and ABE, N. (2011). Group orthogonal matching pursuit for logistic regression. In *AISTATS* 452–460.
- [35] MAZUMDER, R., FREUND, R. and GRIGAS, P. (2017). A new perspective on boosting in linear regression via subgradient optimization and relatives. *Ann. Statist.* **45** 2328–2364. [MR3737894](#)
- [36] MIRZASOLEIMAN, B., BADANIDIYURU, A., KARBASI, A., VONDRÁK, J. and KRAUSE, A. (2015). Lazier than lazy greedy. In *AAAI* 1812–1818.
- [37] NATARAJAN, B. K. (1995). Sparse approximate solutions to linear systems. *SIAM J. Comput.* **24** 227–234.
- [38] NEEDELL, D. and TROPP, J. A. (2009). CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.* **3** 301–321.
- [39] NEGAHBAN, S., RAVIKUMAR, P., YU, B. and WAINWRIGHT, M. J. (2010). A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers. Available at [arXiv:1010.2731v1](#).
- [40] NEGAHBAN, S., RAVIKUMAR, P., YU, B. and WAINWRIGHT, M. J. (2012). A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers. *Statist. Sci.* **27** 538–557. [MR3025133](#)
- [41] NEMHAUSER, G. L., WOLSEY, L. A. and FISHER, M. L. (1978). An analysis of approximations for maximizing submodular set functions—I. *Math. Program.* **14** 265–294.
- [42] PAN, X., JEGELKA, S., GONZALEZ, J. E., BRADLEY, J. K. and JORDAN, M. I. (2014). Parallel double greedy submodular maximization. In *NIPS* 118–126.
- [43] QIAN, C., YU, Y. and ZHOU, Z.-H. (2015). Subset selection by Pareto optimization. In *NIPS* 1774–1782.
- [44] RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2010). Restricted eigenvalue properties for correlated Gaussian designs. *J. Mach. Learn. Res.* **11** 2241–2259. [MR2719855](#)
- [45] RIGOLLET, P. and TSYBAKOV, A. (2011). Exponential screening and optimal rates of sparse estimation. *Ann. Statist.* **39** 731–771.
- [46] TEWARI, A., RAVIKUMAR, P. and DHILLON, I. S. (2011). Greedy algorithms for structurally constrained high dimensional problems. In *NIPS* 882–890.
- [47] VAN DE GEER, S. A. (2008). High-dimensional generalized linear models and the Lasso. *Ann. Statist.* **36** 614–645.
- [48] WEI, K., LIU, Y., KIRCHHOFF, K. and BILMES, J. (2013). Using document summarization techniques for speech data subset selection. In *NAACL-HLT* 721–726.
- [49] WEI, K., RISHABH, I. and BILMES, J. (2015). Submodularity in data subset selection and active learning. In *ICML* 1954–1963.
- [50] YANG, Z., WANG, Z., LIU, H., ELДАР, Y. C. and ZHANG, T. (2016). Sparse nonlinear regression: Parameter estimation and asymptotic inference. In *ICML* 2472–2481.
- [51] YUAN, X.-T., LI, P. and ZHANG, T. (2014). Gradient hard thresholding pursuit for sparsity-constrained optimization. In *ICML* 127–135.
- [52] ZHANG, T. (2008). Adaptive forward–backward greedy algorithm for sparse learning with linear models. In *NIPS* 1921–1928.
- [53] ZHANG, T. (2011). Sparse recovery with orthogonal matching pursuit under RIP. *IEEE Trans. Inform. Theory* **57** 6215–6221. [MR2857968](#)

MULTISCALE SCANNING IN INVERSE PROBLEMS¹

BY KATHARINA PROKSCH*, FRANK WERNER[†] AND AXEL MUNK*,²

*University of Göttingen** and *Max Planck Institute for Biophysical Chemistry[†]*

In this paper, we propose a multiscale scanning method to determine active components of a quantity f w.r.t. a dictionary $\mathcal{U}$ from observations Y in an inverse regression model $Y = Tf + \xi$ with linear operator T and general random error ξ . To this end, we provide uniform confidence statements for the coefficients $\langle \varphi, f \rangle$, $\varphi \in \mathcal{U}$, under the assumption that $(T^*)^{-1}(\mathcal{U})$ is of wavelet-type. Based on this, we obtain a multiple test that allows to identify the active components of $\mathcal{U}$, that is, $\langle f, \varphi \rangle \neq 0$, $\varphi \in \mathcal{U}$, at controlled, family-wise error rate. Our results rely on a Gaussian approximation of the underlying multiscale statistic with a novel scale penalty adapted to the ill-posedness of the problem. The scale penalty furthermore ensures convergence of the statistic's distribution towards a Gumbel limit under reasonable assumptions. The important special cases of tomography and deconvolution are discussed in detail. Further, the regression case, when $T = \text{id}$ and the dictionary consists of moving windows of various sizes (scales), is included, generalizing previous results for this setting. We show that our method obeys an oracle optimality, that is, it attains the same asymptotic power as a single-scale testing procedure at the correct scale. Simulations support our theory and we illustrate the potential of the method as an inferential tool for imaging. As a particular application, we discuss super-resolution microscopy and analyze experimental STED data to locate single DNA origami.

REFERENCES

- ABRAMOVICH, F. and SILVERMAN, B. W. (1998). Wavelet decomposition approaches to statistical inverse problems. *Biometrika* **85** 115–129. [MR1627226](#)
- ALBANI, V., ELBAU, P., DE HOOP, M. V. and SCHERZER, O. (2016). Optimal convergence rates results for linear inverse problems in Hilbert spaces. *Numer. Funct. Anal. Optim.* **37** 521–540.
- ANDERSSON, R. S. (1986). The linear functional strategy for improperly posed problems. In *Inverse Problems (Oberwolfach, 1986)*. *Internat. Schriftenreihe Numer. Math.* **77** 11–30. Birkhäuser, Basel. [MR0902807](#)
- ARIAS-CASTRO, E., DONOHO, D. L. and HUO, X. (2005). Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Inform. Theory* **51** 2402–2425. [MR2246369](#)
- ASPELMEIER, T., EGNER, A. and MUNK, A. (2015). Modern statistical challenges in high-resolution fluorescence microscopy. *Annu. Rev. Stat. Appl.* **2** 163–202.
- BERTERO, M., BOCCACCI, P., DESIDERÀ, G. and VICIDOMINI, G. (2009). Image deblurring with Poisson data: From cells to galaxies. *Inverse Probl.* **25** 025004.
- BISSANTZ, N., HOHAGE, T., MUNK, A. and RUYMGAART, F. (2007). Convergence rates of general regularization methods for statistical inverse problems and applications. *SIAM J. Numer. Anal.* **45** 2610–2636.

MSC2010 subject classifications. Primary 62G10; secondary 62G15, 62G20, 62G32.

Key words and phrases. Multiscale analysis, scan statistic, ill-posed problem, deconvolution, super-resolution, Gumbel extreme value limit.

- BISSANTZ, N., CLAESKENS, G., HOLZMANN, H. and MUNK, A. (2009). Testing for lack of fit in inverse regression—With applications to biophotonic imaging. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 25–48. [MR2655522](#)
- BURGER, M., FLEMMING, J. and HOFMANN, B. (2013). Convergence rates in regularization if the sparsity assumption fails. *Inverse Probl.* **29** 025013.
- BUTUCEA, C. (2007). Goodness-of-fit testing and quadratic functional estimation from indirect observations. *Ann. Statist.* **35** 1907–1930.
- BUTUCEA, C. and COMTE, F. (2009). Adaptive estimation of linear functionals in the convolution model and applications. *Bernoulli* **15** 69–98.
- BUTUCEA, C. and INGSTER, Y. I. (2013). Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli* **19** 2652–2688.
- CASTILLO, I. and NICKL, R. (2014). On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. *Ann. Statist.* **42** 1941–1969. [MR3262473](#)
- CAVALIER, L. and GOLUBEV, Y. (2006). Risk hull method and regularization by projections of ill-posed inverse problems. *Ann. Statist.* **34** 1653–1677. [MR2283712](#)
- CAVALIER, L. and TSYBAKOV, A. (2002). Sharp adaptation for inverse problems with random noise. *Probab. Theory Related Fields* **123** 323–354.
- CAVALIER, L., GOLUBEV, Y., LEPSKI, O. and TSYBAKOV, A. (2003). Block thresholding and sharp adaptive estimation in severely ill-posed inverse problems. *Teor. Veroyatn. Primen.* **48** 534–556.
- CHAN, H. P. and WALTHER, G. (2013). Detection with the scan and the average likelihood ratio. *Statist. Sinica* **23** 409–428.
- CHERNOUSOVA, E. and GOLUBEV, Y. (2014). Spectral cut-off regularizations for ill-posed linear models. *Math. Methods Statist.* **23** 116–131.
- CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564–1597. [MR3262461](#)
- COHEN, A., HOFFMANN, M. and REISS, M. (2004). Adaptive wavelet Galerkin methods for linear inverse problems. *SIAM J. Numer. Anal.* **42** 1479–1501.
- DEDECKER, J., MERLEVÈDE, F. and RIO, E. (2014). Strong approximation of the empirical distribution function for absolutely regular sequences in $\mathbb{R}^d$. *Electron. J. Probab.* **19** 1–56.
- DICKHAUS, T. (2014). *Simultaneous Statistical Inference: With Applications in the Life Sciences*. Springer, Heidelberg.
- DONOHO, D. L. (1995). Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Appl. Comput. Harmon. Anal.* **2** 101–126.
- DÜMBGEN, L. and SPOKOINY, V. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29** 124–152.
- DÜMBGEN, L. and WALTHER, G. (2008). Multiscale inference about a density. *Ann. Statist.* **36** 1758–1785.
- ECKLE, K., BISSANTZ, N. and DETTE, H. (2017). Multiscale inference for multivariate deconvolution. *Electron. J. Stat.* **11** 4179–4219. [MR3716498](#)
- ECKLE, K., BISSANTZ, N., DETTE, H., PROKSCH, K. and EINECKE, S. (2018). Multiscale inference for a multivariate density with applications to X-ray astronomy. *Ann. Inst. Statist. Math.* To appear. DOI:[10.1007/s10463-017-0605-1](#).
- FAN, J. (1991). Asymptotic normality for deconvolution kernel density estimators. *Sankhyā, Ser. A* **53** 97–110.
- FRIEDENBERG, D. A. and GENOVESE, C. R. (2013). Straight to the source: Detecting aggregate objects in astronomical images with proper error control. *J. Amer. Statist. Assoc.* **108** 456–468.
- GENOVESE, C. R., PERONE-PACIFICO, M., VERDINELLI, I. and WASSERMAN, L. (2012). The geometry of nonparametric filament estimation. *J. Amer. Statist. Assoc.* **107** 788–799.
- GOLDENSHLUGER, A. (1999). On pointwise adaptive nonparametric deconvolution. *Bernoulli* **5** 907–925.

- HELL, S. (2007). Far-field optical nanoscopy. *Science* **316** 1153–1158.
- HELL, S. W. and WICHMANN, J. (1994). Breaking the diffraction resolution limit by stimulated emission: Stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.* **19** 780–782.
- HOHAGE, T. and WERNER, F. (2016). Inverse problems with Poisson data: Statistical regularization theory, applications and algorithms. *Inverse Probl.* **32** 093001.
- HOLZMANN, H., BISSANTZ, N. and MUNK, A. (2007). Density testing in a contaminated sample. *J. Multivariate Anal.* **98** 57–75. [MR2292917](#)
- INGSTER, Y. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. I–III. *Math. Methods Statist.* **2** 85–114, 171–189, 249–268.
- INGSTER, Y., LAURENT, B. and MARTEAU, C. (2014). Signal detection for inverse problems in a multidimensional framework. *Math. Methods Statist.* **23** 279–305.
- INGSTER, Y. I., SAPATINAS, T. and SUSLINA, I. A. (2012). Minimax signal detection in ill-posed inverse problems. *Ann. Statist.* **40** 1524–1549. [MR3015034](#)
- JOHNSTONE, I. M. and PAUL, D. (2014). Adaptation in some linear inverse problems. *Stat* **3** 187–199.
- JOHNSTONE, I. M. and SILVERMAN, B. W. (1991). Discretization effects in statistical inverse problems. *J. Complexity* **7** 1–34.
- JOHNSTONE, I. M., KERKYACHARIAN, G., PICARD, D. and RAIMONDO, M. (2004). Wavelet deconvolution in a periodic setting. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 547–573.
- KABLUCHKO, Z. (2011). Extremes of the standardized Gaussian noise. *Stochastic Process. Appl.* **121** 515–533. [MR2763094](#)
- KAZANTSEV, I., LEMAHIEU, I., SALOV, G. and DENYS, R. (2002). Statistical detection of defects in radiographic images in nondestructive testing. *Signal Process.* **82** 791–801.
- KERKYACHARIAN, G., KYRIAZIS, G., LE PENNEC, E., PETRUSHEV, P. and PICARD, D. (2010). Inversion of noisy Radon transform by SVD based needlets. *Appl. Comput. Harmon. Anal.* **28** 24–45.
- KLAR, T. A. and HELL, S. W. (1999). Subdiffraction resolution in far-field fluorescence microscopy. *Opt. Lett.* **24** 954–956.
- KNAPIK, B. T., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2011). Bayesian inverse problems with Gaussian priors. *Ann. Statist.* **39** 2626–2657. [MR2906881](#)
- KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131.
- KOU, J. (2017). Identifying the support of rectangular signals in Gaussian noise. Preprint. Available at [arXiv:1703.06226](#).
- LAURENT, B., LOUBES, J.-M. and MARTEAU, C. (2011). Testing inverse problems: A direct or an indirect problem? *J. Statist. Plann. Inference* **141** 1849–1861.
- LAURENT, B., LOUBES, J.-M. and MARTEAU, C. (2012). Non asymptotic minimax rates of testing in signal detection with heterogeneous variances. *Electron. J. Stat.* **6** 91–122.
- LI, H., MUNK, A., SIELING, H. and WALTHER, G. (2016). The essential histogram. Preprint. Available at [arXiv:1612.07216](#).
- LIN, G. D. (2017). Recent developments on the moment problem. Preprint. Available at [arXiv:1703.01027](#).
- MAIR, B. A. and RUYMGAART, F. H. (1996). Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.* **56** 1424–1444.
- MARTEAU, C. and MATHÉ, P. (2014). General regularization schemes for signal detection in inverse problems. *Math. Methods Statist.* **23** 176–200.
- MATHÉ, P. and PEREVERZEV, S. V. (2002). Direct estimation of linear functionals from indirect noisy observations. *J. Complexity* **18** 500–516. [MR1919446](#)
- MEISTER, A. (2009). *Deconvolution Problems in Nonparametric Statistics. Lecture Notes in Statistics* **193**. Springer, Berlin.

- NATTERER, F. (1986). *The Mathematics of Computerized Tomography*. B. G. Teubner, Stuttgart.
- NICKL, R. and REISS, M. (2012). A Donsker theorem for Lévy measures. *J. Funct. Anal.* **263** 3306–3332.
- NIKOL'SKIĬ, S. M. (1951). Inequalities for entire functions of finite degree and their application in the theory of differentiable functions of several variables. In *Trudy Mat. Inst. Steklov.* **38** 244–278. Izdat. Akad. Nauk SSSR, Moscow.
- O'SULLIVAN, F. (1986). A statistical perspective on ill-posed inverse problems. *Statist. Sci.* **1** 502–527.
- PICKANDS, J. III (1969). Upcrossing probabilities for stationary Gaussian processes. *Trans. Amer. Math. Soc.* **145** 51–73.
- PITERBARG, V. I. (1996). *Asymptotic Methods in the Theory of Gaussian Processes and Fields. Translations of Mathematical Monographs* **148**. Amer. Math. Soc., Providence, RI.
- PROKSCH, K., WERNER, F. and MUNK, A. (2018). Supplement to “Multiscale scanning in inverse problems.” DOI:[10.1214/17-AOS1669SUPP](https://doi.org/10.1214/17-AOS1669SUPP).
- RAY, K. (2013). Bayesian inverse problems with non-conjugate priors. *Electron. J. Stat.* **7** 2516–2549.
- RAY, K. (2017). Adaptive Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.* **45** 2511–2536. [MR3737900](#)
- RIO, E. (1993). Strong approximation for set-indexed partial-sum processes, via KMT constructions. II. *Ann. Probab.* **21** 1706–1727. [MR1235436](#)
- ROHDE, A. (2008). Adaptive goodness-of-fit tests based on signed ranks. *Ann. Statist.* **36** 1346–1374. [MR2418660](#)
- RUFIBACH, K. and WALTHER, G. (2010). The block criterion for multiscale inference about a density, with applications to other multiscale problems. *J. Comput. Graph. Statist.* **19** 175–190.
- SCHMIDT-HIEBER, J., MUNK, A. and DÜMBGEN, L. (2013). Multiscale methods for shape constraints in deconvolution: Confidence statements for qualitative features. *Ann. Statist.* **41** 1299–1328. [MR3113812](#)
- SCHWARTZMAN, A., DOUGHERTY, R. F. and TAYLOR, J. E. (2008). False discovery rate analysis of brain diffusion direction maps. *Ann. Appl. Stat.* **2** 153–175. [MR2415598](#)
- SHARPNACK, J. and ARIAS-CASTRO, E. (2016). Exact asymptotics for the scan statistic and fast alternatives. *Electron. J. Stat.* **10** 2641–2684.
- SÖHL, J. and TRABS, M. (2012). A uniform central limit theorem and efficiency for deconvolution estimators. *Electron. J. Stat.* **6** 2486–2518.
- TA, H., KELLER, J., HALTMEIER, M., SAKA, S. K., SCHMIED, J., OPAZO, F., TINNEFELD, P., MUNK, A. and HELL, S. W. (2015). Mapping molecules in scanning far-field fluorescence nanoscopy. *Nat. Commun.* **6** 7977.
- TSYBAKOV, A. (2000). On the best rate of adaptive estimation in some inverse problems. *C. R. Acad. Sci. Paris Sér. I Math.* **330** 835–840. [MR1769957](#)
- TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York.
- WALTHER, G. (2010). Optimal and fast detection of spatial clusters with scan statistics. *Ann. Statist.* **38** 1010–1033. [MR2604703](#)
- WILLER, T. (2009). Optimal bounds for inverse problems with Jacobi-type eigenfunctions. *Statist. Sinica* **19** 785–800. [MR2514188](#)

SLOPE MEETS LASSO: IMPROVED ORACLE BOUNDS AND OPTIMALITY¹

BY PIERRE C. BELLEC^{*,†,§}, GUILLAUME LECUÉ^{*,†,‡} AND
ALEXANDRE B. TSYBAKOV^{*,†}

ENSAE^{}, CREST (UMR CNRS 9194)[†], CNRS[‡]
and Rutgers University[§]*

We show that two polynomial time methods, a Lasso estimator with adaptively chosen tuning parameter and a Slope estimator, adaptively achieve the minimax prediction and ℓ_2 estimation rate $(s/n)\log(p/s)$ in high-dimensional linear regression on the class of s -sparse vectors in $\mathbb{R}^p$. This is done under the Restricted Eigenvalue (RE) condition for the Lasso and under a slightly more constraining assumption on the design for the Slope. The main results have the form of sharp oracle inequalities accounting for the model misspecification error. The minimax optimal bounds are also obtained for the ℓ_q estimation errors with $1 \leq q \leq 2$ when the model is well specified. The results are nonasymptotic, and hold both in probability and in expectation. The assumptions that we impose on the design are satisfied with high probability for a large class of random matrices with independent and possibly anisotropically distributed rows. We give a comparative analysis of conditions, under which oracle bounds for the Lasso and Slope estimators can be obtained. In particular, we show that several known conditions, such as the RE condition and the sparse eigenvalue condition are equivalent if the ℓ_2 -norms of regressors are uniformly bounded.

REFERENCES

- [1] ABRAMOVICH, F. and GRINSHTEIN, V. (2010). MAP model selection in Gaussian regression. *Electron. J. Stat.* **4** 932–949. [MR2721039](#)
- [2] ABRAMOWITZ, M. and STEGUN, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. National Bureau of Standards Applied Mathematics Series **55**. Washington, DC. [MR0167642](#)
- [3] BERTHET, Q. and RIGOLLET, P. (2013). Optimal detection of sparse principal components in high dimension. *Ann. Statist.* **41** 1780–1815.
- [4] BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of Lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. [MR2533469](#)
- [5] BOGDAN, M., VAN DEN BERG, E., SABATTI, C., SU, W. and CANDÈS, E. J. (2015). SLOPE—Adaptive variable selection via convex optimization. *Ann. Appl. Stat.* **9** 1103–1140.
- [6] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford Univ. Press, London.

MSC2010 subject classifications. Primary 60K35, 62G08; secondary 62C20, 62G05, 62G20.

Key words and phrases. Sparse linear regression, minimax rates, high-dimensional statistics, Slope, Lasso.

- [7] CANDÈS, E. J. and DAVENPORT, M. A. (2013). How well can we estimate a sparse vector? *Appl. Comput. Harmon. Anal.* **34** 317–323. [MR3008569](#)
- [8] CANDÈS, E. J. and TAO, T. (2006). Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inform. Theory* **52** 5406–5425.
- [9] CHAFAÏ, D., GUÉDON, O., LECUÉ, G. and PAJOR, A. (2012). *Interactions Between Compressed Sensing Random Matrices and High Dimensional Geometry. Panoramas et Synthèses* **37**. Société Mathématique de France, Paris. [MR3113826](#)
- [10] DALALYAN, A. S., HEBIRI, M., LEDERER, J. et al. (2017). On the prediction performance of the Lasso. *Bernoulli* **23** 552–581.
- [11] DIRKSEN, S. (2015). Tail bounds via generic chaining. *Electron. J. Probab.* **20** no. 53, 29. [MR3354613](#)
- [12] GIRAUD, C. (2015). *Introduction to High-Dimensional Statistics. Monographs on Statistics and Applied Probability*. **139**. CRC Press, Boca Raton, FL.
- [13] KOLTCHINSKII, V. (2011). *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems. Ecole D'Ete de Probabilites de Saint-Flour XXXVIII-2008*. Springer, New York.
- [14] 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](#)
- [15] KOLTCHINSKII, V. and MENDELSON, S. (2015). Bounding the smallest singular value of a random matrix without concentration. *Int. Math. Res. Not. IMRN* **23** 12991–13008. [MR3431642](#)
- [16] LECUÉ, G. and MENDELSON, S. (2015). Regularization and the small-ball method I: Sparse recovery Technical report, CNRS, ENSAE and Technion, I.I.T.
- [17] LECUÉ, G. and MENDELSON, S. (2017). Sparse recovery under weak moment assumptions. *J. Eur. Math. Soc. (JEMS)* **19** 881–904. [MR3612870](#)
- [18] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes. Ergebnisse der Mathematik und Ihrer Grenzgebiete (3)* **23**. Springer, Berlin. [MR1102015](#)
- [19] LOUNICI, K., PONTIL, M., TSYBAKOV, A. B. and VAN DE GEER, S. A. (2011). Oracle inequalities and optimal inference under group sparsity. *Ann. Statist.* **39** 2164–2204.
- [20] MENDELSON, S. (2014). Learning without concentration. In *Proceedings of the 27th Annual Conference on Learning Theory COLT14* 25–39.
- [21] MENDELSON, S. (2016). Upper bounds on product and multiplier empirical processes. *Stochastic Process. Appl.* **126** 3652–3680.
- [22] MENDELSON, S. (2015). Learning without concentration. *J. ACM* **62** Art. 21, 25.
- [23] RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2011). Minimax rates of estimation for high-dimensional linear regression over ℓ_q -balls. *IEEE Trans. Inform. Theory* **57** 6976–6994. [MR2882274](#)
- [24] RIGOLLET, P. and TSYBAKOV, A. (2011). Exponential screening and optimal rates of sparse estimation. *Ann. Statist.* **39** 731–771. [MR2816337](#)
- [25] SU, W. and CANDÈS, E. (2016). SLOPE is adaptive to unknown sparsity and asymptotically minimax. *Ann. Statist.* **44** 1038–1068. [MR3485953](#)
- [26] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York. [MR2724359](#)
- [27] VERZELEN, N. (2012). Minimax risks for sparse regressions: Ultra-high dimensional phenomena. *Electron. J. Stat.* **6** 38–90.
- [28] WITOLD, B. (2013). Concentration via chaining method and its applications. Technical report, Univ. Warsaw. Available at [arXiv:1405.0676](#).
- [29] YE, F. and ZHANG, C.-H. (2010). Rate minimaxity of the Lasso and Dantzig selector for the ℓ_q loss in ℓ_r balls. *J. Mach. Learn. Res.* **11** 3519–3540.

UNIFORMLY VALID POST-REGULARIZATION CONFIDENCE REGIONS FOR MANY FUNCTIONAL PARAMETERS IN Z-ESTIMATION FRAMEWORK

BY ALEXANDRE BELLONI^{*}, VICTOR CHERNOZHUKOV[†],
DENIS CHETVERIKOV[‡] AND YING WEI[§]

Duke University^{}, MIT[†], UCLA[‡] and Columbia University[§]*

In this paper, we develop procedures to construct simultaneous confidence bands for $\tilde{p}$ potentially infinite-dimensional parameters after model selection for general moment condition models where $\tilde{p}$ is potentially much larger than the sample size of available data, n . This allows us to cover settings with functional response data where each of the $\tilde{p}$ parameters is a function. The procedure is based on the construction of score functions that satisfy Neyman orthogonality condition approximately. The proposed simultaneous confidence bands rely on uniform central limit theorems for high-dimensional vectors (and not on Donsker arguments as we allow for $\tilde{p} \gg n$). To construct the bands, we employ a multiplier bootstrap procedure which is computationally efficient as it only involves resampling the estimated score functions (and does not require resolving the high-dimensional optimization problems). We formally apply the general theory to inference on regression coefficient process in the distribution regression model with a logistic link, where two implementations are analyzed in detail. Simulations and an application to real data are provided to help illustrate the applicability of the results.

REFERENCES

- [1] ANDREWS, D. W. K. (1994). Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica* **62** 43–72. [MR1258665](#)
- [2] BELLONI, A., CHEN, D., CHERNOZHUKOV, V. and HANSEN, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* **80** 2369–2429. [MR3001131](#)
- [3] BELLONI, A. and CHERNOZHUKOV, V. (2011). ℓ_1 -Penalized quantile regression for high dimensional sparse models. *Ann. Statist.* **39** 82–130. [MR2797841](#)
- [4] BELLONI, A. and CHERNOZHUKOV, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli* **19** 521–547. Available at [arXiv:1001.0188](#).
- [5] BELLONI, A., CHERNOZHUKOV, V., CHETVERIKOV, D. and WEI, Y. (2018). Supplement to “Uniformly valid post-regularization confidence regions for many functional parameters in z-estimation framework.” DOI:[10.1214/17-AOS1671SUPP](#).
- [6] BELLONI, A., CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. and HANSEN, C. (2013). Program evaluation with high-dimensional data. Available at [arXiv:1311.2645](#).
- [7] BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2010). Lasso methods for Gaussian instrumental variables models. Available at [arXiv:1012.1297](#).

MSC2010 subject classifications. Primary 62-07; secondary 62H99.

Key words and phrases. Inference after model selection, moment condition models with a continuum of target parameters, Lasso and Post-Lasso with functional response data.

- [8] BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2013). Inference for high-dimensional sparse econometric models. In *Advances in Economics and Econometrics. 10th World Congress of Econometric Society, August 2010, Vol. III.* 245–295. Available at [arXiv:1201.0220](#).
- [9] 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](#)
- [10] BELLONI, A., CHERNOZHUKOV, V. and KATO, K. (2013). Valid post-selection inference in high-dimensional approximately sparse quantile regression models. Available at [arXiv:1312.7186](#).
- [11] BELLONI, A., CHERNOZHUKOV, V. and KATO, K. (2015). Uniform post selection inference for LAD regression models and other Z-estimators. *Biometrika* **102** 77–94.
- [12] BELLONI, A., CHERNOZHUKOV, V. and WANG, L. (2011). Square-root-lasso: Pivotal recovery of sparse signals via conic programming. *Biometrika* **98** 791–806.
- [13] BELLONI, A., CHERNOZHUKOV, V. and WANG, L. (2014). Pivotal estimation via square-root Lasso in nonparametric regression. *Ann. Statist.* **42** 757–788. [MR3210986](#)
- [14] CHAMBERLAIN, G. (1992). Efficiency bounds for semiparametric regression. *Econometrica* **60** 567–596.
- [15] 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.
- [16] 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.
- [17] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Anti-concentration and honest, adaptive confidence bands. *Ann. Statist.* **42** 1787–1818.
- [18] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **4** 2309–2352.
- [19] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564–1597.
- [20] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* **162** 47–70.
- [21] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings. Available at [arXiv:1502.00352](#).
- [22] CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. and MELLY, B. (2013). Inference on counterfactual distributions. *Econometrica* **81** 2205–2268.
- [23] CHERNOZHUKOV, V., HANSEN, C. and SPINDLER, M. (2015). Post-selection and post-regularization inference in linear models with very many controls and instruments. *Am. Econ. Rev. Pap. Proc.* **105** 486–490.
- [24] DENG, H. and ZHANG, C.-H. (2017). Beyond Gaussian approximation: Bootstrap for maxima of sums of independent random vectors. Available at [arXiv:1705.09528](#).
- [25] DUDLEY, R. (1999). *Uniform Central Limit Theorems. Cambridge Studies in Advanced Mathematics* **63**. Cambridge Univ. Press, Cambridge.
- [26] HOTHORN, T., KNEIB, T. and BÜHLMANN, P. (2014). Conditional transformation models. *J. Roy. Statist. Soc. Ser. B* **76** 3–27.
- [27] JAVANMARD, A. and MONTANARI, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909.

- [28] JAVANMARD, A. and MONTANARI, A. (2014). Hypothesis testing in high-dimensional regression under the Gaussian random design model: Asymptotic theory. *IEEE Trans. Inform. Theory* **60** 6522–6554.
- [29] KOSOROK, M. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer, Berlin.
- [30] LEEB, H. and PÖTSCHER, B. (2008). Can one estimate the unconditional distribution of post-model-selection estimators? *Econometric Theory* **24** 338–376.
- [31] LEEB, H. and PÖTSCHER, B. (2008). Recent developments in model selection and related areas. *Econometric Theory* **24** 319–322.
- [32] LEEB, H. and PÖTSCHER, B. M. (2008). Sparse estimators and the oracle property, or the return of Hodges’ estimator. *J. Econometrics* **142** 201–211. [MR2394290](#)
- [33] LINTON, O. (1996). Edgeworth approximation for MINPIN estimators in semiparametric regression models. *Econometric Theory* **12** 30–60. Cowles Foundation Discussion Papers 1086 (1994).
- [34] MAMMEN, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Ann. Statist.* **21** 255–285.
- [35] NEWEY, W. (1990). Semiparametric efficiency bounds. *J. Appl. Econometrics* **5** 99–135.
- [36] NEWEY, W. (1994). The asymptotic variance of semiparametric estimators. *Econometrica* **62** 1349–1382.
- [37] NEYMAN, J. (1959). Optimal asymptotic tests of composite statistical hypotheses. In *Probability and Statistics: The Harald Cramér Volume* (U. Grenander, ed.) 213–234. Almqvist & Wiksell, Stockholm. [MR0112201](#)
- [38] NEYMAN, J. (1979). $c(\alpha)$ tests and their use. *Sankhyā* **41** 1–21.
- [39] NING, Y. and LIU, H. (2014). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. Available at [arXiv:1412.8765](#).
- [40] PÖTSCHER, B. and LEEB, H. (2009). On the distribution of penalized maximum likelihood estimators: The LASSO, SCAD, and thresholding. *J. Multivariate Anal.* **100** 2065–2082.
- [41] ROBINS, J. and ROTNITZKY, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *J. Amer. Statist. Assoc.* **90** 122–129.
- [42] STEIN, C. (1956). Efficient nonparametric testing and estimation. In *Proc. 3rd Berkeley Symp. Math. Statist. and Probab.* **1** 187–195. Univ. California Press, Berkeley, CA. [MR0084921](#)
- [43] 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.
- [44] VAN DER VAART, A. (1998). *Asymptotic Statistics*. Cambridge Univ. Press, Cambridge.
- [45] VAN DER VAART, A. and WELLNER, J. (1996). *Weak Convergence and Empirical Processes*.
- [46] ZHANG, C.-H. and ZHANG, S. (2014). Confidence intervals for low-dimensional parameters with high-dimensional data. *J. Roy. Statist. Soc. Ser. B* **76** 217–242.
- [47] ZHAO, T., KOLAR, M. and LIU, H. (2014). A general framework for robust testing and confidence regions in high-dimensional quantile regression. Available at [arXiv:1412.8724](#).

LOCAL ASYMPTOTIC EQUIVALENCE OF PURE STATES ENSEMBLES AND QUANTUM GAUSSIAN WHITE NOISE

BY CRISTINA BUTUCEA, MĂDĂLIN GUȚĂ AND MICHAEL NUSSBAUM¹

Université Paris-Saclay, University of Nottingham and Cornell University

Dedicated to Richard D. Gill on the occasion of his 66th birthday

Quantum technology is increasingly relying on specialised statistical inference methods for analysing quantum measurement data. This motivates the development of “quantum statistics”, a field that is shaping up at the overlap of quantum physics and “classical” statistics. One of the less investigated topics to date is that of statistical inference for infinite dimensional quantum systems, which can be seen as quantum counterpart of nonparametric statistics. In this paper, we analyse the asymptotic theory of quantum statistical models consisting of ensembles of quantum systems which are identically prepared in a pure state. In the limit of large ensembles, we establish the local asymptotic equivalence (LAE) of this i.i.d. model to a quantum Gaussian white noise model. We use the LAE result in order to establish minimax rates for the estimation of pure states belonging to Hermite–Sobolev classes of wave functions. Moreover, for quadratic functional estimation of the same states we note an elbow effect in the rates, whereas for testing a pure state a sharp parametric rate is attained over the nonparametric Hermite–Sobolev class.

REFERENCES

- [1] ARTILES, L. M., GILL, R. D. and GUȚĂ, M. I. (2005). An invitation to quantum tomography. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 109–134. [MR2136642](#)
- [2] AUDENAERT, K. M. R., CALSAMIGLIA, J., MUÑOZ-TAPIA, R., BAGAN, E., MASANES, L., ACIN, A. and VERSTRAETE, F. (2007). Discriminating states: The quantum Chernoff bound. *Phys. Rev. Lett.* **98** 160501.
- [3] AUDENAERT, K. M. R., NUSSBAUM, M., SZKOLA, A. and VERSTRAETE, F. (2008). Asymptotic error rates in quantum hypothesis testing. *Comm. Math. Phys.* **279** 251–283. [MR2377635](#)
- [4] BAGAN, E., BALLESTER, M. A., GILL, R. D., MUÑOZ-TAPIA, R. and ROMERO-ISART, O. (2006). Separable measurement estimation of density matrices and its fidelity gap with collective protocols. *Phys. Rev. Lett.* **97** 130501.
- [5] BAGAN, E., MONRAS, A. and MUÑOZ-TAPIA, R. (2005). Comprehensive analysis of quantum pure-state estimation for two-level systems. *Phys. Rev. A* **71** 062318.
- [6] BARNDORFF-NIELSEN, O. E., GILL, R. and JUPP, P. E. (2003). On quantum statistical inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **65** 775–816. [MR2017871](#)

MSC2010 subject classifications. Primary 62B15; secondary 62G05, 62G10, 81P50.

Key words and phrases. Le Cam distance, local asymptotic equivalence, quantum Gaussian process, quantum Gaussian sequence, quantum states ensemble, nonparametric estimation, quadratic functionals, nonparametric sharp testing rates.

- [7] BONGIOANNI, B. (2011). Sobolev spaces diversification. *Rev. Un. Mat. Argentina* **52** 23–34. [MR2952949](#)
- [8] BONGIOANNI, B. and TORREA, J. L. (2006). Sobolev spaces associated to the harmonic oscillator. *Proc. Indian Acad. Sci. Math. Sci.* **116** 337–360. [MR2256010](#)
- [9] BRAUNSTEIN, S. L. and CAVES, C. M. (1994). Statistical distance and the geometry of quantum states. *Phys. Rev. Lett.* **72** 3439–3443.
- [10] BUTUCEA, C., GUȚĂ, M. and ARTILES, L. (2007). Minimax and adaptive estimation of the Wigner function in quantum homodyne tomography with noisy data. *Ann. Statist.* **35** 465–494. [MR2336856](#)
- [11] BUTUCEA, C., GUȚĂ, M. and KYPRAIOS, T. (2015). Spectral thresholding quantum tomography for low rank states. *New J. Phys.* **17** 113050.
- [12] BUTUCEA, C., GUȚĂ, M. and NUSSBAUM, M. (2018). Supplement to “Local asymptotic equivalence of pure states ensembles and quantum Gaussian white noise”. DOI: [10.1214/17-AOS1672SUPP](#).
- [13] CAI, T., KIM, D., WANG, Y., YUAN, M. and ZHOU, H. H. (2016). Optimal large-scale quantum state tomography with Pauli measurements. *Ann. Statist.* **44** 682–712. [MR3476614](#)
- [14] CAI, T. T. and LOW, M. G. (2005). Nonquadratic estimators of a quadratic functional. *Ann. Statist.* **33** 2930–2956. [MR2253108](#)
- [15] CARPENTIER, A., EISERT, J., GROSS, D. and NICKL, R. Uncertainty quantification for matrix compressed sensing and quantum tomography problems. Available at [arXiv:1504.03234](#).
- [16] CHRISTANDL, M. and RENNER, R. (2012). Reliable quantum state tomography. *Phys. Rev. Lett.* **109** 120403.
- [17] ERMAKOV, M. S. (1990). Minimax detection of a signal in Gaussian white noise. *Teor. Veroyatn. Primen.* **35** 704–715. [MR1090496](#)
- [18] FAIST, P. and RENNER, R. (2016). Practical and reliable error bars in quantum tomography. *Phys. Rev. Lett.* **117** 010404.
- [19] FLAMMIA, S. T., GROSS, D., LIU, Y.-K. and EISERT, J. (2012). Quantum tomography via compressed sensing: Error bounds, sample complexity, and efficient estimators. *New J. Phys.* **14** 095022.
- [20] FUCHS, C. A. and VAN DE GRAAF, J. (1999). Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Trans. Inform. Theory* **45** 1216–1227. [MR1686254](#)
- [21] GILL, R. D. and GUȚĂ, M. I. (2013). On asymptotic quantum statistical inference. In *From Probability to Statistics and Back: High-Dimensional Models and Processes. Inst. Math. Stat. (IMS) Collect.* **9** 105–127. IMS, Beachwood, OH. [MR3186752](#)
- [22] GILL, R. D. and MASSAR, S. (2000). State estimation for large ensembles. *Phys. Rev. A* **61** 042312.
- [23] GROSS, D. (2011). Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inform. Theory* **57** 1548–1566. [MR2815834](#)
- [24] GROSS, D., LIU, Y.-K., FLAMMIA, S. T., BECKER, S. and EISERT, J. (2010). Quantum state tomography via compressed sensing. *Phys. Rev. Lett.* **105** 150401.
- [25] GUȚĂ, M., JANSSENS, B. and KAHN, J. (2008). Optimal estimation of qubit states with continuous time measurements. *Comm. Math. Phys.* **277** 127–160. [MR2357428](#)
- [26] GUȚĂ, M. and JENČOVÁ, A. (2007). Local asymptotic normality in quantum statistics. *Comm. Math. Phys.* **276** 341–379. [MR2346393](#)
- [27] GUȚĂ, M. and KAHN, J. (2006). Local asymptotic normality for qubit states. *Phys. Rev. A* **73** 052108.
- [28] GUTA, M. and KAHN, J. (2008). Local asymptotic normality and optimal estimation for d-dimensional quantum systems. In *Quantum Stochastics and Information: Statistics, Filtering and Control* 300–322. World Scientific, Singapore. [MR2562438](#)

- [29] HÄFFNER, H., HÄNSEL, W., ROOS, C. F., BENHELM, J., CHEK-AL-KAR, D., CHWALLA, M., KÖRBER, T., RAPOL, U. D., RIEBE, M., SCHMIDT, P. O., BECHER, C., GÜHNE, O., DÜR, W. and BLATT, R. (2005). Scalable multiparticle entanglement of trapped ions. *Nature* **438** 643–646.
- [30] HAYASHI, M. (1998). Asymptotic estimation theory for a finite-dimensional pure state model. *J. Phys. A: Math. Gen.* **31** 4633. [MR1631191](#)
- [31] HAYASHI, M. (2005). *Asymptotic Theory of Quantum Statistical Inference: Selected Papers*. World Scientific, Singapore.
- [32] HELSTROM, C. W. (1976). *Quantum Detection and Estimation Theory*. Academic Press, New York.
- [33] HIAI, F. and PETZ, D. (1991). The proper formula for relative entropy and its asymptotics in quantum probability. *Comm. Math. Phys.* **143** 99–114. [MR1139426](#)
- [34] HOLEVO, A. S. (1982). *Probabilistic and Statistical Aspects of Quantum Theory*. North-Holland Series in Statistics and Probability **1**. North-Holland, Amsterdam. [MR0681693](#)
- [35] INGSTER, Y. and STEPANOVA, N. (2011). Estimation and detection of functions from anisotropic Sobolev classes. *Electron. J. Stat.* **5** 484–506. [MR2813552](#)
- [36] INGSTER, Y. I. and SUSLINA, I. A. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*. *Lecture Notes in Statistics* **169**. Springer, New York. [MR1991446](#)
- [37] JENČOVÁ, A. and PETZ, D. (2006). Sufficiency in quantum statistical inference. *Comm. Math. Phys.* **263** 259–276. [MR2207329](#)
- [38] KAHN, J. and GUȚĂ, M. (2009). Local asymptotic normality for finite dimensional quantum systems. *Comm. Math. Phys.* **289** 597–652. [MR2506764](#)
- [39] KARGIN, V. (2005). On the Chernoff bound for efficiency of quantum hypothesis testing. *Ann. Statist.* **33** 959–976. [MR2163165](#)
- [40] KLEMELÄ, J. (2006). Sharp adaptive estimation of quadratic functionals. *Probab. Theory Related Fields* **134** 539–564. [MR2214904](#)
- [41] KOLTCHINSKII, V. (2011). Von Neumann entropy penalization and low-rank matrix estimation. *Ann. Statist.* **39** 2936–2973. [MR3012397](#)
- [42] KOLTCHINSKII, V. and XIA, D. (2015). Optimal estimation of low rank density matrices. *J. Mach. Learn. Res.* **16** 1757–1792. [MR3417798](#)
- [43] LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York. [MR0856411](#)
- [44] LEONHARDT, U. (1997). *Measuring the Quantum State of Light*. Cambridge Univ. Press, Cambridge.
- [45] LI, K. (2016). Discriminating quantum states: The multiple Chernoff distance. *Ann. Statist.* **44** 1661–1679. [MR3519936](#)
- [46] NAIR, R., GUHA, S. and TAN, S.-H. (2014). Realizable receivers for discriminating coherent and multicopy quantum states near the quantum limit. *Phys. Rev. A* **89** 032318.
- [47] NIELSEN, M. A. and CHUANG, I. L. (2000). *Quantum Computation and Quantum Information*. Cambridge Univ. Press, Cambridge. [MR1796805](#)
- [48] NUSSBAUM, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24** 2399–2430. [MR1425959](#)
- [49] NUSSBAUM, M. and SZKOLA, A. (2006). The Chernoff lower bound for symmetric quantum hypothesis testing. *Ann. Statist.* **37** 1040–1057. [MR2502660](#)
- [50] OGAWA, T. and NAGAOKA, H. (2000). Strong converse and Stein’s lemma in quantum hypothesis testing. *IEEE Trans. Inform. Theory* **46** 2428–2433. [MR1806811](#)
- [51] PARIS, M. G. A. and ŘEHÁČEK, J. (2004). Quantum state estimation. In *Quantum State Estimation*. *Lecture Notes in Physics* **649**. Springer, Berlin.
- [52] PARTHASARATHY, K. R. (1992). *An Introduction to Quantum Stochastic Calculus*. *Modern Birkhäuser Classics*. Birkhäuser/Springer Basel AG, Basel. [MR1164866](#)

- [53] RAY, K. and SCHMIDT-HIEBER, J. (2017). The Le Cam distance between density estimation, Poisson processes and Gaussian white noise. Available at [arXiv:1608.01824](#) [math.ST].
- [54] STRASSER, H. (1985). *Mathematical Theory of Statistics*. de Gruyter, Berlin. [MR812467](#)
- [55] TEMME, K. and VERSTRAETE, F. (2015). Quantum chi-squared and goodness of fit testing. *J. Math. Phys.* **56** 012202. [MR3390829](#)
- [56] TORGERSEN, E. (1991). *Comparison of Statistical Experiments*. *Encyclopedia of Mathematics and Its Applications* **36**. Cambridge Univ. Press, Cambridge. [MR1104437](#)
- [57] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York. [MR2724359](#)
- [58] WALTER, M. and RENES, J. M. (2014). Lower bounds for quantum parameter estimation. *IEEE Trans. Inform. Theory* **60** 8007–8023. [MR3285761](#)

EXTREMAL QUANTILE TREATMENT EFFECTS

BY YICHONG ZHANG

Singapore Management University

This paper establishes an asymptotic theory and inference method for quantile treatment effect estimators when the quantile index is close to or equal to zero. Such quantile treatment effects are of interest in many applications, such as the effect of maternal smoking on an infant's adverse birth outcomes. When the quantile index is close to zero, the sparsity of data jeopardizes conventional asymptotic theory and bootstrap inference. When the quantile index is zero, there are no existing inference methods directly applicable in the treatment effect context. This paper addresses both of these issues by proposing new inference methods that are shown to be asymptotically valid as well as having adequate finite sample properties.

REFERENCES

- [1] ABADIE, A., ANGRIST, J. and IMBENS, G. (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica* **70** 91–117. [MR1926256](#)
- [2] ABADIE, A. and IMBENS, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica* **74** 235–267.
- [3] AI, C. and CHEN, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* **71** 1795–1843. [MR2015420](#)
- [4] ANDREWS, D. W. and CHENG, X. (2012). Estimation and inference with weak, semi-strong, and strong identification. *Econometrica* **80** 2153–2211.
- [5] ANDREWS, D. W. and CHENG, X. (2013). Maximum likelihood estimation and uniform inference with sporadic identification failure. *J. Econometrics* **173** 36–56.
- [6] BELLONI, A., CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. and HANSEN, C. (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica* **85** 233–298. [MR3611771](#)
- [7] BERTAIL, P., HAEFKE, C., POLITIS, D. N. and WHITE, H. (2004). Subsampling the distribution of diverging statistics with applications to finance. *J. Econometrics* **120** 295–326.
- [8] BERTAIL, P., POLITIS, D. N. and ROMANO, J. P. (1999). On subsampling estimators with unknown rate of convergence. *J. Amer. Statist. Assoc.* **94** 569–579.
- [9] BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217. [MR0630103](#)
- [10] BICKEL, P. J. and SAKOV, A. (2008). On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statist. Sinica* **18** 967–985. [MR2440400](#)
- [11] BILLINGSLEY, P. (1999). *Convergence of Probability Measures* **493**, 2nd ed. Wiley, Hoboken, NJ.
- [12] BITLER, M., GELBACH, J. and HOYNES, H. (2006). What mean impacts miss: Distributional effects of welfare reform experiments. *Am. Econ. Rev.* **96** 988–1012.

MSC2010 subject classifications. Primary 62E20, 62G32; secondary 62P20.

Key words and phrases. Extreme quantile, intermediate quantile.

- [13] CHEN, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Handb. Econom.* **6** 5549–5632.
- [14] CHEN, X., PONOMAREVA, M. and TAMER, E. (2014). Likelihood inference in some finite mixture models. *J. Econometrics* **182** 87–99.
- [15] CHERNOZHUKOV, V. (2005). Extremal quantile regression. *Ann. Statist.* **33** 806–839. [MR2163160](#)
- [16] CHERNOZHUKOV, V. and FERNÁNDEZ-VAL, I. (2011). Inference for extremal conditional quantile models, with an application to market and birthweight risks. *Rev. Econ. Stud.* **78** 559–589. Supplementary data available online. [MR2808129](#)
- [17] CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. and KAJI, T. (2016). Extremal quantile regression: An overview. ArXiv Preprint [ArXiv:1612.06850](#).
- [18] CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. and MELLY, B. (2013). Inference on counterfactual distributions. *Econometrica* **81** 2205–2268. [MR3138546](#)
- [19] CHERNOZHUKOV, V. and HANSEN, C. (2005). An IV model of quantile treatment effects. *Econometrica* **73** 245–261.
- [20] CHERNOZHUKOV, V. and HANSEN, C. (2008). Instrumental variable quantile regression: A robust inference approach. *J. Econometrics* **142** 379–398. [MR2408741](#)
- [21] CHERNOZHUKOV, V. and HONG, H. (2004). Likelihood estimation and inference in a class of nonregular econometric models. *Econometrica* **72** 1445–1480.
- [22] CHERNOZHUKOV, V. V. (2000). Conditional extremes and near-extremes: Concepts, asymptotic theory, and economic applications. Ph.D. thesis, Stanford Univ., Stanford, CA.
- [23] D’HAULTFOEUILLE, X., MAUREL, A. and ZHANG, Y. (2018). Extremal quantile regressions for selection models and the black-white wage gap. *J. Econometrics*. **203** 129–142.
- [24] DEKKERS, A. L. M. and DE HAAN, L. (1989). On the estimation of the extreme-value index and large quantile estimation. *Ann. Statist.* **17** 1795–1832. [MR1026314](#)
- [25] DOKSUM, K. (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Ann. Statist.* **2** 267–277. [MR356350](#)
- [26] FALK, M. (1991). A note on the inverse bootstrap process for large quantiles. *Stochastic Process. Appl.* **38** 359–363.
- [27] FEIGIN, P. D. and RESNICK, S. I. (1994). Limit distributions for linear programming time series estimators. *Stochastic Process. Appl.* **51** 135–165.
- [28] FIRPO, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica* **75** 259–276.
- [29] FIRPO, S. and ROTHE, C. (2014). Semiparametric estimation and inference using doubly robust moment conditions. Technical Report, IZA Discussion Paper.
- [30] FRÖLICH, M. and MELLY, B. (2013). Unconditional quantile treatment effects under endogeneity. *J. Bus. Econom. Statist.* **31** 346–357. [MR3173686](#)
- [31] HAHN, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica* **66** 315–331.
- [32] HALL, P. and VAN KEILEGOM, I. (2009). Nonparametric “regression” when errors are positioned at end-points. *Bernoulli* **15** 614–633. [MR2555192](#)
- [33] HANSEN, B. E. (2014). Nonparametric sieve regression: Least squares, averaging least squares, and cross-validation. In *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics* 215–248. Oxford Univ. Press, Oxford. [MR3306927](#)
- [34] HIRANO, K., IMBENS, G. W. and RIDDER, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71** 1161–1189.
- [35] HIRANO, K. and PORTER, J. R. (2003). Asymptotic efficiency in parametric structural models with parameter-dependent support. *Econometrica* **71** 1307–1338.
- [36] IMAI, K. and RATKOVIC, M. (2014). Covariate balancing propensity score. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 243–263.

- [37] KHAN, S. and NEKIPELOV, D. (2013). On uniform inference in nonlinear models with endogeneity. Economic Research Initiatives at Duke (ERID) Working Paper 153.
- [38] KNIGHT, K. (2001). Limiting distributions of linear programming estimators. *Extremes* **4** 87–103.
- [39] LEE, S. and SEO, M. H. (2008). Semiparametric estimation of a binary response model with a change-point due to a covariate threshold. *J. Econometrics* **144** 492–499.
- [40] LEHMANN, E. L. (1974). *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, San Francisco, CA.
- [41] NEWEY, W. K. and MCFADDEN, D. (1994). Large sample estimation and hypothesis testing. *Handb. Econom.* **4** 2111–2245.
- [42] PITMAN, E. J. G. (1949). *Lecture Notes on Nonparametric Statistical Inference*. Columbia Univ.
- [43] PORTNOY, S. and JUREČKOVÁ, J. (1999). On extreme regression quantiles. *Extremes* **2** 227–243. [MR1781938](#)
- [44] ROBINS, J. M. and ROTNITZKY, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *J. Amer. Statist. Assoc.* **90** 122–129.
- [45] ROSENBAUM, P. R. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55.
- [46] RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592.
- [47] RUBIN, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Ann. Statist.* **6** 34–58. [MR472152](#)
- [48] SMITH, R. L. (1994). Nonregular regression. *Biometrika* **81** 173–183.
- [49] STAIGER, D. and STOCK, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica* **65** 557–586. [MR1445622](#)
- [50] STOCK, J. H. and YOGO, M. (2005). Testing for weak instruments in linear IV regression. In *Identification and Inference for Econometric Models* 80–108. Cambridge Univ. Press, Cambridge. [MR2232140](#)
- [51] VAN DER VAART, A. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.
- [52] ZAREPOUR, M. and KNIGHT, K. (1999). Bootstrapping point processes with some applications. *Stochastic Process. Appl.* **84** 81–90.
- [53] ZHANG, Y. (2018). Supplement to “Extremal quantile treatment effects.” DOI:[10.1214/17-AOS1673SUPP](#).

OPTIMAL MAXIMIN L_1 -DISTANCE LATIN HYPERCUBE DESIGNS BASED ON GOOD LATTICE POINT DESIGNS¹

BY LIN WANG*, QIAN XIAO[†] AND HONGQUAN XU*

University of California, Los Angeles and University of Georgia[†]*

Maximin distance Latin hypercube designs are commonly used for computer experiments, but the construction of such designs is challenging. We construct a series of maximin Latin hypercube designs via Williams transformations of good lattice point designs. Some constructed designs are optimal under the maximin L_1 -distance criterion, while others are asymptotically optimal. Moreover, these designs are also shown to have small pairwise correlations between columns.

REFERENCES

- BA, S., MYERS, W. R. and BRENNEMAN, W. A. (2015). Optimal sliced Latin hypercube designs. *Technometrics* **57** 479–487. [MR3425485](#)
- BAILEY, R. A. (1982). The decomposition of treatment degrees of freedom in quantitative factorial experiments. *J. Roy. Statist. Soc. Ser. B* **44** 63–70. [MR0655375](#)
- BUTLER, N. A. (2001). Optimal and orthogonal Latin hypercube designs for computer experiments. *Biometrika* **88** 847–857.
- CHEN, R.-B., HSIEH, D.-N., HUNG, Y. and WANG, W. (2013). Optimizing Latin hypercube designs by particle swarm. *Stat. Comput.* **23** 663–676. [MR3094806](#)
- CIOPPA, T. M. and LUCAS, T. W. (2007). Efficient nearly orthogonal and space-filling Latin hypercubes. *Technometrics* **49** 45–55.
- EDMONDSON, R. N. (1993). Systematic row-and-column designs balanced for low order polynomial interactions between rows and columns. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **55** 707–723.
- FANG, K.-T., LI, R. and SUDJANTO, A. (2006). *Design and Modeling for Computer Experiments*. Chapman & Hall/CRC, Boca Raton, FL. [MR2223960](#)
- FANG, K. T. and WANG, Y. (1994). *Number-Theoretic Methods in Statistics*. Chapman & Hall, London.
- GEORGIOU, S. D. and EFTHIMIOU, I. (2014). Some classes of orthogonal Latin hypercube designs. *Statist. Sinica* **24** 101–120.
- HE, Y., CHENG, C. S. and TANG, B. (2018). Strong orthogonal arrays of strength two plus. *Ann. Statist.* **46** 457–468. [MR3782373](#)
- HE, Y. and TANG, B. (2013). Strong orthogonal arrays and associated Latin hypercubes for computer experiments. *Biometrika* **100** 254–260.
- HE, Y. and TANG, B. (2014). A characterization of strong orthogonal arrays of strength three. *Ann. Statist.* **42** 1347–1360. [MR3226159](#)
- JOHNSON, M. E., MOORE, L. M. and YLVISAKER, D. (1990). Minimax and maximin distance designs. *J. Statist. Plann. Inference* **26** 131–148.

MSC2010 subject classifications. Primary 62K99.

Key words and phrases. Computer experiment, correlation, space-filling design, Williams transformation.

- JOSEPH, V. R. and HUNG, Y. (2008). Orthogonal-maximin Latin hypercube designs. *Statist. Sinica* **18** 171–186. [MR2416907](#)
- KOROBOV, N. M. (1959). The approximate computation of multiple integrals. *Dokl. Akad. Nauk SSSR* **124** 1207–1210.
- LIN, C. D., MUKERJEE, R. and TANG, B. (2009). Construction of orthogonal and nearly orthogonal Latin hypercubes. *Biometrika* **96** 243–247.
- LIN, C. D. and TANG, B. (2015). Latin hypercubes and space-filling designs. In *Handbook of Design and Analysis of Experiments* (A. Dean, M. Morris, J. Stufken and D. Bingham, eds.) 593–625. Chapman & Hall/CRC, London.
- MOON, H., DEAN, A. and SANTNER, T. (2011). Algorithms for generating maximin Latin hypercube and orthogonal designs. *J. Statist. Plann. Inference* **5** 81–98.
- MORRIS, M. D. (1991). Factorial plans for preliminary computational experiments. *Technometrics* **33** 161–174.
- MORRIS, M. D. and MITCHELL, T. J. (1995). Exploratory designs for computational experiments. *J. Statist. Plann. Inference* **43** 381–402.
- MORRIS, M. D. and MOORE, L. M. (2015). Design of computer experiments: Introduction and background. In *Handbook of Design and Analysis of Experiments* (A. Dean, M. Morris, J. Stufken and D. Bingham, eds.) 577–591. CRC Press, Boca Raton, FL. [MR3699362](#)
- SACKS, J., SCHILLER, S. B. and WELCH, W. J. (1989). Designs for computer experiments. *Technometrics* **31** 41–47.
- SANTNER, T. J., WILLIAMS, B. J. and NOTZ, W. I. (2003). *The Design and Analysis of Computer Experiments*. Springer, New York.
- STEINBERG, D. M. and LIN, D. K. J. (2006). A construction method for orthogonal Latin hypercube designs. *Biometrika* **93** 279–288.
- SUN, F. S., LIU, M. Q. and LIN, D. K. J. (2009). Construction of orthogonal Latin hypercube designs. *Biometrika* **96** 971–974.
- SUN, F. S. and TANG, B. (2017). A general rotation method for orthogonal Latin hypercubes. *Biometrika* **104** 465–472.
- TANG, B. (1993). Orthogonal array-based Latin hypercubes. *J. Amer. Statist. Assoc.* **88** 1392–1397.
- WILLIAMS, E. J. (1949). Experimental designs balanced for the estimation of residual effects of treatments. *Aust. J. Sci. Res.* **2** 149–168.
- XIAO, Q. and XU, H. (2017). Construction of maximin distance Latin squares and related Latin hypercube designs. *Biometrika* **104** 455–464.
- XIAO, Q. and XU, H. (2018). Construction of maximin distance designs via level permutation and expansion. *Statist. Sinica* **28** 1395–1414.
- XU, H. (1999). Universally optimal designs for computer experiments. *Statist. Sinica* **9** 1083–1088.
- YANG, J. Y. and LIU, M. Q. (2012). Construction of orthogonal and nearly orthogonal Latin hypercube designs from orthogonal designs. *Statist. Sinica* **22** 433–442.
- YE, K. Q. (1998). Orthogonal column Latin hypercubes and their application in computer experiments. *J. Amer. Statist. Assoc.* **93** 1430–1439.
- ZHOU, Y. and XU, H. (2015). Space-filling properties of good lattice point sets. *Biometrika* **102** 959–966.

RHO-ESTIMATORS REVISITED: GENERAL THEORY AND APPLICATIONS

BY YANNICK BARAUD AND LUCIEN BIRGÉ

Université Côte d’Azur and Sorbonne Universités

Following Baraud, Birgé and Sart [*Invent. Math.* **207** (2017) 425–517], we pursue our attempt to design a robust universal estimator of the joint distribution of n independent (but not necessarily i.i.d.) observations for an Hellinger-type loss. Given such observations with an unknown joint distribution $\mathbf{P}$ and a dominated model $\mathcal{Q}$ for $\mathbf{P}$, we build an estimator $\hat{\mathbf{P}}$ based on $\mathcal{Q}$ (a ρ -estimator) and measure its risk by an Hellinger-type distance. When $\mathbf{P}$ does belong to the model, this risk is bounded by some quantity which relies on the local complexity of the model in a vicinity of $\mathbf{P}$. In most situations, this bound corresponds to the minimax risk over the model (up to a possible logarithmic factor). When $\mathbf{P}$ does not belong to the model, its risk involves an additional bias term proportional to the distance between $\mathbf{P}$ and $\mathcal{Q}$, whatever the true distribution $\mathbf{P}$. From this point of view, this new version of ρ -estimators improves upon the previous one described in Baraud, Birgé and Sart [*Invent. Math.* **207** (2017) 425–517] which required that $\mathbf{P}$ be absolutely continuous with respect to some known reference measure. Further additional improvements have been brought as compared to the former construction. In particular, it provides a very general treatment of the regression framework with random design as well as a computationally tractable procedure for aggregating estimators. We also give some conditions for the maximum likelihood estimator to be a ρ -estimator. Finally, we consider the situation where the statistician has at her or his disposal many different models and we build a penalized version of the ρ -estimator for model selection and adaptation purposes. In the regression setting, this penalized estimator not only allows one to estimate the regression function but also the distribution of the errors.

REFERENCES

- AUDIBERT, J.-Y. and CATONI, O. (2011). Robust linear least squares regression. *Ann. Statist.* **39** 2766–2794. [MR2906886](#)
- BARAUD, Y. (2016). Bounding the expectation of the supremum of an empirical process over a (weak) VC-major class. *Electron. J. Stat.* **10** 1709–1728. [MR3522658](#)
- BARAUD, Y. and BIRGÉ, L. (2016). Rho-estimators for shape restricted density estimation. *Stochastic Process. Appl.* **126** 3888–3912.
- BARAUD, Y. and BIRGÉ, L. (2018). Supplement to “Rho-estimators revisited: General theory and applications.” DOI:[10.1214/17-AOS1675SUPP](#).

MSC2010 subject classifications. 62G35, 62G05, 62G07, 62G08, 62C20, 62F99.

Key words and phrases. ρ -estimation, robust estimation, density estimation, regression with random design, statistical models, maximum likelihood estimators, metric dimension, VC-classes.

- BARAUD, Y., BIRGÉ, L. and SART, M. (2017). A new method for estimation and model selection: ρ -estimation. *Invent. Math.* **207** 425–517.
- BIRGÉ, L. (1983). Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsch. Verw. Gebiete* **65** 181–237. [MR0722129](#)
- BIRGÉ, L. (2006). Model selection via testing: An alternative to (penalized) maximum likelihood estimators. *Ann. Inst. Henri Poincaré Probab. Stat.* **42** 273–325.
- BIRGÉ, L. and MASSART, P. (1998). Minimum contrast estimators on sieves: Exponential bounds and rates of convergence. *Bernoulli* **4** 329–375.
- GINÉ, E. and KOLTCHINSKII, V. (2006). Concentration inequalities and asymptotic results for ratio type empirical processes. *Ann. Probab.* **34** 1143–1216. [MR2243881](#)
- GYÖRFI, L., KOHLER, M., KRZYŻAK, A. and WALK, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. Springer, New York.
- KOLTCHINSKII, V. (2006). Local Rademacher complexities and oracle inequalities in risk minimization. *Ann. Statist.* **34** 2593–2656. [MR2329442](#)
- LE CAM, L. (1973). Convergence of estimates under dimensionality restrictions. *Ann. Statist.* **1** 38–53.
- LE CAM, L. (1975). On local and global properties in the theory of asymptotic normality of experiments. In *Stochastic Processes and Related Topics (Proc. Summer Res. Inst. Statist. Inference for Stochastic Processes, Indiana Univ., Bloomington, Ind., 1974, Vol. 1; Dedicated to Jerzy Neyman)* 13–54. Academic Press, New York.
- LE CAM, L. (1990). Maximum likelihood: An introduction. *Int. Stat. Rev.* **58** 153–171.
- POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.
- SART, M. (2017). Estimating the conditional density by histogram type estimators and model selection. *ESAIM, Probab. Stat.* **21** 34–55.
- VAN DE GEER, S. A. (2000). *Applications of Empirical Process Theory. Cambridge Series in Statistical and Probabilistic Mathematics* **6**. Cambridge Univ. Press, Cambridge. [MR1739079](#)
- VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes. With Applications to Statistics*. Springer, New York.

THINK GLOBALLY, FIT LOCALLY UNDER THE MANIFOLD SETUP: ASYMPTOTIC ANALYSIS OF LOCALLY LINEAR EMBEDDING

BY HAU-TIENG WU^{*,†,1} AND NAN WU[‡]

Duke University^{*}, *National Center for Theoretical Sciences*[†] and
University of Toronto[‡]

Since its introduction in 2000, Locally Linear Embedding (LLE) has been widely applied in data science. We provide an asymptotical analysis of LLE under the manifold setup. We show that for a general manifold, asymptotically we may not obtain the Laplace–Beltrami operator, and the result may depend on nonuniform sampling unless a correct regularization is chosen. We also derive the corresponding kernel function, which indicates that LLE is not a Markov process. A comparison with other commonly applied nonlinear algorithms, particularly a diffusion map, is provided and its relationship with locally linear regression is also discussed.

REFERENCES

- [1] BELKIN, M. and NIYOGI, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **15** 1373–1396.
- [2] BELKIN, M. and NIYOGI, P. (2005). Towards a theoretical foundation for Laplacian-based manifold methods. In *Learning Theory. Lecture Notes in Computer Science* **3559** 486–500. Springer, Berlin. [MR2203282](#)
- [3] BELKIN, M. and NIYOGI, P. (2007). Convergence of Laplacian eigenmaps. In *Advances in Neural Information Processing Systems 19 (NIPS 2006)* 129–136. MIT Press, Cambridge, MA.
- [4] BÉRARD, P., BESSON, G. and GALLOT, S. (1994). Embedding Riemannian manifolds by their heat kernel. *Geom. Funct. Anal.* **4** 373–398.
- [5] BÉRARD, P. H. (1986). *Spectral Geometry: Direct and Inverse Problems. Lecture Notes in Math.* **1207**. Springer, Berlin. [MR0861271](#)
- [6] CHEEGER, J., GROMOV, M. and TAYLOR, M. (1982). Finite propagation speed, kernel estimates for functions of the Laplace operator, and the geometry of complete Riemannian manifolds. *J. Differential Geom.* **17** 15–53.
- [7] CHENG, M.-Y. and WU, H.-T. (2013). Local linear regression on manifolds and its geometric interpretation. *J. Amer. Statist. Assoc.* **108** 1421–1434.
- [8] COIFMAN, R. R. and LAFON, S. (2006). Diffusion maps. *Appl. Comput. Harmon. Anal.* **21** 5–30. [MR2238665](#)
- [9] DEVROYE, L. P. and WAGNER, T. J. (1977). The strong uniform consistency of nearest neighbor density estimates. *Ann. Statist.* **5** 536–540. [MR436442](#)
- [10] DO CARMO, M. P. and FLAHERTY, F. (1992). *Riemannian Geometry*. Birkhäuser, Boston, MA.

MSC2010 subject classifications. 60K35.

Key words and phrases. Locally linear embedding, diffusion maps, dimension reduction, locally linear regression, measurement error.

- [11] DONOHO, D. L., GAVISH, M. and JOHNSTONE, I. M. (2013). Optimal shrinkage of eigenvalues in the spiked covariance model. Available at [arXiv:1311.0851](#).
- [12] DONOHO, D. L. and GRIMES, C. (2003). Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proc. Natl. Acad. Sci. USA* **100** 5591–5596.
- [13] EL KAROUI, N. (2010). On information plus noise kernel random matrices. *Ann. Statist.* **38** 3191–3216. [MR2722468](#)
- [14] EL KAROUI, N. and WU, H.-T. (2016). Graph connection Laplacian methods can be made robust to noise. *Ann. Statist.* **44** 346–372. [MR3449771](#)
- [15] FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall/CRC, Boca Raton, FL.
- [16] GAO, T. (2016). The diffusion geometry of fibre bundles. Available at [arXiv:1602.02330](#).
- [17] GARCIA TRILLOS, N. and SLEPCEV, D. (2018). A variational approach to the consistency of spectral clustering. *Appl. Comput. Harmon. Anal.* To appear.
- [18] GINÉ, E. and KOLTCHINSKII, V. (2006). Empirical graph Laplacian approximation of Laplace–Beltrami operators: Large sample results. In *High Dimensional Probability. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **51** 238–259. IMS, Beachwood, OH. [MR2387773](#)
- [19] HEIN, M., AUDIBERT, J.-Y. and VON LUXBURG, U. (2005). From graphs to manifolds—Weak and strong pointwise consistency of graph Laplacians. In *Learning Theory. Lecture Notes in Computer Science* **3559** 470–485. Springer, Berlin. [MR2203281](#)
- [20] JOHNSTONE, I. M. (2006). High dimensional statistical inference and random matrices. Available at [arXiv:math/0611589v1](#).
- [21] MOORE, D. S. and YACKEL, J. W. (1977). Consistency properties of nearest neighbor density function estimators. *Ann. Statist.* **5** 143–154. [MR426275](#)
- [22] ROWEIS, S. T. and SAUL, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science* **290** 2323–2326.
- [23] SINGER, A. (2006). From graph to manifold Laplacian: The convergence rate. *Appl. Comput. Harmon. Anal.* **21** 128–134.
- [24] SINGER, A. and WU, H.-T. (2012). Vector diffusion maps and the connection Laplacian. *Comm. Pure Appl. Math.* **65** 1067–1144.
- [25] SINGER, A. and WU, H.-T. (2013). 2-D tomography from noisy projections taken at unknown random directions. *SIAM J. Imaging Sci.* **6** 136–175.
- [26] SINGER, A. and WU, H.-T. (2017). Spectral convergence of the connection Laplacian from random samples. *Inf. Inference* **6** 58–123. [MR3636868](#)
- [27] SMOLYANOV, O., WEIZSACKER, H. v. and WITTICH, O. (2007). Chernoff’s theorem and discrete time approximations of Brownian motion on manifolds. *Potential Anal.* **26** 1–29.
- [28] STEIN, E. M. and WEISS, G. (2016). *Introduction to Fourier Analysis on Euclidean Spaces. PMS* **32**. Princeton Univ. Press, Princeton, NJ.
- [29] TENENBAUM, J. B., DE SILVA, V. and LANGFORD, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science* **290** 2319–2323.
- [30] VAN DER MAATEN, L. and HINTON, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9** 2579–2605.
- [31] VON LUXBURG, U., BELKIN, M. and BOUSQUET, O. (2008). Consistency of spectral clustering. *Ann. Statist.* **36** 555–586. [MR2396807](#)
- [32] WANG, X. (2015). Spectral convergence rate of graph Laplacian. Available at [arXiv:1510.08110](#).
- [33] WEINBERGER, K. Q. and SAUL, L. K. (2006). An introduction to nonlinear dimensionality reduction by maximum variance unfolding. In *AAAI* 1683–1686.
- [34] WU, H.-T. and WU, N. (2018). Supplement to “Think globally, fit locally under the manifold setup: Asymptotic analysis of locally linear embedding.” DOI:10.1214/17-AOS1676SUPP.

- [35] ZHANG, Z. and WANG, J. (2006). MLE: Modified locally linear embedding using multiple weights. In *Advances in Neural Information Processing Systems 19 (NIPS 2006)* 1593–1600. MIT Press, Cambridge, MA.
- [36] ZHANG, Z. and ZHA, H. (2004). Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM J. Sci. Comput.* **26** 313–338.

NONPARAMETRIC COVARIATE-ADJUSTED RESPONSE-ADAPTIVE DESIGN BASED ON A FUNCTIONAL URN MODEL

BY GIACOMO ALETTI^{*,†}, ANDREA GHIGLIETTI[‡] AND
WILLIAM F. ROSENBERGER[§]

ADAMSS Center^{} and Università degli Studi di Milano,[†]
Università Cattolica del Sacro Cuore[‡]
and George Mason University[§]*

In this paper, we propose a general class of covariate-adjusted response-adaptive (CARA) designs based on a new functional urn model. We prove strong consistency concerning the functional urn proportion and the proportion of subjects assigned to the treatment groups, in the whole study and for each covariate profile, allowing the distribution of the responses conditioned on covariates to be estimated nonparametrically. In addition, we establish joint central limit theorems for the above quantities and the sufficient statistics of features of interest, which allow to construct procedures to make inference on the conditional response distributions. These results are then applied to typical situations concerning Gaussian and binary responses.

REFERENCES

- [1] ALETTI, G., CRIMALDI, I. and GHIGLIETTI, A. (2017). Synchronization of reinforced stochastic processes with a network-based interaction. *Ann. Appl. Probab.* **27** 3787–3844. [MR3737938](#)
- [2] ALETTI, G. and GHIGLIETTI, A. (2017). Interacting generalized Friedman’s urn systems. *Stochastic Process. Appl.* **127** 2650–2678. [MR3660886](#)
- [3] ALETTI, G., GHIGLIETTI, A. and PAGANONI, A. M. (2013). Randomly reinforced urn designs with prespecified allocations. *J. Appl. Probab.* **50** 486–498. [MR3102495](#)
- [4] ALETTI, G., GHIGLIETTI, A. and ROSENBERGER, W. F. (2018). Supplement to “Nonparametric covariate-adjusted response-adaptive design based on a functional urn model.” DOI:[10.1214/17-AOS1677SUPP](#).
- [5] ATHREYA, K. B. and KARLIN, S. (1970). Branching processes with random environments. *Bull. Amer. Math. Soc.* **76** 865–870.
- [6] ATKINSON, A. C. (1982). Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika* **69** 61–67.
- [7] ATKINSON, A. C. and BISWAS, A. (2005). Adaptive biased-coin designs for skewing the allocation proportion in clinical trials with normal responses. *Stat. Med.* **24** 2477–2492.
- [8] ATKINSON, A. C. and BISWAS, A. (2013). *Randomised Response-Adaptive Designs in Clinical Trials*. CRC Press, New York.
- [9] BAI, Z.-D. and HU, F. (2005). Asymptotics in randomized urn models. *Ann. Appl. Probab.* **15** 914–940. [MR2114994](#)

MSC2010 subject classifications. Primary 62L20, 60F05, 62E20, 62L05; secondary 62F12, 62P10.

Key words and phrases. Clinical trials, covariate-adjusted analysis, inference, large sample theory, personalized medicine, randomization.

- [10] BAI, Z.-D., HU, F. and ZHANG, L.-X. (2002). Gaussian approximation theorems for urn models and their applications. *Ann. Appl. Probab.* **12** 1149–1173. [MR1936587](#)
- [11] BAI, Z. D. and HU, F. (1999). Asymptotic theorems for urn models with nonhomogeneous generating matrices. *Stochastic Process. Appl.* **80** 87–101. [MR1670107](#)
- [12] BALDI ANTOGNINI, A. and ZAGORAIOU, M. (2012). Multi-objective optimal designs in comparative clinical trials with covariates: The reinforced doubly adaptive biased coin design. *Ann. Statist.* **40** 1315–1345. [MR3015027](#)
- [13] BALDI ANTOGNINI, A. and ZAGORAIOU, M. (2017). Estimation accuracy under covariate-adaptive randomization procedures. *Electron. J. Stat.* **11** 1180–1206.
- [14] BANDYOPADHYAY, U. and BISWAS, A. (1999). Allocation by randomized play-the-winner rule in the presence of prognostic factors. *Sankhyā Ser. B* **61** 397–412.
- [15] BANDYOPADHYAY, U. and BISWAS, A. (2001). Adaptive designs for normal responses with prognostic factors. *Biometrika* **88** 409–419.
- [16] BANDYOPADHYAY, U., BISWAS, A. and BHATTACHARYA, R. (2007). A covariate adjusted two-stage allocation design for binary responses in randomized clinical trials. *Stat. Med.* **26** 4386–4399.
- [17] BEGGS, A. W. (2005). On the convergence of reinforcement learning. *J. Econom. Theory* **122** 1–36.
- [18] BENAÏM, M., SCHREIBER, S. J. and TARRÈS, P. (2004). Generalized URN models of evolutionary processes. *Ann. Appl. Probab.* **14** 1455–1478. [MR2071430](#)
- [19] CHEUNG, S. H., ZHANG, L.-X., HU, F. and CHAN, W. S. (2014). Covariate-adjusted response-adaptive designs for generalized linear models. *J. Statist. Plann. Inference* **149** 152–161.
- [20] DURHAM, S. D., FLOURNOY, N. and LI, W. (1998). A sequential design for maximizing the probability of a favourable response. *Canad. J. Statist.* **26** 479–495.
- [21] EGGENBERGER, F. and PÓLYA, G. (1923). Über die Statistik verketteter Vorgänge. *ZAMM Z. Angew. Math. Mech.* **3** 279–289.
- [22] GHIGLIETTI, A., VIDYASHANKAR, A. N. and ROSENBERGER, W. F. (2017). Central limit theorem for an adaptive randomly reinforced urn model. *Ann. Appl. Probab.* **27** 2956–3003. [MR3719951](#)
- [23] HU, F. and ROSENBERGER, W. F. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*. Wiley, New York.
- [24] HU, J., ZHU, H. and HU, F. (2015). A unified family of covariate-adjusted response-adaptive designs based on efficiency and ethics. *J. Amer. Statist. Assoc.* **110** 357–367. [MR3338508](#)
- [25] HU, Y. and HU, F. (2012). Asymptotic properties of covariate-adaptive randomization. *Ann. Statist.* **40** 1794–1815. [MR3015044](#)
- [26] JANSON, S. (2004). Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stochastic Process. Appl.* **110** 177–245.
- [27] KADANE, J. B. (1996). *Bayesian Methods and Ethics in a Clinical Trial Design*. Wiley, New York.
- [28] KALISH, L. A. and HARRINGTON, D. P. (1988). Efficiency of balanced treatment allocation for survival analysis. *Biometrics* **44** 815–821.
- [29] LARUELLE, S. and PAGÈS, G. (2013). Randomized urn models revisited using stochastic approximation. *Ann. Appl. Probab.* **23** 1409–1436. [MR3098437](#)
- [30] MAY, C. and FLOURNOY, N. (2009). Asymptotics in response-adaptive designs generated by a two-color, randomly reinforced urn. *Ann. Statist.* **37** 1058–1078. [MR2502661](#)
- [31] POCOCK, S. J. and SIMON, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics* **31** 103–115.
- [32] ROSENBERGER, W. F. (1996). New directions in adaptive designs. *Statist. Sci.* **11** 137–149. [MR1477667](#)

- [33] ROSENBERGER, W. F., FLOURNOY, N. and DURHAM, S. D. (1997). Asymptotic normality of maximum likelihood estimators from multiparameter response-driven designs. *J. Statist. Plann. Inference* **60** 69–76.
- [34] ROSENBERGER, W. F. and LACHIN, J. M. (2016). *Randomization in Clinical Trials*, 2nd ed. Wiley, New York.
- [35] ROSENBERGER, W. F. and SVERDLOV, O. (2008). Handling covariates in the design of clinical trials. *Statist. Sci.* **23** 404–419. [MR2483911](#)
- [36] ROSENBERGER, W. F., VIDYASHANKAR, A. N. and AGARWAL, D. K. (2001). Covariate-adjusted response-adaptive designs for binary response. *J. Biopharm. Statist.* **11** 227–236.
- [37] SCHORTGEN, F., CLABAULT, K., KATSAHIAN, S., DEVAQUET, J., MERCAT, A., DEYE, N., DELLAMONICA, J., BOUADMA, L., COOK, F., BEJI, O., BRUN-BUISSON, C., LEMAIRE, F. and BROCHARD, L. (2012). Fever control using external cooling in septic shock. *Am. J. Respir. Crit. Care Med.* **185** 1088–1095.
- [38] SMYTHE, R. T. (1996). Central limit theorems for urn models. *Stochastic Process. Appl.* **65** 115–137.
- [39] TAVES, D. R. (1974). Minimization: A new method of assigning patients to treatment and control groups. *Clin. Pharmacol. Ther.* **15** 443–453.
- [40] TOULIS, P., RENNIE, J. and AIROLDI, E. M. (2014). Statistical analysis of stochastic gradient methods for generalized linear models. In *31st International Conference on Machine Learning, ICML 2014* 2128–2136. JMLR: W&CP, Beijing.
- [41] VILLAR, S. and ROSENBERGER, W. F. (2018). Covariate-adjusted response-adaptive randomization for multi-arm clinical trials using a modified forward looking gittins index rule. *Biometrics* **74** 49–57.
- [42] WEI, L. J. (1978). An application of an urn model to the design of sequential controlled clinical trials. *J. Amer. Statist. Assoc.* **73** 559–563.
- [43] WEI, L. J. and DURHAM, S. (1978). The randomized play-the-winner rule in medical trials. *J. Amer. Statist. Assoc.* **73** 840–843.
- [44] ZELEN, M. (1969). Play the winner rule and the controlled clinical trial. *J. Amer. Statist. Assoc.* **64** 131–146.
- [45] ZHANG, L. and ROSENBERGER, W. F. (2006). Response-adaptive randomization for clinical trials with continuous outcomes. *Biometrics* **62** 562–569.
- [46] ZHANG, L.-X. (2016). Central limit theorems of a recursive stochastic algorithm with applications to adaptive designs. *Ann. Appl. Probab.* **26** 3630–3658. [MR3582813](#)
- [47] ZHANG, L.-X., HU, F. and CHEUNG, S. H. (2006). Asymptotic theorems of sequential estimation-adjusted urn models. *Ann. Appl. Probab.* **16** 340–369. [MR2209345](#)
- [48] ZHANG, L.-X., HU, F., CHEUNG, S. H. and CHAN, W. S. (2007). Asymptotic properties of covariate-adjusted response-adaptive designs. *Ann. Statist.* **35** 1166–1182. [MR2341702](#)
- [49] ZHANG, L.-X., HU, F., CHEUNG, S. H. and CHAN, W. S. (2011). Immigrated urn models—Theoretical properties and applications. *Ann. Stat.* **39** 643–671. [MR2797859](#)
- [50] ZHU, H. (2015). Covariate-adjusted response adaptive designs incorporating covariates with and without treatment interactions. *Canad. J. Statist.* **43** 534–553.

THE ANNALS
of
STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

VOLUME 46

2018

CONTENTS OF VOLUME 46

Articles

ALETTI, GIACOMO, GHIGLIETTI, ANDREA AND ROSENBERGER, WILLIAM F. Nonparametric covariate-adjusted response-adaptive design based on a functional urn model	3838–3866
AMINI, ARASH A. AND LEVINA, ELIZAVETA. On semidefinite relaxations for the block model	149–179
BAI, YU, MONTANARI, ANDREA AND MEI, SONG. The landscape of empirical risk for nonconvex losses	2747–2774
BAI, ZHIDONG, CHOI, KWOK PUI AND FUJIKOSHI, YASUNORI. Consistency of AIC and BIC in estimating the number of significant components in high-dimensional principal component analysis	1050–1076
BAI, ZHIDONG, CHOI, KWOK PUI AND FUJIKOSHI, YASUNORI. Limiting behavior of eigenvalues in high-dimensional MANOVA via RMT	2985–3013
BANDEIRA, AFONSO S., MOITRA, ANKUR, PERRY, AMELIA AND WEIN, ALEXANDER S. Optimality and sub-optimality of PCA I: Spiked random matrix models	2416–2451
BARANIUK, RICHARD G., MOUSAVI, ALI AND MALEKI, ARIAN. Consistent parameter estimation for LASSO and approximate message passing	119–148
BARAUD, YANNICK AND BIRGÉ, LUCIEN. Rho-estimators revisited: General theory and applications	3767–3804
BARBER, RINA FOYGEL AND KOLAR, MLADEN. ROCKET: Robust confidence intervals via <u>Kendall's tau</u> for transelliptical graphical models	3422–3450
BATTEY, HEATHER, FAN, JIANQING, LIU, HAN, LU, JUNWEI AND ZHU, ZIWEI. Distributed testing and estimation under sparse high dimensional models	1352–1382
BEHR, MERLE, HOLMES, CHRIS AND MUNK, AXEL. Multiscale blind source separation	711–744
BELLEÇ, PIERRE C. Optimal bounds for aggregation of affine estimators	30–59
BELLEÇ, PIERRE C. Sharp oracle inequalities for Least Squares estimators in shape restricted regression	745–780
BELLEÇ, PIERRE C., LECUÉ, GUILLAUME AND TSYBAKOV, ALEXANDRE B. Slope meets Lasso: Improved oracle bounds and optimality	3603–3642
BELLONI, ALEXANDRE, CHERNOZHUKOV, VICTOR, CHETVERIKOV, DENIS AND WEI, YING. Uniformly valid post-	

regularization confidence regions for many functional parameters in z-estimation framework.....	3643–3675
BERGER, JAMES O., GU, MENG YANG AND WANG, XIAOJING. Robust Gaussian stochastic process emulation.....	3038–3066
BERGHAUS, BETINA AND BÜCHER, AXEL. Weak convergence of a pseudo maximum likelihood estimator for the extremal index	2307–2335
BI, XUAN, QU, ANNIE AND SHEN, XIAOTONG. Multilayer tensor factorization with applications to recommender systems.....	3308–3333
BIRGÉ, LUCIEN AND BARAUD, YANNICK. Rho-estimators revis- ited: General theory and applications.....	3767–3804
BOSE, KOUSHIKI, FAN, JIANQING, LIU, HAN AND ZHOU, WEN- XIN. A new perspective on robust M -estimation: Finite sam- ple theory and applications to dependence-adjusted multiple testing	1904–1931
BROUSTE, ALEXANDRE AND FUKASAWA, MASA AKI. Local asymptotic normality property for fractional Gaussian noise un- der high-frequency observations	2045–2061
BROWN, LAWRENCE D., MUKHERJEE, GOURAB AND WEIN- STEIN, ASAF. Empirical Bayes estimates for a two-way cross- classified model	1693–1720
BÜCHER, AXEL AND BERGHAUS, BETINA. Weak convergence of a pseudo maximum likelihood estimator for the extremal index	2307–2335
BÜHLMANN, PETER, ROTHENHÄUSLER, DOMINIK AND ERNEST, JAN. Causal inference in partially linear structural equation models.....	2904–2938
BUTUCEA, CRISTINA, GUȚĂ, MĂDĂLIN AND NUSSBAUM, MICHAEL. Local asymptotic equivalence of pure states en- sembles and quantum Gaussian white noise	3676–3706
BUTUCEA, CRISTINA, NDAOUD, MOHAMED, STEPANOVA, NA- TALIA A. AND TSYBAKOV, ALEXANDRE B. Variable selec- tion with Hamming loss.....	1837–1875
CAI, T. TONY, GUNTUBOYINA, ADITYANAND AND WEI, YUTING. Adaptive estimation of planar convex sets.....	1018–1049
CAI, T. TONY AND GUO, ZIJIAN. Accuracy assessment for high- dimensional linear regression.....	1807–1836
CAI, T. TONY AND ZHANG, ANRU. Rate-optimal perturbation bounds for singular subspaces with applications to high- dimensional statistics	60–89
CAI, TIANXI AND CHAKRABORTTY, ABHISHEK. Efficient and adaptive linear regression in semi-supervised settings	1541–1572

CHAKRABORTTY, ABHISHEK AND CAI, TIANXI. Efficient and adaptive linear regression in semi-supervised settings	1541–1572
CHAN, KWUN CHUEN GARY, LING, HOK KAN, SIT, TONY AND YAM, SHEUNG CHI PHILLIP. Estimation of a monotone density in s -sample biased sampling models	2125–2152
CHANG, JINYUAN, GUO, BIN AND YAO, QIWEI. Principal component analysis for second-order stationary vector time series . .	2094–2124
CHANG, JINYUAN, TANG, CHENG YONG AND WU, TONG TONG. A new scope of penalized empirical likelihood with high-dimensional estimating equations	3185–3216
CHANG, MING-CHUNG AND CHENG, CHING-SHUI. A Bayesian approach to the selection of two-level multi-stratum factorial designs	1779–1806
CHEN, MENGJIE, GAO, CHAO AND REN, ZHAO. Robust covariance and scatter matrix estimation under Huber’s contamination model	1932–1960
CHEN, SONG XI, NETTLETON, DAN AND QIU, YUMOU. Detecting rare and faint signals via thresholding maximum likelihood estimators	895–923
CHEN, XI AND LIU, WEIDONG. Testing independence with high-dimensional correlated samples	866–894
CHEN, XIAOHUI. Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications	642–678
CHEN, YUDONG, LI, XIAODONG AND XU, JIAMING. Convexified modularity maximization for degree-corrected stochastic block models	1573–1602
CHENG, CHING-SHUI AND CHANG, MING-CHUNG. A Bayesian approach to the selection of two-level multi-stratum factorial designs	1779–1806
CHENG, CHING-SHUI, TANG, BOXIN AND HE, YUANZHEN. Strong orthogonal arrays of strength two plus	457–468
CHERNOZHUKOV, VICTOR, CHETVERIKOV, DENIS, WEI, YING AND BELLONI, ALEXANDRE. Uniformly valid post-regularization confidence regions for many functional parameters in z-estimation framework	3643–3675
CHETVERIKOV, DENIS, WEI, YING, BELLONI, ALEXANDRE AND CHERNOZHUKOV, VICTOR. Uniformly valid post-regularization confidence regions for many functional parameters in z-estimation framework	3643–3675
CHOI, KWOK PUI, FUJIKOSHI, YASUNORI AND BAI, ZHIDONG. Consistency of AIC and BIC in estimating the number of sig-	

nificant components in high-dimensional principal component analysis	1050–1076
CHOI, KWOK PUI, FUJIKOSHI, YASUNORI AND BAI, ZHIDONG. Limiting behavior of eigenvalues in high-dimensional MANOVA via RMT	2985–3013
COLLIER, OLIVIER, COMMINGES, LAËTITIA, TSYBAKOV, ALEXANDRE B. AND VERZELEN, NICOLAS. Optimal adaptive estimation of linear functionals under sparsity	3130–3150
COMMINGES, LAËTITIA, TSYBAKOV, ALEXANDRE B., VERZELEN, NICOLAS AND COLLIER, OLIVIER. Optimal adaptive estimation of linear functionals under sparsity	3130–3150
CUI, HENGJIAN, GUO, WENWEN AND ZHONG, WEI. Test for high-dimensional regression coefficients using refitted cross-validation variance estimation	958–988
DAI, XIONGTAO AND MÜLLER, HANS-GEORG. Principal component analysis for functional data on Riemannian manifolds and spheres	3334–3361
DALALYAN, ARNAK S., GRAPPIN, EDWIN AND PARIS, QUENTIN. On the exponentially weighted aggregate with the Laplace prior	2452–2478
DASGUPTA, TIRTHANKAR, ZHAO, ANQI, DING, PENG AND MUKERJEE, RAHUL. Randomization-based causal inference from split-plot designs	1876–1903
DIMAKIS, ALEXANDROS G., NEGAHBAN, SAHAND, ELENBERG, ETHAN R. AND KHANNA, RAJIV. Restricted strong convexity implies weak submodularity	3539–3568
DING, PENG, MUKERJEE, RAHUL, DASGUPTA, TIRTHANKAR AND ZHAO, ANQI. Randomization-based causal inference from split-plot designs	1876–1903
DINH, VU, HO, LAM SI TUNG, SUCHARD, MARC A. AND MATSEN IV, FREDERICK A. Consistency and convergence rate of phylogenetic inference via regularization	1481–1512
DOBRIAN, EDGAR AND WAGER, STEFAN. High-dimensional asymptotics of prediction: Ridge regression and classification ..	247–279
DONOHU, DAVID, GAVISH, MATAN AND JOHNSTONE, IAIN. Optimal shrinkage of eigenvalues in the spiked covariance model .	1742–1778
DOSS, HANI AND PARK, YEONHEE. An MCMC approach to empirical Bayes inference and Bayesian sensitivity analysis via empirical processes	1630–1663
DRTON, MATHIAS AND LEUNG, DENNIS. Testing independence in high dimensions with sums of rank correlations	280–307

DUCHI, JOHN, KHOSRAVI, KHASHAYAR AND RUAN, FENG. Multiclass classification, information, divergence and surrogate risk	3246–3275
EFRON, BRADLEY. Curvature and inference for maximum likelihood estimates	1664–1692
ELENBERG, ETHAN R., KHANNA, RAJIV, DIMAKIS, ALEXANDROS G. AND NEGAHBAN, SAHAND. Restricted strong convexity implies weak submodularity	3539–3568
ELSENER, ANDREAS AND VAN DE GEER, SARA. Robust low-rank matrix estimation	3481–3509
ELTZNER, BENJAMIN AND HUCKEMANN, STEPHAN F. Backward nested descriptors asymptotics with inference on stem cell differentiation	1994–2019
ERNEST, JAN, BÜHLMANN, PETER AND ROTHENHÄUSLER, DOMINIK. Causal inference in partially linear structural equation models	2904–2938
ESCANCIANO, JUAN CARLOS, PARDO-FERNÁNDEZ, JUAN CARLOS AND VAN KEILEGOM, INGRID. Asymptotic distribution-free tests for semiparametric regressions with dependent data . .	1167–1196
ESCOBAR-BACH, MIKAEL, GOEGEBEUR, YURI AND GUILLOU, ARMELLE. Local robust estimation of the Pickands dependence function	2806–2843
EVANS, ROBIN J. Margins of discrete Bayesian networks	2623–2656
FAN, AILIN, SONG, RUI, LU, WENBIN AND SHI, CHENGCHUN. High-dimensional A -learning for optimal dynamic treatment regimes	925–957
FAN, JIANQING, LIU, HAN, LU, JUNWEI, ZHU, ZIWEI AND BATTLEY, HEATHER. Distributed testing and estimation under sparse high dimensional models	1352–1382
FAN, JIANQING, LIU, HAN, SUN, QIANG AND ZHANG, TONG. I-LAMM for sparse learning: Simultaneous control of algorithmic complexity and statistical error	814–841
FAN, JIANQING, LIU, HAN AND WANG, WEICHEN. Large covariance estimation through elliptical factor models	1383–1414
FAN, JIANQING, LIU, HAN, ZHOU, WEN-XIN AND BOSE, KOUSHIKI. A new perspective on robust M -estimation: Finite sample theory and applications to dependence-adjusted multiple testing	1904–1931
FAN, JIANQING, SHAO, QI-MAN AND ZHOU, WEN-XIN. Are discoveries spurious? Distributions of maximum spurious correlations and their applications	989–1017

FAN, ZHOU AND GUAN, LEYING. Approximate ℓ_0 -penalized estimation of piecewise-constant signals on graphs.....	3217–3245
FINNER, HELMUT AND GONTSCHARUK, VERONIKA. Two-sample Kolmogorov–Smirnov-type tests revisited: Old and new tests in terms of local levels	3014–3037
FRYZLEWICZ, PIOTR. Tail-greedy bottom-up data decompositions and fast multiple change-point detection	3390–3421
FUJIKOSHI, YASUNORI, BAI, ZHIDONG AND CHOI, KWOK PUI. Consistency of AIC and BIC in estimating the number of significant components in high-dimensional principal component analysis	1050–1076
FUJIKOSHI, YASUNORI, BAI, ZHIDONG AND CHOI, KWOK PUI. Limiting behavior of eigenvalues in high-dimensional MANOVA via RMT	2985–3013
FUKASAWA, MASAAKI AND BROUSTE, ALEXANDRE. Local asymptotic normality property for fractional Gaussian noise under high-frequency observations	2045–2061
GAMARNIK, DAVID AND LI, QUAN. Finding a large submatrix of a Gaussian random matrix.....	2511–2561
GAO, CHAO, MA, ZONGMING, ZHANG, ANDERSON Y. AND ZHOU, HARRISON H. Community detection in degree-corrected block models	2153–2185
GAO, CHAO, REN, ZHAO AND CHEN, MENGJIE. Robust covariance and scatter matrix estimation under Huber’s contamination model	1932–1960
GAO, FUQING, XIONG, JIE AND ZHAO, XINGQIU. Moderate deviations and nonparametric inference for monotone functions....	1225–1254
GAO, JITI, ZHANG, BO AND PAN, GUANGMING. CLT for largest eigenvalues and unit root testing for high-dimensional nonstationary time series	2186–2215
GAVISH, MATAN, JOHNSTONE, IAIN AND DONOHO, DAVID. Optimal shrinkage of eigenvalues in the spiked covariance model .	1742–1778
GHIGLIETTI, ANDREA, ROSENBERGER, WILLIAM F. AND ALETTI, GIACOMO. Nonparametric covariate-adjusted response-adaptive design based on a functional urn model.....	3838–3866
GLOTER, ARNAUD, LOUKIANOVA, DASHA AND MAI, HILMAR. Jump filtering and efficient drift estimation for Lévy-driven SDEs	1445–1480
GODOLPHIN, JANET. Designs with blocks of size two and applications to microarray experiments.....	2775–2805

GOEGBEUR, YURI, GUILLOU, ARMELLE AND ESCOBAR-BACH, MIKAEL. Local robust estimation of the Pickands dependence function	2806–2843
GONTSCHARUK, VERONIKA AND FINNER, HELMUT. Two-sample Kolmogorov–Smirnov-type tests revisited: Old and new tests in terms of local levels	3014–3037
GRAPPIN, EDWIN, PARIS, QUENTIN AND DALALYAN, ARNAK S. On the exponentially weighted aggregate with the Laplace prior	2452–2478
GREGORY, KARL B., LAHIRI, SOUMENDRA N. AND NORDMAN, DANIEL J. A smooth block bootstrap for quantile regression with time series	1138–1166
GROENEBOOM, PIET AND HENDRICKX, KIM. Current status linear regression	1415–1444
GU, MENG YANG, WANG, XIAOJING AND BERGER, JAMES O. Robust Gaussian stochastic process emulation	3038–3066
GUȚĂ, MĂDĂLIN, NUSSBAUM, MICHAEL AND BUTUCEA, CRISTINA. Local asymptotic equivalence of pure states ensembles and quantum Gaussian white noise	3676–3706
GUAN, LEYING AND FAN, ZHOU. Approximate ℓ_0 -penalized estimation of piecewise-constant signals on graphs	3217–3245
GUILLOU, ARMELLE, ESCOBAR-BACH, MIKAEL AND GOEGBEUR, YURI. Local robust estimation of the Pickands dependence function	2806–2843
GUNTUBOYINA, ADITYANAND, SAMWORTH, RICHARD J. AND KIM, ARLENE K. H. Adaptation in log-concave density estimation	2279–2306
GUNTUBOYINA, ADITYANAND, WEI, YUTING AND CAI, T. TONY. Adaptive estimation of planar convex sets	1018–1049
GUO, BIN, YAO, QIWEI AND CHANG, JINYUAN. Principal component analysis for second-order stationary vector time series ..	2094–2124
GUO, WENWEN, ZHONG, WEI AND CUI, HENGJIAN. Test for high-dimensional regression coefficients using refitted cross-validation variance estimation	958–988
GUO, ZIJIAN AND CAI, T. TONY. Accuracy assessment for high-dimensional linear regression	1807–1836
HAN, KYUNGHEE AND PARK, BYEONG U. Smooth backfitting for errors-in-variables additive models	2216–2250
HAUSER, ALAIN, MAATHUIS, MARLOES H. AND NANDY, PREETAM. High-dimensional consistency in score-based and hybrid structure learning	3151–3183

HE, XUMING, ZHENG, QI AND PENG, LIMIN. High dimensional censored quantile regression	308–343
HE, YUANZHEN, CHENG, CHING-SHUI AND TANG, BOXIN. Strong orthogonal arrays of strength two plus	457–468
HEINRICH, PHILIPPE AND KAHN, JONAS. Strong identifiability and optimal minimax rates for finite mixture estimation	2844–2870
HENDRICKX, KIM AND GROENEBOOM, PIET. Current status linear regression	1415–1444
HO, LAM SI TUNG, SUCHARD, MARC A., MATSEN IV, FREDERICK A. AND DINH, VU. Consistency and convergence rate of phylogenetic inference via regularization	1481–1512
HOLMES, CHRIS, MUNK, AXEL AND BEHR, MERLE. Multiscale blind source separation	711–744
HÖRMANN, SIEGFRIED, KOKOSZKA, PIOTR AND NISOL, GILLES. Testing for periodicity in functional time series	2960–2984
HUANG, JIAN, LV, SHAOGAO, LIN, HUAZHEN AND LIAN, HENG. Oracle inequalities for sparse additive quantile regression in reproducing kernel Hilbert space	781–813
HUCKEMANN, STEPHAN F. AND ELTZNER, BENJAMIN. Backward nested descriptors asymptotics with inference on stem cell differentiation	1994–2019
JANKOVÁ, JANA AND VAN DE GEER, SARA. Semiparametric efficiency bounds for high-dimensional models	2336–2359
JAVANMARD, ADEL AND MONTANARI, ANDREA. Debiasing the lasso: Optimal sample size for Gaussian designs	2593–2622
JAVANMARD, ADEL AND MONTANARI, ANDREA. Online rules for control of false discovery rate and false discovery exceedance .	526–554
JIANG, BAI, WU, TUNG-YU, JIN, YIFAN AND WONG, WING H. Convergence of contrastive divergence algorithm in exponential family	3067–3098
JIN, BAISUO AND LI, JIALIANG. Multi-threshold accelerated failure time model	2657–2682
JIN, YIFAN, WONG, WING H., JIANG, BAI AND WU, TUNG-YU. Convergence of contrastive divergence algorithm in exponential family	3067–3098
JOHNSTONE, IAIN, DONOHO, DAVID AND GAVISH, MATAN. Optimal shrinkage of eigenvalues in the spiked covariance model .	1742–1778
JOSEPH, V. ROSHAN AND MAK, SIMON. Support points	2562–2592
JUDITSKY, ANATOLI AND NEMIROVSKI, ARKADI. Near-optimality of linear recovery in Gaussian observation scheme under $\ \cdot\ _2^2$ -loss	1603–1629

KAHN, JONAS AND HEINRICH, PHILIPPE. Strong identifiability and optimal minimax rates for finite mixture estimation	2844–2870
KAUFMANN, EMILIE. On Bayesian index policies for sequential resource allocation	842–865
KHANNA, RAJIV, DIMAKIS, ALEXANDROS G., NEGAHBAN, SAHAND AND ELENBERG, ETHAN R. Restricted strong convexity implies weak submodularity	3539–3568
KHOSRAVI, KHASHAYAR, RUAN, FENG AND DUCHI, JOHN. Multiclass classification, information, divergence and surrogate risk	3246–3275
KIM, ARLENE K. H., GUNTUBOYINA, ADITYANAND AND SAMWORTH, RICHARD J. Adaptation in log-concave density estimation	2279–2306
KOKOSZKA, PIOTR, NISOL, GILLES AND HÖRMANN, SIEGFRIED. Testing for periodicity in functional time series	2960–2984
KOLAR, MLADEN AND BARBER, RINA FOYGEL. ROCKET: <u>Robust confidence intervals via Kendall’s tau</u> for transelliptical graphical models	3422–3450
KONG, XIN-BING. On the systematic and idiosyncratic volatility with large panel high-frequency data	1077–1108
KOUL, HIRA L., SONG, WEIXING AND ZHU, XIAOQING. Goodness-of-fit testing of error distribution in linear measurement error models	2479–2510
KUNERT, JOACHIM AND NEUMANN, CHRISTOPH. On MSE-optimal crossover design	2939–2959
LAHIRI, PARTHA AND YOSHIMORI HIROSE, MASAYO. Estimating variance of random effects to solve multiple problems simultaneously	1721–1741
LAHIRI, SOUMENDRA N., NORDMAN, DANIEL J. AND GREGORY, KARL B. A smooth block bootstrap for quantile regression with time series	1138–1166
LAURITZEN, STEFFEN AND SADEGHI, KAYVAN. Unifying Markov properties for graphical models	2251–2278
LECUÉ, GUILLAUME AND MENDELSON, SHAHAR. Regularization and the small-ball method I: Sparse recovery	611–641
LECUÉ, GUILLAUME, TSYBAKOV, ALEXANDRE B. AND BELLEC, PIERRE C. Slope meets Lasso: Improved oracle bounds and optimality	3603–3642
LENKOSKI, ALEX, RICHARDS, DONALD AND UHLER, CAROLINE. Exact formulas for the normalizing constants of Wishart distributions for graphical models	90–118

LEUNG, DENNIS AND DRTON, MATHIAS. Testing independence in high dimensions with sums of rank correlations	280–307
LEVINA, ELIZAVETA AND AMINI, ARASH A. On semidefinite relaxations for the block model	149–179
LI, JIALIANG AND JIN, BAISUO. Multi-threshold accelerated failure time model	2657–2682
LI, QUAN AND GAMARNIK, DAVID. Finding a large submatrix of a Gaussian random matrix	2511–2561
LI, XIAODONG, XU, JIAMING AND CHEN, YUDONG. Convexified modularity maximization for degree-corrected stochastic block models	1573–1602
LI, XIAOOU, LIU, JINGCHEN AND YING, ZHILIANG. Chernoff index for Cox test of separate parametric families	1–29
LIAN, HENG, HUANG, JIAN, LV, SHAOGAO AND LIN, HUAZHEN. Oracle inequalities for sparse additive quantile regression in reproducing kernel Hilbert space	781–813
LIN, HUAZHEN, LIAN, HENG, HUANG, JIAN AND LV, SHAOGAO. Oracle inequalities for sparse additive quantile regression in reproducing kernel Hilbert space	781–813
LIN, QIAN, ZHAO, ZHIGEN AND LIU, JUN S. On consistency and sparsity for sliced inverse regression in high dimensions	580–610
LIN, ZHENHUA, MÜLLER, HANS-GEORG AND YAO, FANG. Mixture inner product spaces and their application to functional data analysis	370–400
LING, HOK KAN, SIT, TONY, YAM, SHEUNG CHI PHILLIP AND CHAN, KWUN CHUEN GARY. Estimation of a monotone density in s -sample biased sampling models	2125–2152
LIU, HAN, LU, JUNWEI, ZHU, ZIWEI, BATTEY, HEATHER AND FAN, JIANQING. Distributed testing and estimation under sparse high dimensional models	1352–1382
LIU, HAN, SUN, QIANG, ZHANG, TONG AND FAN, JIANQING. I-LAMM for sparse learning: Simultaneous control of algorithmic complexity and statistical error	814–841
LIU, HAN, WANG, WEICHEN AND FAN, JIANQING. Large covariance estimation through elliptical factor models	1383–1414
LIU, HAN, ZHANG, TONG AND ZHAO, TUO. Pathwise coordinate optimization for sparse learning: Algorithm and theory	180–218
LIU, HAN, ZHOU, WEN-XIN, BOSE, KOUSHIKI AND FAN, JIANQING. A new perspective on robust M -estimation: Finite sample theory and applications to dependence-adjusted multiple testing	1904–1931

LIU, JINGCHEN, YING, ZHILIANG AND LI, XIAOOU. Chernoff index for Cox test of separate parametric families	1–29
LIU, JUN S., LIN, QIAN AND ZHAO, ZHIGEN. On consistency and sparsity for sliced inverse regression in high dimensions	580–610
LIU, WEIDONG AND CHEN, XI. Testing independence with high-dimensional correlated samples	866–894
LIU, YUFENG, ZHAO, JUNLONG AND YU, GUAN. Assessing robustness of classification using an angular breakdown point....	3362–3389
LOBATO, IGNACIO N. AND VELASCO, CARLOS. Frequency domain minimum distance inference for possibly noninvertible and noncausal ARMA models	555–579
LOUKIANOVA, DASHA, MAI, HILMAR AND GLOTER, ARNAUD. Jump filtering and efficient drift estimation for Lévy-driven SDEs	1445–1480
LOUNICI, KARIM, MEZIANI, KATIA AND PEYRÉ, GABRIEL. Adaptive sup-norm estimation of the Wigner function in noisy quantum homodyne tomography	1318–1351
LU, JUNWEI, ZHU, ZIWEI, BATTEY, HEATHER, FAN, JIANQING AND LIU, HAN. Distributed testing and estimation under sparse high dimensional models	1352–1382
LU, WENBIN, SHI, CHENGCHUN, FAN, AILIN AND SONG, RUI. High-dimensional A-learning for optimal dynamic treatment regimes	925–957
LV, SHAOGAO, LIN, HUAZHEN, LIAN, HENG AND HUANG, JIAN. Oracle inequalities for sparse additive quantile regression in reproducing kernel Hilbert space	781–813
MA, ZONGMING, ZHANG, ANDERSON Y., ZHOU, HARRISON H. AND GAO, CHAO. Community detection in degree-corrected block models	2153–2185
MAATHUIS, MARLOES H., NANDY, PREETAM AND HAUSER, ALAIN. High-dimensional consistency in score-based and hybrid structure learning	3151–3183
MAI, HILMAR, GLOTER, ARNAUD AND LOUKIANOVA, DASHA. Jump filtering and efficient drift estimation for Lévy-driven SDEs	1445–1480
MAK, SIMON AND JOSEPH, V. ROSHAN. Support points.....	2562–2592
MALEKI, ARIAN, BARANIUK, RICHARD G. AND MOUSAVI, ALI. Consistent parameter estimation for LASSO and approximate message passing	119–148
MALEKI, ARIAN, ZHENG, LE AND WENG, HAOLEI. Overcoming the limitations of phase transition by higher order analysis of regularization techniques.....	3099–3129

MARTÍN, NIRIAN AND MOLINA, ISABEL. Empirical best prediction under a nested error model with log transformation	1961–1993
MATSEN IV, FREDERICK A., DINH, VU, HO, LAM SI TUNG AND SUCHARD, MARC A. Consistency and convergence rate of phylogenetic inference via regularization	1481–1512
MEI, SONG, BAI, YU AND MONTANARI, ANDREA. The landscape of empirical risk for nonconvex losses	2747–2774
MENDELSON, SHAHAR AND LECUÉ, GUILLAUME. Regularization and the small-ball method I: Sparse recovery	611–641
MEZIANI, KATIA, PEYRÉ, GABRIEL AND LOUNICI, KARIM. Adaptive sup-norm estimation of the Wigner function in noisy quantum homodyne tomography	1318–1351
MINSKER, STANISLAV. Sub-Gaussian estimators of the mean of a random matrix with heavy-tailed entries	2871–2903
MOITRA, ANKUR, PERRY, AMELIA, WEIN, ALEXANDER S. AND BANDEIRA, AFONSO S. Optimality and sub-optimality of PCA I: Spiked random matrix models	2416–2451
MOLINA, ISABEL AND MARTÍN, NIRIAN. Empirical best prediction under a nested error model with log transformation	1961–1993
MONTANARI, ANDREA AND JAVANMARD, ADEL. Debiasing the lasso: Optimal sample size for Gaussian designs	2593–2622
MONTANARI, ANDREA AND JAVANMARD, ADEL. Online rules for control of false discovery rate and false discovery exceedance	526–554
MONTANARI, ANDREA, MEI, SONG AND BAI, YU. The landscape of empirical risk for nonconvex losses	2747–2774
MOUSAVI, ALI, MALEKI, ARIAN AND BARANIUK, RICHARD G. Consistent parameter estimation for LASSO and approximate message passing	119–148
MUKERJEE, RAHUL, DASGUPTA, TIRTHANKAR, ZHAO, ANQI AND DING, PENG. Randomization-based causal inference from split-plot designs	1876–1903
MUKHERJEE, GOURAB, WEINSTEIN, ASAF AND BROWN, LAWRENCE D. Empirical Bayes estimates for a two-way cross-classified model	1693–1720
MUKHERJEE, RAJARSHI, MUKHERJEE, SUMIT AND SEN, SUBHABRATA. Detection thresholds for the β -model on sparse graphs	1288–1317
MUKHERJEE, RAJARSHI, MUKHERJEE, SUMIT AND YUAN, MING. Global testing against sparse alternatives under Ising models	2062–2093
MUKHERJEE, SUMIT, SEN, SUBHABRATA AND MUKHERJEE, RAJARSHI. Detection thresholds for the β -model on sparse graphs	1288–1317

MUKHERJEE, SUMIT, YUAN, MING AND MUKHERJEE, RAJARSHI. Global testing against sparse alternatives under Ising models...	2062–2093
MÜLLER, HANS-GEORG AND DAI, XIONGTAO. Principal component analysis for functional data on Riemannian manifolds and spheres.....	3334–3361
MÜLLER, HANS-GEORG, YAO, FANG AND LIN, ZHENHUA. Mixture inner product spaces and their application to functional data analysis	370–400
MUNK, AXEL, BEHR, MERLE AND HOLMES, CHRIS. Multiscale blind source separation	711–744
MUNK, AXEL, PROKSCH, KATHARINA AND WERNER, FRANK. Multiscale scanning in inverse problems	3569–3602
NANDY, PREETAM, HAUSER, ALAIN AND MAATHUIS, MARLOES H. High-dimensional consistency in score-based and hybrid structure learning	3151–3183
NDAOUD, MOHAMED, STEPANOVA, NATALIA A., TSYBAKOV, ALEXANDRE B. AND BUTUCEA, CRISTINA. Variable selection with Hamming loss.....	1837–1875
NEGAHBAN, SAHAND, ELENBERG, ETHAN R., KHANNA, RAJIV AND DIMAKIS, ALEXANDROS G. Restricted strong convexity implies weak submodularity.....	3539–3568
NEMIROVSKI, ARKADI AND JUDITSKY, ANATOLI. Near-optimality of linear recovery in Gaussian observation scheme under $\ \cdot\ _2^2$ -loss.....	1603–1629
NETTLETON, DAN, QIU, YUMOU AND CHEN, SONG XI. Detecting rare and faint signals via thresholding maximum likelihood estimators	895–923
NEUMANN, CHRISTOPH AND KUNERT, JOACHIM. On MSE-optimal crossover design.....	2939–2959
NISOL, GILLES, HÖRMANN, SIEGFRIED AND KOKOSZKA, PIOTR. Testing for periodicity in functional time series.....	2960–2984
NORDMAN, DANIEL J., GREGORY, KARL B. AND LAHIRI, SOUMENDRA N. A smooth block bootstrap for quantile regression with time series	1138–1166
NUSSBAUM, MICHAEL, BUTUCEA, CRISTINA AND GUȚĂ, MĂDĂLIN. Local asymptotic equivalence of pure states ensembles and quantum Gaussian white noise	3676–3706
OTSU, TAISUKE AND SEO, MYUNG HWAN. Local M-estimation with discontinuous criterion for dependent and limited observations	344–369

PAINDAVEINE, DAVY AND VAN BEVER, GERMAIN. Halfspace depths for scatter, concentration and shape matrices.....	3276–3307
PAN, GUANGMING, GAO, JITI AND ZHANG, BO. CLT for largest eigenvalues and unit root testing for high-dimensional nonstationary time series.....	2186–2215
PAN, WENLIANG, TIAN, YUAN, WANG, XUEQIN AND ZHANG, HEPING. Ball Divergence: Nonparametric two sample test ...	1109–1137
PAPARODITIS, EFSTATHIOS. Sieve bootstrap for functional time series.....	3510–3538
PARDO-FERNÁNDEZ, JUAN CARLOS, VAN KEILEGOM, INGRID AND ESCANCIANO, JUAN CARLOS. Asymptotic distribution-free tests for semiparametric regressions with dependent data ..	1167–1196
PARIS, QUENTIN, DALALYAN, ARNAK S. AND GRAPPIN, EDWIN. On the exponentially weighted aggregate with the Laplace prior	2452–2478
PARK, BYEONG U. AND HAN, KYUNGHEE. Smooth backfitting for errors-in-variables additive models.....	2216–2250
PARK, YEONHEE AND DOSS, HANI. An MCMC approach to empirical Bayes inference and Bayesian sensitivity analysis via empirical processes	1630–1663
PENG, LIMIN, HE, XUMING AND ZHENG, QI. High dimensional censored quantile regression.....	308–343
PENNEC, XAVIER. Barycentric subspace analysis on manifolds ...	2711–2746
PERRY, AMELIA, WEIN, ALEXANDER S., BANDEIRA, AFONSO S. AND MOITRA, ANKUR. Optimality and sub-optimality of PCA I: Spiked random matrix models	2416–2451
PEYRÉ, GABRIEL, LOUNICI, KARIM AND MEZIANI, KATIA. Adaptive sup-norm estimation of the Wigner function in noisy quantum homodyne tomography	1318–1351
PRETORIUS, CHARL AND SWANEPOEL, JAN W. H. On the asymptotic theory of new bootstrap confidence bounds.....	438–456
PRIEBE, CAREY E. AND TANG, MINH. Limit theorems for eigenvectors of the normalized Laplacian for random graphs.....	2360–2415
PROKSCH, KATHARINA, WERNER, FRANK AND MUNK, AXEL. Multiscale scanning in inverse problems.....	3569–3602
QIU, YUMOU, CHEN, SONG XI AND NETTLETON, DAN. Detecting rare and faint signals via thresholding maximum likelihood estimators	895–923
QU, ANNIE, SHEN, XIAOTONG AND BI, XUAN. Multilayer tensor factorization with applications to recommender systems.....	3308–3333
REN, ZHAO, CHEN, MENGJIE AND GAO, CHAO. Robust covariance and scatter matrix estimation under Huber’s contamination model.....	1932–1960

RICHARDS, DONALD, UHLER, CAROLINE AND LENKOSKI, ALEX. Exact formulas for the normalizing constants of Wishart distributions for graphical models	90–118
RINALDO, ALESSANDRO, TIBSHIRANI, ROB, WASSERMAN, LARRY AND TIBSHIRANI, RYAN J. Uniform asymptotic inference and the bootstrap after model selection	1255–1287
ROČKOVÁ, VERONIKA. Bayesian estimation of sparse signals with a continuous spike-and-slab prior	401–437
ROSENBERGER, WILLIAM F., ALETTI, GIACOMO AND GHIGLI- ETTI, ANDREA. Nonparametric covariate-adjusted response- adaptive design based on a functional urn model	3838–3866
ROTHENHÄUSLER, DOMINIK, ERNEST, JAN AND BÜHLMANN, PE- TER. Causal inference in partially linear structural equation models	2904–2938
RUAN, FENG, DUCHI, JOHN AND KHOSRAVI, KHASHAYAR. Mul- ticlass classification, information, divergence and surrogate risk	3246–3275
SADEGHI, KAYVAN AND LAURITZEN, STEFFEN. Unifying Markov properties for graphical models	2251–2278
SAMWORTH, RICHARD J., KIM, ARLENE K. H. AND GUNTUBOY- INA, ADITYANAND. Adaptation in log-concave density esti- mation	2279–2306
SEN, SUBHABRATA, MUKHERJEE, RAJARSHI AND MUKHERJEE, SUMIT. Detection thresholds for the β -model on sparse graphs	1288–1317
SEO, MYUNG HWAN AND OTSU, TAISUKE. Local M-estimation with discontinuous criterion for dependent and limited observa- tions	344–369
SHAO, QI-MAN, ZHOU, WEN-XIN AND FAN, JIANQING. Are dis- coveries spurious? Distributions of maximum spurious correla- tions and their applications	989–1017
SHAO, XIAOFENG, ZHANG, XIANYANG AND YAO, SHUN. Condi- tional mean and quantile dependence testing in high dimension	219–246
SHEN, XIAOTONG, BI, XUAN AND QU, ANNIE. Multilayer tensor factorization with applications to recommender systems	3308–3333
SHI, CHENGCHUN, FAN, AILIN, SONG, RUI AND LU, WENBIN. High-dimensional A -learning for optimal dynamic treatment regimes	925–957
SIENKIEWICZ, ELA AND WANG, HAONAN. Pareto quantiles of un- labeled tree objects	1513–1540
SIT, TONY, YAM, SHEUNG CHI PHILLIP, CHAN, KWUN CHUEN GARY AND LING, HOK KAN. Estimation of a monotone den- sity in s -sample biased sampling models	2125–2152

SONG, RUI, LU, WENBIN, SHI, CHENGCHUN AND FAN, AILIN. High-dimensional A-learning for optimal dynamic treatment regimes	925–957
SONG, WEIXING, ZHU, XIAOQING AND KOUL, HIRA L. Good- ness-of-fit testing of error distribution in linear measurement er- ror models	2479–2510
STEPANOVA, NATALIA A., TSYBAKOV, ALEXANDRE B., BU- TUCEA, CRISTINA AND NDAOUD, MOHAMED. Variable se- lection with Hamming loss	1837–1875
STRAUCH, CLAUDIA. Adaptive invariant density estimation for er- godic diffusions over anisotropic classes	3451–3480
SUBBA RAO, SUHASINI. Statistical inference for spatial statistics defined in the Fourier domain	469–499
SUCHARD, MARC A., MATSEN IV, FREDERICK A., DINH, VU AND HO, LAM SI TUNG. Consistency and convergence rate of phy- logenetic inference via regularization	1481–1512
SUN, QIANG, ZHANG, TONG, FAN, JIANQING AND LIU, HAN. I-LAMM for sparse learning: Simultaneous control of algorithm- mic complexity and statistical error	814–841
SWANEPOEL, JAN W. H. AND PRETORIUS, CHARL. On the asymp- totic theory of new bootstrap confidence bounds	438–456
TANG, BOXIN, HE, YUANZHEN AND CHENG, CHING-SHUI. Strong orthogonal arrays of strength two plus	457–468
TANG, CHENG YONG, WU, TONG TONG AND CHANG, JINYUAN. A new scope of penalized empirical likelihood with high- dimensional estimating equations	3185–3216
TANG, MINH AND PRIEBE, CAREY E. Limit theorems for eigen- vectors of the normalized Laplacian for random graphs	2360–2415
TAYLOR, JONATHAN AND TIAN, XIAOYING. Selective inference with a randomized response	679–710
TIAN, XIAOYING AND TAYLOR, JONATHAN. Selective inference with a randomized response	679–710
TIAN, YUAN, WANG, XUEQIN, ZHANG, HEPING AND PAN, WEN- LIANG. Ball Divergence: Nonparametric two sample test	1109–1137
TIBSHIRANI, ROB, WASSERMAN, LARRY, TIBSHIRANI, RYAN J. AND RINALDO, ALESSANDRO. Uniform asymptotic infer- ence and the bootstrap after model selection	1255–1287
TIBSHIRANI, RYAN J., RINALDO, ALESSANDRO, TIBSHIRANI, ROB AND WASSERMAN, LARRY. Uniform asymptotic infer- ence and the bootstrap after model selection	1255–1287

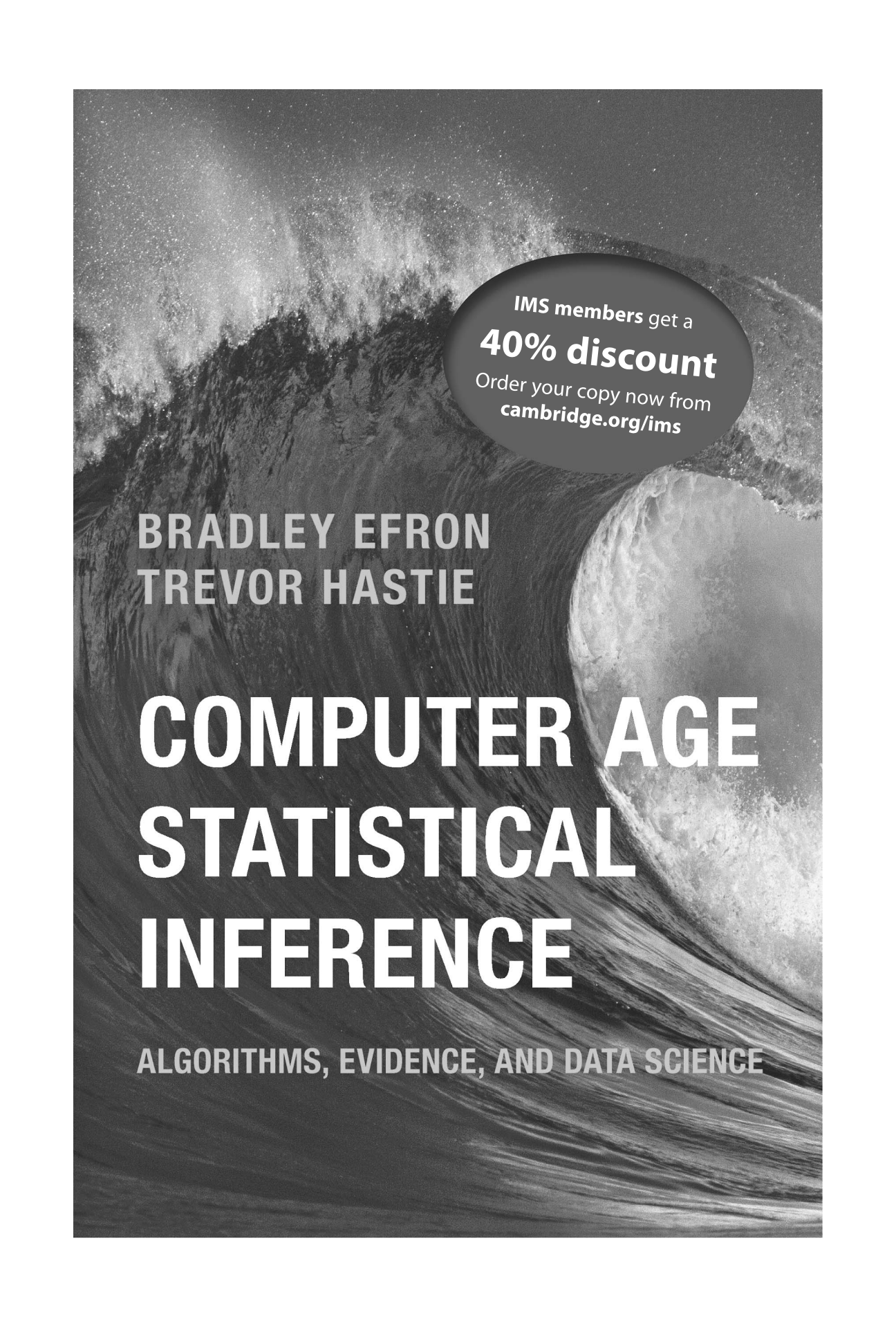
TSYBAKOV, ALEXANDRE B., BELLEC, PIERRE C. AND LECUÉ, GUILLAUME. Slope meets Lasso: Improved oracle bounds and optimality	3603–3642
TSYBAKOV, ALEXANDRE B., BUTUCEA, CRISTINA, NDAOUD, MOHAMED AND STEPANOVA, NATALIA A. Variable selection with Hamming loss.....	1837–1875
TSYBAKOV, ALEXANDRE B., VERZELEN, NICOLAS, COLLIER, OLIVIER AND COMMINGES, LAËTITIA. Optimal adaptive estimation of linear functionals under sparsity	3130–3150
UHLER, CAROLINE, LENKOSKI, ALEX AND RICHARDS, DONALD. Exact formulas for the normalizing constants of Wishart distributions for graphical models.....	90–118
VAN BEVER, GERMAIN AND PAINDAVEINE, DAVY. Halfspace depths for scatter, concentration and shape matrices.....	3276–3307
VAN DE GEER, SARA AND ELSENER, ANDREAS. Robust low-rank matrix estimation.....	3481–3509
VAN DE GEER, SARA AND JANKOVÁ, JANA. Semiparametric efficiency bounds for high-dimensional models.....	2336–2359
VAN KEILEGOM, INGRID, ESCANCIANO, JUAN CARLOS AND PARDO-FERNÁNDEZ, JUAN CARLOS. Asymptotic distribution-free tests for semiparametric regressions with dependent data	1167–1196
VELASCO, CARLOS AND LOBATO, IGNACIO N. Frequency domain minimum distance inference for possibly noninvertible and noncausal ARMA models	555–579
VERZELEN, NICOLAS, COLLIER, OLIVIER, COMMINGES, LAËTITIA AND TSYBAKOV, ALEXANDRE B. Optimal adaptive estimation of linear functionals under sparsity	3130–3150
WAGER, STEFAN AND DOBRIBAN, EDGAR. High-dimensional asymptotics of prediction: Ridge regression and classification..	247–279
WANG, GUANGHUI, ZOU, CHANGLIANG AND YIN, GUOSHENG. Change-point detection in multinomial data with a large number of categories	2020–2044
WANG, HAONAN AND SIENKIEWICZ, ELA. Pareto quantiles of unlabeled tree objects	1513–1540
WANG, LIN, XIAO, QIAN AND XU, HONGQUAN. Optimal maximin L_1 -distance Latin hypercube designs based on good lattice point design	3741–3766
WANG, WEICHEN, FAN, JIANQING AND LIU, HAN. Large covariance estimation through elliptical factor models	1383–1414

WANG, XIAOJING, BERGER, JAMES O. AND GU, MENG YANG.	
Robust Gaussian stochastic process emulation	3038–3066
WANG, XUEQIN, ZHANG, HEPING, PAN, WENLIANG AND TIAN, YUAN.	
Ball Divergence: Nonparametric two sample test	1109–1137
WASSERMAN, LARRY, TIBSHIRANI, RYAN J., RINALDO, ALESSANDRO AND TIBSHIRANI, ROB.	
Uniform asymptotic inference and the bootstrap after model selection	1255–1287
WEI, YING, BELLONI, ALEXANDRE, CHERNOZHUKOV, VICTOR AND CHETVERIKOV, DENIS.	
Uniformly valid post-regularization confidence regions for many functional parameters in z-estimation framework	3643–3675
WEI, YUTING, CAI, T. TONY AND GUNTUBOYINA, ADITYANAND.	
Adaptive estimation of planar convex sets	1018–1049
WEIN, ALEXANDER S., BANDEIRA, AFONSO S., MOITRA, ANKUR AND PERRY, AMELIA.	
Optimality and sub-optimality of PCA I: Spiked random matrix models	2416–2451
WEINSTEIN, ASAF, BROWN, LAWRENCE D. AND MUKHERJEE, GOURAB.	
Empirical Bayes estimates for a two-way cross-classified model	1693–1720
WENG, HAOLEI, MALEKI, ARIAN AND ZHENG, LE.	
Overcoming the limitations of phase transition by higher order analysis of regularization techniques	3099–3129
WERNER, FRANK, MUNK, AXEL AND PROKSCH, KATHARINA.	
Multiscale scanning in inverse problems	3569–3602
WONG, WING H., JIANG, BAI, WU, TUNG-YU AND JIN, YIFAN.	
Convergence of contrastive divergence algorithm in exponential family	3067–3098
WU, HAU-TIENG AND WU, NAN.	
Think globally, fit locally under the manifold setup: Asymptotic analysis of locally linear embedding	3805–3837
WU, NAN AND WU, HAU-TIENG.	
Think globally, fit locally under the manifold setup: Asymptotic analysis of locally linear embedding	3805–3837
WU, TONG TONG, CHANG, JINYUAN AND TANG, CHENG YONG.	
A new scope of penalized empirical likelihood with high-dimensional estimating equations	3185–3216
WU, TUNG-YU, JIN, YIFAN, WONG, WING H. AND JIANG, BAI.	
Convergence of contrastive divergence algorithm in exponential family	3067–3098
WU, WEICHI AND ZHOU, ZHOU.	
Gradient-based structural change detection for nonstationary time series M-estimation	1197–1224

XIA, NINGNING AND ZHENG, XINGHUA. On the inference about the spectral distribution of high-dimensional covariance matrix based on high-frequency noisy observations	500–525
XIAO, QIAN, XU, HONGQUAN AND WANG, LIN. Optimal maximin L_1 -distance Latin hypercube designs based on good lattice point design	3741–3766
XIONG, JIE, ZHAO, XINGQIU AND GAO, FUQING. Moderate deviations and nonparametric inference for monotone functions . . .	1225–1254
XU, HONGQUAN, WANG, LIN AND XIAO, QIAN. Optimal maximin L_1 -distance Latin hypercube designs based on good lattice point design	3741–3766
XU, JIAMING, CHEN, YUDONG AND LI, XIAODONG. Convexified modularity maximization for degree-corrected stochastic block models	1573–1602
XU, KAI, ZHU, LIPING AND ZHANG, YAOWU. Measuring and testing for interval quantile dependence	2683–2710
YAM, SHEUNG CHI PHILLIP, CHAN, KWUN CHUEN GARY, LING, HOK KAN AND SIT, TONY. Estimation of a monotone density in s -sample biased sampling models	2125–2152
YAO, FANG, LIN, ZHENHUA AND MÜLLER, HANS-GEORG. Mixture inner product spaces and their application to functional data analysis	370–400
YAO, QIWEI, CHANG, JINYUAN AND GUO, BIN. Principal component analysis for second-order stationary vector time series . .	2094–2124
YAO, SHUN, SHAO, XIAOFENG AND ZHANG, XIANYANG. Conditional mean and quantile dependence testing in high dimension	219–246
YIN, GUOSHENG, WANG, GUANGHUI AND ZOU, CHANGLIANG. Change-point detection in multinomial data with a large number of categories	2020–2044
YING, ZHILIANG, LI, XIAOOU AND LIU, JINGCHEN. Chernoff index for Cox test of separate parametric families	1–29
YOSHIMORI HIROSE, MASAYO AND LAHIRI, PARTHA. Estimating variance of random effects to solve multiple problems simultaneously	1721–1741
YU, GUAN, LIU, YUFENG AND ZHAO, JUNLONG. Assessing robustness of classification using an angular breakdown point . . .	3362–3389
YUAN, MING, MUKHERJEE, RAJARSHI AND MUKHERJEE, SUMIT. Global testing against sparse alternatives under Ising models . . .	2062–2093
ZHANG, ANDERSON Y., ZHOU, HARRISON H., GAO, CHAO AND MA, ZONGMING. Community detection in degree-corrected block models	2153–2185

ZHANG, ANRU AND CAI, T. TONY. Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics	60–89
ZHANG, BO, PAN, GUANGMING AND GAO, JITI. CLT for largest eigenvalues and unit root testing for high-dimensional nonstationary time series	2186–2215
ZHANG, HEPING, PAN, WENLIANG, TIAN, YUAN AND WANG, XUEQIN. Ball Divergence: Nonparametric two sample test...	1109–1137
ZHANG, TONG, FAN, JIANQING, LIU, HAN AND SUN, QIANG. I-LAMM for sparse learning: Simultaneous control of algorithmic complexity and statistical error	814–841
ZHANG, TONG, ZHAO, TUO AND LIU, HAN. Pathwise coordinate optimization for sparse learning: Algorithm and theory	180–218
ZHANG, XIANYANG, YAO, SHUN AND SHAO, XIAOFENG. Conditional mean and quantile dependence testing in high dimension	219–246
ZHANG, YAOWU, XU, KAI AND ZHU, LIPING. Measuring and testing for interval quantile dependence	2683–2710
ZHANG, YICHONG. Extremal quantile treatment effects	3707–3740
ZHAO, ANQI, DING, PENG, MUKERJEE, RAHUL AND DASGUPTA, TIRTHANKAR. Randomization-based causal inference from split-plot designs	1876–1903
ZHAO, JUNLONG, YU, GUAN AND LIU, YUFENG. Assessing robustness of classification using an angular breakdown point...	3362–3389
ZHAO, TUO, LIU, HAN AND ZHANG, TONG. Pathwise coordinate optimization for sparse learning: Algorithm and theory	180–218
ZHAO, XINGQIU, GAO, FUQING AND XIONG, JIE. Moderate deviations and nonparametric inference for monotone functions...	1225–1254
ZHAO, ZHIGEN, LIU, JUN S. AND LIN, QIAN. On consistency and sparsity for sliced inverse regression in high dimensions	580–610
ZHENG, LE, WENG, HAOLEI AND MALEKI, ARIAN. Overcoming the limitations of phase transition by higher order analysis of regularization techniques	3099–3129
ZHENG, QI, PENG, LIMIN AND HE, XUMING. High dimensional censored quantile regression	308–343
ZHENG, XINGHUA AND XIA, NINGNING. On the inference about the spectral distribution of high-dimensional covariance matrix based on high-frequency noisy observations	500–525
ZHONG, WEI, CUI, HENGJIAN AND GUO, WENWEN. Test for high-dimensional regression coefficients using refitted cross-validation variance estimation	958–988

ZHOU, HARRISON H., GAO, CHAO, MA, ZONGMING AND ZHANG, ANDERSON Y. Community detection in degree-corrected block models	2153–2185
ZHOU, WEN-XIN, BOSE, KOUISHIKI, FAN, JIANQING AND LIU, HAN. A new perspective on robust M -estimation: Finite sample theory and applications to dependence-adjusted multiple testing	1904–1931
ZHOU, WEN-XIN, FAN, JIANQING AND SHAO, QI-MAN. Are discoveries spurious? Distributions of maximum spurious correlations and their applications	989–1017
ZHOU, ZHOU AND WU, WEICHI. Gradient-based structural change detection for nonstationary time series M -estimation	1197–1224
ZHU, LIPING, ZHANG, YAOWU AND XU, KAI. Measuring and testing for interval quantile dependence	2683–2710
ZHU, XIAOQING, KOUL, HIRA L. AND SONG, WEIXING. Goodness-of-fit testing of error distribution in linear measurement error models	2479–2510
ZHU, ZIWEI, BATTEY, HEATHER, FAN, JIANQING, LIU, HAN AND LU, JUNWEI. Distributed testing and estimation under sparse high dimensional models	1352–1382
ZOU, CHANGLIANG, YIN, GUOSHENG AND WANG, GUANGHUI. Change-point detection in multinomial data with a large number of categories	2020–2044



IMS members get a
40% discount

Order your copy now from
cambridge.org/ims

BRADLEY EFRON
TREVOR HASTIE

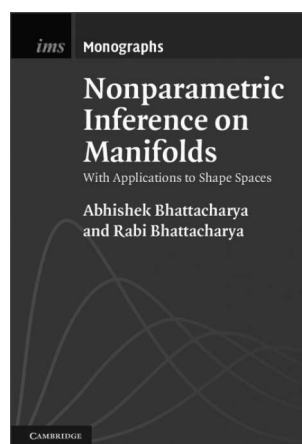
COMPUTER AGE STATISTICAL INFERENCE

ALGORITHMS, EVIDENCE, AND DATA SCIENCE



The Institute of Mathematical Statistics presents

IMS MONOGRAPHS



Nonparametric Inference on Manifolds *With Applications to Shape Spaces*

Abhishek Bhattacharya, Rabi Bhattacharya

This book introduces in a systematic manner a general nonparametric theory of statistics on manifolds, with emphasis on manifolds of shapes. The theory has important and varied applications in medical diagnostics, image analysis, and machine vision. An early chapter of examples establishes the effectiveness of the new methods and demonstrates how they outperform their parametric counterparts. Inference is developed for both intrinsic and extrinsic Fréchet means of probability distributions on manifolds, then applied to shape spaces defined as orbits of landmarks under a Lie group of transformations—in particular, similarity, reflection similarity, affine and projective transformations. In addition, nonparametric Bayesian theory is adapted and extended to manifolds for the purposes of density estimation, regression and classification. Ideal for statisticians who analyze manifold data and wish to develop their own methodology, this book is also of interest to probabilists, mathematicians, computer scientists and morphometricians with mathematical training.

IMS member? Claim
your 40% discount:
www.cambridge.org/ims

Hardback price
US\$51.00
(non-member price
\$85.00)

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.