

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Chernoff index for Cox test of separate parametric families	
XIAOOU LI, JINGCHEN LIU AND ZHILIANG YING	1
Optimal bounds for aggregation of affine estimators	PIERRE C. BELLEC 30
Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics	T. TONY CAI AND ANRU ZHANG 60
Exact formulas for the normalizing constants of Wishart distributions for graphical models	CAROLINE UHLER, ALEX LENKOSKI AND DONALD RICHARDS 90
Consistent parameter estimation for LASSO and approximate message passing	ALI MOUSAVI, ARIAN MALEKI AND RICHARD G. BARANIUK 119
On semidefinite relaxations for the block model	ARASH A. AMINI AND ELIZAVETA LEVINA 149
Pathwise coordinate optimization for sparse learning: Algorithm and theory	TUO ZHAO, HAN LIU AND TONG ZHANG 180
Conditional mean and quantile dependence testing in high dimension	XIANYANG ZHANG, SHUN YAO AND XIAOFENG SHAO 219
High-dimensional asymptotics of prediction: Ridge regression and classification	EDGAR DOBRIBAN AND STEFAN WAGER 247
Testing independence in high dimensions with sums of rank correlations	DENNIS LEUNG AND MATHIAS DRTON 280
High dimensional censored quantile regression	QI ZHENG, LIMIN PENG AND XUMING HE 308
Local M-estimation with discontinuous criterion for dependent and limited observations	MYUNG HWAN SEO AND TAISUKE OTSU 344
Mixture inner product spaces and their application to functional data analysis	ZHENHUA LIN, HANS-GEORG MÜLLER AND FANG YAO 370
Bayesian estimation of sparse signals with a continuous spike-and-slab prior	VERONIKA ROČKOVÁ 401
On the asymptotic theory of new bootstrap confidence bounds	CHARL PRETORIUS AND JAN W. H. SWANEPOEL 438

THE ANNALS OF STATISTICS

Vol. 46, No. 1, pp. 1–456 February 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:** Alison Etheridge, Department of Statistics, University of Oxford, Oxford, OX1 3LB, United Kingdom
- President-Elect:** Xiao-Li Meng, Department of Statistics, Harvard University, Cambridge, Massachusetts 02138-2901, USA
- Past President:** Jon Wellner, Department of Statistics, University of Washington, Seattle, Washington 98195-4322, USA
- Executive Secretary:** Edsel Peña, Department of Statistics, University of South Carolina, Columbia, South Carolina 29208-001, USA
- Treasurer:** Zhengjun Zhang, Department of Statistics, University of Wisconsin, Madison, Wisconsin 53706-1510, USA
- Program Secretary:** Judith Rousseau, Université Paris Dauphine, Place du Maréchal DeLattre de Tassigny, 75016 Paris, France

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 1, February 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.

CHERNOFF INDEX FOR COX TEST OF SEPARATE PARAMETRIC FAMILIES

BY XIAOOU LI*, JINGCHEN LIU^{1,†} AND ZHILIANG YING^{1,†}

University of Minnesota and Columbia University[†]*

The asymptotic efficiency of a generalized likelihood ratio test proposed by Cox is studied under the large deviations framework for error probabilities developed by Chernoff. In particular, two separate parametric families of hypotheses are considered [In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob.* (1961) 105–123; *J. Roy. Statist. Soc. Ser. B* **24** (1962) 406–424]. The significance level is set such that the maximal type I and type II error probabilities for the generalized likelihood ratio test decay exponentially fast with the same rate. We derive the analytic form of such a rate that is also known as the Chernoff index [*Ann. Math. Stat.* **23** (1952) 493–507], a relative efficiency measure when there is no preference between the null and the alternative hypotheses. We further extend the analysis to approximate error probabilities when the two families are not completely separated. Discussions are provided concerning the implications of the present result on model selection.

REFERENCES

- ADLER, R. J., BLANCHET, J. H. and LIU, J. (2012). Efficient Monte Carlo for high excursions of Gaussian random fields. *Ann. Appl. Probab.* **22** 1167–1214. [MR2977989](#)
- ANDRESEN, A. and SPOKOINY, V. (2014). Critical dimension in profile semiparametric estimation. *Electron. J. Stat.* **8** 3077–3125. [MR3301302](#)
- ARCONES, M. A. (2006). Large deviations for M-estimators. *Ann. Inst. Statist. Math.* **58** 21–52. [MR2281205](#)
- BAHADUR, R. R. (1960). Stochastic comparison of tests. *Ann. Math. Stat.* **31** 276–295. [MR0116413](#)
- BAHADUR, R. R. (1967). Rates of convergence of estimates and test statistics. *Ann. Math. Stat.* **38** 303–324. [MR0207085](#)
- BERRINGTON DE GONZÁLEZ, A. and COX, D. R. (2007). Interpretation of interaction: A review. *Ann. Appl. Stat.* **1** 371–385. [MR2415740](#)
- BRAGANCA PEREIRA, B. (2005). Separate families of hypotheses. In *Encyclopedia of Biostatistics*.
- CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Stat.* **23** 493–507. [MR0057518](#)
- COX, D. R. (1961). Tests of separate families of hypotheses. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I* 105–123. Univ. California Press, Berkeley, CA. [MR0131927](#)
- COX, D. R. (1962). Further results on tests of separate families of hypotheses. *J. Roy. Statist. Soc. Ser. B* **24** 406–424. [MR0156409](#)
- COX, D. R. (2013). A return to an old paper: ‘Tests of separate families of hypotheses’. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 207–215. [MR3021385](#)

MSC2010 subject classifications. Primary 62F03; secondary 62J12, 62F12.

Key words and phrases. Asymptotic relative efficiency, generalized likelihood ratio, generalized linear models, large deviation, model selection, nonnested hypotheses, variable selection.

- DAVIDSON, R. and MACKINNON, J. G. (1981). Several tests for model specification in the presence of alternative hypotheses. *Econometrica* **49** 781–793. [MR0619482](#)
- FINE, J. P. (2002). Comparing nonnested Cox models. *Biometrika* **89** 635–647. [MR1929168](#)
- HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability* (Berkeley, Calif., 1965/66), Vol. I: *Statistics* 221–233. Univ. California Press, Berkeley, CA. [MR0216620](#)
- KALLENBERG, W. C. M. (1983). Intermediate efficiency, theory and examples. *Ann. Statist.* **11** 170–182. [MR0684874](#)
- LI, X. and LIU, J. (2015). Rare-event simulation and efficient discretization for the supremum of Gaussian random fields. *Adv. in Appl. Probab.* **47** 787–816. [MR3406608](#)
- LI, X., LIU, J. and YING, Z. (2017). Supplement to “Chernoff Index for Cox Test of Separate Parametric Families.” DOI:[10.1214/16-AOS1532SUPP](#).
- LIU, J., LU, J. and ZHOU, X. (2015). Efficient rare event simulation for failure problems in random media. *SIAM J. Sci. Comput.* **37** A609–A624. [MR3317386](#)
- LIU, J. and XU, G. (2012). Some asymptotic results of Gaussian random fields with varying mean functions and the associated processes. *Ann. Statist.* **40** 262–293. [MR3014307](#)
- LIU, J. and XU, G. (2014a). On the conditional distributions and the efficient simulations of exponential integrals of Gaussian random fields. *Ann. Appl. Probab.* **24** 1691–1738. [MR3211008](#)
- LIU, J. and XU, G. (2014b). Efficient simulations for the exponential integrals of Hölder continuous Gaussian random fields. *ACM Trans. Model. Comput. Simul.* **24** Art. 9, 24. [MR3195228](#)
- MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models*. Chapman & Hall, London. [MR3223057](#)
- PESARAN, M. H. (1974). On the general problem of model selection. *Rev. Econ. Stud.* **41** 153–171.
- PESARAN, M. H. (1984). Asymptotic power comparisons of tests of separate parametric families by Bahadur’s approach. *Biometrika* **71** 245–252. [MR0767152](#)
- PESARAN, M. H. and DEATON, A. S. (1978). Testing non-nested nonlinear regression models. *Econometrica* **46** 677–694. [MR0501711](#)
- RUKHIN, A. L. (1993). Bahadur efficiency of tests of separate hypotheses and adaptive test statistics. *J. Amer. Statist. Assoc.* **88** 161–165. [MR1212484](#)
- SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464. [MR0468014](#)
- SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. [MR0595165](#)
- SPOKOINY, V. (2012). Parametric estimation. Finite sample theory. *Ann. Statist.* **40** 2877–2909. [MR3097963](#)
- VUONG, Q. H. (1989). Likelihood ratio tests for model selection and nonnested hypotheses. *Econometrica* **57** 307–333. [MR0996939](#)
- WHITE, H. (1982a). Maximum likelihood estimation of misspecified models. *Econometrica* **50** 1–25. [MR0640163](#)
- WHITE, H. (1982b). Regularity conditions for Cox’s test of nonnested hypothesis. *J. Econometrics* **19** 301–318. [MR0672058](#)

OPTIMAL BOUNDS FOR AGGREGATION OF AFFINE ESTIMATORS

BY PIERRE C. BELLEC¹

ENSAE and Rutgers University

We study the problem of aggregation of estimators when the estimators are not independent of the data used for aggregation and no sample splitting is allowed. If the estimators are deterministic vectors, it is well known that the minimax rate of aggregation is of order $\log(M)$, where M is the number of estimators to aggregate. It is proved that for affine estimators, the minimax rate of aggregation is unchanged: it is possible to handle the linear dependence between the affine estimators and the data used for aggregation at no extra cost. The minimax rate is not impacted either by the variance of the affine estimators, or any other measure of their statistical complexity. The minimax rate is attained with a penalized procedure over the convex hull of the estimators, for a penalty that is inspired from the Q -aggregation procedure. The results follow from the interplay between the penalty, strong convexity and concentration.

REFERENCES

- [1] ARLOT, S. and BACH, F. R. (2009). Data-driven calibration of linear estimators with minimal penalties. In *Advances in Neural Information Processing Systems* 46–54.
- [2] AUDIBERT, J.-Y. (2007). No fast exponential deviation inequalities for the progressive mixture rule. Preprint. Available at [arXiv:math/0703848](https://arxiv.org/abs/math/0703848).
- [3] BARAUD, Y., GIRAUD, C. and HUET, S. (2014). Estimator selection in the Gaussian setting. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** 1092–1119. [MR3224300](#)
- [4] BELLEC, P. C. (2017). Optimal exponential bounds for aggregation of density estimators. *Bernoulli* **23** 219–248. [MR3556772](#)
- [5] BELLONI, A. and CHERNOZHUKOV, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli* **19** 521–547. [MR3037163](#)
- [6] BELLONI, A., CHERNOZHUKOV, V. and WANG, L. (2014). Pivotal estimation via square-root lasso in nonparametric regression. *Ann. Statist.* **42** 757–788. [MR3210986](#)
- [7] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford Univ. Press, London.
- [8] BOYD, S. and VANDENBERGHE, L. (2009). *Convex Optimization*. Cambridge Univ. Press, Cambridge. [MR2061575](#)
- [9] COHEN, A. (1966). All admissible linear estimates of the mean vector. *Ann. Math. Stat.* **37** 458–463. [MR0189164](#)
- [10] DAI, D., RIGOLLET, P., XIA, L. and ZHANG, T. (2014). Aggregation of affine estimators. *Electron. J. Stat.* **8** 302–327. [MR3192554](#)

MSC2010 subject classifications. Primary 62G05; secondary 62J07.

Key words and phrases. Affine estimator, aggregation, sequence model, sharp oracle inequality, concentration inequality, Hanson–Wright.

- [11] DAI, D., RIGOLLET, P. and ZHANG, T. (2012). Deviation optimal learning using greedy Q -aggregation. *Ann. Statist.* **40** 1878–1905. [MR3015047](#)
- [12] DALALYAN, A. S. and SALMON, J. (2012). Sharp oracle inequalities for aggregation of affine estimators. *Ann. Statist.* **40** 2327–2355. [MR3059085](#)
- [13] DETTE, H., MUNK, A. and WAGNER, T. (1998). Estimating the variance in nonparametric regression—what is a reasonable choice? *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 751–764. [MR1649480](#)
- [14] DONOHO, D. L., LIU, R. C. and MACGIBBON, B. (1990). Minimax risk over hyperrectangles, and implications. *Ann. Statist.* **18** 1416–1437. [MR1062717](#)
- [15] EFRONOVICH, S. Y. and PINSKER, M. S. (1984). A self-training algorithm for nonparametric filtering. *Avtomat. i Telemekh.* **11** 58–65. [MR0797991](#)
- [16] GERCHINOVITZ, S. (2011). Prediction of individual sequences and prediction in the statistical framework: Some links around sparse regression and aggregation techniques. Ph.D. thesis, Univ. Paris Sud-Paris XI. Available at <https://tel.archives-ouvertes.fr/tel-00653550>.
- [17] GIRAUD, C. (2008). Mixing least-squares estimators when the variance is unknown. *Bernoulli* **14** 1089–1107. [MR2543587](#)
- [18] GIRAUD, C., HUET, S. and VERZELEN, N. (2012). High-dimensional regression with unknown variance. *Statist. Sci.* **27** 500–518. [MR3025131](#)
- [19] HALL, P., KAY, J. W. and TITTERINGTON, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* **77** 521–528. [MR1087842](#)
- [20] HANSON, D. L. and WRIGHT, F. T. (1971). A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Stat.* **42** 1079–1083.
- [21] IMMERKAER, J. (1996). Fast noise variance estimation. *Comput. Vis. Image Underst.* **64** 300–302.
- [22] JOHNSTONE, I. M. (2002). Function estimation and gaussian sequence models. Unpublished manuscript 2 (5.3): 2.
- [23] LATAŁA, R. (1999). Tail and moment estimates for some types of chaos. *Studia Math.* **1** 39–53.
- [24] LAURENT, B. and MASSART, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* **28** 1302–1338. [MR1805785](#)
- [25] LECUÉ, G. (2013). Empirical risk minimization is optimal for the convex aggregation problem. *Bernoulli* **19** 2153–2166.
- [26] LECUÉ, G. and RIGOLLET, P. (2014). Optimal learning with Q -aggregation. *Ann. Statist.* **42** 211–224. [MR3178462](#)
- [27] LEUNG, G. and BARRON, A. R. (2006). Information theory and mixing least-squares regressions. *IEEE Trans. Inform. Theory* **52** 3396–3410.
- [28] MALLOWS, C. L. (1973). Some comments on Cp. *Technometrics* **15** 661–675.
- [29] MUNK, A., BISSANTZ, N., WAGNER, T. and FREITAG, G. (2005). On difference-based variance estimation in nonparametric regression when the covariate is high dimensional. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 19–41. [MR2136637](#)
- [30] NEMIROVSKI, A. (2000). Topics in non-parametric statistics. In *Lectures on Probability Theory and Statistics (Saint-Flour, 1998)*. *Lecture Notes in Math.* **1738**. Springer, Berlin.
- [31] RIGOLLET, P. (2012). Kullback–Leibler aggregation and misspecified generalized linear models. *Ann. Statist.* **40** 639–665. [MR2933661](#)
- [32] RIGOLLET, P. and TSYBAKOV, A. (2011). Exponential screening and optimal rates of sparse estimation. *Ann. Statist.* **39** 731–771. [MR2816337](#)
- [33] RIGOLLET, P. and TSYBAKOV, A. B. (2007). Linear and convex aggregation of density estimators. *Math. Methods Statist.* **16** 260–280. [MR2356821](#)
- [34] RIGOLLET, P. and TSYBAKOV, A. B. (2012). Sparse estimation by exponential weighting. *Statist. Sci.* **27** 558–575. [MR3025134](#)

- [35] RUDELSON, M. and VERSHYNIN, R. (2013). Hanson–Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.* **18** 1–9.
- [36] SUN, T. and ZHANG, C.-H. (2012). Scaled sparse linear regression. *Biometrika* **99** 879–898. [MR2999166](#)
- [37] TSYBAKOV, A. B. (2003). Optimal rates of aggregation. In *Learning Theory and Kernel Machines* 303–313. Springer, Berlin.
- [38] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York.
- [39] TSYBAKOV, A. B. (2014). Aggregation and minimax optimality in high-dimensional estimation. In *Proceedings of the International Congress of Mathematicians, Vol. 3* 225–246, Seoul.
- [40] VERSHYNIN, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge Univ. Press, Cambridge. [MR2963170](#)
- [41] WRIGHT, F. T. (1973). A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric. *Ann. Probab.* **1** 1068–1070. [MR353419](#)

RATE-OPTIMAL PERTURBATION BOUNDS FOR SINGULAR SUBSPACES WITH APPLICATIONS TO HIGH-DIMENSIONAL STATISTICS

BY T. TONY CAI^{1,*} AND ANRU ZHANG[†]

*University of Pennsylvania** and *University of Wisconsin–Madison*[†]

Perturbation bounds for singular spaces, in particular Wedin’s $\sin \Theta$ theorem, are a fundamental tool in many fields including high-dimensional statistics, machine learning and applied mathematics. In this paper, we establish separate perturbation bounds, measured in both spectral and Frobenius $\sin \Theta$ distances, for the left and right singular subspaces. Lower bounds, which show that the individual perturbation bounds are rate-optimal, are also given.

The new perturbation bounds are applicable to a wide range of problems. In this paper, we consider in detail applications to low-rank matrix denoising and singular space estimation, high-dimensional clustering and canonical correlation analysis (CCA). In particular, separate matching upper and lower bounds are obtained for estimating the left and right singular spaces. To the best of our knowledge, this is the first result that gives different optimal rates for the left and right singular spaces under the same perturbation.

REFERENCES

- ALZHRANI, T. and HORADAM, K. J. (2016). Community detection in bipartite networks: Algorithms and case studies. In *Complex Systems and Networks. Underst. Complex Syst.* 25–50. Springer, Heidelberg. [MR3586347](#)
- ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley, Hoboken, NJ.
- ARGYRIOU, A., EVGENIOU, T. and PONTIL, M. (2008). Convex multi-task feature learning. *Mach. Learn.* **73** 243–272.
- AZIZYAN, M., SINGH, A. and WASSERMAN, L. (2013). Minimax theory for high-dimensional Gaussian mixtures with sparse mean separation. In *Advances in Neural Information Processing Systems* 2139–2147.
- BALAKRISHNAN, S., XU, M., KRISHNAMURTHY, A. and SINGH, A. (2011). Noise thresholds for spectral clustering. In *Advances in Neural Information Processing Systems* 954–962.
- BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2012). The singular values and vectors of low rank perturbations of large rectangular random matrices. *J. Multivariate Anal.* **111** 120–135. [MR2944410](#)
- BORG, I. and GROENEN, P. J. (2005). *Modern Multidimensional Scaling: Theory and Applications*. Springer, Berlin.
- BOUTSIDIS, C., ZOUZIAS, A., MAHONEY, M. W. and DRINEAS, P. (2015). Randomized dimensionality reduction for k -means clustering. *IEEE Trans. Inform. Theory* **61** 1045–1062.

MSC2010 subject classifications. Primary 62H12, 62C20; secondary 62H25.

Key words and phrases. Canonical correlation analysis, clustering, high-dimensional statistics, low-rank matrix denoising, perturbation bound, singular value decomposition, $\sin \Theta$ distances, spectral method.

- BURA, E. and PFEIFFER, R. (2008). On the distribution of the left singular vectors of a random matrix and its applications. *Statist. Probab. Lett.* **78** 2275–2280.
- CAI, T. T., LI, X. and MA, Z. (2016). Optimal rates of convergence for noisy sparse phase retrieval via thresholded Wirtinger flow. *Ann. Statist.* **44** 2221–2251. [MR3546449](#)
- CAI, T. T., MA, Z. and WU, Y. (2013). Sparse PCA: Optimal rates and adaptive estimation. *Ann. Statist.* **41** 3074–3110. [MR3161458](#)
- CAI, T. T., MA, Z. and WU, Y. (2015). Optimal estimation and rank detection for sparse spiked covariance matrices. *Probab. Theory Related Fields* **161** 781–815.
- CAI, T. T. and ZHANG, A. (2017). Supplement to “Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics.” DOI:[10.1214/17-AOS1541SUPP](#).
- CAI, T., CAI, T. T., LIAO, K. and LIU, W. (2015). Large-scale simultaneous testing of cross-covariance matrix with applications to PheWAS. Technical report.
- CANDES, E. J. and PLAN, Y. (2010). Matrix completion with noise. *Proc. IEEE* **98** 925–936.
- CANDES, E. J. and RECHT, B. (2009). Exact matrix completion via convex optimization. *Found. Comput. Math.* **9** 717–772. [MR2565240](#)
- CANDES, E. J., SING-LONG, C. A. and TRZASKO, J. D. (2013). Unbiased risk estimates for singular value thresholding and spectral estimators. *IEEE Trans. Signal Process.* **61** 4643–4657. [MR3105401](#)
- CANDES, E. J. and TAO, T. (2010). The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inform. Theory* **56** 2053–2080.
- CAPITAINE, M., DONATI-MARTIN, C. and FÉRAL, D. (2009). The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations. *Ann. Probab.* **37** 1–47. [MR2489158](#)
- CHATTERJEE, S. (2014). Matrix estimation by universal singular value thresholding. *Ann. Statist.* **43** 177–214. [MR3285604](#)
- CHEN, M., GAO, C., REN, Z. and ZHOU, H. H. (2013). Sparse CCA via precision adjusted iterative thresholding. Preprint. Available at [arXiv:1311.6186](#).
- CHO, J., KIM, D. and ROHE, K. (2015). Asymptotic theory for estimating the singular vectors and values of a partially-observed low rank matrix with noise. Preprint. Available at [arXiv:1508.05431](#).
- DAVIS, C. and KAHAN, W. M. (1970). The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.* **7** 1–46.
- DONOHOO, D. and GAVISH, M. (2014). Minimax risk of matrix denoising by singular value thresholding. *Ann. Statist.* **42** 2413–2440. [MR3269984](#)
- DOPICO, F. M. (2000). A note on $\sin \Theta$ theorems for singular subspace variations. *BIT Numerical Mathematics* **40** 395–403.
- FAN, J., WANG, W. and ZHONG, Y. (2016). An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. Preprint. Available at [arXiv:1603.03516](#).
- FELDMAN, D., SCHMIDT, M. and SOHLER, C. (2013). Turning big data into tiny data: Constant-size coresets for k -means, PCA and projective clustering. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms* 1434–1453. SIAM, Philadelphia, PA.
- GAO, C., MA, Z. and ZHOU, H. H. (2014). Sparse CCA: Adaptive estimation and computational barriers. Preprint. Available at [arXiv:1409.8565](#).
- GAO, C., MA, Z., REN, Z. and ZHOU, H. H. (2015). Minimax estimation in sparse canonical correlation analysis. *Ann. Statist.* **43** 2168–2197. [MR3396982](#)
- GAVISH, M. and DONOHOO, D. L. (2014). The optimal hard threshold for singular values is $4/\sqrt{3}$. *IEEE Trans. Inform. Theory* **60** 5040–5053. [MR3245370](#)
- GOLDBERG, D., NICHOLS, D., OKI, B. M. and TERRY, D. (1992). Using collaborative filtering to weave an information tapestry. *Commun. ACM* **35** 61–70.
- GROSS, D. (2011). Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inform. Theory* **57** 1548–1566.

- HARDOON, D. R., SZEDMAK, S. and SHAW-ETAYLOR, J. (2004). Canonical correlation analysis: An overview with application to learning methods. *Neural Comput.* **16** 2639–2664.
- HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The Elements of Statistical Learning*, 2nd ed. *Springer Series in Statistics*. Springer, New York. [MR2722294](#)
- HOTELLING, H. (1936). Relations between two sets of variates. *Biometrika* **28** 321–377.
- JIN, J., KE, Z. T. and WANG, W. (2015). Phase transitions for high dimensional clustering and related problems. Preprint. Available at [arXiv:1502.06952](#).
- JIN, J. and WANG, W. (2016). Influential features PCA for high dimensional clustering (with discussion). *Ann. Statist.* **44** 2323–2359. [MR3576548](#)
- JOHNSTONE, I. M. and LU, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* **104** 682–693.
- KESHAVAN, R. H., MONTANARI, A. and OH, S. (2010). Matrix completion from noisy entries. *J. Mach. Learn. Res.* **11** 2057–2078.
- LEI, J. and RINALDO, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann. Statist.* **43** 215–237. [MR3285605](#)
- LIU, Z. and VANDENBERGHE, L. (2009). Interior-point method for nuclear norm approximation with application to system identification. *SIAM J. Matrix Anal. Appl.* **31** 1235–1256. [MR2558821](#)
- MA, Z. and LI, X. (2016). Subspace perspective on canonical correlation analysis: Dimension reduction and minimax rates. Preprint. Available at [arXiv:1605.03662](#).
- MELAMED, D. (2014). Community structures in bipartite networks: A dual-projection approach. *PLoS ONE* **9** e97823.
- O’ROURKE, S., VU, V. and WANG, K. (2013). Random perturbation of low rank matrices: Improving classical bounds. Preprint. Available at [arXiv:1311.2657](#).
- ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. [MR2893856](#)
- SHABALIN, A. A. and NOBEL, A. B. (2013). Reconstruction of a low-rank matrix in the presence of Gaussian noise. *J. Multivariate Anal.* **118** 67–76.
- SINGER, A. and CUCURINGU, M. (2010). Uniqueness of low-rank matrix completion by rigidity theory. *SIAM J. Matrix Anal. Appl.* **31** 1621–1641.
- STEWART, G. W. (1991). Perturbation theory for the singular value decomposition. In *SVD and Signal Processing II: Algorithms, Analysis and Applications* 99–109. Elsevier, Amsterdam.
- STEWART, M. (2006). Perturbation of the SVD in the presence of small singular values. *Linear Algebra Appl.* **419** 53–77.
- SUN, R. and LUO, Z.-Q. (2015). Guaranteed matrix completion via nonconvex factorization. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science—FOCS 2015* 270–289. IEEE Computer Soc., Los Alamitos, CA. [MR3473312](#)
- TAO, T. (2012). *Topics in Random Matrix Theory. Graduate Studies in Mathematics* **132**. Amer. Math. Soc., Providence, RI. [MR2906465](#)
- VERSHYNIN, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge Univ. Press, Cambridge.
- VON LUXBURG, U., BELKIN, M. and BOUSQUET, O. (2008). Consistency of spectral clustering. *Ann. Statist.* **36** 555–586. [MR2396807](#)
- VU, V. (2011). Singular vectors under random perturbation. *Random Structures Algorithms* **39** 526–538.
- WANG, R. (2015). Singular vector perturbation under Gaussian noise. *SIAM J. Matrix Anal. Appl.* **36** 158–177.
- WEDIN, P. (1972). Perturbation bounds in connection with singular value decomposition. *BIT* **12** 99–111.
- WEYL, H. (1912). Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung). *Math. Ann.* **71** 441–479.

- WITTEN, D. M., TIBSHIRANI, R. and HASTIE, T. (2009). A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* 515–534.
- YANG, D., MA, Z. and BUJA, A. (2016). Rate optimal denoising of simultaneously sparse and low rank matrices. *J. Mach. Learn. Res.* **17** Paper No. 92, 27. [MR3543498](#)
- YU, Y., WANG, T. and SAMWORTH, R. J. (2015). A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika* **102** 315–323.

EXACT FORMULAS FOR THE NORMALIZING CONSTANTS OF WISHART DISTRIBUTIONS FOR GRAPHICAL MODELS

BY CAROLINE UHLER¹, ALEX LENKOSKI² AND DONALD RICHARDS³

*Massachusetts Institute of Technology, Norwegian Computing Center and
 Penn State University*

Gaussian graphical models have received considerable attention during the past four decades from the statistical and machine learning communities. In Bayesian treatments of this model, the G -Wishart distribution serves as the conjugate prior for inverse covariance matrices satisfying graphical constraints. While it is straightforward to posit the unnormalized densities, the normalizing constants of these distributions have been known only for graphs that are chordal, or decomposable. Up until now, it was unknown whether the normalizing constant for a general graph could be represented explicitly, and a considerable body of computational literature emerged that attempted to avoid this apparent intractability. We close this question by providing an explicit representation of the G -Wishart normalizing constant for general graphs.

REFERENCES

- [1] ANDREWS, G. E., ASKEY, R. and ROY, R. (1999). *Special Functions. Encyclopedia of Mathematics and Its Applications* **71**. Cambridge Univ. Press, Cambridge. [MR1688958](#)
- [2] ATAY-KAYIS, A. and MASSAM, H. (2005). A Monte Carlo method for computing the marginal likelihood in nondecomposable Gaussian graphical models. *Biometrika* **92** 317–335. [MR2201362](#)
- [3] BOJDECKI, T. and GOROSTIZA, L. G. (1999). Fractional Brownian motion via fractional Laplacian. *Statist. Probab. Lett.* **44** 107–108. [MR1706362](#)
- [4] CHENG, Y. and LENKOSKI, A. (2012). Hierarchical Gaussian graphical models: Beyond reversible jump. *Electron. J. Stat.* **6** 2309–2331. [MR3020264](#)
- [5] DAWID, A. P. and LAURITZEN, S. L. (1993). Hyper-Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.* **21** 1272–1317. [MR1241267](#)
- [6] DEMPSTER, A. P. (1972). Covariance selection. *Biometrics* **28** 157–175.
- [7] DOBRA, A. and LENKOSKI, A. (2011). Copula Gaussian graphical models and their application to modeling functional disability data. *Ann. Appl. Stat.* **5** 969–993. [MR2840183](#)
- [8] DOBRA, A., LENKOSKI, A. and RODRIGUEZ, A. (2011). Bayesian inference for general Gaussian graphical models with application to multivariate lattice data. *J. Amer. Statist. Assoc.* **106** 1418–1433. [MR2896846](#)
- [9] GÄRDING, L. (1947). The solution of Cauchy’s problem for two totally hyperbolic linear differential equations by means of Riesz integrals. *Ann. of Math. (2)* **48** 785–826. [MR0022648](#)

MSC2010 subject classifications. Primary 62H05, 60E05; secondary 62E15.

Key words and phrases. Bartlett decomposition, bipartite graph, Cholesky decomposition, chordal graph, directed acyclic graph, G -Wishart distribution, Gaussian graphical model, generalized hypergeometric function of matrix argument, moral graph, normalizing constant, Wishart distribution.

- [10] GIRI, N. C. (2004). *Multivariate Statistical Analysis*. Dekker, New York. [MRMR0468025](#)
- [11] GIUDICI, P. and GREEN, P. J. (1999). Decomposable graphical Gaussian model determination. *Biometrika* **86** 785–801. [MR1741977](#)
- [12] GROSS, K. I. and RICHARDS, D. ST. P. (1987). Special functions of matrix argument. I. Algebraic induction, zonal polynomials, and hypergeometric functions. *Trans. Amer. Math. Soc.* **301** 781–811. [MR0882715](#)
- [13] HAFF, L. R., KIM, P. T., KOO, J.-Y. and RICHARDS, D. ST. P. (2011). Minimax estimation for mixtures of Wishart distributions. *Ann. Statist.* **39** 3417–3440. [MR3012414](#)
- [14] HERZ, C. S. (1955). Bessel functions of matrix argument. *Ann. of Math.* (2) **61** 474–523. [MR0069960](#)
- [15] HILLE, E. and PHILLIPS, R. S. (1957). *Functional Analysis and Semi-Groups*, rev. ed. *American Mathematical Society Colloquium Publications* **31**. Amer. Math. Soc., Providence, RI. [MR0089373](#)
- [16] INGHAM, A. E. (1933). An integral which occurs in statistics. *Math. Proc. Cambridge Philos. Soc.* **29** 271–276.
- [17] JAMES, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Stat.* **35** 475–501. [MR0181057](#)
- [18] JONES, B., CARVALHO, C., DOBRA, A., HANS, C., CARTER, C. and WEST, M. (2005). Experiments in stochastic computation for high-dimensional graphical models. *Statist. Sci.* **20** 388–400. [MR2210226](#)
- [19] KARLIN, S. (1968). *Total Positivity, Vol. I*. Stanford Univ. Press, Stanford, CA. [MR0230102](#)
- [20] KIM, P. T. and RICHARDS, D. ST. P. (2011). Deconvolution density estimation on the space of positive definite symmetric matrices. In *Nonparametric Statistics and Mixture Models* 147–168. World Sci. Publ., Hackensack, NJ. [MR2838725](#)
- [21] LAURITZEN, S. L. (1996). *Graphical Models. Oxford Statistical Science Series* **17**. Clarendon Press, Oxford. [MR1419991](#)
- [22] LENKOSKI, A. (2013). A direct sampler for G -Wishart variates. *Stat* **2** 119–128.
- [23] LENKOSKI, A. and DOBRA, A. (2011). Computational aspects related to inference in Gaussian graphical models with the G -Wishart prior. *J. Comput. Graph. Statist.* **20** 140–157. Supplementary material available online. [MR2816542](#)
- [24] LETAC, G. and MASSAM, H. (2007). Wishart distributions for decomposable graphs. *Ann. Statist.* **35** 1278–1323. [MR2341706](#)
- [25] MAASS, H. (1971). *Siegel's Modular Forms and Dirichlet Series. Lecture Notes in Mathematics* **216**. Springer, Berlin. Dedicated to the last great representative of a passing epoch Carl Ludwig Siegel on the occasion of his seventy-fifth birthday. [MR0344198](#)
- [26] MITSAKAKIS, N., MASSAM, H. and ESCOBAR, M. D. (2011). A Metropolis–Hastings based method for sampling from the G -Wishart distribution in Gaussian graphical models. *Electron. J. Stat.* **5** 18–30. [MR2763796](#)
- [27] MUIRHEAD, R. J. (1975). Expressions for some hypergeometric functions of matrix argument with applications. *J. Multivariate Anal.* **5** 283–293. [MR0381137](#)
- [28] MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York. [MR0652932](#)
- [29] OLKIN, I. (2002). The 70th anniversary of the distribution of random matrices: A survey. *Linear Algebra Appl.* **354** 231–243. Ninth special issue on linear algebra and statistics. [MR1927659](#)
- [30] PICCIONI, M. (2000). Independence structure of natural conjugate densities to exponential families and the Gibbs' sampler. *Scand. J. Stat.* **27** 111–127. [MR1774047](#)
- [31] ROVERATO, A. (2002). Hyper inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models. *Scand. J. Stat.* **29** 391–411. [MR1925566](#)

- [32] SIEGEL, C. L. (1935). Über die analytische Theorie der quadratischen Formen. *Ann. of Math.* (2) **36** 527–606. [MR1503238](#)
- [33] SPEED, T. P. and KIIVERI, H. T. (1986). Gaussian Markov distributions over finite graphs. *Ann. Statist.* **14** 138–150. [MR0829559](#)
- [34] UHLER, C., LENKOSKI, A. and RICHARDS, D. (2017). Supplement to “Exact formulas for the normalizing constants of Wishart distributions for graphical models.” DOI:[10.1214/17-AOS1543SUPP](#).
- [35] WANG, H. and CARVALHO, C. M. (2010). Simulation of hyper-inverse Wishart distributions for non-decomposable graphs. *Electron. J. Stat.* **4** 1470–1475. [MR2741209](#)
- [36] WANG, H. and LI, S. Z. (2012). Efficient Gaussian graphical model determination under G -Wishart prior distributions. *Electron. J. Stat.* **6** 168–198. [MR2879676](#)
- [37] WISHART, J. (1928). The generalised product moment distribution in samples from a normal multivariate population. *Biometrika* **20A** 32–52.
- [38] WISHART, J. and BARTLETT, M. S. (1933). The generalised product moment distribution in a normal system. *Math. Proc. Cambridge Philos. Soc.* **29** 260–270.

CONSISTENT PARAMETER ESTIMATION FOR LASSO AND APPROXIMATE MESSAGE PASSING

BY ALI MOUSAVI*, ARIAN MALEKI[†] AND RICHARD G. BARANIUK*

*Rice University** and *Columbia University[†]*

This paper studies the optimal tuning of the regularization parameter in LASSO or the threshold parameters in approximate message passing (AMP). Considering a model in which the design matrix and noise are zero-mean i.i.d. Gaussian, we propose a data-driven approach for estimating the regularization parameter of LASSO and the threshold parameters in AMP. Our estimates are consistent, that is, they converge to their asymptotically optimal values in probability as n , the number of observations, and p , the ambient dimension of the sparse vector, grow to infinity, while n/p converges to a fixed number δ . As a byproduct of our analysis, we will shed light on the asymptotic properties of the solution paths of LASSO and AMP.

REFERENCES

- [1] AMELUNXEN, D., LOTZ, M., MCCOY, M. B. and TROPP, J. A. (2013). Living on the edge: A geometric theory of phase transitions in convex optimization. ArXiv preprint. Available at [arXiv:1303.6672](https://arxiv.org/abs/1303.6672).
- [2] BAYATI, M., LELARGE, M. and MONTANARI, A. (2015). Universality in polytope phase transitions and message passing algorithms. *Ann. Appl. Probab.* **25** 753–822. [MR3313755](#)
- [3] BAYATI, M. and MONTANARI, A. (2011). The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inform. Theory* **57** 764–785. [MR2810285](#)
- [4] BAYATI, M. and MONTANARI, A. (2012). The LASSO risk for Gaussian matrices. *IEEE Trans. Inform. Theory* **58** 1997–2017. [MR2951312](#)
- [5] BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. [MR2533469](#)
- [6] BUNEA, F., TSYBAKOV, A. and WEGKAMP, M. (2007). Sparsity oracle inequalities for the Lasso. *Electron. J. Stat.* **1** 169–194. [MR2312149](#)
- [7] CAI, T. T., WANG, L. and XU, G. (2010). Shifting inequality and recovery of sparse signals. *IEEE Trans. Signal Process.* **58** 1300–1308. [MR2730209](#)
- [8] CAI, T. T., WANG, L. and XU, G. (2010). Stable recovery of sparse signals and an oracle inequality. *IEEE Trans. Inform. Theory* **56** 3516–3522. [MR2799010](#)
- [9] CAI, T. T., XU, G. and ZHANG, J. (2009). On recovery of sparse signals via ℓ_1 minimization. *IEEE Trans. Inform. Theory* **55** 3388–3397. [MR2598028](#)
- [10] CANDÈS, E. J., ROMBERG, J. K. and TAO, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.* **59** 1207–1223. [MR2230846](#)
- [11] CHATTERJEE, S. and JAFAROV, J. (2015). Prediction error of cross-validated Lasso. ArXiv preprint. Available at [arXiv:1502.06291](https://arxiv.org/abs/1502.06291).

MSC2010 subject classifications. 62G05, 62J05.

Key words and phrases. LASSO, estimation, sparsity, approximate message passing.

- [12] CHEN, S. S., DONOHO, D. L. and SAUNDERS, M. A. (1998). Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **20** 33–61. [MR1639094](#)
- [13] DONOHO, D. and TANNER, J. (2009). Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **367** 4273–4293. [MR2546388](#)
- [14] DONOHO, D. L. (2006). High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. *Discrete Comput. Geom.* **35** 617–652. [MR2225676](#)
- [15] DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2009). Message-passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci. USA* **106** 18914–18919.
- [16] DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2011). The noise-sensitivity phase transition in compressed sensing. *IEEE Trans. Inform. Theory* **57** 6920–6941. [MR2882271](#)
- [17] DONOHO, D. L. and TANNER, J. (2005). Neighborliness of randomly projected simplices in high dimensions. *Proc. Natl. Acad. Sci. USA* **102** 9452–9457. [MR2168716](#)
- [18] DONOHO, D. L. and TANNER, J. (2009). Counting faces of randomly projected polytopes when the projection radically lowers dimension. *J. Amer. Math. Soc.* **22** 1–53. [MR2449053](#)
- [19] DONOHO, D. L. and TANNER, J. (2010). Precise undersampling theorems. *Proc. IEEE* **98** 913–924.
- [20] EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499. [MR2060166](#)
- [21] HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J. and FRANKLIN, J. (2001). The elements of statistical learning: Data mining, inference, and prediction. *Math. Intelligencer* **27** 83–85.
- [22] HOMRIGHAUSEN, D. and McDONALD, D. J. (2014). Leave-one-out cross-validation is risk consistent for lasso. *Mach. Learn.* **97** 65–78. [MR3252827](#)
- [23] KNIGHT, K. and FU, W. (2000). Asymptotics for lasso-type estimators. *Ann. Statist.* **28** 1356–1378. [MR1805787](#)
- [24] LOCKHART, R., TAYLOR, J., TIBSHIRANI, R. J. and TIBSHIRANI, R. (2014). A significance test for the lasso. *Ann. Statist.* **42** 413–468. [MR3210970](#)
- [25] MALEKI, A. (2010). Approximate message passing algorithm for compressed sensing. Ph.D. thesis, Stanford Univ., Stanford, CA.
- [26] MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. [MR2278363](#)
- [27] MEINSHAUSEN, N. and YU, B. (2009). Lasso-type recovery of sparse representations for high-dimensional data. *Ann. Statist.* **37** 246–270. [MR2488351](#)
- [28] MOUSAVI, A., MALEKI, A. and BARANIUK, R. G. (2013). Parameterless optimal approximate message passing. ArXiv preprint. Available at [arXiv:1311.0035](#).
- [29] MOUSAVI, A., MALEKI, A. and BARANIUK, R. G. (2018). Supplement to “Consistent parameter estimation for LASSO and approximate message passing.” DOI:[10.1214/17-AOS1544SUPP](#).
- [30] OYMAK, S. and HASSIBI, B. (2012). On a relation between the minimax risk and the phase transitions of compressed recovery. In *50th Annual Allerton Conference on Communication, Control, and Computing* 1018–1025. IEEE, New York. DOI:[10.1109/Allerton.2012.6483330](#).
- [31] 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](#)
- [32] STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. [MR0630098](#)
- [33] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **58** 267–288. [MR1379242](#)

- [34] TIBSHIRANI, R. J. (2013). The lasso problem and uniqueness. *Electron. J. Stat.* **7** 1456–1490. [MR3066375](#)
- [35] 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](#)
- [36] WASSERMAN, L. and ROEDER, K. (2009). High-dimensional variable selection. *Ann. Statist.* **37** 2178–2201. [MR2543689](#)
- [37] ZHANG, C.-H. and HUANG, J. (2008). The sparsity and bias of the LASSO selection in high-dimensional linear regression. *Ann. Statist.* **36** 1567–1594. [MR2435448](#)
- [38] ZHANG, T. (2009). Some sharp performance bounds for least squares regression with L_1 regularization. *Ann. Statist.* **37** 2109–2144. [MR2543687](#)
- [39] ZHAO, P. and YU, B. (2006). On model selection consistency of Lasso. *J. Mach. Learn. Res.* **7** 2541–2563. [MR2274449](#)
- [40] ZHENG, L., MALEKI, A., WANG, X. and LONG, T. (2015). Does ℓ_p -minimization outperform ℓ_1 -minimization? ArXiv preprint. Available at [arXiv:1501.03704](#).
- [41] ZOU, H., HASTIE, T. and TIBSHIRANI, R. (2007). On the “degrees of freedom” of the lasso. *Ann. Statist.* **35** 2173–2192. [MR2363967](#)

ON SEMIDEFINITE RELAXATIONS FOR THE BLOCK MODEL¹

BY ARASH A. AMINI AND ELIZAVETA LEVINA

University of California, Los Angeles and University of Michigan

The stochastic block model (SBM) is a popular tool for community detection in networks, but fitting it by maximum likelihood (MLE) involves a computationally infeasible optimization problem. We propose a new semidefinite programming (SDP) solution to the problem of fitting the SBM, derived as a relaxation of the MLE. We put ours and previously proposed SDPs in a unified framework, as relaxations of the MLE over various subclasses of the SBM, which also reveals a connection to the well-known problem of sparse PCA. Our main relaxation, which we call SDP-1, is tighter than other recently proposed SDP relaxations, and thus previously established theoretical guarantees carry over. However, we show that SDP-1 exactly recovers true communities over a wider class of SBMs than those covered by current results. In particular, the assumption of strong assortativity of the SBM, implicit in consistency conditions for previously proposed SDPs, can be relaxed to weak assortativity for our approach, thus significantly broadening the class of SBMs covered by the consistency results. We also show that strong assortativity is indeed a necessary condition for exact recovery for previously proposed SDP approaches and not an artifact of the proofs. Our analysis of SDPs is based on primal-dual witness constructions, which provides some insight into the nature of the solutions of various SDPs. In particular, we show how to combine features from SDP-1 and already available SDPs to achieve the most flexibility in terms of both assortativity and block-size constraints, as our relaxation has the tendency to produce communities of similar sizes. This tendency makes it the ideal tool for fitting network histograms, a method gaining popularity in the graphon estimation literature, as we illustrate on an example of a social networks of dolphins. We also provide empirical evidence that SDPs outperform spectral methods for fitting SBMs with a large number of blocks.

REFERENCES

- [1] ABBE, E., BANDEIRA, A. S. and HALL, G. (2016). Exact recovery in the stochastic block model. *IEEE Trans. Inform. Theory* **62** 471–487. [MR3447993](#)
- [2] ABBE, E. and SANDON, C. (2015). Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science—FOCS 2015* 670–688. IEEE Computer Soc., Los Alamitos, CA. [MR3473334](#)
- [3] AGARWAL, N., BANDEIRA, A. S., KOILIARIS, K. and KOLLA, A. (2017). Multisection in the stochastic block model using semidefinite programming. In *Applied and Numerical*

MSC2010 subject classifications. 62G20, 90C22, 62H99.

Key words and phrases. Community detection, network, semidefinite programming, stochastic block model.

Harmonic Analysis (H. Boche, G. Caire, R. Calderbank, et al., eds.) 125–162. Birkhäuser, Cham.

- [4] AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P. (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.
- [5] AIROLDI, E. M., COSTA, T. B. and CHAN, S. H. (2013). Stochastic blockmodel approximation of a graphon: Theory and consistent estimation. In *Advances in NIPS* 26, 692–700.
- [6] ALDOUS, D. J. (1981). Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* **11** 581–598. [MR0637937](#)
- [7] AMES, B. P. W. and VAVASIS, S. A. (2014). Convex optimization for the planted k -disjoint-clique problem. *Math. Program.* **143** 299–337. [MR3152071](#)
- [8] AMINI, A. A., CHEN, A., BICKEL, P. J. and LEVINA, E. (2013). Pseudo-likelihood methods for community detection in large sparse networks. *Ann. Statist.* **41** 2097–2122. [MR3127859](#)
- [9] AMINI, A. A. and LEVINA, E. (2018). Supplement to “On semidefinite relaxations for the block model.” DOI:[10.1214/17-AOS1545SUPP](#).
- [10] AMINI, A. A. and WAINWRIGHT, M. J. (2009). High-dimensional analysis of semidefinite relaxations for sparse principal components. *Ann. Statist.* **37** 2877–2921. [MR2541450](#)
- [11] AWASTHI, P., BANDEIRA, A. S., CHARIKAR, M., KRISHNASWAMY, R., VILLAR, S. and WARD, R. (2015). Relax, no need to round: Integrality of clustering formulations. In *ITCS’15—Proceedings of the 6th Innovations in Theoretical Computer Science* 191–200. ACM, New York. [MR3419009](#)
- [12] BANDEIRA, A. S. (2015). Random Laplacian matrices and convex relaxations. Available at [arXiv:1504.03987](#).
- [13] BANDEIRA, A. S. (2016). A note on probably certifiably correct algorithms. *C. R. Math. Acad. Sci. Paris* **354** 329–333. [MR3463033](#)
- [14] BICKEL, P., CHOI, D., CHANG, X. and ZHANG, H. (2013). Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *Ann. Statist.* **41** 1922–1943. [MR3127853](#)
- [15] BICKEL, P. J. and CHEN, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA* **106** 21068–21073.
- [16] BURER, S. and MONTEIRO, R. D. C. (2003). A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Math. Program.* **95** 329–357. [MR1976484](#)
- [17] CAI, T. T. and LI, X. (2015). Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. *Ann. Statist.* **43** 1027–1059. [MR3346696](#)
- [18] CELISSE, A., DAUDIN, J.-J. and PIERRE, L. (2012). Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electron. J. Stat.* **6** 1847–1899. [MR2988467](#)
- [19] CHAUDHURI, K., CHUNG, F. and TSIATAS, A. (2012). Spectral clustering of graphs with general degrees in the extended planted partition model. In *JMLR Workshop and Conference Proceedings* 23, 35.1–35.23.
- [20] CHEN, Y., SANGHAVI, S. and XU, H. (2014). Improved graph clustering. *IEEE Trans. Inform. Theory* **60** 6440–6455. [MR3265033](#)
- [21] CHEN, Y. and XU, J. (2016). Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *J. Mach. Learn. Res.* **17** 882–938.
- [22] D’ASPREMONT, A., EL GHAOU, L., JORDAN, M. I. and LANCKRIET, G. R. G. (2007). A direct formulation for sparse PCA using semidefinite programming. *SIAM Rev.* **49** 434–448. [MR2353806](#)

- [23] DECELLE, A., KRZAKALA, F., MOORE, C. and ZDEBOROVÁ, L. (2011). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* **84** 066106.
- [24] DECELLE, A., KRZAKALA, F., MOORE, C. and ZDEBOROVÁ, L. (2012). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* **84** 066106.
- [25] FEIGE, U. and OFEK, E. (2005). Spectral techniques applied to sparse random graphs. *Random Structures Algorithms* **27** 251–275. [MR2155709](#)
- [26] GAO, C., LU, Y. and ZHOU, H. H. (2015). Rate-optimal graphon estimation. *Ann. Statist.* **43** 2624–2652. [MR3405606](#)
- [27] GUÉDON, O. and VERSHYNIN, R. (2016). Community detection in sparse networks via Grothendieck’s inequality. *Probab. Theory Related Fields* **165** 1025–1049. [MR3520025](#)
- [28] HAJEK, B., WU, Y. and XU, J. (2016). Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Trans. Inform. Theory* **62** 2788–2797. [MR3493879](#)
- [29] HAJEK, B., WU, Y. and XU, J. (2016). Achieving exact cluster recovery threshold via semidefinite programming: Extensions. *IEEE Trans. Inform. Theory* **62** 5918–5937. [MR3552431](#)
- [30] HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. [MR0718088](#)
- [31] JAVANMARD, A., MONTANARI, A. and RICCI-TERSENGHI, F. (2016). Phase transitions in semidefinite relaxations. *Proc. Natl. Acad. Sci. USA* **113** E2218–E2223. [MR3494080](#)
- [32] JOSEPH, A. and YU, B. (2016). Impact of regularization on spectral clustering. *Ann. Statist.* **44** 1765–1791.
- [33] KLOPP, O., TSYBAKOV, A. B., VERZELEN, N. et al. (2017). Oracle inequalities for network models and sparse graphon estimation. *Ann. Statist.* **45** 316–354. [MR3611494](#)
- [34] LE, C. M., LEVINA, E. and VERSHYNIN, R. (2015). Sparse random graphs: Regularization and concentration of the Laplacian. Preprint. Available at [arXiv:1502.03049](#).
- [35] LEI, J. and RINALDO, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann. Statist.* **43** 215–237. [MR3285605](#)
- [36] LUSSEAU, D. and NEWMAN, M. E. J. (2004). Identifying the role that animals play in their social networks. *Proc. R. Soc. Lond., B Biol. Sci.* **271** S477–S481.
- [37] MASSOULIÉ, L. (2014). Community detection thresholds and the weak Ramanujan property. In *STOC’14—Proceedings of the 2014 ACM Symposium on Theory of Computing* 694–703. ACM, New York. [MR3238997](#)
- [38] MATHIEU, C. and SCHUDY, W. (2010). Correlation clustering with noisy input. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms* 712–728. SIAM, Philadelphia, PA. [MR2768627](#)
- [39] MOITRA, A., PERRY, W. and WEIN, A. S. (2016). How robust are reconstruction thresholds for community detection? In *STOC’16—Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing* 828–841. ACM, New York. [MR3536617](#)
- [40] MONTANARI, A. (2016). A Grothendieck-type inequality for local maxima. Preprint. Available at [arXiv:1603.04064](#).
- [41] MONTANARI, A. and SEN, S. (2016). Semidefinite programs on sparse random graphs and their application to community detection. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing* 814–827. ACM, New York.
- [42] MOSSEL, E., NEEMAN, J. and SLY, A. (2012). Stochastic block models and reconstruction. Available at [arXiv:1202.1499](#).
- [43] MOSSEL, E., NEEMAN, J. and SLY, A. (2015). Consistency thresholds for the planted bisection model. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing. STOC ’15* 69–75. ACM, New York.
- [44] MOSSEL, E., NEEMAN, J. and SLY, A. (2017). A proof of the block model threshold conjecture. *Combinatorica*. DOI:[10.1007/s00493-016-3238-8](#).

- [45] NOWICKI, K. and SNIJDERS, T. A. B. (2001). Estimation and prediction for stochastic block-structures. *J. Amer. Statist. Assoc.* **96** 1077–1087. [MR1947255](#)
- [46] OLHEDE, S. C. and WOLFE, P. J. (2014). Network histograms and universality of blockmodel approximation. *Proc. Natl. Acad. Sci. USA* **111** 14722–14727.
- [47] PENG, J. and WEI, Y. (2007). Approximating K -means-type clustering via semidefinite programming. *SIAM J. Optim.* **18** 186–205. [MR2299680](#)
- [48] PERRY, A. and WEIN, A. S. (2017). A semidefinite program for unbalanced multisection in the stochastic block model. In *Sampling Theory and Applications (SampTA)*, 2017 *International Conference on* 64–67. IEEE, New York.
- [49] QIN, T. and ROHE, K. (2013). Regularized spectral clustering under the degree-corrected stochastic blockmodel. In *Advances in Neural Information Processing Systems* 3120–3128.
- [50] ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. [MR2893856](#)
- [51] SNIJDERS, T. A. B. and NOWICKI, K. (1997). Estimation and prediction for stochastic block-models for graphs with latent block structure. *J. Classification* **14** 75–100. [MR1449742](#)
- [52] TOMOZEI, D.-C. and MASSOULIÉ, L. (2014). Distributed user profiling via spectral methods. *Stoch. Syst.* **4** 1–43. [MR3353213](#)
- [53] VU, V. Q., CHO, J., LEI, J. and ROHE, K. (2013). Fantope projection and selection: A near-optimal convex relaxation of sparse PCA. In *Advances in Neural Information Processing Systems* 2670–2678.
- [54] WAINWRIGHT, M. J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1 -constrained quadratic programming (Lasso). *IEEE Trans. Inform. Theory* **55** 2183–2202. [MR2729873](#)
- [55] WOLFE, P. J. and OLHEDE, S. C. (2013). Nonparametric graphon estimation. Preprint. Available at [arXiv:1309.5936](#).
- [56] XING, E. P. and JORDAN, M. I. (2003). On semidefinite relaxations for normalized k -cut and connections to spectral clustering. Technical report, Univ. California, Berkeley.
- [57] YAN, B. and SARKAR, P. (2016). Convex relaxation for community detection with covariates. Preprint. Available at [arXiv:1607.02675](#).
- [58] ZHANG, Y., LEVINA, E. and ZHU, J. (2015). Estimating network edge probabilities by neighborhood smoothing. Available at [arXiv:1509.08588](#).
- [59] ZHAO, Y., LEVINA, E. and ZHU, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *Ann. Statist.* **40** 2266–2292. [MR3059083](#)

PATHWISE COORDINATE OPTIMIZATION FOR SPARSE LEARNING: ALGORITHM AND THEORY¹

BY TUO ZHAO, HAN LIU AND TONG ZHANG

Georgia Institute of Technology, Princeton University and Tencent AI Lab

The pathwise coordinate optimization is one of the most important computational frameworks for high dimensional convex and nonconvex sparse learning problems. It differs from the classical coordinate optimization algorithms in three salient features: *warm start initialization*, *active set updating* and *strong rule for coordinate preselection*. Such a complex algorithmic structure grants superior empirical performance, but also poses significant challenge to theoretical analysis. To tackle this long lasting problem, we develop a new theory showing that these three features play pivotal roles in guaranteeing the outstanding statistical and computational performance of the pathwise coordinate optimization framework. Particularly, we analyze the existing pathwise coordinate optimization algorithms and provide new theoretical insights into them. The obtained insights further motivate the development of several modifications to improve the pathwise coordinate optimization framework, which guarantees linear convergence to a unique sparse local optimum with optimal statistical properties in parameter estimation and support recovery. This is the first result on the computational and statistical guarantees of the pathwise coordinate optimization framework in high dimensions. Thorough numerical experiments are provided to support our theory.

REFERENCES

- BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. [MR2533469](#)
- DUCHI, J. (2015). Lecture Notes for Statistics and Information Theory. Available at http://stanford.edu/class/stats311/Lectures/full_notes.pdf.
- ELOYAN, A., MUSCHELLI, J., NEBEL, M. B., LIU, H., HAN, F., ZHAO, T., BARBER, A. D., JOEL, S., PEKAR, J. J., MOSTOFISKY, S. H. and CAFFO, B. (2012). Automated diagnoses of attention deficit hyperactive disorder using magnetic resonance imaging. *Front. Syst. Neurosci.* **6** 61.
- FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. [MR1946581](#)
- FAN, J., XUE, L. and ZOU, H. (2014). Strong oracle optimality of folded concave penalized estimation. *Ann. Statist.* **42** 819–849. [MR3210988](#)
- FRIEDMAN, J., HASTIE, T., HÖFLING, H. and TIBSHIRANI, R. (2007). Pathwise coordinate optimization. *Ann. Appl. Stat.* **1** 302–332. [MR2415737](#)
- FU, W. J. (1998). Penalized regressions: The bridge versus the lasso. *J. Comput. Graph. Statist.* **7** 397–416. [MR1646710](#)

MSC2010 subject classifications. Primary 62F30, 90C26; secondary 62J12, 90C52.

Key words and phrases. Nonconvex sparse learning, pathwise coordinate optimization, global linear convergence, optimal statistical rates of convergence, oracle property, active set, strong rule.

- HASTIE, T. (2009). Fast regularization paths via coordinate descent. In *The 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Denver 2008*. Plenary Talk.
- HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models. Monographs on Statistics and Applied Probability* **43**. Chapman & Hall, London. [MR1082147](#)
- LI, X., ZHAO, T., ARORA, R., LIU, H. and HONG, M. (2016). An improved convergence analysis of cyclic block coordinate descent-type methods for strongly convex minimization. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*.
- LIU, H., WANG, L. and ZHAO, T. (2015). Calibrated multivariate regression with application to neural semantic basis discovery. *J. Mach. Learn. Res.* **16** 1579–1606. [MR3417792](#)
- 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](#)
- LU, Z. and XIAO, L. (2015). On the complexity analysis of randomized block-coordinate descent methods. *Math. Program.* **152** 615–642. [MR3369495](#)
- LUO, Z. Q. and TSENG, P. (1992). On the convergence of the coordinate descent method for convex differentiable minimization. *J. Optim. Theory Appl.* **72** 7–35. [MR1141764](#)
- MAZUMDER, R., FRIEDMAN, J. H. and HASTIE, T. (2011). *SparseNet*: Coordinate descent with nonconvex penalties. *J. Amer. Statist. Assoc.* **106** 1125–1138. [MR2894769](#)
- MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. [MR2278363](#)
- NEALE, B. M., KOU, Y., LIU, L., MA'AYAN, A., SAMOCHA, K. E., SABO, A., LIN, C.-F., STEVENS, C., WANG, L.-S., MAKAROV, V. et al. (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485** 242–245.
- NEGAHBAN, S. N., RAVIKUMAR, P., WAINWRIGHT, M. J. and YU, B. (2012). A unified framework for high-dimensional analysis of M -estimators with decomposable regularizers. *Statist. Sci.* **27** 538–557. [MR3025133](#)
- NESTEROV, YU. (2013). Gradient methods for minimizing composite functions. *Math. Program.* **140** 125–161. [MR3071865](#)
- RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2010). Restricted eigenvalue properties for correlated Gaussian designs. *J. Mach. Learn. Res.* **11** 2241–2259. [MR2719855](#)
- 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](#)
- RAZAVIYAYN, M., HONG, M. and LUO, Z.-Q. (2013). A unified convergence analysis of block successive minimization methods for nonsmooth optimization. *SIAM J. Optim.* **23** 1126–1153. [MR3063152](#)
- RICHTÁRIK, P. and TAKÁČ, M. (2014). Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function. *Math. Program.* **144** 1–38. [MR3179953](#)
- SHALEV-SHWARTZ, S. and TEWARI, A. (2011). Stochastic methods for ℓ_1 -regularized loss minimization. *J. Mach. Learn. Res.* **12** 1865–1892. [MR2819020](#)
- TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. [MR1379242](#)
- TIBSHIRANI, R., BIEN, J., FRIEDMAN, J., HASTIE, T., SIMON, N., TAYLOR, J. and TIBSHIRANI, R. J. (2012). Strong rules for discarding predictors in lasso-type problems. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 245–266. [MR2899862](#)
- WANG, L., KIM, Y. and LI, R. (2013). Calibrating nonconvex penalized regression in ultra-high dimension. *Ann. Statist.* **41** 2505–2536. [MR3127873](#)
- WANG, Z., LIU, H. and ZHANG, T. (2014). Optimal computational and statistical rates of convergence for sparse nonconvex learning problems. *Ann. Statist.* **42** 2164–2201. [MR3269977](#)
- ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. [MR2604701](#)

- ZHANG, T. (2013). Multi-stage convex relaxation for feature selection. *Bernoulli* **19** 2277–2293. [MR3160554](#)
- ZHANG, C.-H. and HUANG, J. (2008). The sparsity and bias of the LASSO selection in high-dimensional linear regression. *Ann. Statist.* **36** 1567–1594. [MR2435448](#)
- ZHAO, T. and LIU, H. (2016). Accelerated path-following iterative shrinkage thresholding algorithm with application to semiparametric graph estimation. *J. Comput. Graph. Statist.* **25** 1272–1296. [MR3572040](#)
- ZHAO, T., LIU, H. and ZHANG, T. (2018). Supplement to “Pathwise coordinate optimization for sparse learning: Algorithm and theory.” DOI:[10.1214/17-AOS1547SUPP](#).
- ZHAO, P. and YU, B. (2006). On model selection consistency of Lasso. *J. Mach. Learn. Res.* **7** 2541–2563. [MR2274449](#)
- ZHAO, T., LIU, H., ROEDER, K., LAFFERTY, J. and WASSERMAN, L. (2012). The huge package for high-dimensional undirected graph estimation in R. *J. Mach. Learn. Res.* **13** 1059–1062. [MR2930633](#)
- ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. [MR2279469](#)
- ZOU, H. and LI, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.* **36** 1509–1533. [MR2435443](#)

CONDITIONAL MEAN AND QUANTILE DEPENDENCE TESTING IN HIGH DIMENSION

BY XIANYANG ZHANG^{*,1}, SHUN YAO[†] AND XIAOFENG SHAO^{†,2}

Texas A&M University^{} and University of Illinois at Urbana-Champaign[†]*

Motivated by applications in biological science, we propose a novel test to assess the conditional mean dependence of a response variable on a large number of covariates. Our procedure is built on the martingale difference divergence recently proposed in Shao and Zhang [*J. Amer. Statist. Assoc.* **109** (2014) 1302–1318], and it is able to detect certain type of departure from the null hypothesis of conditional mean independence without making any specific model assumptions. Theoretically, we establish the asymptotic normality of the proposed test statistic under suitable assumption on the eigenvalues of a Hermitian operator, which is constructed based on the characteristic function of the covariates. These conditions can be simplified under banded dependence structure on the covariates or Gaussian design. To account for heterogeneity within the data, we further develop a testing procedure for conditional quantile independence at a given quantile level and provide an asymptotic justification. Empirically, our test of conditional mean independence delivers comparable results to the competitor, which was constructed under the linear model framework, when the underlying model is linear. It significantly outperforms the competitor when the conditional mean admits a nonlinear form.

REFERENCES

- CHEN, S. X. and QIN, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. [MR2604697](#)
- EFRON, B. and TIBSHIRANI, R. (2007). On testing the significance of sets of genes. *Ann. Appl. Stat.* **1** 107–129. [MR2393843](#)
- FAN, J. and LV, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 849–911. [MR2530322](#)
- FANG, K. T., KOTZ, S. and NG, K. W. (1990). *Symmetric Multivariate and Related Distributions. Monographs on Statistics and Applied Probability* **36**. Chapman & Hall, London. [MR1071174](#)
- FENG, L., ZOU, C., WANG, Z. and CHEN, B. (2013). Rank-based score tests for high-dimensional regression coefficients. *Electron. J. Stat.* **7** 2131–2149. [MR3104951](#)
- GOEMAN, J. J., VAN DE GEER, S. A. and VAN HOUWELINGEN, H. C. (2006). Testing against a high dimensional alternative. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 477–493. [MR2278336](#)
- HALL, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.* **14** 1–16. [MR0734096](#)
- HE, X., WANG, L. and HONG, H. G. (2013). Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *Ann. Statist.* **41** 342–369. [MR3059421](#)

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

Key words and phrases. Large- p -small- n , martingale difference divergence, simultaneous test, U -statistics.

- KOENKER, R. and BASSETT, G. JR. (1978). Regression quantiles. *Econometrica* **46** 33–50. [MR0474644](#)
- LAN, W., WANG, H. and TSAI, C.-L. (2014). Testing covariates in high-dimensional regression. *Ann. Inst. Statist. Math.* **66** 279–301. [MR3171406](#)
- LI, Q., HSIAO, C. and ZINN, J. (2003). Consistent specification tests for semiparametric/nonparametric models based on series estimation methods. *J. Econometrics* **112** 295–325. [MR1951146](#)
- LIU, J., ZHONG, W. and LI, R. (2015). A selective overview of feature screening for ultrahigh-dimensional data. *Sci. China Math.* **58** 2033–2054. [MR3400642](#)
- MCKEAGUE, I. W. and QIAN, M. (2015). An adaptive resampling test for detecting the presence of significant predictors. *J. Amer. Statist. Assoc.* **110** 1422–1433. [MR3449037](#)
- NEWTON, M. A., QUINTANA, F. A., DEN BOON, J. A., SENGUPTA, S. and AHLQUIST, P. (2007). Random-set methods identify distinct aspects of the enrichment signal in gene-set analysis. *Ann. Appl. Stat.* **1** 85–106. [MR2393842](#)
- PARK, T., SHAO, X. and YAO, S. (2015). Partial martingale difference correlation. *Electron. J. Stat.* **9** 1492–1517. [MR3367668](#)
- SEJDINOVIC, D., SRIPERUMBUDUR, B., GRETTON, A. and FUKUMIZU, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.* **41** 2263–2291. [MR3127866](#)
- SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. [MR0595165](#)
- SHAO, X. and ZHANG, J. (2014). Martingale difference correlation and its use in high-dimensional variable screening. *J. Amer. Statist. Assoc.* **109** 1302–1318. [MR3265698](#)
- SUBRAMANIAN, A., MOOTHA, V. K., MUKHERJEE, S., EBERT, B. L., GILLETTE, M. A., PAULOVICH, A., POMEROY, S. L., GOLUB, T. R., LANDER, E. S. and MESIROV, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102** 15545–15550.
- SZÉKELY, G. J. and RIZZO, M. L. (2013). The distance correlation t -test of independence in high dimension. *J. Multivariate Anal.* **117** 193–213. [MR3053543](#)
- SZÉKELY, G. J. and RIZZO, M. L. (2014). Partial distance correlation with methods for dissimilarities. *Ann. Statist.* **42** 2382–2412. [MR3269983](#)
- SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. [MR2382665](#)
- WANG, S. and CUI, H. (2013). Generalized F test for high dimensional linear regression coefficients. *J. Multivariate Anal.* **117** 134–149. [MR3053539](#)
- WANG, H. and HE, X. (2007). Detecting differential expressions in GeneChip microarray studies: A quantile approach. *J. Amer. Statist. Assoc.* **102** 104–112. [MR2293303](#)
- WANG, H. and HE, X. (2008). An enhanced quantile approach for assessing differential gene expressions. *Biometrics* **64** 449–457, 666. [MR2432415](#)
- WANG, L., WU, Y. and LI, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *J. Amer. Statist. Assoc.* **107** 214–222. [MR2949353](#)
- WESTFALL, P. H. and YOUNG, S. S. (1993). *Resampling-Based Multiple Testing: Examples and Methods for P-Value Adjustment*. Wiley, New York.
- WU, C. F. J. and HAMADA, M. S. (2009). *Experiments: Planning, Analysis, and Optimization*, 2nd ed. Wiley, Hoboken, NJ. [MR2583259](#)
- YATA, K. and AOSHIMA, M. (2013). Correlation tests for high-dimensional data using extended cross-data-matrix methodology. *J. Multivariate Anal.* **117** 313–331. [MR3053550](#)
- ZHANG, X., YAO, S. and SHAO, X. (2018). Supplement to “Conditional mean and quantile dependence testing in high dimension.” DOI:[10.1214/17-AOS1548SUPP](#).
- ZHONG, P.-S. and CHEN, S. X. (2011). Tests for high-dimensional regression coefficients with factorial designs. *J. Amer. Statist. Assoc.* **106** 260–274. [MR2816719](#)

HIGH-DIMENSIONAL ASYMPTOTICS OF PREDICTION: RIDGE REGRESSION AND CLASSIFICATION

BY EDGAR DOBRIBAN¹ AND STEFAN WAGER

University of Pennsylvania and Stanford University

We provide a unified analysis of the predictive risk of ridge regression and regularized discriminant analysis in a dense random effects model. We work in a high-dimensional asymptotic regime where $p, n \rightarrow \infty$ and $p/n \rightarrow \gamma > 0$, and allow for arbitrary covariance among the features. For both methods, we provide an explicit and efficiently computable expression for the limiting predictive risk, which depends only on the spectrum of the feature-covariance matrix, the signal strength and the aspect ratio γ . Especially in the case of regularized discriminant analysis, we find that predictive accuracy has a nuanced dependence on the eigenvalue distribution of the covariance matrix, suggesting that analyses based on the operator norm of the covariance matrix may not be sharp. Our results also uncover an exact *inverse* relation between the limiting predictive risk and the limiting estimation risk in high-dimensional linear models. The analysis builds on recent advances in random matrix theory.

REFERENCES

- ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley, New York. [MR1990662](#)
- BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer, Berlin. [MR2567175](#)
- BARTLETT, P. L. and MENDELSON, S. (2003). Rademacher and Gaussian complexities: Risk bounds and structural results. *J. Mach. Learn. Res.* **3** 463–482.
- BAYATI, M. and MONTANARI, A. (2012). The LASSO risk for Gaussian matrices. *IEEE Trans. Inform. Theory* **58** 1997–2017.
- BEAN, D., BICKEL, P. J., EL KAROUI, N. and YU, B. (2013). Optimal M-estimation in high-dimensional regression. *Proc. Natl. Acad. Sci. USA* **110** 14563–14568.
- BERNAU, C., RIESTER, M., BOULESTEIX, A.-L., PARMIGIANI, G., HUTTENHOWER, C., WALDRON, L. and TRIPPA, L. (2014). Cross-study validation for the assessment of prediction algorithms. *Bioinformatics* **30** i105–i112.
- BICKEL, P. J. and LEVINA, E. (2004). Some theory of Fisher’s linear discriminant function, ‘naive Bayes’, and some alternatives when there are many more variables than observations. *Bernoulli* **10** 989–1010. [MR2108040](#)
- CANDÈS, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n . *Ann. Statist.* **35** 2313–2351. [MR2382644](#)
- CHEN, L. S., PAUL, D., PRENTICE, R. L. and WANG, P. (2011). A regularized Hotelling’s T^2 test for pathway analysis in proteomic studies. *J. Amer. Statist. Assoc.* **106**. 1345–1360. [MR2896840](#)

MSC2010 subject classifications. Primary 62H99; secondary 62J05, 62H30.

Key words and phrases. High-dimensional asymptotics, ridge regression, regularized discriminant analysis, prediction error, random matrix theory.

- COUILLET, R. and DEBBAH, M. (2011). *Random Matrix Methods for Wireless Communications*. Cambridge Univ. Press, Cambridge.
- DEEV, A. (1970). Representation of statistics of discriminant analysis and asymptotic expansion when space dimensions are comparable with sample size. *Sov. Math., Dokl.* **11** 1547–1550.
- DICKER, L. (2013). Optimal equivariant prediction for high-dimensional linear models with arbitrary predictor covariance. *Electron. J. Stat.* **7** 1806–1834.
- DICKER, L. (2016). Ridge regression and asymptotic minimax estimation over spheres of growing dimension. *Bernoulli* **22** 1–37.
- DOBRIAN, E. (2017). Sharp detection in PCA under correlations: All eigenvalues matter. *Ann. Statist.* **45** 1810–1833. [MR3670197](#)
- DOBRIAN, E. and WAGER, S. (2018). Supplement to “High-dimensional asymptotics of prediction: Ridge regression and classification.” DOI:[10.1214/17-AOS1549SUPP](#).
- DONOHU, D. L. and MONTANARI, A. (2015). Variance breakdown of Huber (M)-estimators. $n/p \rightarrow m \in (1, \infty)$. Preprint. Available at [arXiv:1503.02106](#).
- DONOHU, D. L., JOHNSTONE, I. M., HOCH, J. C. and STERN, A. S. (1992). Maximum entropy and the nearly black object. *J. Roy. Statist. Soc. Ser. B* **54** 41–81. [MR1157714](#)
- EFRON, B. (1975). The efficiency of logistic regression compared to normal discriminant analysis. *J. Amer. Statist. Assoc.* **70** 892–898.
- EL KAROUI, N. (2013). Asymptotic behavior of unregularized and ridge-regularized high-dimensional robust regression estimators: Rigorous results. Preprint. Available at [arXiv:1311.2445](#).
- EL KAROUI, N. (2015). On the impact of predictor geometry on the performance on high-dimensional ridge-regularized generalized robust regression estimators. Technical report, Univ. California, Berkeley.
- EL KAROUI, N. and KÖSTERS, H. (2011). Geometric sensitivity of random matrix results: Consequences for shrinkage estimators of covariance and related statistical methods. Preprint. Available at [arXiv:1105.1404](#).
- FAN, J., FAN, Y. and WU, Y. (2011). High dimensional classification. In *High-Dimensional Data Analysis* (T. Cai and X. Shen, eds.) 3–37. World Sci. Publ., Singapore. [MR2848198](#)
- FRIEDMAN, J. H. (1989). Regularized discriminant analysis. *J. Amer. Statist. Assoc.* **84** 165–175.
- FUJIKOSHI, Y. and SEO, T. (1998). Asymptotic approximations for EPMCs of the linear and the quadratic discriminant functions when the sample sizes and the dimension are large. *Random Oper. Stochastic Equations* **6** 269–280. [MR1631003](#)
- FUJIKOSHI, Y., ULYANOV, V. V. and SHIMIZU, R. (2011). *Multivariate Statistics: High-Dimensional and Large-Sample Approximations*. Wiley, New York.
- GRENNANDER, U. and SZEGŐ, G. (1984). *Toeplitz Forms and Their Applications*, 2nd ed. Chelsea Publishing Co., New York. [MR0890515](#)
- HACHEM, W., LOUBATON, P. and NAJIM, J. (2007). Deterministic equivalents for certain functionals of large random matrices. *Ann. Appl. Probab.* **17** 875–930. [MR2326235](#)
- HACHEM, W., KHORUNZHIY, O., LOUBATON, P., NAJIM, J. and PASTUR, L. (2008). A new approach for mutual information analysis of large dimensional multi-antenna channels. *IEEE Trans. Inform. Theory* **54** 3987–4004.
- HASTIE, T., TIBSHIRANI, R. and WAINWRIGHT, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press, Boca Raton, FL.
- HSU, D., KAKADE, S. M. and ZHANG, T. (2014). Random design analysis of ridge regression. *Found. Comput. Math.* **14** 569–600.
- KLEINBERG, J., LUDWIG, J., MULLAINATHAN, S., OBERMEYER, Z. et al. (2015). Prediction policy problems. *Am. Econ. Rev.* **105** 491–495.
- KUBOKAWA, T., HYODO, M. and SRIVASTAVA, M. S. (2013). Asymptotic expansion and estimation of EPMC for linear classification rules in high dimension. *J. Multivariate Anal.* **115** 496–515.

- LEDOIT, O. and PÉCHÉ, S. (2011). Eigenvectors of some large sample covariance matrix ensembles. *Probab. Theory Related Fields* **151** 233–264.
- LIANG, P. and SREBRO, N. (2010). On the interaction between norm and dimensionality: Multiple regimes in learning. In *ICML*.
- MARCHENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mat. Sb.* **114** 507–536.
- NG, A. and JORDAN, M. (2001). On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes. In *NIPS*.
- PICKRELL, J. K. and PRITCHARD, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8** e1002967.
- RAUDYS, Š. (1967). On determining training sample size of linear classifier. *Comput. Syst.* **28** 79–87 (in Russian).
- RAUDYS, Š. (1972). On the amount of a priori information in designing the classification algorithm. *Technical Cybernetics* **4** 168–174 (in Russian).
- RAUDYS, Š. (2001). *Statistical and Neural Classifiers: An Integrated Approach to Design*. Springer Science & Business Media, Berlin.
- RAUDYS, Š. and SAUDARGIENE, A. (1998). Structures of the covariance matrices in the classifier design. In *Joint IAPR Intl Workshops on SPR and SSPR* 583–592. Springer, Berlin.
- RAUDYS, Š. and SKURICHINA, M. (1995). Small sample properties of ridge estimate of the covariance matrix in statistical and neural net classification. *New Trends Probab. Stat.* **3** 237–245.
- RAUDYS, Š. and YOUNG, D. M. (2004). Results in statistical discriminant analysis: A review of the former Soviet Union literature. *J. Multivariate Anal.* **89** 1–35.
- RIFAI, S., DAUPHIN, Y., VINCENT, P., BENGIO, Y. and MULLER, X. (2011). The manifold tangent classifier. *Adv. Neural Inf. Process. Syst.* **24** 2294–2302.
- RUBIO, F., MESTRE, X. and PALOMAR, D. P. (2012). Performance analysis and optimal selection of large minimum variance portfolios under estimation risk. *IEEE J. Sel. Top. Signal Process.* **6** 337–350.
- RUSSAKOVSKY, O., DENG, J., SU, H., KRAUSE, J., SATHEESH, S., MA, S., HUANG, Z., KARPATY, A., KHOSLA, A., BERNSTEIN, M. et al. (2015). ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115** 211–252.
- SARANADASA, H. (1993). Asymptotic expansion of the misclassification probabilities of D- and A-criteria for discrimination from two high dimensional populations using the theory of large dimensional random matrices. *J. Multivariate Anal.* **46** 154–174.
- SERDOBOLSKII, V. I. (1983). On minimum error probability in discriminant analysis. *Dokl. Akad. Nauk SSSR* **27** 720–725.
- SERDOBOLSKII, V. I. (2007). *Multiparametric Statistics*. Elsevier, Amsterdam.
- SILVERSTEIN, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *J. Multivariate Anal.* **55** 331–339.
- SIMARD, P. Y., LE CUN, Y. A., DENKER, J. S. and VICTORRI, B. (2000). Transformation invariance in pattern recognition: Tangent distance and propagation. *Int. J. Imaging Syst. Technol.* **11** 181–197.
- SUTTON, C. and MCCALLUM, A. (2006). An introduction to conditional random fields for relational learning. In *Introduction to Statistical Relational Learning* 93–128.
- TOUTANOVA, K., KLEIN, D., MANNING, C. D. and SINGER, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. In *NAACL*.
- TULINO, A. M. and VERDÚ, S. (2004). Random matrix theory and wireless communications. *Commun. Inf. Theory* **1** 1–182.
- VAPNIK, V. N. and CHERVONENKIS, A. Y. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.* **16** 264–280.

- WANG, S. and MANNING, C. D. (2012). Baselines and bigrams: Simple, good sentiment and topic classification. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers, Vol. 2* 90–94. Association for Computational Linguistics, Stroudsburg PA .
- WRAY, N. R., GODDARD, M. E. and VISSCHER, P. M. (2007). Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res.* **17** 1520–1528.
- YAO, J., BAI, Z. and ZHENG, S. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis*. Cambridge Univ. Press, Cambridge.
- ZHANG, M., RUBIO, F., PALOMAR, D. P. and MESTRE, X. (2013). Finite-sample linear filter optimization in wireless communications and financial systems. *IEEE Trans. Signal Process.* **61** 5014–5025.
- ZOLLANVARI, A., BRAGA-NETO, U. M. and DOUGHERTY, E. R. (2011). Analytic study of performance of error estimators for linear discriminant analysis. *IEEE Trans. Signal Process.* **59** 4238–4255.
- ZOLLANVARI, A. and DOUGHERTY, E. R. (2015). Generalized consistent error estimator of linear discriminant analysis. *IEEE Trans. Signal Process.* **63** 2804–2814. [MR3345588](#)

TESTING INDEPENDENCE IN HIGH DIMENSIONS WITH SUMS OF RANK CORRELATIONS

BY DENNIS LEUNG AND MATHIAS DRTON¹

Chinese University of Hong Kong and University of Washington

We treat the problem of testing independence between m continuous variables when m can be larger than the available sample size n . We consider three types of test statistics that are constructed as sums or sums of squares of pairwise rank correlations. In the asymptotic regime where both m and n tend to infinity, a martingale central limit theorem is applied to show that the null distributions of these statistics converge to Gaussian limits, which are valid with no specific distributional or moment assumptions on the data. Using the framework of U-statistics, our result covers a variety of rank correlations including Kendall's tau and a dominating term of Spearman's rank correlation coefficient (ρ), but also degenerate U-statistics such as Hoeffding's D , or the τ^* of Bergsma and Dassios [Bernoulli **20** (2014) 1006–1028]. As in the classical theory for U-statistics, the test statistics need to be scaled differently when the rank correlations used to construct them are degenerate U-statistics. The power of the considered tests is explored in rate-optimality theory under a Gaussian equicorrelation alternative as well as in numerical experiments for specific cases of more general alternatives.

REFERENCES

- ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley, Hoboken, NJ. [MR1990662](#)
- BERGSMA, W. and DASSIOS, A. (2014). A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli* **20** 1006–1028. [MR3178526](#)
- CAI, T. T. and JIANG, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39** 1496–1525. [MR2850210](#)
- CAI, T. T. and MA, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** 2359–2388. [MR3160557](#)
- CHEN, X. (2016). Gaussian approximation for the sup-norm of high-dimensional matrix-variate U-statistics and its applications. ArXiv preprint. Available at [arXiv:1602.00199](#).
- CHEN, Y. and SHAO, Q.-M. (2012). Berry–Esseen inequality for unbounded exchangeable pairs. In *Probability Approximations and Beyond. Lecture Notes in Statistics* **205** 13–30. Springer, New York. [MR3289373](#)
- CHEN, S. X., ZHANG, L.-X. and ZHONG, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** 810–819. [MR2724863](#)
- CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. [MR3161448](#)

MSC2010 subject classifications. 60K35.

Key words and phrases. High-dimensional statistics, independence, U-statistics, minimax optimality, rank correlations.

- CHRISTENSEN, D. (2005). Fast algorithms for the calculation of Kendall's τ . *Comput. Statist.* **20** 51–62. [MR2162534](#)
- HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York. [MR0624435](#)
- HAN, F. and LIU, H. (2014). Distribution-free tests of independence with applications to testing more structures. Available at [arXiv:1410.4179](#).
- HELLER, Y. and HELLER, R. (2016). Computing the Bergsma Dassios sign-covariance. Available at [arXiv:1605.08732](#).
- HOEFFDING, W. (1948a). A non-parametric test of independence. *Ann. Math. Stat.* **19** 546–557. [MR0029139](#)
- HOEFFDING, W. (1948b). A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* **19** 293–325. [MR0026294](#)
- JIANG, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **14** 865–880. [MR2052906](#)
- JIANG, T. and QI, Y. (2015). Likelihood ratio tests for high-dimensional normal distributions. *Scand. J. Stat.* **42** 988–1009. [MR3426306](#)
- JOHN, S. (1972). The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika* **59** 169–173. [MR0312619](#)
- KRUSKAL, W. H. (1958). Ordinal measures of association. *J. Amer. Statist. Assoc.* **53** 814–861. [MR0100941](#)
- LEDOIT, O. and WOLF, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30** 1081–1102. [MR1926169](#)
- LEUNG, D. and DRTON, M. (2018). Supplement to “Testing independence in high dimensions with sums of rank correlations.” DOI:[10.1214/17-AOS1550SUPP](#).
- LI, D., LIU, W.-D. and ROSALSKY, A. (2010). Necessary and sufficient conditions for the asymptotic distribution of the largest entry of a sample correlation matrix. *Probab. Theory Related Fields* **148** 5–35. [MR2653220](#)
- LI, D., QI, Y. and ROSALSKY, A. (2012). On Jiang's asymptotic distribution of the largest entry of a sample correlation matrix. *J. Multivariate Anal.* **111** 256–270. [MR2944420](#)
- LINDSKOG, F., MCNEIL, A. and SCHMOCK, U. (2003). Credit risk. In *Kendall's Tau for Elliptical Distributions* 149–156. Physica-Verlag, Heidelberg.
- LIU, W.-D., LIN, Z. and SHAO, Q.-M. (2008). The asymptotic distribution and Berry–Esseen bound of a new test for independence in high dimension with an application to stochastic optimization. *Ann. Appl. Probab.* **18** 2337–2366. [MR2474539](#)
- MAO, G. (2014). A new test of independence for high-dimensional data. *Statist. Probab. Lett.* **93** 14–18. [MR3244549](#)
- NAGAO, H. (1973). On some test criteria for covariance matrix. *Ann. Statist.* **1** 700–709. [MR0339405](#)
- NANDY, P., WEIHS, L. and DRTON, M. (2016). Large-sample theory for the Bergsma–Dassios sign covariance. *Electron. J. Stat.* **10** 2287–2311. [MR3541972](#)
- OJA, H. (2010). *Multivariate Nonparametric Methods with R. An Approach Based on Spatial Signs and Ranks. Lecture Notes in Statistics* **199**. Springer, New York. [MR2598854](#)
- SCHOTT, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika* **92** 951–956. [MR2234197](#)
- SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. [MR0595165](#)
- VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. [MR1652247](#)
- WANG, G., ZOU, C. and WANG, Z. (2013). A necessary test for complete independence in high dimensions using rank-correlations. *J. Multivariate Anal.* **121** 224–232. [MR3090478](#)
- WEIHS, L., DRTON, M. and LEUNG, D. (2016). Efficient computation of the Bergsma–Dassios sign covariance. *Comput. Statist.* **31** 315–328. [MR3481807](#)

ZHOU, W. (2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices.
Trans. Amer. Math. Soc. **359** 5345–5363. [MR2327033](#)

HIGH DIMENSIONAL CENSORED QUANTILE REGRESSION

BY QI ZHENG¹, LIMIN PENG² AND XUMING HE³

University of Louisville, Emory University and University of Michigan

Censored quantile regression (CQR) has emerged as a useful regression tool for survival analysis. Some commonly used CQR methods can be characterized by stochastic integral-based estimating equations in a sequential manner across quantile levels. In this paper, we analyze CQR in a high dimensional setting where the regression functions over a continuum of quantile levels are of interest. We propose a two-step penalization procedure, which accommodates stochastic integral based estimating equations and address the challenges due to the recursive nature of the procedure. We establish the uniform convergence rates for the proposed estimators, and investigate the properties on weak convergence and variable selection. We conduct numerical studies to confirm our theoretical findings and illustrate the practical utility of our proposals.

REFERENCES

- BELLONI, A. and CHERNOZHUKOV, V. (2011). ℓ_1 -penalized quantile regression in high-dimensional sparse models. *Ann. Statist.* **39** 82–130. [MR2797841](#)
- BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. [MR2533469](#)
- BRADIC, J., FAN, J. and JIANG, J. (2011). Regularization for Cox’s proportional hazards model with NP-dimensionality. *Ann. Statist.* **39** 3092–3120. [MR3012402](#)
- EFRON, B. (1967). The two sample problem with censored data. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability IV* 831–853. Prentice Hall, Engewood Cliffs, NJ.
- FAN, J., FAN, Y. and BARUT, E. (2014). Adaptive robust variable selection. *Ann. Statist.* **42** 324–351. [MR3189488](#)
- FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. [MR1946581](#)
- FLEMING, T. R. and HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York. [MR1100924](#)
- HE, X. and SHAO, Q.-M. (2000). On parameters of increasing dimensions. *J. Multivariate Anal.* **73** 120–135. [MR1766124](#)
- HUANG, J. and MA, S. (2010). Variable selection in the accelerated failure time model via the bridge method. *Lifetime Data Anal.* **16** 176–195. [MR2608284](#)
- HUANG, J., MA, S. and XIE, H. (2006). Regularized estimation in the accelerated failure time model with high-dimensional covariates. *Biometrics* **62** 813–820. [MR2247210](#)
- HUANG, J., MA, S. and ZHANG, C.-H. (2008). Adaptive Lasso for sparse high-dimensional regression models. *Statist. Sinica* **18** 1603–1618. [MR2469326](#)

MSC2010 subject classifications. Primary 62J07; secondary 62H12.

Key words and phrases. High dimensional survival data, varying covariate effects, censored quantile regression.

- JOHNSON, B. A. (2009). On lasso for censored data. *Electron. J. Stat.* **3** 485–506. [MR2507457](#)
- KOENKER, R. (2005). *Quantile Regression*. Cambridge Univ. Press, Cambridge.
- KOENKER, R. (2016). quantreg: Quantile regression. R package version 5.29.
- KOENKER, R. and BASSETT, G. JR. (1978). Regression quantiles. *Econometrica* **46** 33–50. [MR0474644](#)
- NEOCLEOUS, T., BRANDEN, K. V. and PORTNOY, S. (2006). Correction to: “Censored regression quantiles” [*J. Amer. Statist. Assoc.* **98** (2003) 1001–1012; [MR2041488](#)] by Portnoy. *J. Amer. Statist. Assoc.* **101** 860–861. [MR2281250](#)
- PENG, L. (2012). Self-consistent estimation of censored quantile regression. *J. Multivariate Anal.* **105** 368–379. [MR2877523](#)
- PENG, L. and HUANG, Y. (2008). Survival analysis with quantile regression models. *J. Amer. Statist. Assoc.* **103** 637–649. [MR2435468](#)
- PENG, L., XU, J. and KUTNER, N. (2014). Shrinkage estimation of varying covariate effects based on quantile regression. *Stat. Comput.* **24** 853–869. [MR3229701](#)
- PORTNOY, S. (2003). Censored regression quantiles. *J. Amer. Statist. Assoc.* **98** 1001–1012. [MR2041488](#)
- PORTNOY, S. and LIN, G. (2010). Asymptotics for censored regression quantiles. *J. Nonparametr. Stat.* **22** 115–130. [MR2598957](#)
- POWELL, J. L. (1986). Censored regression quantiles. *J. Econometrics* **32** 143–155. [MR0853049](#)
- SHEDDEN, K., TAYLOR, J. M. et al. (2008). Gene expression-based survival prediction in lung adenocarcinoma: A multi-site, blinded validation study. *Nature Medicine* **14** 822–827.
- SHOWS, J. H., LU, W. and ZHANG, H. H. (2010). Sparse estimation and inference for censored median regression. *J. Statist. Plann. Inference* **140** 1903–1917. [MR2606727](#)
- TALAGRAND, M. (2005). *The Generic Chaining: Upper and Lower Bounds of Stochastic Processes*. Springer, Berlin. [MR2133757](#)
- TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **58** 267–288. [MR1379242](#)
- VAN DE GEER, S. A. (2008). High-dimensional generalized linear models and the lasso. *Ann. Statist.* **36** 614–645. [MR2396809](#)
- VOLGUSHEV, S., WAGENER, J. and DETTE, H. (2014). Censored quantile regression processes under dependence and penalization. *Electron. J. Stat.* **8** 2405–2447. [MR3278338](#)
- WANG, H. J. and WANG, L. (2009). Locally weighted censored quantile regression. *J. Amer. Statist. Assoc.* **104** 1117–1128. [MR2562007](#)
- WANG, H. J., ZHOU, J. and LI, Y. (2013). Variable selection for censored quantile regression. *Statist. Sinica* **23** 145–167. [MR3076162](#)
- WANG, L., WU, Y. and LI, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *J. Amer. Statist. Assoc.* **107** 214–222. [MR2949353](#)
- YING, Z., JUNG, S. H. and WEI, L. J. (1995). Survival analysis with median regression models. *J. Amer. Statist. Assoc.* **90** 178–184. [MR1325125](#)
- ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. [MR2604701](#)
- ZHANG, C.-H. and HUANG, J. (2008). The sparsity and bias of the LASSO selection in high-dimensional linear regression. *Ann. Statist.* **36** 1567–1594. [MR2435448](#)
- ZHENG, Q., GALLAGHER, C. and KULASEKERA, K. B. (2013). Adaptive penalized quantile regression for high dimensional data. *J. Statist. Plann. Inference* **143** 1029–1038. [MR3029230](#)
- ZHENG, Q., PENG, L. and HE, X. (2015). Globally adaptive quantile regression with ultra-high dimensional data. *Ann. Statist.* **43** 2225–2258. [MR3396984](#)
- ZHENG, Q., PENG, L. and HE, X. (2018). Supplement to “High dimensional censored quantile regression.” DOI:[10.1214/17-AOS1551SUPP](#).
- ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. [MR2279469](#)

LOCAL M-ESTIMATION WITH DISCONTINUOUS CRITERION FOR DEPENDENT AND LIMITED OBSERVATIONS

BY MYUNG HWAN SEO¹ AND TAISUKE OTSU²

Seoul National University and London School of Economics

We examine the asymptotic properties of local M-estimators under three sets of high-level conditions. These conditions are sufficiently general to cover the minimum volume predictive region, the conditional maximum score estimator for a panel data discrete choice model and many other widely used estimators in statistics and econometrics. Specifically, they allow for discontinuous criterion functions of weakly dependent observations which may be localized by kernel smoothing and contain nuisance parameters with growing dimension. Furthermore, the localization can occur around parameter values rather than around a fixed point and the observations may take limited values which lead to set estimators. Our theory produces three different nonparametric cube root rates for local M-estimators and enables valid inference building on novel maximal inequalities for weakly dependent observations. The standard cube root asymptotics is included as a special case. The results are illustrated by various examples such as the Hough transform estimator with diminishing bandwidth, the maximum score-type set estimator and many others.

REFERENCES

- ABREVAYA, J. and HUANG, J. (2005). On the bootstrap of the maximum score estimator. *Econometrica* **73** 1175–1204. [MR2149245](#)
- ADAMCZAK, R. (2008). A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.* **13** 1000–1034. [MR2424985](#)
- ANDREWS, D. F., BICKEL, P. J., HAMPEL, F. R., HUBER, P. J., ROGERS, W. H. and TUKEY, J. W. (1972). *Robust Estimates of Location: Survey and Advances*. Princeton Univ. Press, Princeton, NJ. [MR0331595](#)
- ANEVSKI, D. and HÖSSJER, O. (2006). A general asymptotic scheme for inference under order restrictions. *Ann. Statist.* **34** 1874–1930. [MR2283721](#)
- BANERJEE, M. and MCKEAGUE, I. W. (2007). Confidence sets for split points in decision trees. *Ann. Statist.* **35** 543–574. [MR2336859](#)
- BARAUD, Y. (2010). A Bernstein-type inequality for suprema of random processes with applications to model selection in non-Gaussian regression. *Bernoulli* **16** 1064–1085. [MR2759169](#)
- 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](#)
- BÜHLMANN, P. and YU, B. (2002). Analyzing bagging. *Ann. Statist.* **30** 927–961. [MR1926165](#)

MSC2010 subject classifications. Primary 62F12; secondary 60F17, 60G15, 62G20.

Key words and phrases. Cube root asymptotics, maximal inequality, mixing process, partial identification, parameter-dependent localization.

- CARRASCO, M. and CHEN, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* **18** 17–39. [MR1885348](#)
- CHAN, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Ann. Statist.* **21** 520–533. [MR1212191](#)
- CHEN, X. (2007). Chapter 76. Large sample sieve estimation of semi-nonparametric models. In *Handbook of Econometrics* **6B** 5549–5632. Elsevier, Amsterdam.
- CHEN, X., HANSEN, L. P. and CARRASCO, M. (2010). Nonlinearity and temporal dependence. *J. Econometrics* **155** 155–169. [MR2607192](#)
- CHERNOFF, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.* **16** 31–41. [MR0172382](#)
- CHERNOZHUKOV, V., HONG, H. and TAMER, E. (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica* **75** 1243–1284. [MR2347346](#)
- DE JONG, R. M. and WOUTERSEN, T. (2011). Dynamic time series binary choice. *Econometric Theory* **27** 673–702. [MR2822362](#)
- DOUKHAN, P., MASSART, P. and RIO, E. (1995). Invariance principles for absolutely regular empirical processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **31** 393–427. [MR1324814](#)
- GAUTIER, E. and KITAMURA, Y. (2013). Nonparametric estimation in random coefficients binary choice models. *Econometrica* **81** 581–607. [MR3043343](#)
- GOLDENSHLUGER, A. and ZEEVI, A. (2004). The Hough transform estimator. *Ann. Statist.* **32** 1908–1932. [MR2102497](#)
- HIRANO, K. and PORTER, J. R. (2003). Asymptotic efficiency in parametric structural models with parameter-dependent support. *Econometrica* **71** 1307–1338. [MR2000249](#)
- HONORÉ, B. E. and KYRIAZIDOU, E. (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica* **68** 839–874. [MR1771585](#)
- KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219. [MR1041391](#)
- KOO, B. and SEO, M. H. (2015). Structural-break models under mis-specification: Implications for forecasting. *J. Econometrics* **188** 166–181. [MR3371666](#)
- KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York. [MR2724368](#)
- LEE, M.-J. (1989). Mode regression. *J. Econometrics* **42** 337–349. [MR1040748](#)
- MANSKI, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *J. Econometrics* **3** 205–228. [MR0436905](#)
- MANSKI, C. F. and TAMER, E. (2002). Inference on regressions with interval data on a regressor or outcome. *Econometrica* **70** 519–546. [MR1913822](#)
- MERLEVÈDE, F., PELIGRAD, M. and RIO, E. (2009). Bernstein inequality and moderate deviations under strong mixing conditions. In *High Dimensional Probability V: The Luminy Volume*. *Inst. Math. Stat. (IMS) Collect.* **5** 273–292. IMS, Beachwood, OH. [MR2797953](#)
- MERLEVÈDE, F., PELIGRAD, M. and RIO, E. (2011). A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probab. Theory Related Fields* **151** 435–474. [MR2851689](#)
- NICKL, R. and SÖHL, J. (2016). Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions. Working paper, [arXiv:1510.05526v2](#).
- PAULIN, D. (2015). Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electron. J. Probab.* **20** 1–32. [MR3383563](#)
- POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling*. Springer, New York. [MR1707286](#)
- POLLARD, D. (1989). Asymptotics via empirical processes. *Statist. Sci.* **4** 341–366. [MR1041762](#)
- POLONIK, W. and YAO, Q. (2000). Conditional minimum volume predictive regions for stochastic processes. *J. Amer. Statist. Assoc.* **95** 509–519. [MR1803169](#)
- PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36. [MR0267677](#)

- RIO, E. (1997). About the Lindeberg method for strongly mixing sequences. *ESAIM Probab. Stat.* **1** 35–61. [MR1382517](#)
- ROMANO, J. P. and SHAIKH, A. M. (2008). Inference for identifiable parameters in partially identified econometric models. *J. Statist. Plann. Inference* **138** 2786–2807. [MR2422399](#)
- ROUSSEEUW, P. J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79** 871–880. [MR0770281](#)
- SEN, B., BANERJEE, M. and WOODROOFE, M. (2010). Inconsistency of bootstrap: The Grenander estimator. *Ann. Statist.* **38** 1953–1977. [MR2676880](#)
- SEO, M. H and OTSU, T. (2018). Supplement to “Local M-estimation with discontinuous criterion for dependent and limited observations.” DOI:[10.1214/17-AOS1552SUPP](#).
- TALAGRAND, M. (2005). *The Generic Chaining*. Springer, Berlin. [MR2133757](#)
- VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. [MR1385671](#)
- VAN DER VAART, A. W. and WELLNER, J. A. (2007). Empirical processes indexed by estimated functions. In *Asymptotics: Particles, Processes and Inverse Problems. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **55** 234–252. IMS, Beachwood, OH. [MR2459942](#)
- YAO, W., LINDSAY, B. G. and LI, R. (2012). Local modal regression. *J. Nonparametr. Stat.* **24** 647–663. [MR2968894](#)
- ZINDE-WALSH, V. (2002). Asymptotic theory for some high breakdown point estimators. *Econometric Theory* **18** 1172–1196. [MR1926018](#)

MIXTURE INNER PRODUCT SPACES AND THEIR APPLICATION TO FUNCTIONAL DATA ANALYSIS

BY ZHENHUA LIN^{*,1,2}, HANS-GEORG MÜLLER^{†,3} AND FANG YAO^{*,‡,1}

University of Toronto^{*}, *University of California, Davis*[†] and *Peking University*[‡]

We introduce the concept of mixture inner product spaces associated with a given separable Hilbert space, which feature an infinite-dimensional mixture of finite-dimensional vector spaces and are dense in the underlying Hilbert space. Any Hilbert valued random element can be arbitrarily closely approximated by mixture inner product space valued random elements. While this concept can be applied to data in any infinite-dimensional Hilbert space, the case of functional data that are random elements in the L^2 space of square integrable functions is of special interest. For functional data, mixture inner product spaces provide a new perspective, where each realization of the underlying stochastic process falls into one of the component spaces and is represented by a finite number of basis functions, the number of which corresponds to the dimension of the component space. In the mixture representation of functional data, the number of included mixture components used to represent a given random element in L^2 is specifically adapted to each random trajectory and may be arbitrarily large. Key benefits of this novel approach are, first, that it provides a new perspective on the construction of a probability density in function space under mild regularity conditions, and second, that individual trajectories possess a trajectory-specific dimension that corresponds to a latent random variable, making it possible to use a larger number of components for less smooth and a smaller number for smoother trajectories. This enables flexible and parsimonious modeling of heterogeneous trajectory shapes. We establish estimation consistency of the functional mixture density and introduce an algorithm for fitting the functional mixture model based on a modified expectation-maximization algorithm. Simulations confirm that in comparison to traditional functional principal component analysis the proposed method achieves similar or better data recovery while using fewer components on average. Its practical merits are also demonstrated in an analysis of egg-laying trajectories for medflies.

REFERENCES

- BENAGLIA, T., CHAUVEAU, D. and HUNTER, D. R. (2009). An EM-like algorithm for semi- and nonparametric estimation in multivariate mixtures. *J. Comput. Graph. Statist.* **18** 505–526. [MR2749842](#)
- BESSE, P. and RAMSAY, J. O. (1986). Principal components analysis of sampled functions. *Psychometrika* **51** 285–311. [MR0848110](#)

MSC2010 subject classifications. 62G05, 62G08.

Key words and phrases. Basis, functional data analysis, infinite mixture, probability density, trajectory representation.

- BOENTE, G. and FRAIMAN, R. (2000). Kernel-based functional principal components. *Statist. Probab. Lett.* **48** 335–345. [MR1771495](#)
- BONGIORNO, E. G. and GOIA, A. (2016). Some insights about the small ball probability factorization for Hilbert random elements. Preprint. Available at [arXiv:1501.04308v2](#).
- CAREY, J. R., LIEDO, P., MÜLLER, H.-G., WANG, J.-L. and CHIOU, J.-M. (1998). Relationship of age patterns of fecundity to mortality, longevity, and lifetime reproduction in a large, cohort of Mediterranean fruit fly females. *J. Gerontol., Ser. A, Biol. Sci. Med. Sci.* **53** B245–B251.
- 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.
- CHEN, K. and LEI, J. (2015). Localized functional principal component analysis. *J. Amer. Statist. Assoc.* **110** 1266–1275. [MR3420700](#)
- CHIOU, J.-M. and LI, P.-L. (2007). Functional clustering and identifying substructures of longitudinal data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 679–699. [MR2370075](#)
- DABO-NIANG, S. (2002). Estimation de la densité dans un espace de dimension infinie: Application aux diffusions. *C. R. Math. Acad. Sci. Paris* **334** 213–216. [MR1891061](#)
- DAUXOIS, J., POUSSE, A. and ROMAIN, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *J. Multivariate Anal.* **12** 136–154. [MR0650934](#)
- DELAIGLE, A. and HALL, P. (2010). Defining probability density for a distribution of random functions. *Ann. Statist.* **38** 1171–1193. [MR2604709](#)
- GASSER, T., HALL, P. and PRESNELL, B. (1998). Nonparametric estimation of the mode of a distribution of random curves. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 681–691. [MR1649539](#)
- GIKHMAN, I. I. and SKOROKHOD, A. V. (1969). *Introduction to the Theory of Random Processes*. W. B. Saunders Co., Philadelphia, PA. [MR0247660](#)
- GRENANDER, U. (1950). Stochastic processes and statistical inference. *Ark. Mat.* **1** 195–277. [MR0039202](#)
- HALL, P. and HOROWITZ, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35** 70–91. [MR2332269](#)
- HALL, P. and HOSSEINI-NASAB, M. (2006). On properties of functional principal components analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 109–126. [MR2212577](#)
- HALL, P. and HOSSEINI-NASAB, M. (2009). Theory for high-order bounds in functional principal components analysis. *Math. Proc. Cambridge Philos. Soc.* **146** 225–256. [MR2461880](#)
- HALL, P., MÜLLER, H.-G. and WANG, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34** 1493–1517. [MR2278365](#)
- HALL, P. and VIAL, C. (2006). Assessing the finite dimensionality of functional data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 689–705. [MR2301015](#)
- HSING, T. and EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley, Chichester. [MR3379106](#)
- JACQUES, J. and PREDÀ, C. (2014). Model-based clustering for multivariate functional data. *Comput. Statist. Data Anal.* **71** 92–106. [MR3131956](#)
- KNEIP, A. and UTIKAL, K. J. (2001). Inference for density families using functional principal component analysis. *J. Amer. Statist. Assoc.* **96** 519–542. [MR1946423](#)
- LEVINE, M., HUNTER, D. R. and CHAUVEAU, D. (2011). Maximum smoothed likelihood for multivariate mixtures. *Biometrika* **98** 403–416. [MR2806437](#)
- LI, Y. and GUAN, Y. (2014). Functional principal component analysis of spatiotemporal point processes with applications in disease surveillance. *J. Amer. Statist. Assoc.* **109** 1205–1215. [MR3265691](#)
- LI, Y. and HSING, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist.* **38** 3321–3351. [MR2766854](#)
- LI, W. V. and LINDE, W. (1999). Approximation, metric entropy and small ball estimates for Gaussian measures. *Ann. Probab.* **27** 1556–1578. [MR1733160](#)

- LI, Y., WANG, N. and CARROLL, R. J. (2013). Selecting the number of principal components in functional data. *J. Amer. Statist. Assoc.* **108** 1284–1294. [MR3174708](#)
- LIU, X. and MÜLLER, H.-G. (2003). Modes and clustering for time-warped gene expression profile data. *Bioinformatics* **19** 1937–1944.
- RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. Springer, New York. [MR2168993](#)
- RAO, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics* **14** 1–17.
- RICE, J. A. and SILVERMAN, B. W. (1991). Estimating the mean and covariance structure non-parametrically when the data are curves. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **53** 233–243. [MR1094283](#)
- SHIRYAYEV, A. N. (1984). *Probability. Graduate Texts in Mathematics* **95**. Springer, New York. [MR0737192](#)
- SILVERMAN, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *Ann. Statist.* **24** 1–24. [MR1389877](#)
- SLAETS, L., CLAESKENS, G. and HUBERT, M. (2012). Phase and amplitude-based clustering for functional data. *Comput. Statist. Data Anal.* **56** 2360–2374. [MR2897585](#)
- SUNG, S. H. (1999). Weak law of large numbers for arrays of random variables. *Statist. Probab. Lett.* **42** 293–298. [MR1688146](#)
- VAKHANIA, N. N., TARIELADZE, V. I. and CHOBANYAN, S. A. (1987). *Probability Distributions on Banach Spaces. Mathematics and Its Applications (Soviet Series)* **14**. D. Reidel, Dordrecht. [MR1435288](#)
- YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. [MR2160561](#)
- ZHANG, X. and WANG, J.-L. (2016). From sparse to dense functional data and beyond. *Ann. Statist.* **44** 2281–2321. [MR3546451](#)

BAYESIAN ESTIMATION OF SPARSE SIGNALS WITH A CONTINUOUS SPIKE-AND-SLAB PRIOR¹

BY VERONIKA ROČKOVÁ

University of Chicago

We introduce a new framework for estimation of sparse normal means, bridging the gap between popular frequentist strategies (LASSO) and popular Bayesian strategies (spike-and-slab). The main thrust of this paper is to introduce the family of Spike-and-Slab LASSO (SS-LASSO) priors, which form a continuum between the Laplace prior and the point-mass spike-and-slab prior. We establish several appealing frequentist properties of SS-LASSO priors, contrasting them with these two limiting cases. First, we adopt the penalized likelihood perspective on Bayesian modal estimation and introduce the framework of Bayesian penalty mixing with spike-and-slab priors. We show that the SS-LASSO global posterior mode is (near) minimax rate-optimal under squared error loss, similarly as the LASSO. Going further, we introduce an adaptive two-step estimator which can achieve provably sharper performance than the LASSO. Second, we show that the whole posterior keeps pace with the global mode and concentrates at the (near) minimax rate, a property that is known *not to hold* for the single Laplace prior. The minimax-rate optimality is obtained with a suitable class of independent product priors (for known levels of sparsity) as well as with dependent mixing priors (adapting to the unknown levels of sparsity). Up to now, the rate-optimal posterior concentration has been established only for spike-and-slab priors with a point mass at zero. Thus, the SS-LASSO priors, despite being continuous, possess similar optimality properties as the “theoretically ideal” point-mass mixtures. These results provide valuable theoretical justification for our proposed class of priors, underpinning their intuitive appeal and practical potential.

REFERENCES

- [1] ABRAMOVICH, F., GRINSHTEIN, V. and PENSKEY, M. (2007). On optimality of Bayesian estimation in the normal means problem. *Ann. Statist.* **35** 2261–2286. [MR2363971](#)
- [2] ANTONIADIS, A. and FAN, J. (2001). Regularization of wavelet approximations. *J. Amer. Statist. Assoc.* **96** 939–967. [MR1946364](#)
- [3] BHATTACHARYA, A., PATI, D., PILLAI, N. S. and DUNSON, D. B. (2015). Dirichlet–Laplace priors for optimal shrinkage. *J. Amer. Statist. Assoc.* **110** 1479–1490. [MR3449048](#)
- [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] BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer, Heidelberg. [MR2807761](#)

MSC2010 subject classifications. Primary 62J99; secondary 62F15.

Key words and phrases. Asymptotic minimaxity, LASSO, posterior concentration, spike-and-slab.

- [6] CARLIN, B. P. and CHIB, S. (1995). Bayesian model choice via Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **57** 473–484.
- [7] CARVALHO, C. M., POLSON, N. G. and SCOTT, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika* **97** 465–480. [MR2650751](#)
- [8] CASTILLO, I. (2014). Bayesian nonparametrics, convergence and limiting shape of posterior distributions. Univ. Paris Diderot Paris 7.
- [9] CASTILLO, I., SCHMIDT-HIEBER, J. and VAN DER VAART, A. W. (2015). Bayesian linear regression with sparse priors. *Ann. Statist.* **43** 1986–2018. [MR3375874](#)
- [10] CASTILLO, I. and VAN DER VAART, A. W. (2012). Needles and straw in a haystack: Posterior concentration for possibly sparse sequences. *Ann. Statist.* **40** 2069–2101. [MR3059077](#)
- [11] CHIPMAN, H., GEORGE, E. I. and McCULLOCH, R. E. (2001). The practical implementation of Bayesian model selection. In *Model Selection. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **38** 65–134. IMS, Beachwood, OH. [MR2000752](#)
- [12] CLYDE, M., DESIMONE, H. and PARMIGIANI, G. (1994). Prediction via orthogonalized model mixing. *J. Amer. Statist. Assoc.* **91** 1197–1208.
- [13] DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455. [MR1311089](#)
- [14] DONOHO, D. L., JOHNSTONE, I. M., HOCH, J. C. and STERN, A. S. (1992). Maximum entropy and the nearly black object. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **54** 41–81. [MR1157714](#)
- [15] FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360.
- [16] GEORGE, E. I. and FOSTER, D. P. (2000). Calibration and empirical Bayes variable selection. *Biometrika* **87** 731–747.
- [17] GEORGE, E. I. and McCULLOCH, R. E. (1993). Variable selection via Gibbs sampling. *J. Amer. Statist. Assoc.* **88** 881–889.
- [18] ISHWARAN, H. and RAO, J. S. (2003). Detecting differentially expressed genes in microarrays using Bayesian model selection. *J. Amer. Statist. Assoc.* **98** 438–455. [MR1995720](#)
- [19] ISHWARAN, H. and RAO, J. S. (2005). Spike and slab variable selection: Frequentist and Bayesian strategies. *Ann. Statist.* **33** 730–773. [MR2163158](#)
- [20] ISHWARAN, H. and RAO, J. S. (2011). Consistency of spike and slab regression. *Statist. Probab. Lett.* **81** 1920–1928.
- [21] JOHNSTONE, I. M. and SILVERMAN, B. W. (2004). Needles and straw in haystacks: Empirical Bayes estimates of possibly sparse sequences. *Ann. Statist.* **32** 1594–1649.
- [22] JOHNSTONE, I. M. and SILVERMAN, B. W. (2005). Empirical Bayes estimates selection of wavelet thresholds. *Ann. Statist.* **33** 1700–1752.
- [23] LEMPERS, F. B. (1971). *Posterior Probabilities of Alternative Linear Models: Some Theoretical Considerations and Empirical Experiments*. Rotterdam Univ. Press, Rotterdam.
- [24] 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](#)
- [25] MARTIN, R. and WALKER, S. G. (2014). Asymptotically minimax empirical Bayes estimation of a sparse normal mean vector. *Electron. J. Stat.* **8** 2188–2206. [MR3273623](#)
- [26] MITCHELL, T. J. and BEAUCHAMP, J. J. (1988). Bayesian variable selection in linear regression. *J. Amer. Statist. Assoc.* **83** 1023–1032.
- [27] NARISSETTY, N. and HE, X. (2014). Bayesian variable selection with shrinking and diffusing priors. *Ann. Statist.* **42** 789–817. [MR3210987](#)
- [28] PATI, D., BHATTACHARYA, A., PILLAI, N. and DUNSON, D. (2014). Posterior contraction in sparse Bayesian factor models for massive covariance matrices. *Ann. Statist.* **42** 1102–1130. [MR3210997](#)
- [29] ROČKOVÁ, V. (2018). Supplement to “Bayesian estimation of sparse signals with a continuous spike-and-slab prior.” DOI:[10.1214/17-AOS1554SUPP](#).

- [30] ROČKOVÁ, V. and GEORGE, E. I. (2014). EMVS: The EM approach to Bayesian variable selection. *J. Amer. Statist. Assoc.* **109** 828–846. [MR3223753](#)
- [31] ROCKOVA, V. and GEORGE, E. I. (2016). Fast Bayesian factor analysis via automatic rotations to sparsity. *J. Amer. Statist. Assoc.* **111** 160–1622.
- [32] ROCKOVA, V. and GEORGE, E. I. (2017). The Spike-and-Slab LASSO. *J. Amer. Statist. Assoc.* To appear.
- [33] SU, W. and CANDÉS, E. (2016). SLOPE is adaptive to unknown sparsity and asymptotically minimax. *Ann. Statist.* **44** 1038–1068. [MR3485953](#)
- [34] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **58** 267–288. [MR1379242](#)
- [35] 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.
- [36] VAN DER PAS, S. L., KLEIJN, B. J. K. and VAN DER VAART, A. W. (2014). The horseshoe estimator: Posterior concentration around nearly black vectors. *Electron. J. Stat.* **8** 2585–2618. [MR3285877](#)
- [37] ZHANG, C. H. and HUANG, J. (2008). The sparsity and bias of the LASSO selection in high-dimensional regression. *Ann. Statist.* **36** 1567–1594. [MR2435448](#)
- [38] ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. [MR2279469](#)
- [39] ZOU, H. and HASTIE, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 301–320. [MR2137327](#)

ON THE ASYMPTOTIC THEORY OF NEW BOOTSTRAP CONFIDENCE BOUNDS

BY CHARL PRETORIUS AND JAN W. H. SWANEPOEL¹

North-West University

We propose a new method, based on sample splitting, for constructing bootstrap confidence bounds for a parameter appearing in the regular smooth function model. It has been demonstrated in the literature, for example, by Hall [*Ann. Statist.* **16** (1988) 927–985; *The Bootstrap and Edgeworth Expansion* (1992) Springer], that the well-known *percentile-t* method for constructing bootstrap confidence bounds typically incurs a coverage error of order $O(n^{-1})$, with n being the sample size. Our version of the *percentile-t* bound reduces this coverage error to order $O(n^{-3/2})$ and in some cases to $O(n^{-2})$. Furthermore, whereas the standard *percentile* bounds typically incur coverage error of $O(n^{-1/2})$, the new bounds have reduced error of $O(n^{-1})$. In the case where the parameter of interest is the population mean, we derive for each confidence bound the exact coefficient of the leading term in an asymptotic expansion of the coverage error, although similar results may be obtained for other parameters such as the variance, the correlation coefficient, and the ratio of two means. We show that equal-tailed confidence intervals with coverage error at most $O(n^{-2})$ may be obtained from the newly proposed bounds, as opposed to the typical error $O(n^{-1})$ of the standard intervals. It is also shown that the good properties of the new *percentile-t* method carry over to regression problems. Results of independent interest are derived, such as a generalisation of a delta method by Cramér [*Mathematical Methods of Statistics* (1946) Princeton Univ. Press] and Hurt [*Apl. Mat.* **21** (1976) 444–456], and an expression for a polynomial appearing in an Edgeworth expansion of the distribution of a Studentised statistic for the slope parameter in a regression model. A small simulation study illustrates the behavior of the confidence bounds for small to moderate sample sizes.

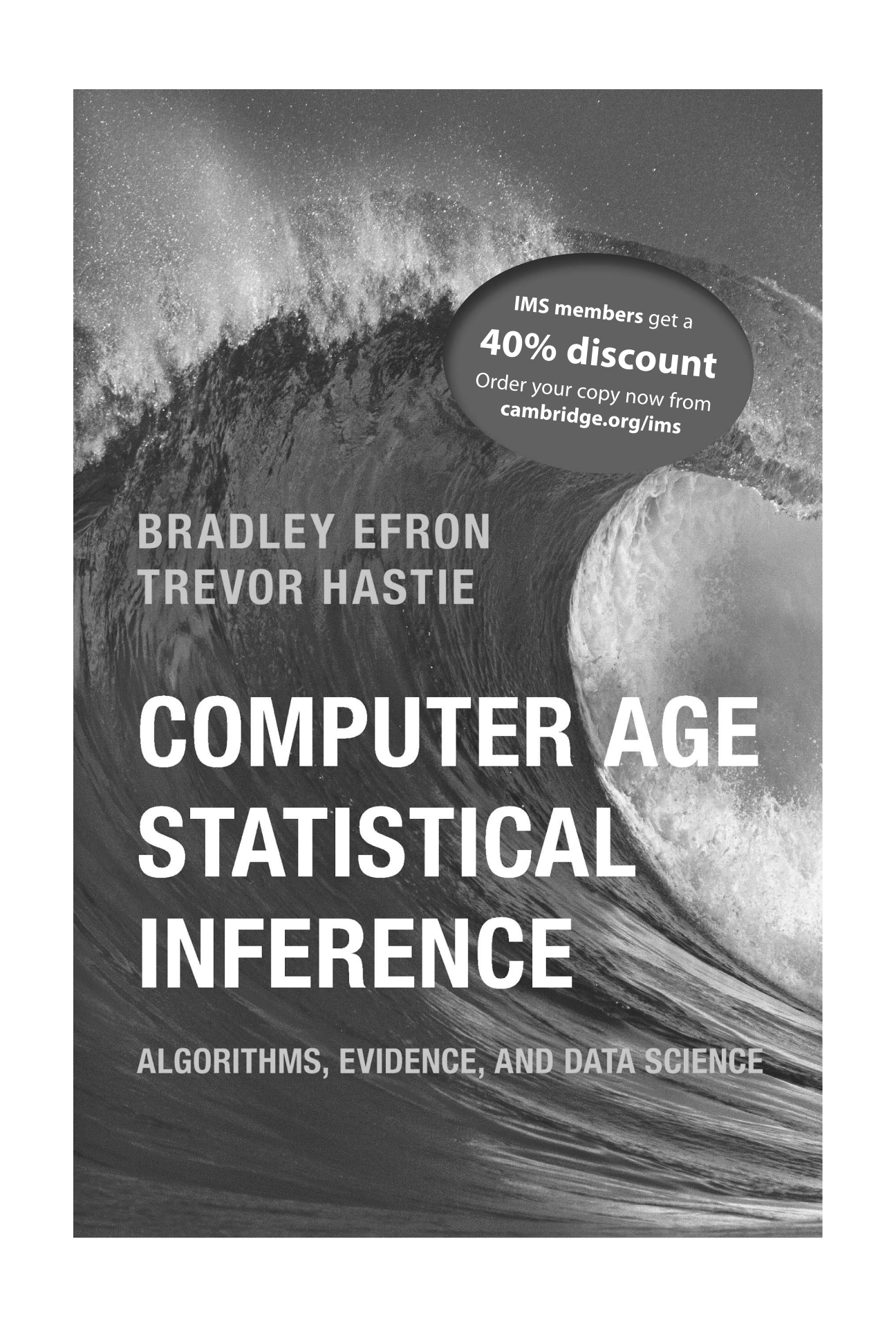
REFERENCES

- [1] BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434–451. [MR0471142](#)
- [2] BICKEL, P. J., GÖTZE, F. and VAN ZWET, W. R. (1997). Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statist. Sinica* **7** 1–31. [MR1441142](#)
- [3] CHANG, C. C. and POLITIS, D. N. (2011). Bootstrap with larger resample size for root- n consistent density estimation with time series data. *Statist. Probab. Lett.* **81** 652–661. [MR2783862](#)
- [4] CHUNG, K.-H. and LEE, S. M. S. (2001). Optimal bootstrap sample size in construction of percentile confidence bounds. *Scand. J. Stat.* **28** 225–239. [MR1844358](#)

MSC2010 subject classifications. Primary 62G09, 62G20; secondary 62G15.

Key words and phrases. Confidence bounds, sample splitting, coverage error, smooth function model, Edgeworth polynomials, Cornish–Fisher expansion, regression.

- [5] CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton Mathematical Series **9**. Princeton Univ. Press, Princeton, NJ. [MR0016588](#)
- [6] EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26. [MR0515681](#)
- [7] EFRON, B. and TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap*. Monographs on Statistics and Applied Probability **57**. Chapman & Hall, New York. [MR1270903](#)
- [8] FINNER, H. and DICKHAUS, T. (2010). Edgeworth expansions and rates of convergence for normalized sums: Chung’s 1946 method revisited. *Statist. Probab. Lett.* **80** 1875–1880. [MR2734254](#)
- [9] HALL, P. (1988). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist.* **16** 927–985. [MR0959185](#)
- [10] HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer, New York. [MR1145237](#)
- [11] HURT, J. (1976). Asymptotic expansions of functions of statistics. *Apl. Mat.* **21** 444–456. [MR0418309](#)
- [12] PRETORIUS, C. and SWANEPOEL, J. W. H. (2018). Supplement to “On the asymptotic theory of new bootstrap confidence bounds”. DOI:[10.1214/17-AOS1557SUPP](#).
- [13] SWANEPOEL, J. W. H. (1986). A note on proving that the (modified) bootstrap works. *Comm. Statist. Theory Methods* **15** 3193–3203. [MR0860478](#)



IMS members get a
40% discount

Order your copy now from
cambridge.org/ims

BRADLEY EFRON
TREVOR HASTIE

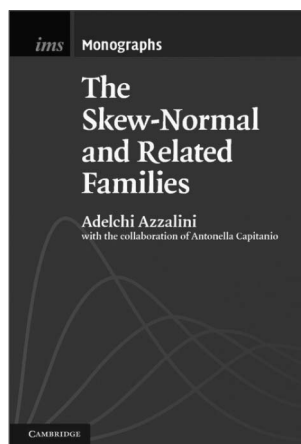
COMPUTER AGE STATISTICAL INFERENCE

ALGORITHMS, EVIDENCE, AND DATA SCIENCE



The Institute of Mathematical Statistics presents

IMS MONOGRAPHS



The Skew-Normal and Related Families

Adelchi Azzalini

in collaboration with Antonella Capitanio

Interest in the skew-normal and related families of distributions has grown enormously over recent years, as theory has advanced, challenges of data have grown, and computational tools have made substantial progress. This comprehensive treatment, blending theory and practice, will be the standard resource for statisticians and applied researchers. Assuming only basic knowledge of (non-measure-theoretic) probability and statistical inference, the book is accessible to the wide range of researchers who use statistical modelling techniques. Guiding readers through the main concepts and results, it covers both the probability and the statistics sides of the subject, in the univariate and multivariate settings. The theoretical development is complemented by numerous illustrations and applications to a range of fields including quantitative finance, medical statistics, environmental risk studies, and industrial and business efficiency.

The author's freely available R package *sn*, available from CRAN, equips readers to put the methods into action with their own data.

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

**Hardback price
US\$48.00
(non-member price
\$80.00)**

Cambridge University Press, in conjunction with the Institute of Mathematical Statistics, established the IMS Monographs and IMS Textbooks series of high-quality books. The Series Editors are Xiao-Li Meng, Susan Holmes, Ben Hambly, D. R. Cox and Alan Agresti.