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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

Vol. 50, No. 1—February 2022

THE ANNALS OF STATISTICS Vol. 50, No. 1, pp. 1–613 February 2022

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.

The Annals of Statistics **[ISSN 0090-5364 (print); ISSN 2168-8966 (online)],** Volume 50, Number 1, February 2022. Published bimonthly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, Ohio 44094, USA. Periodicals postage paid at Cleveland, Ohio, and at additional mailing offices.

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

The Annals of Statistics
2022, Vol. 50, No. 1, 1–29
https://doi.org/10.1214/21-AOS2061

AN OPTIMAL STATISTICAL AND COMPUTATIONAL FRAMEWORK FOR GENERALIZED TENSOR ESTIMATION

BY RUNGANG HAN[1,*], REBECCA WILLETT[2] AND ANRU R. ZHANG[1,†]

[1]*Department of Statistics, University of Wisconsin-Madison,* [*]*rhan32@stat.wisc.edu;* [†]*anruzhang@stat.wisc.edu*
[2]*Departments of Statistics and Computer Science, University of Chicago, willett@uchicago.edu*

This paper describes a flexible framework for generalized low-rank tensor estimation problems that includes many important instances arising from applications in computational imaging, genomics, and network analysis. The proposed estimator consists of finding a low-rank tensor fit to the data under generalized parametric models. To overcome the difficulty of nonconvexity in these problems, we introduce a unified approach of projected gradient descent that adapts to the underlying low-rank structure. Under mild conditions on the loss function, we establish both an upper bound on statistical error and the linear rate of computational convergence through a general deterministic analysis. Then we further consider a suite of generalized tensor estimation problems, including sub-Gaussian tensor PCA, tensor regression, and Poisson and binomial tensor PCA. We prove that the proposed algorithm achieves the minimax optimal rate of convergence in estimation error. Finally, we demonstrate the superiority of the proposed framework via extensive experiments on both simulated and real data.

REFERENCES

AHMED, A., RECHT, B. and ROMBERG, J. (2014). Blind deconvolution using convex programming. *IEEE Trans. Inf. Theory* **60** 1711–1732. MR3168432 https://doi.org/10.1109/TIT.2013.2294644

ANANDKUMAR, A., GE, R., HSU, D., KAKADE, S. M. and TELGARSKY, M. (2014). Tensor decompositions for learning latent variable models. *J. Mach. Learn. Res.* **15** 2773–2832. MR3270750

ARROYO, J., ATHREYA, A., CAPE, J., CHEN, G., PRIEBE, C. E. and VOGELSTEIN, J. T. (2021). Inference for multiple heterogeneous networks with a common invariant subspace. *J. Mach. Learn. Res.* **22** Paper No. 142. MR4318498

BAHADORI, M. T., YU, Q. R. and LIU, Y. (2014). Fast multivariate spatio-temporal analysis via low rank tensor learning. In *Advances in Neural Information Processing Systems* 3491–3499.

BAKER, Y., TANG, T. M. and ALLEN, G. I. (2020). Feature selection for data integration with mixed multiview data. *Ann. Appl. Stat.* **14** 1676–1698. MR4194243 https://doi.org/10.1214/20-AOAS1389

BARAK, B. and MOITRA, A. (2016). Noisy tensor completion via the sum-of-squares hierarchy. In *Conference on Learning Theory* 417–445.

BI, X., QU, A. and SHEN, X. (2018). Multilayer tensor factorization with applications to recommender systems. *Ann. Statist.* **46** 3308–3333. MR3852653 https://doi.org/10.1214/17-AOS1659

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 https://doi.org/10.1214/16-AOS1443

CAI, C., LI, G., POOR, H. V. and CHEN, Y. (2019). Nonconvex low-rank tensor completion from noisy data. In *Advances in Neural Information Processing Systems* 1861–1872.

CANDÈS, E. J., LI, X. and SOLTANOLKOTABI, M. (2015). Phase retrieval via Wirtinger flow: Theory and algorithms. *IEEE Trans. Inf. Theory* **61** 1985–2007. MR3332993 https://doi.org/10.1109/TIT.2015.2399924

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

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

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

MSC2020 subject classifications. Primary 62H12, 62H25; secondary 62C20.
Key words and phrases. Generalize tensor estimation, gradient descent, image denoising, low-rank tensor, minimax optimality, nonconvex optimization.

CAO, Y. and XIE, Y. (2016). Poisson matrix recovery and completion. *IEEE Trans. Signal Process.* **64** 1609–1620. MR3548877 https://doi.org/10.1109/TSP.2015.2500192

CAO, Y., ZHANG, A. and LI, H. (2020). Multisample estimation of bacterial composition matrices in metagenomics data. *Biometrika* **107** 75–92. MR4064141 https://doi.org/10.1093/biomet/asz062

CHEN, W.-K. (2019). Phase transition in the spiked random tensor with Rademacher prior. *Ann. Statist.* **47** 2734–2756. MR3988771 https://doi.org/10.1214/18-AOS1763

CHEN, Y. and CANDES, E. (2015). Solving random quadratic systems of equations is nearly as easy as solving linear systems. In *Advances in Neural Information Processing Systems* 739–747.

CHEN, Y. and CHI, Y. (2018). Harnessing structures in big data via guaranteed low-rank matrix estimation. ArXiv preprint. Available at arXiv:1802.08397.

CHEN, H., RASKUTTI, G. and YUAN, M. (2019). Non-convex projected gradient descent for generalized low-rank tensor regression. *J. Mach. Learn. Res.* **20** Paper No. 5. MR3911412

CHI, E. C. and KOLDA, T. G. (2012). On tensors, sparsity, and nonnegative factorizations. *SIAM J. Matrix Anal. Appl.* **33** 1272–1299. MR3023474 https://doi.org/10.1137/110859063

CHI, Y., LU, Y. M. and CHEN, Y. (2019). Nonconvex optimization meets low-rank matrix factorization: An overview. *IEEE Trans. Signal Process.* **67** 5239–5269. MR4016283 https://doi.org/10.1109/TSP.2019.2937282

CHI, E. C., GAINES, B. R., SUN, W. W., ZHOU, H. and YANG, J. (2020). Provable convex co-clustering of tensors. *J. Mach. Learn. Res.* **21** Paper No. 214. MR4209500

DE LATHAUWER, L., DE MOOR, B. and VANDEWALLE, J. (2000a). A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.* **21** 1253–1278. MR1780272 https://doi.org/10.1137/S0895479896305696

DE LATHAUWER, L., DE MOOR, B. and VANDEWALLE, J. (2000b). On the best rank-1 and rank-$(R_1, R_2, \dots, R_N)$ approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl.* **21** 1324–1342. MR1780276 https://doi.org/10.1137/S0895479898346995

FAN, J., GONG, W. and ZHU, Z. (2019). Generalized high-dimensional trace regression via nuclear norm regularization. *J. Econometrics* **212** 177–202. MR3994013 https://doi.org/10.1016/j.jeconom.2019.04.026

FAUST, K., SATHIRAPONGSASUTI, J. F., IZARD, J., SEGATA, N., GEVERS, D., RAES, J. and HUTTENHOWER, C. (2012). Microbial co-occurrence relationships in the human microbiome. *PLoS Comput. Biol.* **8** e1002606.

FAZEL, M. (2002). Matrix rank minimization with applications.

FLORES, G. E., CAPORASO, J. G., HENLEY, J. B., RIDEOUT, J. R., DOMOGALA, D., CHASE, J., LEFF, J. W., VÁZQUEZ-BAEZA, Y., GONZALEZ, A. et al. (2014). Temporal variability is a personalized feature of the human microbiome. *Genome Biol.* **15** 531. https://doi.org/10.1186/s13059-014-0531-y

FRIEDLAND, S. and LIM, L.-H. (2018). Nuclear norm of higher-order tensors. *Math. Comp.* **87** 1255–1281. MR3766387 https://doi.org/10.1090/mcom/3239

GANDY, S., RECHT, B. and YAMADA, I. (2011). Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Probl.* **27** 025010. MR2765628 https://doi.org/10.1088/0266-5611/27/2/025010

GUHANIYOGI, R., QAMAR, S. and DUNSON, D. B. (2017). Bayesian tensor regression. *J. Mach. Learn. Res.* **18** Paper No. 79. MR3714242

GUO, W., KOTSIA, I. and PATRAS, I. (2012). Tensor learning for regression. *IEEE Trans. Image Process.* **21** 816–827. MR2932176 https://doi.org/10.1109/TIP.2011.2165291

HALL, E. C., RASKUTTI, G. and WILLETT, R. (2016). Inference of high-dimensional autoregressive generalized linear models. ArXiv preprint. Available at arXiv:1605.02693.

HAN, R., WILLETT, R. and ZHANG, A. R. (2022). Supplement to "An optimal statistical and computational framework for generalized tensor estimation." https://doi.org/10.1214/21-AOS2061SUPP

HAN, Q., XU, K. and AIROLDI, E. (2015). Consistent estimation of dynamic and multi-layer block models. In *International Conference on Machine Learning* 1511–1520.

HAO, B., ZHANG, A. and CHENG, G. (2020). Sparse and low-rank tensor estimation via cubic sketchings. *IEEE Trans. Inf. Theory* **66** 5927–5964. MR4158653 https://doi.org/10.1109/TIT.2020.2982499

HENRIQUES, R. and MADEIRA, S. C. (2019). Triclustering algorithms for three-dimensional data analysis: A comprehensive survey. *ACM Computing Surveys* (*CSUR*) **51** 95.

HILLAR, C. J. and LIM, L.-H. (2013). Most tensor problems are NP-hard. *J. ACM* **60** Art. 45. MR3144915 https://doi.org/10.1145/2512329

HOFF, P. D. (2015). Multilinear tensor regression for longitudinal relational data. *Ann. Appl. Stat.* **9** 1169–1193. MR3418719 https://doi.org/10.1214/15-AOAS839

HONG, D., KOLDA, T. G. and DUERSCH, J. A. (2020). Generalized canonical polyadic tensor decomposition. *SIAM Rev.* **62** 133–163. MR4064532 https://doi.org/10.1137/18M1203626

HOPKINS, S. B., SHI, J. and STEURER, D. (2015). Tensor principal component analysis via sum-of-square proofs. In *Proceedings of the* 28*th Conference on Learning Theory*, *COLT* 3–6.

JAVANMARD, A. and MONTANARI, A. (2018). Debiasing the Lasso: Optimal sample size for Gaussian designs. *Ann. Statist.* **46** 2593–2622. MR3851749 https://doi.org/10.1214/17-AOS1630

JIANG, X., RASKUTTI, G. and WILLETT, R. (2015). Minimax optimal rates for Poisson inverse problems with physical constraints. *IEEE Trans. Inf. Theory* **61** 4458–4474. MR3372365 https://doi.org/10.1109/TIT.2015.2441072

JOHNDROW, J. E., BHATTACHARYA, A. and DUNSON, D. B. (2017). Tensor decompositions and sparse log-linear models. *Ann. Statist.* **45** 1–38. MR3611485 https://doi.org/10.1214/15-AOS1414

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

KESHAVAN, R. H., MONTANARI, A. and OH, S. (2010). Matrix completion from noisy entries. *J. Mach. Learn. Res.* **11** 2057–2078. MR2678022

KOLDA, T. G. and BADER, B. W. (2009). Tensor decompositions and applications. *SIAM Rev.* **51** 455–500. MR2535056 https://doi.org/10.1137/07070111X

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

KRIVANEK, O., DELLBY, N. and LUPINI, A. (1999). Towards sub-A electron beams. *Ultramicroscopy* **78** 1–11.

KROONENBERG, P. M. (2008). *Applied Multiway Data Analysis. Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR2378349 https://doi.org/10.1002/9780470238004

LEI, J., CHEN, K. and LYNCH, B. (2020). Consistent community detection in multi-layer network data. *Biometrika* **107** 61–73. MR4064140 https://doi.org/10.1093/biomet/asz068

LESIEUR, T., KRZAKALA, F. and ZDEBOROVÁ, L. (2017). Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications. *J. Stat. Mech. Theory Exp.* **7** 073403. MR3683819 https://doi.org/10.1088/1742-5468/aa7284

LI, N. and LI, B. (2010). Tensor completion for on-board compression of hyperspectral images. In 2010 *IEEE International Conference on Image Processing* 517–520. IEEE, New York.

LI, L. and ZHANG, X. (2017). Parsimonious tensor response regression. *J. Amer. Statist. Assoc.* **112** 1131–1146. MR3735365 https://doi.org/10.1080/01621459.2016.1193022

LI, X., XU, D., ZHOU, H. and LI, L. (2018). Tucker tensor regression and neuroimaging analysis. *Stat. Biosci.* **10** 520–545.

LIU, J., MUSIALSKI, P., WONKA, P. and YE, J. (2013). Tensor completion for estimating missing values in visual data. *IEEE Trans. Pattern Anal. Mach. Intell.* **35** 208–220.

LUBICH, C., ROHWEDDER, T., SCHNEIDER, R. and VANDEREYCKEN, B. (2013). Dynamical approximation by hierarchical Tucker and tensor-train tensors. *SIAM J. Matrix Anal. Appl.* **34** 470–494. MR3053575 https://doi.org/10.1137/120885723

MA, Z. and MA, Z. (2017). Exploration of large networks via fast and universal latent space model fitting. ArXiv preprint. Available at arXiv:1705.02372.

MCMAHAN, H. B., HOLT, G., SCULLEY, D., YOUNG, M., EBNER, D., GRADY, J., NIE, L., PHILLIPS, T., DAVYDOV, E. et al. (2013). Ad click prediction: A view from the trenches. In *Proceedings of the* 19*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 1222–1230. ACM, New York.

MONTANARI, A., REICHMAN, D. and ZEITOUNI, O. (2017). On the limitation of spectral methods: From the Gaussian hidden clique problem to rank one perturbations of Gaussian tensors. *IEEE Trans. Inf. Theory* **63** 1572–1579. MR3625981 https://doi.org/10.1109/TIT.2016.2637959

MONTANARI, A. and SUN, N. (2018). Spectral algorithms for tensor completion. *Comm. Pure Appl. Math.* **71** 2381–2425. MR3862094 https://doi.org/10.1002/cpa.21748

OYMAK, S., JALALI, A., FAZEL, M., ELDAR, Y. C. and HASSIBI, B. (2015). Simultaneously structured models with application to sparse and low-rank matrices. *IEEE Trans. Inf. Theory* **61** 2886–2908. MR3342310 https://doi.org/10.1109/TIT.2015.2401574

PARK, D., KYRILLIDIS, A., CARAMANIS, C. and SANGHAVI, S. (2018). Finding low-rank solutions via nonconvex matrix factorization, efficiently and provably. *SIAM J. Imaging Sci.* **11** 2165–2204. MR3860118 https://doi.org/10.1137/17M1150189

PENSKY, M. and ZHANG, T. (2019). Spectral clustering in the dynamic stochastic block model. *Electron. J. Stat.* **13** 678–709. MR3914178 https://doi.org/10.1214/19-ejs1533

PERRY, A., WEIN, A. S. and BANDEIRA, A. S. (2020). Statistical limits of spiked tensor models. *Ann. Inst. Henri Poincaré Probab. Stat.* **56** 230–264. MR4058987 https://doi.org/10.1214/19-AIHP960

RASKUTTI, G., YUAN, M. and CHEN, H. (2019). Convex regularization for high-dimensional multiresponse tensor regression. *Ann. Statist.* **47** 1554–1584. MR3911122 https://doi.org/10.1214/18-AOS1725

RAUHUT, H., SCHNEIDER, R. and STOJANAC, Ž. (2015). Tensor completion in hierarchical tensor representations. In *Compressed Sensing and Its Applications. Appl. Numer. Harmon. Anal.* 419–450. Birkhäuser/Springer, Cham. MR3382114

RAUHUT, H., SCHNEIDER, R. and STOJANAC, Ž. (2017). Low rank tensor recovery via iterative hard thresholding. *Linear Algebra Appl*. **523** 220–262. MR3624675 https://doi.org/10.1016/j.laa.2017.02.028

RICHARD, E. and MONTANARI, A. (2014). A statistical model for tensor PCA. In *Advances in Neural Information Processing Systems* 2897–2905.

SALMON, J., HARMANY, Z., DELEDALLE, C.-A. and WILLETT, R. (2014). Poisson noise reduction with non-local PCA. *J. Math. Imaging Vision* **48** 279–294. MR3152105 https://doi.org/10.1007/s10851-013-0435-6

SEWELL, D. K. and CHEN, Y. (2015). Latent space models for dynamic networks. *J. Amer. Statist. Assoc*. **110** 1646–1657. MR3449061 https://doi.org/10.1080/01621459.2014.988214

SHAH, D. (2018). Matrix estimation, latent variable model and collaborative filtering. In 37*th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science*. *LIPIcs. Leibniz Int. Proc. Inform*. **93** Art. No. 4. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR3774048

SHAN, L., LIN, L., SUN, C. and WANG, X. (2016). Predicting ad click-through rates via feature-based fully coupled interaction tensor factorization. *Electronic Commerce Research and Applications* **16** 30–42.

SIGNORETTO, M., VAN DE PLAS, R., DE MOOR, B. and SUYKENS, J. A. (2011). Tensor versus matrix completion: A comparison with application to spectral data. *IEEE Signal Process. Lett*. **18** 403–406.

SUN, W. W. and LI, L. (2017). STORE: Sparse tensor response regression and neuroimaging analysis. *J. Mach. Learn. Res*. **18** Paper No. 135. MR3763769

SUN, R. and LUO, Z.-Q. (2015). Guaranteed matrix completion via nonconvex factorization. In 2015 *IEEE* 56*th Annual Symposium on Foundations of Computer Science—FOCS* 2015 270–289. IEEE Computer Soc., Los Alamitos, CA. MR3473312 https://doi.org/10.1109/FOCS.2015.25

SUN, W. W., LU, J., LIU, H. and CHENG, G. (2017). Provable sparse tensor decomposition. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **79** 899–916. MR3641413 https://doi.org/10.1111/rssb.12190

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

TIMMERMANN, K. E. and NOWAK, R. D. (1999). Multiscale modeling and estimation of Poisson processes with application to photon-limited imaging. *IEEE Trans. Inf. Theory* **45** 846–862. MR1682515 https://doi.org/10.1109/18.761328

TOMIOKA, R. and SUZUKI, T. (2013). Convex tensor decomposition via structured Schatten norm regularization. In *Advances in Neural Information Processing Systems* 1331–1339.

TOMIOKA, R., SUZUKI, T., HAYASHI, K. and KASHIMA, H. (2011). Statistical performance of convex tensor decomposition. In *Advances in Neural Information Processing Systems* 972–980.

TU, S., BOCZAR, R., SIMCHOWITZ, M., SOLTANOLKOTABI, M. and RECHT, B. (2016). Low-rank solutions of linear matrix equations via procrustes flow. In *International Conference on Machine Learning* 964–973.

WANG, M., FISCHER, J. and SONG, Y. S. (2017). Three-way clustering of multi-tissue multi-individual gene expression data using constrained tensor decomposition. *BioRxiv* 229245.

WANG, M. and LI, L. (2020). Learning from binary multiway data: Probabilistic tensor decomposition and its statistical optimality. *J. Mach. Learn. Res*. **21** Paper No. 154. MR4209440

WANG, M. and ZENG, Y. (2019). Multiway clustering via tensor block models. In *Advances in Neural Information Processing Systems* 713–723.

WEDDERBURN, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* **61** 439–447. MR0375592 https://doi.org/10.1093/biomet/61.3.439

WEN, Z., YIN, W. and ZHANG, Y. (2012). Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm. *Math. Program. Comput*. **4** 333–361. MR3006618 https://doi.org/10.1007/s12532-012-0044-1

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

WILMOTH, J. R. and SHKOLNIKOV, V. (2006). Human mortality database. Available at http://www.mortality.org.

XIA, D. and YUAN, M. (2019). On polynomial time methods for exact low-rank tensor completion. *Found. Comput. Math*. **19** 1265–1313. MR4029842 https://doi.org/10.1007/s10208-018-09408-6

XIA, D., YUAN, M. and ZHANG, C.-H. (2021). Statistically optimal and computationally efficient low rank tensor completion from noisy entries. *Ann. Statist*. **49** 76–99. MR4206670 https://doi.org/10.1214/20-AOS1942

XU, Z., HU, J. and WANG, M. (2019). Generalized tensor regression with covariates on multiple modes. ArXiv preprint. Available at arXiv:1910.09499.

YANKOVICH, A. B., ZHANG, C., OH, A., SLATER, T. J., AZOUGH, F., FREER, R., HAIGH, S. J., WILLETT, R. and VOYLES, P. M. (2016). Non-rigid registration and non-local principle component analysis to improve electron microscopy spectrum images. *Nanotechnology* **27** 364001.

YOKOTA, T., LEE, N. and CICHOCKI, A. (2017). Robust multilinear tensor rank estimation using higher order singular value decomposition and information criteria. *IEEE Trans. Signal Process*. **65** 1196–1206. MR3584316 https://doi.org/10.1109/TSP.2016.2620965

YONEL, B. and YAZICI, B. (2020). A deterministic theory for exact non-convex phase retrieval. *IEEE Trans. Signal Process*. **68** 4612–4626. MR4144925 https://doi.org/10.1109/TSP.2020.3007967

YU, M., GUPTA, V. and KOLAR, M. (2020). Recovery of simultaneous low rank and two-way sparse coefficient matrices, a nonconvex approach. *Electron. J. Stat*. **14** 413–457. MR4054252 https://doi.org/10.1214/19-EJS1658

YUAN, M. and ZHANG, C.-H. (2016). On tensor completion via nuclear norm minimization. *Found. Comput. Math*. **16** 1031–1068. MR3529132 https://doi.org/10.1007/s10208-015-9269-5

ZHANG, A. (2019). Cross: Efficient low-rank tensor completion. *Ann. Statist*. **47** 936–964. MR3909956 https://doi.org/10.1214/18-AOS1694

ZHANG, A., CAI, T. T. and WU, Y. (2018). Heteroskedastic PCA: Algorithm, optimality, and applications. ArXiv preprint. Available at arXiv:1810.08316.

ZHANG, A. and HAN, R. (2019). Optimal sparse singular value decomposition for high-dimensional high-order data. *J. Amer. Statist. Assoc*. **114** 1708–1725. MR4047294 https://doi.org/10.1080/01621459.2018.1527227

ZHANG, A. and XIA, D. (2018). Tensor SVD: Statistical and computational limits. *IEEE Trans. Inf. Theory* **64** 7311–7338. MR3876445 https://doi.org/10.1109/TIT.2018.2841377

ZHANG, A. R., LUO, Y., RASKUTTI, G. and YUAN, M. (2020). ISLET: Fast and optimal low-rank tensor regression via importance sketching. *SIAM J. Math. Data Sci*. **2** 444–479. MR4106613 https://doi.org/10.1137/19M126476X

ZHAO, T., WANG, Z. and LIU, H. (2015). A nonconvex optimization framework for low rank matrix estimation. In *Advances in Neural Information Processing Systems* 559–567.

ZHOU, H. (2017). Matlab TensorReg Toolbox Version 1.0. Available at https://hua-zhou.github.io/TensorReg/.

ZHOU, H., LI, L. and ZHU, H. (2013). Tensor regression with applications in neuroimaging data analysis. *J. Amer. Statist. Assoc*. **108** 540–552. MR3174640 https://doi.org/10.1080/01621459.2013.776499

ZHU, Z., LI, Q., TANG, G. and WAKIN, M. B. (2017). The global optimization geometry of nonsymmetric matrix factorization and sensing. ArXiv preprint. Available at arXiv:1703.01256.

The Annals of Statistics
2022, Vol. 50, No. 1, 30–52
https://doi.org/10.1214/21-AOS2067

SPATIAL DEPENDENCE AND SPACE–TIME TREND IN EXTREME EVENTS

BY JOHN H. J. EINMAHL[1], ANA FERREIRA[2], LAURENS DE HAAN[3,4], CLÁUDIA NEVES[5] AND CHEN ZHOU[6,7]

[1]*Department of Econometrics and Operations Research, Tilburg University, j.h.j.einmahl@uvt.nl*
[2]*Instituto Superior Técnico (IST), University of Lisbon, anafh@tecnico.ulisboa.pt*
[3]*Department of Statistics, University of Lisbon*
[4]*Department of Economics, Erasmus University Rotterdam, ldehaan@ese.eur.nl*
[5]*Department of Mathematics and Statistics, University of Reading, c.neves@reading.ac.uk*
[6]*Econometric Institute, Erasmus University Rotterdam, zhou@ese.eur.nl*
[7]*Economics and Research Division, Bank of The Netherlands*

The statistical theory of extremes is extended to independent multivariate observations that are non-stationary both over time and across space. The non-stationarity over time and space is controlled via the scedasis (tail scale) in the marginal distributions. Spatial dependence stems from multivariate extreme value theory. We establish asymptotic theory for both the weighted sequential tail empirical process and the weighted tail quantile process based on all observations, taken over time and space. The results yield two statistical tests for homoscedasticity in the tail, one in space and one in time. Further, we show that the common extreme value index can be estimated via a pseudo-maximum likelihood procedure based on pooling all (non-stationary and dependent) observations. Our leading example and application is rainfall in Northern Germany.

REFERENCES

BÜCHER, A., JÄSCHKE, S. and WIED, D. (2015). Nonparametric tests for constant tail dependence with an application to energy and finance. *J. Econometrics* **187** 154–168. MR3347300 https://doi.org/10.1016/j.jeconom.2015.02.002

COLES, S. G. and TAWN, J. A. (1996). Modelling extremes of the areal rainfall process. *J. Roy. Statist. Soc. Ser. B* **58** 329–347. MR1377836

DAVISON, A. C., PADOAN, S. A. and RIBATET, M. (2012). Statistical modeling of spatial extremes. *Statist. Sci.* **27** 161–186. MR2963980 https://doi.org/10.1214/11-STS376

DAVISON, A. C. and SMITH, R. L. (1990). Models for exceedances over high thresholds. *J. Roy. Statist. Soc. Ser. B* **52** 393–442. MR1086795

DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory: An introduction. Springer Series in Operations Research and Financial Engineering*. Springer, New York. MR2234156 https://doi.org/10.1007/0-387-34471-3

DE HAAN, L., KLEIN TANK, A. and NEVES, C. (2015). On tail trend detection: Modeling relative risk. *Extremes* **18** 141–178. MR3351812 https://doi.org/10.1007/s10687-014-0207-8

DOMBRY, C. and FERREIRA, A. (2019). Maximum likelihood estimators based on the block maxima method. *Bernoulli* **25** 1690–1723. MR3961227 https://doi.org/10.3150/18-BEJ1032

DREES, H., DE HAAN, L. and LI, D. (2006). Approximations to the tail empirical distribution function with application to testing extreme value conditions. *J. Statist. Plann. Inference* **136** 3498–3538. MR2284668 https://doi.org/10.1016/j.jspi.2005.02.017

EINMAHL, J. H. J., DE HAAN, L. and ZHOU, C. (2016). Statistics of heteroscedastic extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 31–51. MR3453645 https://doi.org/10.1111/rssb.12099

EINMAHL, J. H. J., KRAJINA, A. and SEGERS, J. (2012). An M-estimator for tail dependence in arbitrary dimensions. *Ann. Statist.* **40** 1764–1793. MR3015043 https://doi.org/10.1214/12-AOS1023

MSC2020 subject classifications. Primary 62G32, 62G30, 62G05, 62G10, 62G20; secondary 60F17, 60G70.

Key words and phrases. Multivariate extreme value statistics, non-identical distributions, sequential tail empirical process, testing.

EINMAHL, J. H, FERREIRA, A., DE HAAN, L., NEVES, C. and ZHOU, C. (2022). Supplement to "Spatial dependence and space–time trend in extreme events." https://doi.org/10.1214/21-AOS2067SUPP

FERGER, D. and VOGEL, D. (2015). Weak convergence of the empirical process and the rescaled empirical distribution function in the Skorokhod product space. *Theory Probab. Appl.* **4** 609–625.

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

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

The Annals of Statistics
2022, Vol. 50, No. 1, 53–80
https://doi.org/10.1214/21-AOS2074

HETEROSKEDASTIC PCA: ALGORITHM, OPTIMALITY, AND APPLICATIONS

BY ANRU R. ZHANG[1,2], T. TONY CAI[3] AND YIHONG WU[4]

[1]*Department of Statistics, University of Wisconsin-Madison*
[2]*Department of Biostatistics & Bioinformatics, Duke University, anru.zhang@duke.edu*
[3]*Department of Statistics, The Wharton School, University of Pennsylvania, tcai@wharton.upenn.edu*
[4]*Department of Statistics and Data Science, Yale University, yihong.wu@yale.edu*

A general framework for principal component analysis (PCA) in the presence of heteroskedastic noise is introduced. We propose an algorithm called HeteroPCA, which involves iteratively imputing the diagonal entries of the sample covariance matrix to remove estimation bias due to heteroskedasticity. This procedure is computationally efficient and provably optimal under the generalized spiked covariance model. A key technical step is a deterministic robust perturbation analysis on singular subspaces, which can be of independent interest. The effectiveness of the proposed algorithm is demonstrated in a suite of problems in high-dimensional statistics, including singular value decomposition (SVD) under heteroskedastic noise, Poisson PCA, and SVD for heteroskedastic and incomplete data.

REFERENCES

[1] BAI, J. and LI, K. (2012). Statistical analysis of factor models of high dimension. *Ann. Statist.* **40** 436–465. MR3014313 https://doi.org/10.1214/11-AOS966

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

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

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

[5] BARTLETT, M. S. (1937). The statistical conception of mental factors. *Br. J. Health Psychol.* **28** 97.

[6] BUBECK, S. (2015). Convex optimization: Algorithms and complexity. *Found. Trends Mach. Learn.* **8** 231–357.

[7] CAI, J.-F. and WEI, K. (2018). Exploiting the structure effectively and efficiently in low-rank matrix recovery. In *Processing, Analyzing and Learning of Images, Shapes, and Forms. Part* 1. *Handb. Numer. Anal.* **19** 21–51. North-Holland, Amsterdam. MR3839040

[8] CAI, T. T., HAN, R. and ZHANG, A. R. (2020). On the non-asymptotic concentration of heteroskedastic Wishart-type matrix. ArXiv preprint. Available at arXiv:2008.12434.

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

[10] CAI, T. T. and ZHANG, A. (2018). Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *Ann. Statist.* **46** 60–89. MR3766946 https://doi.org/10.1214/17-AOS1541

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

[12] CANDÈS, 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 https://doi.org/10.1109/TSP.2013.2270464

MSC2020 subject classifications. Primary 62H12, 62H25; secondary 62C20.
Key words and phrases. Factor analysis model, heteroskedasticity, perturbation bound, principal component analysis, singular value decomposition.

[13] CANDÈS, E. J. and TAO, T. (2010). The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inf. Theory* **56** 2053–2080. MR2723472 https://doi.org/10.1109/TIT.2010.2044061
[14] CAO, Y., ZHANG, A. and LI, H. (2020). Multisample estimation of bacterial composition matrices in metagenomics data. *Biometrika* **107** 75–92. MR4064141 https://doi.org/10.1093/biomet/asz062
[15] CHATTERJEE, S. (2015). Matrix estimation by universal singular value thresholding. *Ann. Statist.* **43** 177–214. MR3285604 https://doi.org/10.1214/14-AOS1272
[16] COLLINS, M., DASGUPTA, S. and SCHAPIRE, R. E. (2002). A generalization of principal components analysis to the exponential family. In *Advances in Neural Information Processing Systems* 617–624.
[17] DAVIS, C. and KAHAN, W. M. (1970). The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.* **7** 1–46. MR0264450 https://doi.org/10.1137/0707001
[18] DE LATHAUWER, L., DE MOOR, B. and VANDEWALLE, J. (2000). On the best rank-1 and rank-$(R_1, R_2, \dots, R_N)$ approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl.* **21** 1324–1342. MR1780276 https://doi.org/10.1137/S0895479898346995
[19] DONOHO, D. and GAVISH, M. (2014). Minimax risk of matrix denoising by singular value thresholding. *Ann. Statist.* **42** 2413–2440. MR3269984 https://doi.org/10.1214/14-AOS1257
[20] DONOHO, D., GAVISH, M. and JOHNSTONE, I. (2018). Optimal shrinkage of eigenvalues in the spiked covariance model. *Ann. Statist.* **46** 1742–1778. MR3819116 https://doi.org/10.1214/17-AOS1601
[21] FLORESCU, L. and PERKINS, W. (2016). Spectral thresholds in the bipartite stochastic block model. In *Conference on Learning Theory* 943–959.
[22] FORTUNATO, S. (2010). Community detection in graphs. *Phys. Rep.* **486** 75–174. MR2580414 https://doi.org/10.1016/j.physrep.2009.11.002
[23] GAVISH, M. and DONOHO, D. L. (2017). Optimal shrinkage of singular values. *IEEE Trans. Inf. Theory* **63** 2137–2152. MR3626861 https://doi.org/10.1109/TIT.2017.2653801
[24] GHOSH, J. and DUNSON, D. B. (2009). Default prior distributions and efficient posterior computation in Bayesian factor analysis. *J. Comput. Graph. Statist.* **18** 306–320. MR2749834 https://doi.org/10.1198/jcgs.2009.07145
[25] HONG, D., BALZANO, L. and FESSLER, J. A. (2016). Towards a theoretical analysis of PCA for heteroscedastic data. In *Communication, Control, and Computing (Allerton)*, 2016 54*th Annual Allerton Conference on* 496–503. IEEE, Los Alamitos.
[26] HONG, D., BALZANO, L. and FESSLER, J. A. (2018). Asymptotic performance of PCA for high-dimensional heteroscedastic data. *J. Multivariate Anal.* **167** 435–452. MR3830656 https://doi.org/10.1016/j.jmva.2018.06.002
[27] HONG, D., BALZANO, L. and FESSLER, J. A. (2018). Asymptotic performance of PCA for high-dimensional heteroscedastic data. *J. Multivariate Anal.* **167** 435–452. MR3830656 https://doi.org/10.1016/j.jmva.2018.06.002
[28] HOTELLING, H. (1936). Relations between two sets of variates. *Biometrika* **28** 321–377.
[29] JAIN, P., NETRAPALLI, P. and SANGHAVI, S. (2013). Low-rank matrix completion using alternating minimization (extended abstract). In *STOC'13—Proceedings of the* 2013 *ACM Symposium on Theory of Computing* 665–674. ACM, New York. MR3210828 https://doi.org/10.1145/2488608.2488693
[30] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961 https://doi.org/10.1214/aos/1009210544
[31] KESHAVAN, R. H. (2012). Efficient algorithms for collaborative filtering Ph.D. thesis Stanford Univ.
[32] KESHAVAN, R. H., MONTANARI, A. and OH, S. (2010). Matrix completion from a few entries. *IEEE Trans. Inf. Theory* **56** 2980–2998. MR2683452 https://doi.org/10.1109/TIT.2010.2046205
[33] KESHAVAN, R. H., MONTANARI, A. and OH, S. (2010). Matrix completion from noisy entries. *J. Mach. Learn. Res.* **11** 2057–2078. MR2678022
[34] LAWLEY, D. N. and MAXWELL, A. E. (1962). Factor analysis as a statistical method. *J. R. Stat. Soc., Ser. D, Stat.* **12** 209–229.
[35] LIU, L. T., DOBRIBAN, E. and SINGER, A. (2018). *e*PCA: High dimensional exponential family PCA. *Ann. Appl. Stat.* **12** 2121–2150. MR3875695 https://doi.org/10.1214/18-AOAS1146
[36] LOUNICI, K. (2014). High-dimensional covariance matrix estimation with missing observations. *Bernoulli* **20** 1029–1058. MR3217437 https://doi.org/10.3150/12-BEJ487
[37] LUO, Y., HAN, R. and ZHANG, A. R. (2020). A Schatten-q matrix perturbation theory via perturbation projection error bound. ArXiv preprint. Available at arXiv:2008.01312.
[38] MARTIN, A. D., QUINN, K. M. and PARK, J. H. (2011). MCMCpack: Markov chain Monte Carlo in R. *J. Stat. Softw.* **42**.
[39] MAZUMDER, R., HASTIE, T. and TIBSHIRANI, R. (2010). Spectral regularization algorithms for learning large incomplete matrices. *J. Mach. Learn. Res.* **11** 2287–2322. MR2719857
[40] MELAMED, D. (2014). Community structures in bipartite networks: A dual-projection approach. *PLoS ONE* **9** e97823. https://doi.org/10.1371/journal.pone.0097823

[41] MOHAMED, S., GHAHRAMANI, Z. and HELLER, K. A. (2009). Bayesian exponential family PCA. In *Advances in Neural Information Processing Systems* 1089–1096.

[42] NADAKUDITI, R. R. (2014). OptShrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage. *IEEE Trans. Inf. Theory* **60** 3002–3018. MR3200641 https://doi.org/10.1109/TIT.2014.2311661

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

[44] OWEN, A. B. and WANG, J. (2016). Bi-cross-validation for factor analysis. *Statist. Sci.* **31** 119–139. MR3458596 https://doi.org/10.1214/15-STS539

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

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

[47] ROBIN, G., JOSSE, J., MOULINES, É. and SARDY, S. (2019). Low-rank model with covariates for count data with missing values. *J. Multivariate Anal.* **173** 416–434. MR3946894 https://doi.org/10.1016/j.jmva.2019.04.004

[48] SALMON, J., HARMANY, Z., DELEDALLE, C.-A. and WILLETT, R. (2014). Poisson noise reduction with non-local PCA. *J. Math. Imaging Vision* **48** 279–294. MR3152105 https://doi.org/10.1007/s10851-013-0435-6

[49] THOMSON, G. (1939). The factorial analysis of human ability. *Br. J. Educ. Psychol.* **9** 188–195.

[50] TIPPING, M. E. and BISHOP, C. M. (1999). Probabilistic principal component analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **61** 611–622. MR1707864 https://doi.org/10.1111/1467-9868.00196

[51] VASWANI, N. and GUO, H. (2016). Correlated-PCA: Principal components' analysis when data and noise are correlated. In *Advances in Neural Information Processing Systems* 1768–1776.

[52] VASWANI, N. and NARAYANAMURTHY, P. (2017). Finite sample guarantees for PCA in non-isotropic and data-dependent noise. In 2017 55*th Annual Allerton Conference on Communication, Control, and Computing* (*Allerton*) 783–789. IEEE, Los Alamitos.

[53] VASWANI, N. and NARAYANAMURTHY, P. (2018). PCA in sparse data-dependent noise. In 2018 *IEEE International Symposium on Information Theory* (*ISIT*) 641–645. IEEE, Los Alamitos.

[54] VASWANI, N. and NARAYANAMURTHY, P. (2020). Fast robust subspace tracking via PCA in sparse data-dependent noise. *IEEE J. Sel. Areas Inf. Theory* **1** 723–744.

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

[56] WANG, W. and FAN, J. (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Ann. Statist.* **45** 1342–1374. MR3662457 https://doi.org/10.1214/16-AOS1487

[57] WEDIN, P. (1972). Perturbation bounds in connection with singular value decomposition. *BIT* **12** 99–111. MR0309968 https://doi.org/10.1007/bf01932678

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

[59] YU, Y., WANG, T. and SAMWORTH, R. J. (2015). A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika* **102** 315–323. MR3371006 https://doi.org/10.1093/biomet/asv008

[60] ZHANG, A., CAI, T. T. and WU, Y. (2022). Supplement to "Heteroskedastic PCA: Algorithm, optimality, and applications." https://doi.org/10.1214/21-AOS2074SUPP

[61] ZHANG, A. and XIA, D. (2018). Tensor SVD: Statistical and computational limits. *IEEE Trans. Inf. Theory* **64** 7311–7338. MR3876445 https://doi.org/10.1109/TIT.2018.2841377

[62] ZHOU, Z. and AMINI, A. A. (2020). Optimal bipartite network clustering. *J. Mach. Learn. Res.* **21** 40. MR4073773

The Annals of Statistics
2022, Vol. 50, No. 1, 81–106
https://doi.org/10.1214/21-AOS2086

ON MINIMAX OPTIMALITY OF SPARSE BAYES PREDICTIVE DENSITY ESTIMATES

BY GOURAB MUKHERJEE[1] AND IAIN M. JOHNSTONE[2]

[1]*Department of Data Sciences and Operations, University of Southern California, gourab@usc.edu*
[2]*Departments of Statistics and of Biomedical Data Science, Stanford University, imj@stanford.edu*

We study predictive density estimation under Kullback–Leibler loss in ℓ_0-sparse Gaussian sequence models. We propose proper Bayes predictive density estimates and establish asymptotic minimaxity in sparse models. Fundamental for this is a new risk decomposition for sparse, or spike-and-slab priors.

A surprise is the existence of a phase transition in the future-to-past variance ratio r. For $r < r_0 = (\sqrt{5} - 1)/4$, the natural discrete prior ceases to be asymptotically optimal. Instead, for subcritical r, a 'bi-grid' prior with a central region of reduced grid spacing recovers asymptotic minimaxity. This phenomenon seems to have no analog in the otherwise parallel theory of point estimation of a multivariate normal mean under quadratic loss.

For spike-and-uniform slab priors to have any prospect of minimaxity, we show that the sparse parameter space needs also to be magnitude constrained. Within a substantial range of magnitudes, such spike-and-slab priors can attain asymptotic minimaxity.

REFERENCES

[1] AITCHISON, J. and DUNSMORE, I. R. (1975). *Statistical Prediction Analysis*. Cambridge Univ. Press, Cambridge-Melbourne. MR0408097

[2] ASLAN, M. (2006). Asymptotically minimax Bayes predictive densities. *Ann. Statist.* **34** 2921–2938. MR2329473 https://doi.org/10.1214/009053606000000885

[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 https://doi.org/10.1080/01621459.2014.960967

[4] BROWN, L. D., GEORGE, E. I. and XU, X. (2008). Admissible predictive density estimation. *Ann. Statist.* **36** 1156–1170. MR2418653 https://doi.org/10.1214/07-AOS506

[5] CARVALHO, C. M., POLSON, N. G. and SCOTT, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika* **97** 465–480. MR2650751 https://doi.org/10.1093/biomet/asq017

[6] CASTILLO, I. and VAN DER VAART, A. (2012). Needles and straw in a haystack: Posterior concentration for possibly sparse sequences. *Ann. Statist.* **40** 2069–2101. MR3059077 https://doi.org/10.1214/12-AOS1029

[7] FOURDRINIER, D., MARCHAND, É., RIGHI, A. and STRAWDERMAN, W. E. (2011). On improved predictive density estimation with parametric constraints. *Electron. J. Stat.* **5** 172–191. MR2792550 https://doi.org/10.1214/11-EJS603

[8] GANGOPADHYAY, U. and MUKHERJEE, G. (2021). On discrete priors and sparse minimax optimal predictive densities. *Electron. J. Stat.* **15**. MR4255304 https://doi.org/10.1214/21-ejs1818

[9] GEISSER, S. (1993). *Predictive Inference: An Introduction. Monographs on Statistics and Applied Probability* **55**. CRC Press, New York. MR1252174 https://doi.org/10.1007/978-1-4899-4467-2

[10] GEORGE, E. I., LIANG, F. and XU, X. (2006). Improved minimax predictive densities under Kullback-Leibler loss. *Ann. Statist.* **34** 78–91. MR2275235 https://doi.org/10.1214/009053606000000155

[11] GEORGE, E. I., LIANG, F. and XU, X. (2012). From minimax shrinkage estimation to minimax shrinkage prediction. *Statist. Sci.* **27** 82–94. MR2953497 https://doi.org/10.1214/11-STS383

[12] GEORGE, E. I. and MCCULLOCH, R. E. (1997). Approaches for Bayesian variable selection. *Statist. Sinica* 339–373.

MSC2020 subject classifications. Primary 62C12; secondary 62C25, 62F10, 62J07.
Key words and phrases. Predictive density, asymptotic minimaxity, proper Bayes rule, sparsity, high dimensional, least favorable prior, spike and slab.

[13] GEORGE, E. I. and XU, X. (2008). Predictive density estimation for multiple regression. *Econometric Theory* **24** 528–544. MR2391619 https://doi.org/10.1017/S0266466608080213

[14] GHOSH, M., MERGEL, V. and DATTA, G. S. (2008). Estimation, prediction and the Stein phenomenon under divergence loss. *J. Multivariate Anal.* **99** 1941–1961. MR2466545 https://doi.org/10.1016/j.jmva.2008.02.002

[15] HARTIGAN, J. A. (1998). The maximum likelihood prior. *Ann. Statist.* **26** 2083–2103. MR1700222 https://doi.org/10.1214/aos/1024691462

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

[17] ISHWARAN, H. and RAO, J. S. (2005). Spike and slab gene selection for multigroup microarray data. *J. Amer. Statist. Assoc.* **100** 764–780. MR2201009 https://doi.org/10.1198/016214505000000051

[18] JOHNSTONE, I. M. (1994). On minimax estimation of a sparse normal mean vector. *Ann. Statist.* **22** 271–289. MR1272083 https://doi.org/10.1214/aos/1176325368

[19] JOHNSTONE, I. M. (2013). Gaussian estimation: Sequence and wavelet models. Available at https://imjohnstone.su.domains.

[20] 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. MR2089135 https://doi.org/10.1214/009053604000000030

[21] KOBAYASHI, K. and KOMAKI, F. (2008). Bayesian shrinkage prediction for the regression problem. *J. Multivariate Anal.* **99** 1888–1905. MR2466542 https://doi.org/10.1016/j.jmva.2008.01.014

[22] KOMAKI, F. (1996). On asymptotic properties of predictive distributions. *Biometrika* **83** 299–313. MR1439785 https://doi.org/10.1093/biomet/83.2.299

[23] KOMAKI, F. (2001). A shrinkage predictive distribution for multivariate normal observables. *Biometrika* **88** 859–864. MR1859415 https://doi.org/10.1093/biomet/88.3.859

[24] KOMAKI, F. (2006). Shrinkage priors for Bayesian prediction. *Ann. Statist.* **34** 808–819. MR2283393 https://doi.org/10.1214/009053606000000010

[25] KUBOKAWA, T., MARCHAND, É. and STRAWDERMAN, W. E. (2015). On predictive density estimation for location families under integrated squared error loss. *J. Multivariate Anal.* **142** 57–74. MR3412739 https://doi.org/10.1016/j.jmva.2015.07.013

[26] KUBOKAWA, T., MARCHAND, É. and STRAWDERMAN, W. E. (2017). On predictive density estimation for location families under integrated absolute error loss. *Bernoulli* **23** 3197–3212. MR3654804 https://doi.org/10.3150/16-BEJ842

[27] KUBOKAWA, T., MARCHAND, É., STRAWDERMAN, W. E. and TURCOTTE, J.-P. (2013). Minimaxity in predictive density estimation with parametric constraints. *J. Multivariate Anal.* **116** 382–397. MR3049911 https://doi.org/10.1016/j.jmva.2013.01.001

[28] MALLOWS, C. (1978). Minimizing an integral. *SIAM Rev.* **20** 183–183. https://doi.org/10.1137/1020016

[29] MARUYAMA, Y., MATSUDA, T. and OHNISHI, T. (2019). Harmonic Bayesian prediction under α-divergence. *IEEE Trans. Inf. Theory* **65** 5352–5366. MR4009238 https://doi.org/10.1109/TIT.2019.2915245

[30] MATSUDA, T. and KOMAKI, F. (2015). Singular value shrinkage priors for Bayesian prediction. *Biometrika* **102** 843–854. MR3431557 https://doi.org/10.1093/biomet/asv036

[31] MITCHELL, T. J. and BEAUCHAMP, J. J. (1988). Bayesian variable selection in linear regression. *J. Amer. Statist. Assoc.* **83** 1023–1036. MR0997578

[32] MUKHERJEE, G. (2013). *Sparsity and Shrinkage in Predictive Density Estimation*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)–Stanford University. MR4187552

[33] MUKHERJEE, G. (2021). Minimax adaptive predictive density estimation for nonparametric regression. In preparation.

[34] MUKHERJEE, G. and JOHNSTONE, I. M. (2015). Exact minimax estimation of the predictive density in sparse Gaussian models. *Ann. Statist.* **43** 937–961. MR3346693 https://doi.org/10.1214/14-AOS1251

[35] MUKHERJEE, G. and JOHNSTONE, I. M. (2022). Supplement to "On Minimax Optimality of Sparse Bayes Predictive Density Estimates." https://doi.org/10.1214/21-AOS2086SUPP

[36] O'HARA, R. B. and SILLANPÄÄ, M. J. (2009). A review of Bayesian variable selection methods: What, how and which. *Bayesian Anal.* **4** 85–117. MR2486240 https://doi.org/10.1214/09-BA403

[37] PARK, T. and CASELLA, G. (2008). The Bayesian lasso. *J. Amer. Statist. Assoc.* **103** 681–686. MR2524001 https://doi.org/10.1198/016214508000000337

[38] ROČKOVÁ, V. (2018). Bayesian estimation of sparse signals with a continuous spike-and-slab prior. *Ann. Statist.* **46** 401–437. MR3766957 https://doi.org/10.1214/17-AOS1554

[39] ROČKOVÁ, V. and GEORGE, E. I. (2014). Negotiating multicollinearity with spike-and-slab priors. *Metron* **72** 217–229. MR3233150 https://doi.org/10.1007/s40300-014-0047-y

[40] Ročková, V. and George, E. I. (2018). The spike-and-slab LASSO. *J. Amer. Statist. Assoc.* **113** 431–444. MR3803476 https://doi.org/10.1080/01621459.2016.1260469

[41] van Erven, T., Szabo, B. et al. (2021). Fast exact Bayesian inference for sparse signals in the normal sequence model. *Bayesian Anal.* **16** 933–960. MR4303874 https://doi.org/10.1214/20-BA1227

[42] Xu, X. and Liang, F. (2010). Asymptotic minimax risk of predictive density estimation for non-parametric regression. *Bernoulli* **16** 543–560. MR2668914 https://doi.org/10.3150/09-BEJ222

[43] Xu, X. and Zhou, D. (2011). Empirical Bayes predictive densities for high-dimensional normal models. *J. Multivariate Anal.* **102** 1417–1428. MR2819959 https://doi.org/10.1016/j.jmva.2011.05.008

[44] Yano, K. and Komaki, F. (2017). Information criteria for prediction when the distributions of current and future observations differ. *Statist. Sinica* **27** 1205–1223. MR3699701

The Annals of Statistics
2022, Vol. 50, No. 1, 107–128
https://doi.org/10.1214/21-AOS2091

DIMENSION REDUCTION FOR FUNCTIONAL DATA BASED ON WEAK CONDITIONAL MOMENTS

BY BING LI[1] AND JUN SONG[2]

[1]*Department of Statistics, The Pennsylvania State University, bxl9@psu.edu*
[2]*Department of Statistics, Korea University, junsong19@korea.ac.kr*

We develop a general theory and estimation methods for functional linear sufficient dimension reduction, where both the predictor and the response can be random functions, or even vectors of functions. Unlike the existing dimension reduction methods, our approach does not rely on the estimation of conditional mean and conditional variance. Instead, it is based on a new statistical construction—the weak conditional expectation, which is based on Carleman operators and their inducing functions. Weak conditional expectation is a generalization of conditional expectation. Its key advantage is to replace the projection on to an L_2-space—which defines conditional expectation—by projection on to an arbitrary Hilbert space, while still maintaining the unbiasedness of the related dimension reduction methods. This flexibility is particularly important for functional data, because attempting to estimate a full-fledged conditional mean or conditional variance by slicing or smoothing over the space of vector-valued functions may be inefficient due to the curse of dimensionality. We evaluated the performances of the our new methods by simulation and in several applied settings.

REFERENCES

[1] BILLINGSLEY, P. (1979). *Probability and Measure*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0534323

[2] BONACCORSI, S. and PRIOLA, E. (2006). From Brownian motion to stochastic differential equations. In 10*th Internet Seminar*.

[3] BOSQ, D. (2000). *Linear Processes in Function Spaces*: *Theory and Applications*. *Lecture Notes in Statistics* **149**. Springer, New York. MR1783138 https://doi.org/10.1007/978-1-4612-1154-9

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

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

[6] COOK, R. D. (1998). *Regression Graphics*: *Ideas for Studying Regressions Through Graphics*. *Wiley Series in Probability and Statistics*: *Probability and Statistics*. Wiley, New York. MR1645673 https://doi.org/10.1002/9780470316931

[7] COOK, R. D. and LI, B. (2002). Dimension reduction for conditional mean in regression. *Ann. Statist.* **30** 455–474. MR1902895 https://doi.org/10.1214/aos/1021379861

[8] FANAEE-T, H. and GAMA, J. (2014). Event labeling combining ensemble detectors and background knowledge. *Prog. Artif. Intell.* **2** 113–127.

[9] FERRATY, F., VAN KEILEGOM, I. and VIEU, P. (2012). Regression when both response and predictor are functions. *J. Multivariate Anal.* **109** 10–28. MR2922850 https://doi.org/10.1016/j.jmva.2012.02.008

[10] FERRÉ, L. and YAO, A.-F. (2005). Smoothed functional inverse regression. *Statist. Sinica* **15** 665–683. MR2233905

[11] FERRÉ, L. and YAO, A. F. (2003). Functional sliced inverse regression analysis. *Statistics* **37** 475–488. MR2022235 https://doi.org/10.1080/0233188031000112845

[12] FUKUMIZU, K., BACH, F. R. and GRETTON, A. (2007). Statistical consistency of kernel canonical correlation analysis. *J. Mach. Learn. Res.* **8** 361–383. MR2320675

MSC2020 subject classifications. Primary 62G08; secondary 62H99, 62B05.
Key words and phrases. Carleman operator, weak conditional moment, weak inverse regression estimate, weak average variance estimate, weak directional regression, function-on-function regression.

[13] FUNG, W. K., HE, X., LIU, L. and SHI, P. (2002). Dimension reduction based on canonical correlation. *Statist. Sinica* **12** 1093–1113. MR1947065

[14] HSING, T. and EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators. Wiley Series in Probability and Statistics*. Wiley, Chichester. MR3379106 https://doi.org/10.1002/9781118762547

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

[16] IVANESCU, A. E., STAICU, A.-M., SCHEIPL, F. and GREVEN, S. (2015). Penalized function-on-function regression. *Comput. Statist.* **30** 539–568. MR3357075 https://doi.org/10.1007/s00180-014-0548-4

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

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

[19] LEE, K.-Y., LI, B. and ZHAO, H. (2016). Variable selection via additive conditional independence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 1037–1055. MR3557188 https://doi.org/10.1111/rssb.12150

[20] LI, B. (2018). *Sufficient Dimension Reduction: Methods and Applications with R. Monographs on Statistics and Applied Probability* **161**. CRC Press, Boca Raton, FL. MR3838449 https://doi.org/10.1201/9781315119427

[21] LI, B. (2018). Linear operator-based statistical analysis: A useful paradigm for big data. *Canad. J. Statist.* **46** 79–103. MR3767167 https://doi.org/10.1002/cjs.11329

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

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

[24] LI, B. and SONG, J. (2022). Supplement to "Dimension reduction for functional data based on weak conditional moments." https://doi.org/10.1214/21-AOS2091SUPP

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

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

[27] COOK, R. D. and WEISBERG, S. (1991). Sliced inverse regression for dimension reduction: Comment. **86** 328–332.

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

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

[30] LIAN, H. and LI, G. (2014). Series expansion for functional sufficient dimension reduction. *J. Multivariate Anal.* **124** 150–165. MR3147317 https://doi.org/10.1016/j.jmva.2013.10.019

[31] MÜLLER, H.-G. and STADTMÜLLER, U. (2005). Generalized functional linear models. *Ann. Statist.* **33** 774–805. MR2163159 https://doi.org/10.1214/009053604000001156

[32] QI, X. and LUO, R. (2019). Nonlinear function-on-function additive model with multiple predictor curves. *Statist. Sinica* **29** 719–739. MR3931385

[33] RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2168993

[34] WANG, G., LIN, N. and ZHANG, B. (2013). Functional contour regression. *J. Multivariate Anal.* **116** 1–13. MR3049886 https://doi.org/10.1016/j.jmva.2012.11.005

[35] WANG, G., ZHOU, Y., FENG, X.-N. and ZHANG, B. (2015). The hybrid method of FSIR and FSAVE for functional effective dimension reduction. *Comput. Statist. Data Anal.* **91** 64–77. MR3368006 https://doi.org/10.1016/j.csda.2015.05.011

[36] WEIDMANN, J. (1980). *Linear Operators in Hilbert Spaces. Graduate Texts in Mathematics* **68**. Springer, New York. MR0566954

[37] 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 https://doi.org/10.1198/016214504000001745

[38] YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33** 2873–2903. MR2253106 https://doi.org/10.1214/009053605000000660

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

The Annals of Statistics
2022, Vol. 50, No. 1, 129–146
https://doi.org/10.1214/21-AOS2094

PATTERN GRAPHS: A GRAPHICAL APPROACH TO NONMONOTONE MISSING DATA

BY YEN-CHI CHEN

Department of Statistics, University of Washington, yenchic@uw.edu

We introduce the concept of pattern graphs–directed acyclic graphs representing how response patterns are associated. A pattern graph represents an identifying restriction that is nonparametrically identified/saturated and is often a missing not at random restriction. We introduce a selection model and a pattern mixture model formulations using the pattern graphs and show that they are equivalent. A pattern graph leads to an inverse probability weighting estimator as well as an imputation-based estimator. We also study the semiparametric efficiency theory and derive a multiply-robust estimator using pattern graphs.

REFERENCES

BHATTACHARYA, R., MALINSKY, D. and SHPITSER, I. (2020). Causal inference under interference and network uncertainty. In *Uncertainty in Artificial Intelligence* 1028–1038. PMLR.

CHEN, Y.-C. (2022). Supplement to "Pattern graphs: A graphical approach to nonmonotone missing data." https://doi.org/10.1214/21-AOS2094SUPP

CHEN, Y.-C. and SADINLE, M. (2019). Nonparametric pattern-mixture models for inference with missing data. arXiv preprint. Available at arXiv:1904.11085.

DANIELS, M. J. and HOGAN, J. W. (2008). *Missing Data in Longitudinal Studies*: *Strategies for Bayesian Modeling and Sensitivity Analysis*. *Monographs on Statistics and Applied Probability* **109**. CRC Press/CRC, Boca Raton, FL. MR2459796 https://doi.org/10.1201/9781420011180

DIGGLE, P. J., HEAGERTY, P. J., LIANG, K.-Y. and ZEGER, S. L. (2002). *Analysis of Longitudinal Data*, 2nd ed. *Oxford Statistical Science Series* **25**. Oxford Univ. Press, Oxford. MR2049007

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

EFRON, B. and TIBSHIRANI, R. J. (1994). *An Introduction to the Bootstrap*. CRC press, Boca Raton.

FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2001). *The Elements of Statistical Learning* **1**. Springer series in statistics, New York.

GILL, R. D., VAN DER LAAN, M. J. and ROBINS, J. M. (1997). Coarsening at random: Characterizations, conjectures, counter-examples. In *Proceedings of the First Seattle Symposium in Biostatistics*: *Survival Analysis* 255–294.

HALL, P. (2013). *The Bootstrap and Edgeworth Expansion*. Springer, Berlin.

HOETING, J. A., MADIGAN, D., RAFTERY, A. E. and VOLINSKY, C. T. (1999). Bayesian model averaging: A tutorial. *Statist*. *Sci*. **14** 382–417. MR1765176 https://doi.org/10.1214/ss/1009212519

HOONHOUT, P. and RIDDER, G. (2019). Nonignorable attrition in multi-period panels with refreshment samples. *J*. *Bus*. *Econom*. *Statist*. **37** 377–390. MR3968379 https://doi.org/10.1080/07350015.2017.1345744

HOROWITZ, J. L. and MANSKI, C. F. (2000). Nonparametric analysis of randomized experiments with missing covariate and outcome data. *J*. *Amer*. *Statist*. *Assoc*. **95** 77–88. MR1803142 https://doi.org/10.2307/2669526

LINERO, A. R. (2017). Bayesian nonparametric analysis of longitudinal studies in the presence of informative missingness. *Biometrika* **104** 327–341. MR3698257 https://doi.org/10.1093/biomet/asx015

LITTLE, R. J. (1993a). Pattern-mixture models for multivariate incomplete data. *J*. *Amer*. *Statist*. *Assoc*. **88** 125–134.

LITTLE, R. J. A. (1993b). Pattern-mixture models for multivariate incomplete data. *J*. *Amer*. *Statist*. *Assoc*. **88** 125–134.

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

MSC2020 subject classifications. Primary 62F30; secondary 62H05, 65D18.

Key words and phrases. Missing data, nonignorable missingness, nomonotone missing, inverse probability weighting, pattern graphs, selection models.

MALINSKY, D., SHPITSER, I. and TCHETGEN, E. J. T. (2019). Semiparametric inference for non-monotone missing-not-at-random data: The no self-censoring model. arXiv preprint. Available at arXiv:1909.01848.

MANSKI, C. F. (1990). Nonparametric bounds on treatment effects. *Am. Econ. Rev.* **80** 319–323.

MOHAN, K. and PEARL, J. (2014). Graphical models for recovering probabilistic and causal queries from missing data. In *Advances in Neural Information Processing Systems* 1520–1528.

MOHAN, K. and PEARL, J. (2021). Graphical models for processing missing data. *J. Amer. Statist. Assoc.* **116** 1023–1037. MR4270041 https://doi.org/10.1080/01621459.2021.1874961

MOHAN, K., PEARL, J. and TIAN, J. (2013). Graphical models for inference with missing data. In *Advances in Neural Information Processing Systems* 1277–1285.

MOLENBERGHS, G., MICHIELS, B., KENWARD, M. G. and DIGGLE, P. J. (1998). Monotone missing data and pattern-mixture models. *Stat. Neerl.* **52** 153–161. MR1649081 https://doi.org/10.1111/1467-9574.00075

MOLENBERGHS, G., FITZMAURICE, G., KENWARD, M. G., TSIATIS, A. and VERBEKE, G. (2014). *Handbook of Missing Data Methodology*. CRC Press/CRC, Boca Raton.

NABI, R., BHATTACHARYA, R. and SHPITSER, I. (2020). Full law identification in graphical models of missing data: Completeness results. arXiv preprint. Available at arXiv:2004.04872.

ROBINS, J. M. (1997). Non-response models for the analysis of non-monotone non-ignorable missing data. *Stat. Med.* **16** 21–37.

ROBINS, J. M. and GILL, R. D. (1997). Non-response models for the analysis of non-monotone ignorable missing data. *Stat. Med.* **16** 39–56.

ROBINS, J. M., ROTNITZKY, A. and SCHARFSTEIN, D. O. (2000). Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials* (*Minneapolis, MN*, 1997). *IMA Vol. Math. Appl.* **116** 1–94. Springer, New York. MR1731681 https://doi.org/10.1007/978-1-4612-1284-3_1

RUBIN, D. B. (2004). *Multiple Imputation for Nonresponse in Surveys*. *Wiley Classics Library*. Wiley Interscience, Hoboken, NJ. MR2117498

SADINLE, M. and REITER, J. P. (2017). Itemwise conditionally independent nonresponse modelling for incomplete multivariate data. *Biometrika* **104** 207–220. MR3626477 https://doi.org/10.1093/biomet/asw063

SEAMAN, S. R. and VANSTEELANDT, S. (2018). Introduction to double robust methods for incomplete data. *Statist. Sci.* **33** 184–197. MR3797709 https://doi.org/10.1214/18-STS647

SHPITSER, I. (2016). Consistent estimation of functions of data missing non-monotonically and not at random. In *Advances in Neural Information Processing Systems* 3144–3152.

SHPITSER, I., MOHAN, K. and PEARL, J. (2015). Missing data as a causal and probabilistic problem Technical report, California Univ. Los Angeles Dept. of Computer Science.

SUN, B. and TCHETGEN TCHETGEN, E. J. (2018). On inverse probability weighting for nonmonotone missing at random data. *J. Amer. Statist. Assoc.* **113** 369–379. MR3803471 https://doi.org/10.1080/01621459.2016.1256814

TCHETGEN TCHETGEN, E. J., WANG, L. and SUN, B. (2018). Discrete choice models for nonmonotone nonignorable missing data: Identification and inference. *Statist. Sinica* **28** 2069–2088. MR3839004

THIJS, H., MOLENBERGHS, G., MICHIELS, B., VERBEKE, G. and CURRAN (2002). Strategies to fit pattern-mixture models. *Biostatistics* **3** 245–265.

TIAN, J. (2015). Missing at random in graphical models. In *Artificial Intelligence and Statistics* 977–985.

TSIATIS, A. (2007). *Semiparametric Theory and Missing Data*. Springer, Berlin.

VANSTEELANDT, S., GOETGHEBEUR, E., KENWARD, M. G. and MOLENBERGHS, G. (2006). Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statist. Sinica* **16** 953–979. MR2281311

The Annals of Statistics
2022, Vol. 50, No. 1, 147–169
https://doi.org/10.1214/21-AOS2099

TESTING COMMUNITY STRUCTURE FOR HYPERGRAPHS

BY MINGAO YUAN[1], RUIQI LIU[2], YANG FENG[3] AND ZUOFENG SHANG[4]

[1]*Department of Statistics, North Dakota State University, mingao.yuan@ndsu.edu*
[2]*Department of Mathematical Sciences, Texas Tech University, ruiqliu@ttu.edu*
[3]*Department of Biostatistics, New York University, yang.feng@nyu.edu*
[4]*Department of Mathematical Sciences, New Jersey Institute of Technology, zshang@njit.edu*

Many complex networks in the real world can be formulated as hypergraphs where community detection has been widely used. However, the fundamental question of whether communities exist or not in an observed hypergraph remains unclear. This work aims to tackle this important problem. Specifically, we systematically study when a hypergraph with community structure can be successfully distinguished from its Erdős–Rényi counterpart, and propose concrete test statistics when the models are distinguishable. The main contribution of this paper is threefold. First, we discover a phase transition in the hyperedge probability for distinguishability. Second, in the bounded-degree regime, we derive a sharp signal-to-noise ratio (SNR) threshold for distinguishability in the special two-community 3-uniform hypergraphs, and derive nearly tight SNR thresholds in the general two-community m-uniform hypergraphs. Third, in the dense regime, we propose a computationally feasible test based on sub-hypergraph counts, obtain its asymptotic distribution, and analyze its power. Our results are further extended to nonuniform hypergraphs in which a new test involving both edge and hyperedge information is proposed. The proofs rely on Janson's contiguity theory (*Combin. Probab. Comput.* **4** (1995) 369–405), a high-moments driven asymptotic normality result by Gao and Wormald (*Probab. Theory Related Fields* **130** (2004) 368–376), and a truncation technique for analyzing the likelihood ratio.

REFERENCES

[1] ABBE, E. (2017). Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** 177. MR3827065
[2] ABBE, E. and SANDON, C. (2016). Detection in the stochastic block model with multiple clusters: Proof of the achievability conjectures, acyclic BP, and the information-computation gap. https://arxiv.org/pdf/1512.09080.pdf.
[3] ABBE, E. and SANDON, C. (2018). Proof of the achievability conjectures for the general stochastic block model. *Comm. Pure Appl. Math.* **71** 1334–1406. MR3812075 https://doi.org/10.1002/cpa.21719
[4] AGARWAL, S., BRANSON, K. and BELONGIE, S. (2006). Higher order learning with graphs. In *Proceedings of the International Conference on Machine Learning* 17–24.
[5] AHN, K., LEE, K. and SUH, C. (2016). Community recovery in hypergraphs. In 54*th Annual Allerton Conference on Communication, Control, and Computing* (*Allerton*). https://doi.org/10.1109/ALLERTON.2016.7852294
[6] AHN, K., LEE, K. and SUH, C. (2018). Hypergraph spectral clustering in the weighted stochastic block model. *IEEE J. Sel. Top. Signal Process.* **12**.
[7] 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 https://doi.org/10.1214/13-AOS1138
[8] ANGELINI, M., CALTAGIRONE, F., KRZAKALA, F. and ZDEBOROVA, L. (2015). Spectral detection on sparse hypergraphs. In *Allerton Conference on Communication, Control, and Computing* 66–73.
[9] BANERJEE, D. (2018). Contiguity and non-reconstruction results for planted partition models: The dense case. *Electron. J. Probab.* **23** 18. MR3771755 https://doi.org/10.1214/17-EJP128

MSC2020 subject classifications. Primary 62G10; secondary 05C80.
Key words and phrases. Hypergraph, stochastic block model, hypothesis testing, contiguity, l-cycle.

[10] BANERJEE, D. and MA, Z. (2017). Optimal hypothesis testing for stochastic block models with growing degrees. Available at https://arxiv.org/pdf/1705.05305.pdf.

[11] BANKDS, J., MOORE, C., NEEMAN, J. and NETRAPALLI, P. (2016). Information-theoretic thresholds for community detection in sparse networks. *J. Mach. Learn. Res. Workshop Conf. Proc.* **49** 1–34.

[12] BICKEL, P. J. and SARKAR, P. (2016). Hypothesis testing for automated community detection in networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 253–273. MR3453655 https://doi.org/10.1111/rssb.12117

[13] BOLLA, M. (1993). Spectra, Euclidean representations and clusterings of hypergraphs. *Discrete Math.* **117** 19–39. MR1226129 https://doi.org/10.1016/0012-365X(93)90322-K

[14] BOLLOBÁS, B. (2001). *Random Graphs*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **73**. Cambridge Univ. Press, Cambridge. MR1864966 https://doi.org/10.1017/CBO9780511814068

[15] BOLLOBÁS, B. and ERDŐS, P. (1976). Cliques in random graphs. *Math. Proc. Cambridge Philos. Soc.* **80** 419–427. MR0498256 https://doi.org/10.1017/S0305004100053056

[16] CHEN, J. and YUAN, B. (2006). Detecting functional modules in the yeast proteinprotein interaction network. *Bioinformatics* **22** 2283–2290.

[17] CHERTOK, M. and KELLER, Y. (2010). Efficient high order matching. *IEEE Trans. Pattern Anal. Mach. Intell.* **32** 2205–2215.

[18] CHIEN, I., LIN, C. and WANG, I. (2018). Community detection in hypergraphs: Optimal statistical limit and efficient algorithms. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics* **84** 871–879.

[19] DALL'AMICO, L. and COUILLET, R. (2019). Community detection in sparse reallistic graphs: Improving the Bethe Hessian. In *ICASSP* 2019 18778248.

[20] DALL'AMICO, L., COUILLET, R. and TREMBLAY, N. (2019). Revisiting the Bethe-Hessian: Improved community detection in sparse heterogeneous graphs. In *NIPS* 2019.

[21] ERDŐS, P. and RÉNYI, A. (1960). On the evolution of random graphs. *Magy. Tud. Akad. Mat. Kut. Intéz. Közl.* **5** 17–61. MR0125031

[22] ESTRADA, E. and RODRIGUEZ-VELASQUEZ, J. (2005). Complex networks as hypergraphs. Available at https://arxiv.org/ftp/physics/papers/0505/0505137.pdf.

[23] FLORESCU, L. and PERKINS, W. (2016). Spectral thresholds in the bipartite stochastic block model. In 29*th Annual Conference on Learning Theory* **49** 943–959.

[24] FORTUNATO, S. (2010). Community detection in graphs. *Phys. Rep.* **486** 75–174. MR2580414 https://doi.org/10.1016/j.physrep.2009.11.002

[25] FRIEZE, A. and KARONSKI, M. (2015). *Introduction to Random Graphs*. Cambridge Univ. Press, Cambridge.

[26] GAO, C. and LAFFERTY, J. (2017). Testing for global network structure using small subgraph statistics. Available at https://arxiv.org/pdf/1710.00862.pdf.

[27] GAO, C. and LAFFERTY, J. (2017). Testing network structure using relations between small subgraph probabilities. Available at https://arxiv.org/pdf/1704.06742.pdf.

[28] GAO, C., MA, Z., ZHANG, A. Y. and ZHOU, H. H. (2018). Community detection in degree-corrected block models. *Ann. Statist.* **46** 2153–2185. MR3845014 https://doi.org/10.1214/17-AOS1615

[29] GAO, Z. and WORMALD, N. C. (2004). Asymptotic normality determined by high moments, and submap counts of random maps. *Probab. Theory Related Fields* **130** 368–376. MR2095934 https://doi.org/10.1007/s00440-004-0356-9

[30] GHOSHAL, G., ZLATIĆ, V., CALDARELLI, G. and NEWMAN, M. E. J. (2009). Random hypergraphs and their applications. *Phys. Rev. E* **79** 066118. MR2551286 https://doi.org/10.1103/PhysRevE.79.066118

[31] GHOSHDASTIDAR, D. and DUKKIPATI, A. (2014). Consistency of spectral partitioning of uniform hypergraphs under planted partition model. *Adv. Neural Inf. Process. Syst.* 397–405.

[32] GHOSHDASTIDAR, D. and DUKKIPATI, A. (2017). Consistency of spectral hypergraph partitioning under planted partition model. *Ann. Statist.* **45** 289–315. MR3611493 https://doi.org/10.1214/16-AOS1453

[33] GIBSON, D., KLEINBERG, J. and RAGHAVAN, P. (2000). Clustering categorical data: An approach based on dynamical systems. *VLDB J.* **8** 222–236.

[34] GOLDENBERG, A., ZHENG, A. X. S., FIENBERG, E. and AIROLDI, E. M. (2010). A survey of statistical network models. *Found. Trends Mach. Learn.* **2** 129–233.

[35] GOVINDU, V. M. (2005). A tensor decomposition for geometric grouping and segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*) 1150–1157.

[36] JANSON, S. (1995). Random regular graphs: Asymptotic distributions and contiguity. *Combin. Probab. Comput.* **4** 369–405. MR1377557 https://doi.org/10.1017/S0963548300001735

[37] KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* **83** 016107. MR2788206 https://doi.org/10.1103/PhysRevE.83.016107

[38] KE, Z., SHI, F. and XIA, D. (2020). Community Detection for Hypergraph Networks via Regularized Tensor Power Iteration. Available at https://arxiv.org/pdf/1909.06503.pdf.

[39] KIM, C., BANDEIRA, A. and GOEMANS, M. (2017). Community detection in hypergraphs, spiked tensor models, and sum-of-squares. In 2017 *International Conference on Sampling Theory and Applications* (*SampTA*) 124–128.

[40] LEI, J. (2016). A goodness-of-fit test for stochastic block models. *Ann*. *Statist*. **44** 401–424. MR3449773 https://doi.org/10.1214/15-AOS1370

[41] LIN, C., CHIEN, I. and WANG, I. (2017). On the fundamental statistical limit of community detection in random hypergraphs. In *Information Theory* (*ISIT*), 2017 *IEEE International Symposium* 2178–2182.

[42] MICHOEL, T. and NACHTERGAELE, B. (2012). Alignment and integration of complex networks by hypergraph-based spectral clustering. *Phys*. *Rev. E* **86**.

[43] MONTANARI, A. and SEN, S. (2016). Semidefinite programs on sparse random graphs and their application to community detection. In *STOC'*16*—Proceedings of the* 48*th Annual ACM SIGACT Symposium on Theory of Computing* 814–827. ACM, New York. MR3536616 https://doi.org/10.1145/2897518.2897548

[44] MOSSEL, E., NEEMAN, J. and SLY, A. (2015). Reconstruction and estimation in the planted partition model. *Probab. Theory Related Fields* **162** 431–461. MR3383334 https://doi.org/10.1007/s00440-014-0576-6

[45] MOSSEL, E., NEEMAN, J. and SLY, A. (2018). A proof of the block model threshold conjecture. *Combinatorica* **38** 665–708. MR3876880 https://doi.org/10.1007/s00493-016-3238-8

[46] NEEMAN, J. and NETRAPALLI, P. (2014). Non-reconstructability in the stochastic block model. Available at https://arxiv.org/abs/1404.6304.

[47] NEWMAN, M. (2001). Scientific collaboration networks. I. Network construction and fundamental results. *Phys*. *Rev. E* **64** 016–131.

[48] OUVRARD, X., GOFF, J. and MARCHAND-MAILLET, S. (2017). Networks of Collaborations: Hypergraph modeling and visualisation. Available at https://arxiv.org/pdf/1707.00115.pdf.

[49] RAMASCO, J., DOROGOVTSEV, S. N. and PASTOR-SATORRAS, R. (2004). Self-organization of collaboration networks. *Phys*. *Rev. E* **70** 036–106.

[50] RODRÍGUEZ, J. A. (2009). Laplacian eigenvalues and partition problems in hypergraphs. *Appl*. *Math*. *Lett*. **22** 916–921. MR2523606 https://doi.org/10.1016/j.aml.2008.07.020

[51] ROTA BULO, S. and PELILLO, M. (2013). A game-theoretic approach to hypergraph clustering. *IEEE Trans*. *Pattern Anal*. *Mach*. *Intell*. **35** 1312–1327.

[52] SHI, J. and MALIK, J. (1997). Normalized cuts and image segmentation. *IEEE Trans*. *Pattern Anal*. *Mach*. *Intell*. **22** 888–905.

[53] TOKUDA, T. (2018). Statistical test for detecting community structure in real-valued edge-weighted graphs. *PLoS ONE* **13** e0194079. https://doi.org/10.1371/journal.pone.0194079

[54] WORMALD, N. C. (1999). Models of random regular graphs. In *Surveys in Combinatorics*, 1999 (*Canterbury*). *London Mathematical Society Lecture Note Series* **267** 239–298. Cambridge Univ. Press, Cambridge. MR1725006

[55] YUAN, M., FENG, Y. and SHANG, Z. (2018). A likelihood-ratio type test for stochastic block models with bounded degrees. Available at https://arxiv.org/pdf/1807.04426.pdf.

[56] YUAN, M., FENG, Y. and SHANG, Z. (2018). Inference on multi-community stochastic block models with bounded degree. Manuscript.

[57] YUAN, M., LIU, R., FENG, Y. and SHANG, Z. (2022). Supplement to "Testing community structure for hypergraphs." https://doi.org/10.1214/21-AOS2099SUPP

[58] ZHAO, Y., LEVINA, E. and ZHU, J. (2011). Community extraction for social networks. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **108** 7321–7326.

[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 https://doi.org/10.1214/12-AOS1036

The Annals of Statistics
2022, Vol. 50, No. 1, 170–196
https://doi.org/10.1214/21-AOS2100

FUNDAMENTAL BARRIERS TO HIGH-DIMENSIONAL REGRESSION WITH CONVEX PENALTIES

BY MICHAEL CELENTANO[*] AND ANDREA MONTANARI[†]

Department of Statistics, Stanford University, [*]*mcelen@stanford.edu;* [†]*montanar@stanford.edu*

In high-dimensional regression, we attempt to estimate a parameter vector $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ from $n \lesssim p$ observations $\{(y_i, \boldsymbol{x}_i)\}_{i \le n}$, where $\boldsymbol{x}_i \in \mathbb{R}^p$ is a vector of predictors and y_i is a response variable. A well-established approach uses convex regularizers to promote specific structures (e.g., sparsity) of the estimate $\widehat{\boldsymbol{\beta}}$ while allowing for practical algorithms. Theoretical analysis implies that convex penalization schemes have nearly optimal estimation properties in certain settings. However, in general the gaps between statistically optimal estimation (with unbounded computational resources) and convex methods are poorly understood.

We show that when the statistican has very simple structural information about the distribution of the entries of $\boldsymbol{\beta}_0$, a large gap frequently exists between the best performance achieved by *any convex regularizer* satisfying a mild technical condition and either: *(i)* the optimal statistical error or *(ii)* the statistical error achieved by optimal approximate message passing algorithms. Remarkably, a gap occurs at high enough signal-to-noise ratio if and only if the distribution of the coordinates of $\boldsymbol{\beta}_0$ is not log-concave. These conclusions follow from an analysis of standard Gaussian designs. Our lower bounds are expected to be generally tight, and we prove tightness under certain conditions.

REFERENCES

[1] ADVANI, M. and GANGULI, S. (2016). Statistical mechanics of optimal convex inference in high dimensions. *Phys. Rev. X* **6** 031034.

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

[3] BA, K. D., INDYK, P., PRICE, E. and WOODRUFF, D. P. (2010). Lower bounds for sparse recovery. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms* 1190–1197. SIAM, Philadelphia, PA. MR2809736

[4] BANDEIRA, A. S., PERRY, A. and WEIN, A. S. (2018). Notes on computational-to-statistical gaps: Predictions using statistical physics. *Port. Math.* **75** 159–186. MR3892753 https://doi.org/10.4171/PM/2014

[5] BANKS, J., MOHANTY, S. and RAGHAVENDRA, P. (2021). Local statistics, semidefinite programming, and community detection. In *Proceedings of the* 2021 *ACM-SIAM Symposium on Discrete Algorithms* (*SODA*) 1298–1316. SIAM, Philadelphia, PA. MR4262512 https://doi.org/10.1137/1.9781611976465.79

[6] BARBIER, J., DIA, M., MACRIS, N. and KRZAKALA, F. (2016). The mutual information in random linear estimation. In 2016 54*th Annual Allerton Conference on Communication*, *Control*, *and Computing* (*Allerton*) 625–632.

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

[8] BARBIER, J., MACRIS, N., DIA, M. and KRZAKALA, F. (2020). Mutual information and optimality of approximate message-passing in random linear estimation. *IEEE Trans. Inf. Theory* **66** 4270–4303. MR4130617 https://doi.org/10.1109/TIT.2020.2990880

MSC2020 subject classifications. Primary 62J05, 62J07; secondary 62F12.

Key words and phrases. High-dimensional regression, M-estimation, computational to statistical gaps, approximate message passing, convex, penalty.

[9] BAYATI, M., LELARGE, M. and MONTANARI, A. (2015). Universality in polytope phase transitions and message passing algorithms. *Ann. Appl. Probab.* **25** 753–822. MR3313755 https://doi.org/10.1214/14-AAP1010

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

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

[12] 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.

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

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

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

[16] BOGDAN, M., VAN DEN BERG, E., SABATTI, C., SU, W. and CANDÈS, E. J. (2015). SLOPE—adaptive variable selection via convex optimization. *Ann. Appl. Stat.* **9** 1103–1140. MR3418717 https://doi.org/10.1214/15-AOAS842

[17] BOLTHAUSEN, E. (2014). An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model. *Comm. Math. Phys.* **325** 333–366. MR3147441 https://doi.org/10.1007/s00220-013-1862-3

[18] BONDELL, H. D. and REICH, B. J. (2008). Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with OSCAR. *Biometrics* **64** 115–123. MR2422825 https://doi.org/10.1111/j.1541-0420.2007.00843.x

[19] BU, Z., KLUSOWSKI, J. M., RUSH, C. and SU, W. J. (2021). Algorithmic analysis and statistical estimation of SLOPE via approximate message passing. *IEEE Trans. Inf. Theory* **67** 506–537. MR4231969 https://doi.org/10.1109/TIT.2020.3025272

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

[21] CANDES, E. J. and TAO, T. (2005). Decoding by linear programming. *IEEE Trans. Inf. Theory* **51** 4203–4215. MR2243152 https://doi.org/10.1109/TIT.2005.858979

[22] CELENTANO, M. and MONTANARI, A. (2022). Supplement to "Fundamental barriers to high-dimensional regression with convex penalties." https://doi.org/10.1214/21-AOS2100SUPP

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

[24] CHEN, S. S. and DONOHO, D. (1995). Examples of basis pursuit. In *Proceedings of Wavelet Applications in Signal and Image Processing III*, San Diego, CA.

[25] DONOHO, D., MALEKI, A. and MONTANARI, A. (2009). Message-passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci. USA* **106** 18914–18919. https://doi.org/10.1073/pnas.0909892106

[26] DONOHO, D., MALEKI, A. and MONTANARI, A. (2010). Message passing algorithms for compressed sensing: I. Motivation and construction. In 2010 *IEEE Information Theory Workshop on Information Theory* (*ITW* 2010, *Cairo*) 1–5. IEEE.

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

[28] EL KAROUI, N. (2013). Asymptotic behavior of unregularized and ridge-regularized high-dimensional robust regression estimators: Rigorous results. Available at arXiv:1311.2445.

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

[30] EL KAROUI, N., BEAN, D., BICKEL, P. J., LIM, C. and YU, B. (2013). On robust regression with high-dimensional predictors. *Proc. Natl. Acad. Sci. USA* **110** 14557–14562.

[31] FIGUEIREDO, M. A. T. and NOWAK, R. D. (2014). Sparse estimation with strongly correlated variables using ordered weighted L1 regularization. Available at arXiv:1409.4005.

[32] GAMARNIK, D. and ILIAS, Z. High dimensional regression with binary coefficients. Estimating squared error and a phase transtition. In *Proceedings of the* 2017 *Conference on Learning Theory* (S. Kale and O. Shamir, eds.). *Proceedings of Machine Learning Research* **65** 948–953.

[33] GAMARNIK, D. and ZADIK, I. (2017). Sparse high-dimensional linear regression. Algorithmic barriers and a local search algorithm. Available at arXiv:1711.04952.

[34] GUO, D., WU, Y., SHAMAI, S. and VERDÚ, S. (2011). Estimation in Gaussian noise: Properties of the minimum mean-square error. *IEEE Trans. Inf. Theory* **57** 2371–2385. MR2809096 https://doi.org/10.1109/TIT.2011.2111010

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

[36] KORADA, S. B. and MONTANARI, A. (2011). Applications of the Lindeberg principle in communications and statistical learning. *IEEE Trans. Inf. Theory* **57** 2440–2450. MR2809100 https://doi.org/10.1109/TIT.2011.2112231

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

[38] MA, J., XU, J. and MALEKI, A. (2019). Optimization-based AMP for phase retrieval: The impact of initialization and ℓ_2 regularization. *IEEE Trans. Inf. Theory* **65** 3600–3629. MR3959008 https://doi.org/10.1109/TIT.2019.2893254

[39] MÉZARD, M. and MONTANARI, A. (2009). *Information, Physics, and Computation*. *Oxford Graduate Texts*. Oxford Univ. Press, Oxford. MR2518205 https://doi.org/10.1093/acprof:oso/9780198570837.001.0001

[40] MIGNACCO, F., KRZAKALA, F., LU, Y., URBANI, P. and ZDEBOROVÁ, L. The role of regularization in classification of high-dimensional noisy Gaussian mixture. In *Proceedings of the* 37*th International Conference on Machine Learning* (H. Daumé, III and A. Singh, eds.). *Proceedings of Machine Learning Research* **119** 6874–6883.

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

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

[43] PARIKH, N. and BOYD, S. (2013). Proximal algorithms. *Found. Trends Optim.* **1** 123–231.

[44] RANGAN, S. (2011). Generalized approximate message passing for estimation with random linear mixing. In *Information Theory Proceedings* (*ISIT*), 2011 *IEEE International Symposium on* 2168–2172. IEEE.

[45] RANGAN, S., SCHNITER, P. and FLETCHER, A. K. (2014). On the convergence of approximate message passing with arbitrary matrices. In *Information Theory Proceedings* (*ISIT*), 2014 *IEEE International Symposium on* 236–240. IEEE.

[46] RANGAN, S., SCHNITER, P. and FLETCHER, A. K. (2019). Vector approximate message passing. In *Information Theory Proceedings* (*ISIT*), 2017 *IEEE International Symposium on* 1588–1592. IEEE.

[47] REEVES, G. and PFISTER, H. D. (2016). The replica-symmetric prediction for compressed sensing with Gaussian matrices is exact. In *Information Theory* (*ISIT*), 2016 *IEEE International Symposium on* 665–669. IEEE.

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

[49] SAUMARD, A. and WELLNER, J. A. (2014). Log-concavity and strong log-concavity: A review. *Stat. Surv.* **8** 45–114. MR3290441 https://doi.org/10.1214/14-SS107

[50] SCHNITER, P. and RANGAN, S. (2015). Compressive phase retrieval via generalized approximate message passing. *IEEE Trans. Signal Process.* **63** 1043–1055. MR3311635 https://doi.org/10.1109/TSP.2014.2386294

[51] STOJNIC, M. (2010). Recovery thresholds for ℓ_1 optimization in binary compressed sensing. In *Information Theory Proceedings* (*ISIT*), 2010 *IEEE International Symposium on* 1593–1597. IEEE.

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

[53] SU, W. and CANDÈS, E. (2016). SLOPE is adaptive to unknown sparsity and asymptotically minimax. *Ann. Statist.* **44** 1038–1068. MR3485953 https://doi.org/10.1214/15-AOS1397

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

[55] TAHERI, H., PEDARSANI, R. and THRAMPOULIDIS, C. Sharp asymptotics and optimal performance for inference in binary models. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics* (S. Chiappa and R. Calandra, eds.). *Proceedings of Machine Learning Research* **108** 3739–3749.

[56] TAHERI, H., PEDARSANI, R. and THRAMPOULIDIS, C. Fundamental limits of ridge-regularized empirical risk minimization in high dimensions. In *Proceedings of the* 24*th International Conference on Artificial Intelligence and Statistics* (A. Banerjee and K. Fukumizu, eds.). *Proceedings of Machine Learning Research* **130** 2773–2781.

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

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

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

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

[61] WAINWRIGHT, M. J. (2009). Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting. *IEEE Trans. Inf. Theory* **55** 5728–5741. MR2597190 https://doi.org/10.1109/TIT.2009.2032816

[62] ZDEBOROVÁ, L. and KRZAKALA, F. (2015). Statistical physics of inference: Thresholds and algorithms. *Adv. Phys.* **65** 453–552.

The Annals of Statistics
2022, Vol. 50, No. 1, 197–224
https://doi.org/10.1214/21-AOS2101

APPROXIMATE MESSAGE PASSING ALGORITHMS FOR ROTATIONALLY INVARIANT MATRICES

BY ZHOU FAN

Department of Statistics and Data Science, Yale University, zhou.fan@yale.edu

Approximate Message Passing (AMP) algorithms have seen widespread use across a variety of applications. However, the precise forms for their Onsager corrections and state evolutions depend on properties of the underlying random matrix ensemble, limiting the extent to which AMP algorithms derived for white noise may be applicable to data matrices that arise in practice.

In this work, we study more general AMP algorithms for random matrices $\mathbf{W}$ that satisfy orthogonal rotational invariance in law, where $\mathbf{W}$ may have a spectral distribution that is different from the semicircle and Marcenko–Pastur laws characteristic of white noise. The Onsager corrections and state evolutions in these algorithms are defined by the free cumulants or rectangular free cumulants of the spectral distribution of $\mathbf{W}$. Their forms were derived previously by Opper, Çakmak and Winther using nonrigorous dynamic functional theory techniques, and we provide rigorous proofs.

Our motivating application is a Bayes-AMP algorithm for Principal Components Analysis, when there is prior structure for the principal components (PCs) and possibly nonwhite noise. For sufficiently large signal strengths and any non-Gaussian prior distributions for the PCs, we show that this algorithm provably achieves higher estimation accuracy than the sample PCs.

REFERENCES

[1] ALAOUI, A. E., MONTANARI, A. and SELLKE, M. (2020). Optimization of mean-field spin glasses. Preprint. Available at arXiv:2001.00904.

[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 https://doi.org/10.1214/14-AAP1010

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

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

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

[6] BENAYCH-GEORGES, F. (2009). Rectangular random matrices, entropy, and Fisher's information. *J. Operator Theory* **62** 371–419. MR2552088

[7] BENAYCH-GEORGES, F. (2011). Rectangular R-transform as the limit of rectangular spherical integrals. *J. Theoret. Probab.* **24** 969–987. MR2851240 https://doi.org/10.1007/s10959-011-0362-7

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

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

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

MSC2020 subject classifications. 62E20.
Key words and phrases. AMP, free probability theory, high-dimensional asymptotics, PCA.

[11] BOLTHAUSEN, E. (2014). An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model. *Comm. Math. Phys.* **325** 333–366. MR3147441 https://doi.org/10.1007/s00220-013-1862-3

[12] BORGERDING, M. and SCHNITER, P. (2016). Onsager-corrected deep learning for sparse linear inverse problems. In 2016 *IEEE Global Conference on Signal and Information Processing* (*GlobalSIP*) 227–231. IEEE, New York.

[13] BORGERDING, M., SCHNITER, P. and RANGAN, S. (2017). AMP-inspired deep networks for sparse linear inverse problems. *IEEE Trans. Signal Process.* **65** 4293–4308. MR3684065 https://doi.org/10.1109/TSP.2017.2708040

[14] BU, Z., KLUSOWSKI, J., RUSH, C. and SU, W. (2019). Algorithmic analysis and statistical estimation of SLOPE via approximate message passing. In *Advances in Neural Information Processing Systems* 9366–9376.

[15] ÇAKMAK, B. and OPPER, M. (2019). Memory-free dynamics for the Thouless–Anderson–Palmer equations of Ising models with arbitrary rotation-invariant ensembles of random coupling matrices. *Phys. Rev. E* **99** 062140. MR3984544

[16] ÇAKMAK, B. and OPPER, M. (2020). A dynamical mean-field theory for learning in restricted Boltzmann machines. *J. Stat. Mech. Theory Exp.* **10** 103303. MR4197533 https://doi.org/10.1088/1742-5468/abb8c9

[17] CAKMAK, B., WINTHER, O. and FLEURY, B. H. (2014). S-AMP: Approximate message passing for general matrix ensembles. In 2014 *IEEE Information Theory Workshop* (*ITW* 2014) 192–196. IEEE, New York.

[18] CALTAGIRONE, F., ZDEBOROVÁ, L. and KRZAKALA, F. (2014). On convergence of approximate message passing. In 2014 *IEEE International Symposium on Information Theory* 1812–1816. IEEE, New York.

[19] CHEN, W.-K. and LAM, W.-K. (2021). Universality of approximate message passing algorithms. *Electron. J. Probab.* **26** Paper No. 36. MR4235487

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

[21] DESHPANDE, Y., ABBE, E. and MONTANARI, A. (2017). Asymptotic mutual information for the balanced binary stochastic block model. *Inf. Inference* **6** 125–170. MR3671474 https://doi.org/10.1093/imaiai/iaw017

[22] DESHPANDE, Y. and MONTANARI, A. (2014). Information-theoretically optimal sparse PCA. In 2014 *IEEE International Symposium on Information Theory* 2197–2201. IEEE, New York.

[23] DIA, M., MACRIS, N., KRZAKALA, F., LESIEUR, T. and ZDEBOROVÁ, L. (2016). Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. In *Advances in Neural Information Processing Systems* 424–432.

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

[25] DONOHO, D. L., JAVANMARD, A. and MONTANARI, A. (2013). Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing. *IEEE Trans. Inf. Theory* **59** 7434–7464. MR3124654 https://doi.org/10.1109/TIT.2013.2274513

[26] DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2009). Message-passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci. USA* **106** 18914–18919.

[27] DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2010a). Message passing algorithms for compressed sensing: I. Motivation and construction. In 2010 *IEEE Information Theory Workshop on Information Theory* (*ITW* 2010, *Cairo*) 1–5. IEEE.

[28] DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2010b). Message passing algorithms for compressed sensing: II. Analysis and validation. In 2010 *IEEE Information Theory Workshop on Information Theory* (*ITW* 2010, *Cairo*) 1–5. IEEE.

[29] FAN, Z. (2022). Supplement to "Approximate Message Passing algorithms for rotationally invariant matrices." https://doi.org/10.1214/21-AOS2101SUPP

[30] FENG, O. Y., VENKATARAMANAN, R., RUSH, C. and SAMWORTH, R. J. (2021). A unifying tutorial on Approximate Message Passing. Preprint. Available at arXiv:2105.02180.

[31] FLETCHER, A., SAHRAEE-ARDAKAN, M., RANGAN, S. and SCHNITER, P. (2016). Expectation consistent approximate inference: Generalizations and convergence. In 2016 *IEEE International Symposium on Information Theory* (*ISIT*) 190–194. IEEE.

[32] GAMARNIK, D. and JAGANNATH, A. (2021). The overlap gap property and approximate message passing algorithms for p-spin models. *Ann. Probab.* **49** 180–205. MR4203336 https://doi.org/10.1214/20-AOP1448

[33] GUIONNET, A. and MAÏDA, M. (2005). A Fourier view on the R-transform and related asymptotics of spherical integrals. *J. Funct. Anal.* **222** 435–490. MR2132396 https://doi.org/10.1016/j.jfa.2004.09.015
[34] JAVANMARD, A. and MONTANARI, A. (2013). State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Inf. Inference* **2** 115–144. MR3311445 https://doi.org/10.1093/imaiai/iat004
[35] KABASHIMA, Y. (2003). A CDMA multiuser detection algorithm on the basis of belief propagation. *J. Phys. A: Math. Gen.* **36** 11111.
[36] KABASHIMA, Y. and VEHKAPERÄ, M. (2014). Signal recovery using expectation consistent approximation for linear observations. In 2014 *IEEE International Symposium on Information Theory* 226–230. IEEE, New York.
[37] MA, J. and PING, L. (2017). Orthogonal AMP. *IEEE Access* **5** 2020–2033.
[38] MAILLARD, A., FOINI, L., LAGE CASTELLANOS, A., KRZAKALA, F., MÉZARD, M. and ZDEBOROVÁ, L. (2019). High-temperature expansions and message passing algorithms. *J. Stat. Mech. Theory Exp.* **11** 113301. MR4059712 https://doi.org/10.1088/1742-5468/ab4bbb
[39] MALEKI, A., ANITORI, L., YANG, Z. and BARANIUK, R. G. (2013). Asymptotic analysis of complex LASSO via complex approximate message passing (CAMP). *IEEE Trans. Inf. Theory* **59** 4290–4308. MR3071330 https://doi.org/10.1109/TIT.2013.2252232
[40] METZLER, C., MOUSAVI, A. and BARANIUK, R. (2017). Learned D-AMP: Principled neural network based compressive image recovery. In *Advances in Neural Information Processing Systems* 1772–1783.
[41] MINKA, T. P. (2001). A family of algorithms for approximate Bayesian inference. Ph.D. thesis, Massachusetts Institute of Technology.
[42] MONTANARI, A. (2019). Optimization of the Sherrington–Kirkpatrick Hamiltonian. In 2019 *IEEE 60th Annual Symposium on Foundations of Computer Science* (*FOCS*) 1417–1433. IEEE, New York.
[43] MONTANARI, A. and RICHARD, E. (2016). Non-negative principal component analysis: Message passing algorithms and sharp asymptotics. *IEEE Trans. Inf. Theory* **62** 1458–1484. MR3472260 https://doi.org/10.1109/TIT.2015.2457942
[44] MONTANARI, A. and VENKATARAMANAN, R. (2021). Estimation of low-rank matrices via approximate message passing. *Ann. Statist.* **49** 321–345. MR4206680 https://doi.org/10.1214/20-AOS1958
[45] MOUSAVI, A., MALEKI, A. and BARANIUK, R. G. (2018). Consistent parameter estimation for LASSO and approximate message passing. *Ann. Statist.* **46** 119–148. MR3766948 https://doi.org/10.1214/17-AOS1544
[46] NOVAK, J. (2014). Three lectures on free probability. *Random matrix theory*, *interacting particle systems*, *and integrable systems* **65** 13.
[47] OPPER, M., ÇAKMAK, B. and WINTHER, O. (2016). A theory of solving TAP equations for Ising models with general invariant random matrices. *J. Phys. A* **49** 114002. MR3462332 https://doi.org/10.1088/1751-8113/49/11/114002
[48] OPPER, M. and WINTHER, O. (2001). Adaptive and self-averaging Thouless–Anderson–Palmer mean-field theory for probabilistic modeling. *Phys. Rev. E* **64** 056131.
[49] OPPER, M. and WINTHER, O. (2001). Tractable approximations for probabilistic models: The adaptive Thouless–Anderson–Palmer mean field approach. *Phys. Rev. Lett.* **86** 3695–3699. https://doi.org/10.1103/PhysRevLett.86.3695
[50] OPPER, M. and WINTHER, O. (2005). Expectation consistent approximate inference. *J. Mach. Learn. Res.* **6** 2177–2204. MR2249885
[51] PERRY, A., WEIN, A. S., BANDEIRA, A. S. and MOITRA, A. (2018). Message-passing algorithms for synchronization problems over compact groups. *Comm. Pure Appl. Math.* **71** 2275–2322. MR3862091 https://doi.org/10.1002/cpa.21750
[52] RANGAN, S. (2011). Generalized approximate message passing for estimation with random linear mixing. In 2011 *IEEE International Symposium on Information Theory Proceedings* 2168–2172. IEEE, New York.
[53] RANGAN, S. and FLETCHER, A. K. (2012). Iterative estimation of constrained rank-one matrices in noise. In 2012 *IEEE International Symposium on Information Theory Proceedings* 1246–1250. IEEE.
[54] RANGAN, S., SCHNITER, P. and FLETCHER, A. K. (2019). Vector approximate message passing. *IEEE Trans. Inf. Theory* **65** 6664–6684. MR4009222 https://doi.org/10.1109/TIT.2019.2916359
[55] RANGAN, S., SCHNITER, P., FLETCHER, A. K. and SARKAR, S. (2019). On the convergence of approximate message passing with arbitrary matrices. *IEEE Trans. Inf. Theory* **65** 5339–5351. MR4009237 https://doi.org/10.1109/TIT.2019.2913109

[56] SCHNITER, P. and RANGAN, S. (2015). Compressive phase retrieval via generalized approximate message passing. *IEEE Trans. Signal Process.* **63** 1043–1055. MR3311635 https://doi.org/10.1109/TSP.2014.2386294

[57] SCHNITER, P., RANGAN, S. and FLETCHER, A. K. (2016). Vector approximate message passing for the generalized linear model. In 2016 50*th Asilomar Conference on Signals, Systems and Computers* 1525–1529. IEEE.

[58] SPEICHER, R. (1998). *Combinatorial Theory of the Free Product with Amalgamation and Operator-Valued Free Probability Theory* **627**. AMS, Providence.

[59] SU, W., BOGDAN, M. and CANDÈS, E. (2017). False discoveries occur early on the Lasso path. *Ann. Statist.* **45** 2133–2150. MR3718164 https://doi.org/10.1214/16-AOS1521

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

[61] SUR, P., CHEN, Y. and CANDÈS, E. J. (2019). The likelihood ratio test in high-dimensional logistic regression is asymptotically a *rescaled* chi-square. *Probab. Theory Related Fields* **175** 487–558. MR4009715 https://doi.org/10.1007/s00440-018-00896-9

[62] TAKEUCHI, K. (2017). Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements. In 2017 *IEEE International Symposium on Information Theory* (*ISIT*) 501–505. IEEE, New York.

[63] TAKEUCHI, K. (2019). A unified framework of state evolution for message-passing algorithms. In 2019 *IEEE International Symposium on Information Theory* (*ISIT*) 151–155. IEEE, New York.

[64] TAKEUCHI, K. (2020). Convolutional approximate message-passing. *IEEE Signal Process. Lett.* **27** 416–420.

[65] TAKEUCHI, K. (2020). Bayes-optimal convolutional AMP. Preprint. Available at arXiv:2003.12245.

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

The Annals of Statistics
2022, Vol. 50, No. 1, 225–251
https://doi.org/10.1214/21-AOS2103

MINIMAX OPTIMALITY OF PERMUTATION TESTS

BY ILMUN KIM[1], SIVARAMAN BALAKRISHNAN[2,*] AND LARRY WASSERMAN[2,†]

[1]*Department of Statistics & Data Science, Department of Applied Statistics, Yonsei University, ilmun@yonsei.ac.kr*
[2]*Department of Statistics & Data Science, Carnegie Mellon University, [*]siva@stat.cmu.edu; [†]larry@stat.cmu.edu*

Permutation tests are widely used in statistics, providing a finite-sample guarantee on the type I error rate whenever the distribution of the samples under the null hypothesis is invariant to some rearrangement. Despite its increasing popularity and empirical success, theoretical properties of the permutation test, especially its power, have not been fully explored beyond simple cases. In this paper, we attempt to partly fill this gap by presenting a general nonasymptotic framework for analyzing the minimax power of the permutation test. The utility of our proposed framework is illustrated in the context of two-sample and independence testing under both discrete and continuous settings. In each setting, we introduce permutation tests based on U-statistics and study their minimax performance. We also develop exponential concentration bounds for permuted U-statistics based on a novel coupling idea, which may be of independent interest. Building on these exponential bounds, we introduce permutation tests, which are adaptive to unknown smoothness parameters without losing much power. The proposed framework is further illustrated using more sophisticated test statistics including weighted U-statistics for multinomial testing and Gaussian kernel-based statistics for density testing. Finally, we provide some simulation results that further justify the permutation approach.

REFERENCES

ACHARYA, J., JAFARPOUR, A., ORLITSKY, A. and SURESH, A. T. (2014). Sublinear algorithms for outlier detection and generalized closeness testing. In 2014 *IEEE International Symposium on Information Theory* 3200–3204. IEEE, Los Alamitos.

ALBERS, W., BICKEL, P. J. and VAN ZWET, W. R. (1976). Asymptotic expansions for the power of distribution free tests in the one-sample problem. *Ann. Statist.* **4** 108–156. MR0391373

ALBERT, M. (2015). Tests of independence by bootstrap and permutation: An asymptotic and non-asymptotic study. Application to neurosciences Ph.D. thesis, Université Nice Sophia Antipolis.

ALBERT, M. (2019). Concentration inequalities for randomly permuted sums. In *High Dimensional Probability VIII* 341–383. Springer, Berlin.

ANDERSON, N. H., HALL, P. and TITTERINGTON, D. M. (1994). Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates. *J. Multivariate Anal.* **50** 41–54. MR1292607 https://doi.org/10.1006/jmva.1994.1033

ANDERSON, M. J. and ROBINSON, J. (2001). Permutation tests for linear models. *Aust. N. Z. J. Stat.* **43** 75–88. MR1837497 https://doi.org/10.1111/1467-842X.00156

ARIAS-CASTRO, E., PELLETIER, B. and SALIGRAMA, V. (2018). Remember the curse of dimensionality: The case of goodness-of-fit testing in arbitrary dimension. *J. Nonparametr. Stat.* **30** 448–471. MR3794401 https://doi.org/10.1080/10485252.2018.1435875

BALAKRISHNAN, S. and WASSERMAN, L. (2018). Hypothesis testing for high-dimensional multinomials: A selective review. *Ann. Appl. Stat.* **12** 727–749. MR3834283 https://doi.org/10.1214/18-AOAS1155SF

BALAKRISHNAN, S. and WASSERMAN, L. (2019a). Hypothesis testing for densities and high-dimensional multinomials: Sharp local minimax rates. *Ann. Statist.* **47** 1893–1927. MR3953439 https://doi.org/10.1214/18-AOS1729

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

MSC2020 subject classifications. Primary 62C20; secondary 62G10.

Key words and phrases. Concentration inequalities, independence testing, minimax optimality, permutation tests, two-sample testing, U-statistics.

BARINGHAUS, L. and FRANZ, C. (2004). On a new multivariate two-sample test. *J. Multivariate Anal.* **88** 190–206. MR2021870 https://doi.org/10.1016/S0047-259X(03)00079-4

BERCU, B., DELYON, B. and RIO, E. (2015). *Concentration Inequalities for Sums and Martingales*. *SpringerBriefs in Mathematics*. Springer, Cham. MR3363542 https://doi.org/10.1007/978-3-319-22099-4

BERRETT, T. B., KONTOYIANNIS, I. and SAMWORTH, R. J. (2020). Optimal rates for independence testing via U-statistic permutation tests. *Annals of Statistics*. Available at arXiv:2001.05513.

BHATTACHARYA, B. and VALIANT, G. (2015). Testing closeness with unequal sized samples. In *Advances in Neural Information Processing Systems* 2611–2619.

BICKEL, P. J. and VAN ZWET, W. R. (1978). Asymptotic expansions for the power of distribution free tests in the two-sample problem. *Ann. Statist.* **6** 937–1004. MR0499567

BLACKFORD, J. U., SALOMON, R. M. and WALLER, N. G. (2009). Detecting change in biological rhythms: A multivariate permutation test approach to Fourier-transformed data. *Chronobiol. Int.* **26** 258–281.

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

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

CHAN, S.-O., DIAKONIKOLAS, I., VALIANT, G. and VALIANT, P. (2014). Optimal algorithms for testing closeness of discrete distributions. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms* 1193–1203. ACM, New York. MR3376448 https://doi.org/10.1137/1.9781611973402.88

CHATTERJEE, S. (2007). Stein's method for concentration inequalities. *Probab. Theory Related Fields* **138** 305–321. MR2288072 https://doi.org/10.1007/s00440-006-0029-y

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

CHUNG, E. and ROMANO, J. P. (2016). Multivariate and multiple permutation tests. *J. Econometrics* **193** 76–91. MR3500177 https://doi.org/10.1016/j.jeconom.2016.01.003

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

DIAKONIKOLAS, I. and KANE, D. M. (2016). A new approach for testing properties of discrete distributions. In 57*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2016 685–694. IEEE Computer Soc., Los Alamitos, CA. MR3631031

DICICCIO, C. J. and ROMANO, J. P. (2017). Robust permutation tests for correlation and regression coefficients. *J. Amer. Statist. Assoc.* **112** 1211–1220. MR3735371 https://doi.org/10.1080/01621459.2016.1202117

DORAN, G., MUANDET, K., ZHANG, K. and SCHÖLKOPF, B. (2014). A permutation-based kernel conditional independence test. In *Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence*, *UAI'*14 132–141. AUAI Press, Arlington, VA.

DUEMBGEN, L. (1998). Symmetrization and decoupling of combinatorial random elements. *Statist. Probab. Lett.* **39** 355–361. MR1646222 https://doi.org/10.1016/S0167-7152(98)00083-2

EFRON, B. and PETROSIAN, V. (1992). A simple test of independence for truncated data with applications to redshift surveys. *Astrophys. J.* **399** 345–352.

EFRON, B. and TIBSHIRANI, R. J. (1994). *An Introduction to the Bootstrap*. CRC press, Boca Raton.

FREEMAN, P. E., KIM, I. and LEE, A. B. (2017). Local two-sample testing: A new tool for analysing high-dimensional astronomical data. *Mon. Not. R. Astron. Soc.* **471** 3273–3282.

FROMONT, M., LAURENT, B. and REYNAUD-BOURET, P. (2013). The two-sample problem for Poisson processes: Adaptive tests with a nonasymptotic wild bootstrap approach. *Ann. Statist.* **41** 1431–1461. MR3113817 https://doi.org/10.1214/13-AOS1114

FUKUMIZU, K., GRETTON, A., SUN, X. and SCHÖLKOPF, B. (2008). Kernel measures of conditional dependence. In *Advances in Neural Information Processing Systems* (J. C. Platt, D. Koller, Y. Singer and S. T. Roweis, eds.) **20** 489–496. Curran Associates, Red Hook.

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

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

HAYASAKA, S. and NICHOLS, T. E. (2004). Combining voxel intensity and cluster extent with permutation test framework. *NeuroImage* **23** 54–63. https://doi.org/10.1016/j.neuroimage.2004.04.035

HEMERIK, J. and GOEMAN, J. (2018). Exact testing with random permutations. *TEST* **27** 811–825. MR3878362 https://doi.org/10.1007/s11749-017-0571-1

HODARA, P. and REYNAUD-BOURET, P. (2019). Exponential inequality for chaos based on sampling without replacement. *Statist. Probab. Lett.* **146** 65–69. MR3880791 https://doi.org/10.1016/j.spl.2018.11.003

HOEFFDING, W. (1952). The large-sample power of tests based on permutations of observations. *Ann. Math. Stat.* **23** 169–192. MR0057521 https://doi.org/10.1214/aoms/1177729436

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

INGSTER, Y. I. (1987). Minimax testing of nonparametric hypotheses on a distribution density in the L_p metrics. *Theory Probab. Appl.* **31** 333–337.

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

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

INGSTER, Y. I. and SUSLINA, I. A. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*. *Lecture Notes in Statistics* **169**. Springer, New York. MR1991446 https://doi.org/10.1007/978-0-387-21580-8

JANSSEN, A. (1997). Studentized permutation tests for non-i.i.d. hypotheses and the generalized Behrens–Fisher problem. *Statist. Probab. Lett.* **36** 9–21. MR1491070 https://doi.org/10.1016/S0167-7152(97)00043-6

JANSSEN, A. (2005). Resampling Student's t-type statistics. *Ann. Inst. Statist. Math.* **57** 507–529. MR2206536 https://doi.org/10.1007/BF02509237

JANSSEN, A. and PAULS, T. (2003). How do bootstrap and permutation tests work? *Ann. Statist.* **31** 768–806. MR1994730 https://doi.org/10.1214/aos/1056562462

KIM, I. (2021). Comparing a large number of multivariate distributions. *Bernoulli* **27** 419–441. MR4177375 https://doi.org/10.3150/20-BEJ1244

KIM, I., BALAKRISHNAN, S. and WASSERMAN, L. (2020). Robust multivariate nonparametric tests via projection averaging. *Ann. Statist.* **48** 3417–3441. MR4185814 https://doi.org/10.1214/19-AOS1936

KIM, I., BALAKRISHNAN, S. and WASSERMAN, L. (2022). Supplement to "Minimax optimality of permutation tests." https://doi.org/10.1214/21-AOS2103SUPP

KIM, I., LEE, A. B. and LEI, J. (2019). Global and local two-sample tests via regression. *Electron. J. Stat.* **13** 5253–5305. MR4043073 https://doi.org/10.1214/19-EJS1648

KIM, I., RAMDAS, A., SINGH, A. and WASSERMAN, L. (2021). Classification accuracy as a proxy for two-sample testing. *Ann. Statist.* **49** 411–434. MR4206684 https://doi.org/10.1214/20-AOS1962

KIRCH, C. and STEINEBACH, J. (2006). Permutation principles for the change analysis of stochastic processes under strong invariance. *J. Comput. Appl. Math.* **186** 64–88. MR2190298 https://doi.org/10.1016/j.cam.2005.03.065

LEE, A. J. (1990). *U-Statistics: Theory and Practice*. *Statistics: Textbooks and Monographs* **110**. Dekker, New York. MR1075417

LEHMANN, E. L. and ROMANO, J. P. (2006). *Testing Statistical Hypotheses*. Springer, Berlin.

LEPSKI, O. V. and SPOKOINY, V. G. (1999). Minimax nonparametric hypothesis testing: The case of an inhomogeneous alternative. *Bernoulli* **5** 333–358. MR1681702 https://doi.org/10.2307/3318439

MAGLIETTA, R., D'ADDABBO, A., PIEPOLI, A., PERRI, F., LIUNI, S., PESOLE, G. and ANCONA, N. (2007). Selection of relevant genes in cancer diagnosis based on their prediction accuracy. *Artif. Intell. Med.* **40** 29–44. https://doi.org/10.1016/j.artmed.2006.06.002

MARIS, E. and OOSTENVELD, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *J. Neurosci. Methods* **164** 177–190.

MEYNAOUI, A., LAURENT, B., ALBERT, M. and MARREL, A. (2019). Aggregated test of independence based on HSIC measures. arXiv preprint. Available at arXiv:1902.06441.

NEYKOV, M., BALAKRISHNAN, S. and WASSERMAN, L. (2020). Minimax optimal conditional independence testing. *Annals of Statistics*. Available at arXiv:2001.03039.

OJALA, M. and GARRIGA, G. C. (2010). Permutation tests for studying classifier performance. *J. Mach. Learn. Res.* **11** 1833–1863. MR2660654 https://doi.org/10.1109/ICDM.2009.108

PANINSKI, L. (2008). A coincidence-based test for uniformity given very sparsely sampled discrete data. *IEEE Trans. Inf. Theory* **54** 4750–4755. MR2591136 https://doi.org/10.1109/TIT.2008.928987

PARK, P. J., MANJOURIDES, J., BONETTI, M. and PAGANO, M. (2009). A permutation test for determining significance of clusters with applications to spatial and gene expression data. *Comput. Statist. Data Anal.* **53** 4290–4300. MR2744325 https://doi.org/10.1016/j.csda.2009.05.031

PAULY, M., BRUNNER, E. and KONIETSCHKE, F. (2015). Asymptotic permutation tests in general factorial designs. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 461–473. MR3310535 https://doi.org/10.1111/rssb.12073

PESARIN, F. (2001a). *Multivariate Permutation Tests: With Applications in Biostatistics*. Wiley, Chichester. MR1855501

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

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

ROBINSON, J. (1973). The large-sample power of permutation tests for randomization models. *Ann*. *Statist*. **1** 291–296. MR0359168

ROMANO, J. P. (1989). Bootstrap and randomization tests of some nonparametric hypotheses. *Ann*. *Statist*. **17** 141–159. MR0981441 https://doi.org/10.1214/aos/1176347007

ROMANO, J. P. (1990). On the behavior of randomization tests without a group invariance assumption. *J*. *Amer*. *Statist*. *Assoc*. **85** 686–692. MR1138350

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

STRANGER, B. E., NICA, A. C., FORREST, M. S., DIMAS, A., BIRD, C. P., BEAZLEY, C., INGLE, C. E., DUNNING, M., FLICEK, P. et al. (2007). Population genomics of human gene expression. *Nat*. *Genet*. **39** 1217–1224.

SZÉKELY, G. J. and RIZZO, M. L. (2004). Testing for equal distributions in high dimension. *Interstate* **5**.

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

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

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

ZHOU, Y.-H., MARRON, J. S. and WRIGHT, F. A. (2018). Eigenvalue significance testing for genetic association. *Biometrics* **74** 439–447. MR3825330 https://doi.org/10.1111/biom.12767

The Annals of Statistics
2022, Vol. 50, No. 1, 252–276
https://doi.org/10.1214/21-AOS2104

POWERFUL KNOCKOFFS VIA MINIMIZING RECONSTRUCTABILITY

BY ASHER SPECTOR* AND LUCAS JANSON†

Department of Statistics, Harvard University, *asherspector@college.harvard.edu; †ljanson@fas.harvard.edu

Model-X knockoffs (*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** (2018) 551–577) allows analysts to perform feature selection using almost any machine learning algorithm while provably controlling the expected proportion of false discoveries. This procedure involves constructing synthetic variables, called knockoffs, which effectively act as controls during feature selection. The gold standard for constructing knockoffs has been to minimize the mean absolute correlation (MAC) between features and their knockoffs, but, surprisingly, we prove this procedure can be powerless in extremely easy settings, including Gaussian linear models with correlated exchangeable features. The key problem is that minimizing the MAC creates joint dependencies between the features and knockoffs, which allow machine learning algorithms to reconstruct the effect of the features on the response using the knockoffs. To improve power, we propose generating knockoffs which minimize the reconstructability (MRC) of the features, and we demonstrate our proposal for Gaussian features by showing it is computationally efficient, robust, and powerful. We also prove that certain MRC knockoffs minimize a notion of estimation error in Gaussian linear models. Through extensive simulations, we show MRC knockoffs often dramatically outperform MAC-minimizing knockoffs, and we find no settings in which MAC-minimizing knockoffs outperform MRC knockoffs by more than a slight margin. We implement our methods and many others from the knockoffs literature in a new python package `knockpy`.

REFERENCES

[1] ANDERSON, T. W. (2009). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

[2] ASKARI, A., REBJOCK, Q., D'ASPREMONT, A. and GHAOUI, L. E. (2020). FANOK: Knockoffs in linear time. Preprint. Available at arXiv:2006.08790.

[3] BARBER, R. F. and CANDÈS, E. J. (2015). Controlling the false discovery rate via knockoffs. *Ann. Statist.* **43** 2055–2085. MR3375876 https://doi.org/10.1214/15-AOS1337

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

[5] BATES, S., CANDÈS, E., JANSON, L. and WANG, W. (2020). Metropolized knockoff sampling. *J. Amer. Statist. Assoc.* **0** 1–15.

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

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

[8] CHEN, J., HOU, A. and HOU, T. Y. (2020). A prototype knockoff filter for group selection with FDR control. *Inf. Inference* **9** 271–288. MR4108970 https://doi.org/10.1093/imaiai/iaz012

[9] DAI, C., LIN, B., XING, X. and LIU, J. S. (2020). False discovery rate control via data splitting. Preprint. Available at arXiv:2002.08542.

[10] DAI, R. and BARBER, R. (2016). The knockoff filter for FDR control in group-sparse and multitask regression (M. F. Balcan and K. Q. Weinberger, eds.). *Proceedings of Machine Learning Research* **48** 1851–1859. PMLR, New York, New York, USA.

MSC2020 subject classifications. Primary 62G10; secondary 62J12.

Key words and phrases. High-dimensional inference, knockoffs, model-X, variable selection, false discovery rate (FDR), power.

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

[12] FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2007). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441.

[13] GIMENEZ, J. R., GHORBANI, A. and ZOU, J. Y. (2019). Knockoffs for the mass: New feature importance statistics with false discovery guarantees. In *The 22nd International Conference on Artificial Intelligence and Statistics*, *AISTATS* 2019, 16–18 *April* 2019, *Naha*, *Okinawa*, *Japan* (K. Chaudhuri and M. Sugiyama, eds.). *Proceedings of Machine Learning Research* **89** 2125–2133. PMLR.

[14] GIMENEZ, J. R. and ZOU, J. (2019). Improving the stability of the knockoff procedure: Multiple simultaneous knockoffs and entropy maximization. In *Proceedings of Machine Learning Research* (K. Chaudhuri and M. Sugiyama, eds.) **89** 2184–2192. PMLR.

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

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

[17] JORDON, J., YOON, J. and VAN DER SCHAAR, M. (2019). KnockoffGAN: Generating knockoffs for feature selection using generative adversarial networks. In *International Conference on Learning Representations*.

[18] KE, Z. T., LIU, J. S. and MA, Y. (2020). Power of FDR control methods: The impact of ranking algorithm, tampered design, and symmetric statistic. Preprint. Available at arXiv:2010.08132.

[19] LEDOIT, O. and WOLF, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivariate Anal.* **88** 365–411. MR2026339 https://doi.org/10.1016/S0047-259X(03)00096-4

[20] LI, J. and MAATHUIS, M. H. (2019). GGM knockoff filter: False discovery rate control for Gaussian graphical models.

[21] LIU, J. and RIGOLLET, P. (2019). Power analysis of knockoff filters for correlated designs. In *Advances in Neural Information Processing Systems* 32 (H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox and R. Garnett, eds.) 15446–15455. Curran Associates, Red Hook.

[22] LU, Y. Y., FAN, Y., LV, J. and NOBLE, W. S. (2018). DeepPINK: Reproducible feature selection in deep neural networks. In *Advances in Neural Information Processing Systems* 31: *Annual Conference on Neural Information Processing Systems* 2018, *NeurIPS* 2018, 3-8 *December* 2018, *Montréal*, *Canada* (S. Bengio, H. M. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi and R. Garnett, eds.) 8690–8700.

[23] PERSSON, H. and CLARKE, R. M. (1986). Algorithm 12: Solving the entropy maximization problem with equality and inequality constraints. *Environment and Planning A*: *Economy and Space* **18** 1665–1676.

[24] ROMANO, Y., SESIA, M. and CANDÈS, E. (2020). Deep knockoffs. *J. Amer. Statist. Assoc.* **115** 1861–1872. MR4189763 https://doi.org/10.1080/01621459.2019.1660174

[25] SESIA, M., BATES, S., CANDÈS, E., MARCHINI, J. and SABATTI, C. (2020). Controlling the false discovery rate in GWAS with population structure. *BioRxiv*.

[26] SESIA, M., KATSEVICH, E., BATES, S., CANDÈS, E. and SABATTI, C. (2019). Multi-resolution localization of causal variants across the genome. *BioRxiv*.

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

[28] SPECTOR, A. and JANSON, L. (2022). Supplement to "Powerful knockoffs via minimizing reconstructability." https://doi.org/10.1214/21-AOS2104SUPP

[29] XING, X., ZHAO, Z. and LIU, J. S. (2019). Controlling false discovery rate using Gaussian mirrors.

The Annals of Statistics
2022, Vol. 50, No. 1, 277–302
https://doi.org/10.1214/21-AOS2105

ON LEAST SQUARES ESTIMATION UNDER HETEROSCEDASTIC AND HEAVY-TAILED ERRORS

BY ARUN K. KUCHIBHOTLA[1] AND ROHIT K. PATRA[2]

[1]*Department of Statistics and Data Science, Carnegie Mellon University, arunku@cmu.edu*
[2]*Department of Statistics, University of Florida, rohitpatra@ufl.edu*

We consider least squares estimation in a general nonparametric regression model where the error is allowed to depend on the covariates. The rate of convergence of the least squares estimator (LSE) for the unknown regression function is well studied when the errors are sub-Gaussian. We find upper bounds on the rates of convergence of the LSE when the error has a uniformly bounded conditional variance and has only finitely many moments. Our upper bound on the rate of convergence of the LSE depends on the moment assumptions on the error, the metric entropy of the class of functions involved and the "local" structure of the function class around the truth. We find sufficient conditions on the error distribution under which the rate of the LSE matches the rate of the LSE under sub-Gaussian error. Our results are finite sample and allow for heteroscedastic and heavy-tailed errors.

REFERENCES

AGMON, S. (2010). *Lectures on Elliptic Boundary Value Problems*. AMS Chelsea Publishing, Providence, RI. Prepared for publication by B. Frank Jones, Jr. with the assistance of George W. Batten, Jr., Revised edition of the 1965 original. MR2589244 https://doi.org/10.1090/chel/369

AUDIBERT, J.-Y. and CATONI, O. (2011). Robust linear least squares regression. *Ann. Statist.* **39** 2766–2794. MR2906886 https://doi.org/10.1214/11-AOS918

BARTLETT, P. L. (2008). Fast rates for estimation error and oracle inequalities for model selection. *Econometric Theory* **24** 545–552. MR2490397 https://doi.org/10.1017/S0266466608080225

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

BIRGÉ, L. (1989). The Grenander estimator: A nonasymptotic approach. *Ann. Statist.* **17** 1532–1549. MR1026298 https://doi.org/10.1214/aos/1176347380

BIRGÉ, L. and MASSART, P. (1993). Rates of convergence for minimum contrast estimators. *Probab. Theory Related Fields* **97** 113–150. MR1240719 https://doi.org/10.1007/BF01199316

BIRGÉ, L. and MASSART, P. (1998). Minimum contrast estimators on sieves: Exponential bounds and rates of convergence. *Bernoulli* **4** 329–375. MR1653272 https://doi.org/10.2307/3318720

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

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

BUJA, A., HASTIE, T. and TIBSHIRANI, R. (1989). Linear smoothers and additive models. *Ann. Statist.* **17** 453–555. MR0994249 https://doi.org/10.1214/aos/1176347115

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

CHATTERJEE, S. and LAFFERTY, J. (2019). Adaptive risk bounds in unimodal regression. *Bernoulli* **25** 1–25. MR3892309 https://doi.org/10.3150/16-bej922

CHEN, X. and SHEN, X. (1998). Sieve extremum estimates for weakly dependent data. *Econometrica* **66** 289–314. MR1612238 https://doi.org/10.2307/2998559

MSC2020 subject classifications. 63GO8, 62E17.

Key words and phrases. Dyadic peeling, finite sample tail probability bounds, interpolation inequality, local envelopes, maximal inequality, heavy tails.

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* **162** 47–70. MR3350040 https://doi.org/10.1007/s00440-014-0565-9

DE LA PEÑA, V. H. and GINÉ, E. (1999). *Decoupling*. *Probability and Its Applications* (*New York*). Springer, New York. From dependence to independence, Randomly stopped processes. U-statistics and processes. Martingales and beyond. MR1666908 https://doi.org/10.1007/978-1-4612-0537-1

DIRKSEN, S. (2015). Tail bounds via generic chaining. *Electron. J. Probab.* **20** no. 53, 29. MR3354613 https://doi.org/10.1214/EJP.v20-3760

DOSS, C. R. (2015). Bracketing Numbers of Convex Functions on Polytopes. Preprint. Available at arXiv:1506.00034.

DÜMBGEN, L., VAN DE GEER, S. A., VERAAR, M. C. and WELLNER, J. A. (2010). Nemirovski's inequalities revisited. *Amer. Math. Monthly* **117** 138–160. MR2590193 https://doi.org/10.4169/000298910X476059

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

GAILLARD, P. and GERCHINOVITZ, S. (2015). A chaining algorithm for online nonparametric regression. In *Conference on Learning Theory* 764–796.

GAO, C., HAN, F. and ZHANG, C.-H. (2020a). On estimation of isotonic piecewise constant signals. Preprint. Available at arXiv:1705.06386.

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

GHOSAL, P. and SEN, B. (2017). On univariate convex regression. *Sankhya A* **79** 215–253. MR3707421 https://doi.org/10.1007/s13171-017-0104-8

GINÉ, E. and KOLTCHINSKII, V. (2006). Concentration inequalities and asymptotic results for ratio type empirical processes. *Ann. Probab.* **34** 1143–1216. MR2243881 https://doi.org/10.1214/009117906000000070

GINÉ, E., LATAŁA, R. and ZINN, J. (2000). Exponential and moment inequalities for U-statistics. In *High Dimensional Probability*, *II* (*Seattle*, *WA*, 1999). *Progress in Probability* **47** 13–38. Birkhäuser, Boston, MA. MR1857312

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

GINÉ, E. and ZINN, J. (1983). Central limit theorems and weak laws of large numbers in certain Banach spaces. *Z. Wahrsch. Verw. Gebiete* **62** 323–354. MR0688642 https://doi.org/10.1007/BF00535258

GOLDENSHLUGER, A. and LEPSKI, O. (2020). Minimax estimation of norms of a probability density: Ii. rate-optimal estimation procedures. Available at arXiv:2008.10987.

GUNTUBOYINA, A. and SEN, B. (2015). Global risk bounds and adaptation in univariate convex regression. *Probab. Theory Related Fields* **163** 379–411. MR3405621 https://doi.org/10.1007/s00440-014-0595-3

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

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

HAN, Q. and WELLNER, J. A. (2017). A sharp multiplier inequality with applications to heavy-tailed regression problems. Preprint. Available at arXiv:1706.02410v1.

HAN, Q. and WELLNER, J. A. (2018). Robustness of shape-restricted regression estimators: An envelope perspective. Preprint. Available at arXiv:1805.02542.

HAN, Q. and WELLNER, J. A. (2019). Convergence rates of least squares regression estimators with heavy-tailed errors. *Ann. Statist.* **47** 2286–2319. MR3953452 https://doi.org/10.1214/18-AOS1748

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

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

HRISTACHE, M., JUDITSKY, A., POLZEHL, J. and SPOKOINY, V. (2001). Structure adaptive approach for dimension reduction. *Ann. Statist.* **29** 1537–1566. MR1891738 https://doi.org/10.1214/aos/1015345954

KOLMOGOROFF, A. (1949). On inequalities between the upper bounds of the successive derivatives of an arbitrary function on an infinite interval. *Amer. Math. Soc. Transl.* **1949** 19. MR0031009

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

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

KUCHIBHOTLA, A. K. and CHAKRABORTTY, A. (2018). Moving beyond sub-gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression. Preprint. Available at arXiv:1804.02605.
KUCHIBHOTLA, A. K. and PATRA, R. K. (2022). Supplement to “On least squares estimation under heteroscedastic and heavy-tailed errors.” https://doi.org/10.1214/21-AOS2105SUPP
KUCHIBHOTLA, A. K., PATRA, R. K. and SEN, B. (2021). Semiparametric efficiency in convexity constrained single index model. *J. Amer. Statist. Assoc.* Preprint. Available at arXiv:1708.00145v3.
KUR, G., DAGAN, Y. and RAKHLIN, A. (2019). Optimality of maximum likelihood for log-concave density estimation and bounded convex regression. Preprint. Available at arXiv:1903.05315.
LECUÉ, G. and LERASLE, M. (2020). Robust machine learning by median-of-means: Theory and practice. *Ann. Statist.* **48** 906–931. MR4102681 https://doi.org/10.1214/19-AOS1828
LECUÉ, G. and MENDELSON, S. (2013). Learning subgaussian classes: Upper and minimax bounds. Preprint. Available at arXiv:1305.4825.
LECUÉ, G. and MENDELSON, S. (2016). Performance of empirical risk minimization in linear aggregation. *Bernoulli* **22** 1520–1534. MR3474824 https://doi.org/10.3150/15-BEJ701
LUGOSI, G. and MENDELSON, S. (2020). Risk minimization by median-of-means tournaments. *J. Eur. Math. Soc.* (*JEMS*) **22** 925–965. MR4055993 https://doi.org/10.4171/jems/937
MASSART, P. and NÉDÉLEC, É. (2006). Risk bounds for statistical learning. *Ann. Statist.* **34** 2326–2366. MR2291502 https://doi.org/10.1214/009053606000000786
MASSART, P. and ROSSIGNOL, R. (2013). Around Nemirovski’s inequality. In *From Probability to Statistics and Back*: *High-Dimensional Models and Processes*. *Inst. Math. Stat.* (*IMS*) *Collect.* **9** 254–265. IMS, Beachwood, OH. MR3202638 https://doi.org/10.1214/12-IMSCOLL918
MENDELSON, S. (2008). On weakly bounded empirical processes. *Math. Ann.* **340** 293–314. MR2368981 https://doi.org/10.1007/s00208-007-0152-9
MENDELSON, S. (2015). Learning without concentration. *J. ACM* **62** Art. 21, 25. MR3367000 https://doi.org/10.1145/2699439
MENDELSON, S. (2016). Upper bounds on product and multiplier empirical processes. *Stochastic Process. Appl.* **126** 3652–3680. MR3565471 https://doi.org/10.1016/j.spa.2016.04.019
MENDELSON, S. (2017). “Local” vs. “global” parameters—breaking the Gaussian complexity barrier. *Ann. Statist.* **45** 1835–1862. MR3718154 https://doi.org/10.1214/16-AOS1510
MENDELSON, S. (2019). An unrestricted learning procedure. *J. ACM* **66** Art. 42, 42. MR4040282 https://doi.org/10.1145/3361699
NICKL, R. and PÖTSCHER, B. M. (2007). Bracketing metric entropy rates and empirical central limit theorems for function classes of Besov- and Sobolev-type. *J. Theoret. Probab.* **20** 177–199. MR2324525 https://doi.org/10.1007/s10959-007-0058-1
NIRENBERG, L. (2011). On elliptic partial differential equations. In *The Principle of Minimum and Its Applications to Functional Equations* 1–48. Springer, Berlin.
POLLARD, D. (1984). *Convergence of Stochastic Processes*. *Springer Series in Statistics*. Springer, New York. MR0762984 https://doi.org/10.1007/978-1-4612-5254-2
RAKHLIN, A., SRIDHARAN, K. and TSYBAKOV, A. B. (2017). Empirical entropy, minimax regret and minimax risk. *Bernoulli* **23** 789–824. MR3606751 https://doi.org/10.3150/14-BEJ679
SHEN, X. and WONG, W. H. (1994). Convergence rate of sieve estimates. *Ann. Statist.* **22** 580–615. MR1292531 https://doi.org/10.1214/aos/1176325486
STONE, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation. *Ann. Statist.* **22** 118–184. MR1272079 https://doi.org/10.1214/aos/1176325361
TALAGRAND, M. (1996). Majorizing measures: The generic chaining. *Ann. Probab.* **24** 1049–1103. MR1411488 https://doi.org/10.1214/aop/1065725175
TALAGRAND, M. (2014). *Upper and Lower Bounds for Stochastic Processes*: *Modern Methods and Classical Problems*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete*. 3. *Folge*. *A Series of Modern Surveys in Mathematics* [*Results in Mathematics and Related Areas*. 3*rd Series*. *A Series of Modern Surveys in Mathematics*] **60**. Springer, Heidelberg. MR3184689 https://doi.org/10.1007/978-3-642-54075-2
VAN DE GEER, S. (1990). Estimating a regression function. *Ann. Statist.* **18** 907–924. MR1056343 https://doi.org/10.1214/aos/1176347632
VAN DE GEER, S. A. (2000). *Applications of Empirical Process Theory*. *Cambridge Series in Statistical and Probabilistic Mathematics* **6**. Cambridge Univ. Press, Cambridge. MR1739079
VAN DE GEER, S. and MURO, A. (2014). On higher order isotropy conditions and lower bounds for sparse quadratic forms. *Electron. J. Stat.* **8** 3031–3061. MR3301300 https://doi.org/10.1214/15-EJS983
VAN DE GEER, S. and WAINWRIGHT, M. J. (2017). On concentration for (regularized) empirical risk minimization. *Sankhya A* **79** 159–200. MR3707417 https://doi.org/10.1007/s13171-017-0111-9
VAN DE GEER, S. and WEGKAMP, M. (1996). Consistency for the least squares estimator in nonparametric regression. *Ann. Statist.* **24** 2513–2523. MR1425964 https://doi.org/10.1214/aos/1032181165

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

VAN DER VAART, A. (2002). Semiparametric statistics. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1999). *Lecture Notes in Math*. **1781** 331–457. Springer, Berlin. MR1915446

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

VAN DER VAART, A. and WELLNER, J. A. (2011). A local maximal inequality under uniform entropy. *Electron*. *J*. *Stat*. **5** 192–203. MR2792551 https://doi.org/10.1214/11-EJS605

YANG, Y. and BARRON, A. (1999). Information-theoretic determination of minimax rates of convergence. *Ann*. *Statist*. **27** 1564–1599. MR1742500 https://doi.org/10.1214/aos/1017939142

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

The Annals of Statistics
2022, Vol. 50, No. 1, 303–323
https://doi.org/10.1214/21-AOS2106

DECONVOLUTION WITH UNKNOWN NOISE DISTRIBUTION IS POSSIBLE FOR MULTIVARIATE SIGNALS

BY ÉLISABETH GASSIAT[1], SYLVAIN LE CORFF[2] AND LUC LEHÉRICY[3]

[1]*Laboratoire de mathématiques d'Orsay, Université Paris-Saclay, CNRS, elisabeth.gassiat@universite-paris-saclay.fr*
[2]*Samovar, Télécom SudParis, département CITI, TIPIC, Institut Polytechnique de Paris, sylvain.le_corff@telecom-sudparis.eu*
[3]*Laboratoire J. A. Dieudonné, Université Côte d'Azur, CNRS, luc.lehericy@univ-cotedazur.fr*

This paper considers the deconvolution problem in the case where the target signal is multidimensional and no information is known about the noise distribution. More precisely, no assumption is made on the noise distribution and no samples are available to estimate it: the deconvolution problem is solved based only on observations of the corrupted signal. We establish the identifiability of the model up to translation when the signal has a Laplace transform with an exponential growth ρ smaller than 2 and when it can be decomposed into two dependent components. Then we propose an estimator of the probability density function of the signal, which is consistent for any unknown noise distribution with finite variance. We also prove rates of convergence and, as the estimator depends on ρ which is usually unknown, we propose a model selection procedure to obtain an adaptive estimator with the same rate of convergence as the estimator with a known tail parameter. This rate of convergence is known to be minimax when $\rho = 1$.

REFERENCES

[1] ATTIAS, H. and SCHREINER, C. E. (1998). Blind source separation and deconvolution: The dynamic component analysis algorithm. *Neural Comput*. **10** 1373–1424.
[2] BATSON, J. and ROYER, L. (2019). Noise2Self: Blind denoising by self-supervision. In *Proceedings of the* 36*th International Conference on Machine Learning* (*ICML*).
[3] BAUDRY, J.-P., MAUGIS, C. and MICHEL, B. (2012). Slope heuristics: Overview and implementation. *Stat*. *Comput*. **22** 455–470. MR2865029 https://doi.org/10.1007/s11222-011-9236-1
[4] BELOMESTNY, D. and GOLDENSHLUGER, A. (2021). Density deconvolution under general assumptions on the distribution of measurement errors. *Ann*. *Statist*. **49** 615–649. MR4255101 https://doi.org/10.1214/20-aos1969
[5] BERTIN, K., LACOUR, C. and RIVOIRARD, V. (2016). Adaptive pointwise estimation of conditional density function. *Ann*. *Inst*. *Henri Poincaré Probab*. *Stat*. **52** 939–980. MR3498017 https://doi.org/10.1214/14-AIHP665
[6] BUTUCEA, C. and TSYBAKOV, A. B. (2007). Sharp optimality in density deconvolution with dominating bias. I. *Teor*. *Veroyatn*. *Primen*. **52** 111–128. MR2354572 https://doi.org/10.1137/S0040585X97982840
[7] BUTUCEA, C. and TSYBAKOV, A. B. (2007). Sharp optimality in density deconvolution with dominating bias. II. *Teor*. *Veroyatn*. *Primen*. **52** 336–349. MR2742504 https://doi.org/10.1137/S0040585X97982992
[8] CAMPISI, P. and EGIAZARIAN, K., eds. (2007). *Blind Image Deconvolution*: *Theory and Applications* CRC Press, Boca Raton, FL. MR2404093 https://doi.org/10.1201/9781420007299
[9] CAPITAO MINICONI, J. (2021). Reconstruction of smooth curves from noisy measurements. Preprint.
[10] CARROLL, R. J. and HALL, P. (1988). Optimal rates of convergence for deconvolving a density. *J*. *Amer*. *Statist*. *Assoc*. **83** 1184–1186. MR0997599
[11] CHAZAL, F., COHEN-STEINER, D. and MÉRIGOT, Q. (2011). Geometric inference for probability measures. *Found*. *Comput*. *Math*. **11** 733–751. MR2859954 https://doi.org/10.1007/s10208-011-9098-0
[12] CHAZAL, F., FASY, B., LECCI, F., MICHEL, B., RINALDO, A. and WASSERMAN, L. (2017). Robust topological inference: Distance to a measure and kernel distance. *J*. *Mach*. *Learn*. *Res*. **18** Paper No. 159, 40. MR3813808

MSC2020 subject classifications. Primary 62G05; secondary 62G07, 62G20, 62H12.
Key words and phrases. Deconvolution, identifiability, nonparametric estimation, minimax rates, adaptivity.

[13] CHAZAL, F. and MICHEL, B. (2017). An introduction to Topological Data Analysis: Fundamental and practical aspects for data scientists. Preprint. Available at arXiv:1710.04019.

[14] COMON, P. (1994). Independent component analysis: A new concept? *Signal Process*. **36** 287–314.

[15] COMTE, F. and LACOUR, C. (2010). Pointwise deconvolution with unknown error distribution. *C. R. Math. Acad. Sci. Paris* **348** 323–326. MR2600132 https://doi.org/10.1016/j.crma.2010.02.012

[16] COMTE, F. and LACOUR, C. (2013). Anisotropic adaptive kernel deconvolution. *Ann. Inst. Henri Poincaré Probab. Stat*. **49** 569–609. MR3088382 https://doi.org/10.1214/11-aihp470

[17] DELAIGLE, A., HALL, P. and MEISTER, A. (2008). On deconvolution with repeated measurements. *Ann. Statist*. **36** 665–685. MR2396811 https://doi.org/10.1214/009053607000000884

[18] DEVROYE, L. (1989). Consistent deconvolution in density estimation. *Canad. J. Statist*. **17** 235–239. MR1033106 https://doi.org/10.2307/3314852

[19] ECKLE, K., BISSANTZ, N. and DETTE, H. (2017). Multiscale inference for multivariate deconvolution. *Electron. J. Stat*. **11** 4179–4219. MR3716498 https://doi.org/10.1214/17-EJS1355

[20] ERIKSSON, J. and KOIVUNEN, V. (2004). Identifiability, separability, uniqueness of linear ICA models. *IEEE Signal Process. Lett*. **11** 601–604.

[21] FAN, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist*. **19** 1257–1272. MR1126324 https://doi.org/10.1214/aos/1176348248

[22] GASSIAT, É., LE CORFF, S. and LEHÉRICY, L. (2020). Identifiability and consistent estimation of nonparametric translation hidden Markov models with general state space. *J. Mach. Learn. Res*. **21** Paper No. 115, 40. MR4119183

[23] GASSIAT, É., LE CORFF, S. and LEHÉRICY, L. (2022). Supplement to "Deconvolution with unknown noise distribution is possible for multivariate signals." https://doi.org/10.1214/21-AOS2106SUPP

[24] GASSIAT, E., LE CORFF, S. and LEHÉRICY, L. (2020). Deconvolution with unknown noise distribution is possible for multivariate signals. Preprint. Available at arXiv:2006.14226.

[25] GASSIAT, E. and ROUSSEAU, J. (2016). Nonparametric finite translation hidden Markov models and extensions. *Bernoulli* **22** 193–212. MR3449780 https://doi.org/10.3150/14-BEJ631

[26] GOLDENSHLUGER, A. and LEPSKI, O. (2008). Universal pointwise selection rule in multivariate function estimation. *Bernoulli* **14** 1150–1190. MR2543590 https://doi.org/10.3150/08-BEJ144

[27] GOLDENSHLUGER, A. V. and LEPSKI, O. V. (2013). General selection rule from a family of linear estimators. *Theory Probab. Appl*. **57** 209–226. MR3201652 https://doi.org/10.1137/S0040585X97985923

[28] HYVARINEN, A., KARHUNEN, J. and OJA, E. (2002). *Independent Component Analysis*. Wiley, New York.

[29] JOHANNES, J. (2009). Deconvolution with unknown error distribution. *Ann. Statist*. **37** 2301–2323. MR2543693 https://doi.org/10.1214/08-AOS652

[30] JUTTEN, C. (1991). Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Process*. **2** 1–10.

[31] KHEMAKHEM, I., KINGMA, D. P., PIO MONTI, R. and HYVARINEN, A. (2020). Variational Autoencoders and Nonlinear ICA: A Unifying Framework. Preprint. Available at arXiv:1907.04809.

[32] KOTLARSKI, I. (1967). On characterizing the gamma and the normal distribution. *Pacific J. Math*. **20** 69–76. MR0203769

[33] KRULL, A., BUCHHOLZ, T. O. and JUG, F. (2019). Noise2Void—learning denoising from single noisy images. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*).

[34] KUNDUR, D. and HATZINAKOS, D. (1996). Blind image deconvolution. *IEEE Signal Process. Mag*. **13** 43–64.

[35] LI, T. and VUONG, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *J. Multivariate Anal*. **65** 139–165. MR1625869 https://doi.org/10.1006/jmva.1998.1741

[36] LIN, W. and KULASEKERA, K. B. (2007). Identifiability of single-index models and additive-index models. *Biometrika* **94** 496–501. MR2380574 https://doi.org/10.1093/biomet/asm029

[37] LIU, M. C. and TAYLOR, R. L. (1989). A consistent nonparametric density estimator for the deconvolution problem. *Canad. J. Statist*. **17** 427–438. MR1047309 https://doi.org/10.2307/3315482

[38] MEISTER, A. (2004). On the effect of misspecifying the error density in a deconvolution problem. *Canad. J. Statist*. **32** 439–449. MR2125855 https://doi.org/10.2307/3316026

[39] MEISTER, A. (2007). Deconvolving compactly supported densities. *Math. Methods Statist*. **16** 63–76. MR2319471 https://doi.org/10.3103/S106653070701005X

[40] MEISTER, A. (2009). *Deconvolution Problems in Nonparametric Statistics*. *Lecture Notes in Statistics* **193**. Springer, Berlin. MR2768576 https://doi.org/10.1007/978-3-540-87557-4

[41] MOULINES, E., CARDOSO, J.-F. and GASSIAT, E. (1997). Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models. In *IEEE International Conference on Acoustics, Speech, and Signal Processing* **5** 3617–3620. IEEE, New York.

[42] OLLION, J., OLLION, C., GASSIAT, E., LEHÉRICY, L. and LE CORFF, S. (2021). Joint self-supervised blind denoising and noise estimation. Preprint. Available at arXiv:2102.08023.

[43] PFISTER, N., WEICHWALD, S., BÜHLMANN, P. and SCHÖLKOPF, B. (2019). Robustifying independent component analysis by adjusting for group-wise stationary noise. *J. Mach. Learn. Res.* **20** Paper No. 147, 50. MR4030161

[44] SARKAR, A., PATI, D., CHAKRABORTY, A., MALLICK, B. K. and CARROLL, R. J. (2018). Bayesian semiparametric multivariate density deconvolution. *J. Amer. Statist. Assoc.* **113** 401–416. MR3803474 https://doi.org/10.1080/01621459.2016.1260467

[45] SCHENNACH, S. M. and HU, Y. (2013). Nonparametric identification and semiparametric estimation of classical measurement error models without side information. *J. Amer. Statist. Assoc.* **108** 177–186. MR3174611 https://doi.org/10.1080/01621459.2012.751872

[46] STARCK, J.-L., PANTIN, E. and MURTAGH, F. (2002). Deconvolution in astronomy: A review. *Publ. Astron. Soc. Pac.* **114** 1051.

[47] STEFANSKI, L. and CARROLL, R. J. (1990). Deconvoluting kernel density estimators. *Statistics* **21** 169–184. MR1054861 https://doi.org/10.1080/02331889008802238

[48] YUAN, M. (2011). On the identifiability of additive index models. *Statist. Sinica* **21** 1901–1911. MR2896004 https://doi.org/10.5705/ss.2008.117

The Annals of Statistics
2022, Vol. 50, No. 1, 324–350
https://doi.org/10.1214/21-AOS2107

ISOTONIC REGRESSION WITH UNKNOWN PERMUTATIONS: STATISTICS, COMPUTATION AND ADAPTATION

BY ASHWIN PANANJADY[1] AND RICHARD J. SAMWORTH[2]

Dedicated to the memory of Matthew Brennan

[1]*Schools of Industrial & Systems Engineering and Electrical & Computer Engineering, Georgia Institute of Technology*
ashwinpm@gatech.edu
[2]*Statistical Laboratory, University of Cambridge*
r.samworth@statslab.cam.ac.uk

Motivated by models for multiway comparison data, we consider the problem of estimating a coordinatewise isotonic function on the domain $[0, 1]^d$ from noisy observations collected on a uniform lattice, but where the design points have been permuted along each dimension. While the univariate and bivariate versions of this problem have received significant attention, our focus is on the multivariate case $d \geq 3$. We study both the minimax risk of estimation (in empirical L_2 loss) and the fundamental limits of adaptation (quantified by the adaptivity index) to a family of piecewise constant functions. We provide a computationally efficient Mirsky partition estimator that is minimax optimal while also achieving the smallest adaptivity index possible for polynomial time procedures. Thus, from a worst-case perspective and in sharp contrast to the bivariate case, the latent permutations in the model do not introduce significant computational difficulties over and above vanilla isotonic regression. On the other hand, the fundamental limits of adaptation are significantly different with and without unknown permutations: Assuming a hardness conjecture from average-case complexity theory, a statistical-computational gap manifests in the former case. In a complementary direction, we show that natural modifications of existing estimators fail to satisfy at least one of the desiderata of optimal worst-case statistical performance, computational efficiency and fast adaptation. Along the way to showing our results, we improve adaptation results in the special case $d = 2$ and establish some properties of estimators for vanilla isotonic regression, both of which may be of independent interest.

REFERENCES

ABBE, E. and MONTANARI, A. (2013). Conditional random fields, planted constraint satisfaction and entropy concentration. In *Approximation, Randomization, and Combinatorial Optimization. Lecture Notes in Computer Science* **8096** 332–346. Springer, Heidelberg. MR3126539 https://doi.org/10.1007/978-3-642-40328-6_24

ABID, A. and ZOU, J. (2018). A stochastic expectation-maximization approach to shuffled linear regression. In 2018 56*th Annual Allerton Conference on Communication, Control, and Computing* (*Allerton*) 470–477. IEEE.

AIROLDI, E. M., COSTA, T. B. and CHAN, S. H. (2013). Stochastic blockmodel approximation of a graphon: Theory and consistent estimation. In *Advances in Neural Information Processing Systems* 692–700.

ANDERSEN, C. M. and BRO, R. (2003). Practical aspects of PARAFAC modeling of fluorescence excitation-emission data. *J. Chemom.* **17** 200–215.

ANGELINI, M. C., CALTAGIRONE, F., KRZAKALA, F. and ZDEBOROVÁ, L. (2015). Spectral detection on sparse hypergraphs. In 2015 53*rd Annual Allerton Conference on Communication, Control, and Computing* (*Allerton*) 66–73. IEEE.

MSC2020 subject classifications. 62G05.
Key words and phrases. Permutation-based model, statistical-computational gap, adaptive estimation, shape-constrained estimation.

BALABDAOUI, F., DOSS, C. R. and DUROT, C. (2020). Unlinked monotone regression. arXiv preprint. Available at arXiv:2007.00830.

BALLINGER, T. P. and WILCOX, N. T. (1997). Decisions, error and heterogeneity. *Econ. J.* **107** 1090–1105.

BALTRUNAS, L., MAKCINSKAS, T. and RICCI, F. (2010). Group recommendations with rank aggregation and collaborative filtering. In *Proceedings of the Fourth ACM Conference on Recommender Systems*, *RecSys '10* 119–126. ACM, New York, USA.

BARAK, B., HOPKINS, S., KELNER, J., KOTHARI, P. K., MOITRA, A. and POTECHIN, A. (2019). A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM J. Comput*. **48** 687–735. MR3945259 https://doi.org/10.1137/17M1138236

BEHR, M. and MUNK, A. (2017). Minimax estimation in linear models with unknown finite alphabet design. arXiv preprint. Available at arXiv:1711.04145.

BELLEC, P. C. and TSYBAKOV, A. B. (2015). Sharp oracle bounds for monotone and convex regression through aggregation. *J. Mach. Learn. Res.* **16** 1879–1892. MR3417801

BEN-AKIVA, M. E., LERMAN, S. R. and LERMAN, S. R. (1985). *Discrete Choice Analysis*: *Theory and Application to Travel Demand* **9**. MIT press, Cambridge.

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

BIRGÉ, L. and MASSART, P. (1993). Rates of convergence for minimum contrast estimators. *Probab. Theory Related Fields* **97** 113–150. MR1240719 https://doi.org/10.1007/BF01199316

BRADLEY, R. A. and TERRY, M. E. (1952). Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika* **39** 324–345. MR0070925 https://doi.org/10.2307/2334029

BRENNAN, M. and BRESLER, G. (2020). Reducibility and statistical-computational gaps from secret leakage. In *Proceedings of* 33*rd Conference on Learning Theory* (J. Abernethy and S. Agarwal, eds.). *Proceedings of Machine Learning Research*, *PMLR* **125** 648–847.

BRENNAN, M., BRESLER, G. and HULEIHEL, W. (2018). Reducibility and computational lower bounds for problems with planted sparse structure. In *Proceedings of the* 31*st Conference on Learning Theory* (S. Bubeck, V. Perchet and P. Rigollet, eds.). *Proceedings of Machine Learning Research*, *PMLR* **75** 48–166.

BRUNK, H. D. (1955). Maximum likelihood estimates of monotone parameters. *Ann. Math. Stat.* **26** 607–616. MR0073894 https://doi.org/10.1214/aoms/1177728420

CARPENTIER, A. and SCHLUETER, T. (2016). Learning relationships between data obtained independently. In *Proceedings of the* 19*th International Conference on Artificial Intelligence and Statistics* 658–666.

CHAN, S. and AIROLDI, E. (2014). A consistent histogram estimator for exchangeable graph models. In *International Conference on Machine Learning* 208–216.

CHANDUKALA, S. R., KIM, J., OTTER, T. and ALLENBY, G. M. (2008). *Choice Models in Marketing*: *Economic Assumptions*, *Challenges and Trends*. Now Publishers, Hanover.

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

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

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

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

CHEN, X., BENNETT, P. N., COLLINS-THOMPSON, K. and HORVITZ, E. (2013). Pairwise ranking aggregation in a crowdsourced setting. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining* 193–202. ACM, New York.

COLLIER, O. and DALALYAN, A. S. (2016). Minimax rates in permutation estimation for feature matching. *J. Mach. Learn. Res.* **17** 6. MR3482926

CRASWELL, N., ZOETER, O., TAYLOR, M. and RAMSEY, B. (2008). An experimental comparison of click position-bias models. In *Proceedings of the* 2008 *International Conference on Web Search and Data Mining* 87–94.

DAWID, A. P. and SKENE, A. M. (1979). Maximum likelihood estimation of observer error-rates using the em algorithm. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **28** 20–28.

DEGROOT, M. H. and GOEL, P. K. (1980). Estimation of the correlation coefficient from a broken random sample. *Ann. Statist.* **8** 264–278. MR0560728

DENG, H. and ZHANG, C.-H. (2020). Isotonic regression in multi-dimensional spaces and graphs. *Ann. Statist.* **48** 3672–3698. MR4185824 https://doi.org/10.1214/20-AOS1947

DWORK, C., KUMAR, R., NAOR, M. and SIVAKUMAR, D. (2001). Rank aggregation methods for the web. In *Proceedings of the* 10*th International Conference on World Wide Web* 613–622. ACM, New York.

DYKSTRA, R. L. and ROBERTSON, T. (1982). An algorithm for isotonic regression for two or more independent variables. *Ann. Statist.* **10** 708–716. MR0663427

FARIAS, V. F., JAGABATHULA, S. and SHAH, D. (2013). A nonparametric approach to modeling choice with limited data. *Manage. Sci.* **59** 305–322.

FEIGE, U. and KRAUTHGAMER, R. (2003). The probable value of the Lovász–Schrijver relaxations for maximum independent set. *SIAM J. Comput.* **32** 345–370. MR1969394 https://doi.org/10.1137/S009753970240118X

FISHBURN, P. C. (1973). Binary choice probabilities: On the varieties of stochastic transitivity. *J. Math. Psych.* **10** 327–352. MR0327330 https://doi.org/10.1016/0022-2496(73)90021-7

FLAMMARION, N., MAO, C. and RIGOLLET, P. (2019). Optimal rates of statistical seriation. *Bernoulli* **25** 623–653. MR3892331 https://doi.org/10.3150/17-bej1000

FOKIANOS, K., LEUCHT, A. and NEUMANN, M. H. (2020). On integrated L^1 convergence rate of an isotonic regression estimator for multivariate observations. *IEEE Trans. Inf. Theory* **66** 6389–6402. MR4173546 https://doi.org/10.1109/TIT.2020.3013390

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

GAO, C., LU, Y. and ZHOU, H. H. (2015). Rate-optimal graphon estimation. *Ann. Statist.* **43** 2624–2652. MR3405606 https://doi.org/10.1214/15-AOS1354

GHOSHDASTIDAR, D. and DUKKIPATI, A. (2017). Consistency of spectral hypergraph partitioning under planted partition model. *Ann. Statist.* **45** 289–315. MR3611493 https://doi.org/10.1214/16-AOS1453

HAN, Q. (2021+). Set structured global empirical risk minimizers are rate optimal in general dimensions. *Ann. Statist.* to appear.

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

HARDOUIN, J.-B. and MESBAH, M. (2004). Clustering binary variables in subscales using an extended Rasch model and Akaike information criterion. *Comm. Statist. Theory Methods* **33** 1277–1294. MR2069568 https://doi.org/10.1081/STA-120030149

HOPKINS, S. (2018). *Statistical Inference and the Sum of Squares Method*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—Cornell University. MR3864930

HSU, D. J., SHI, K. and SUN, X. (2017). Linear regression without correspondence. In *Advances in Neural Information Processing Systems* 1531–1540.

JERRUM, M. (1992). Large cliques elude the Metropolis process. *Random Structures Algorithms* **3** 347–359. MR1179827 https://doi.org/10.1002/rsa.3240030402

KIM, A. K. H. and SAMWORTH, R. J. (2016). Global rates of convergence in log-concave density estimation. *Ann. Statist.* **44** 2756–2779. MR3576560 https://doi.org/10.1214/16-AOS1480

KÖK, A. G., FISHER, M. L. and VAIDYANATHAN, R. (2015). Assortment planning: Review of literature and industry practice. In *Retail Supply Chain Management: Quantitative Models and Empirical Studies* (N. Agrawal and S. A. Smith, eds.) 175–236. Springer US, Boston, MA.

KOLDA, T. G., BADER, B. W. and KENNY, J. P. (2005). *Higher-Order Web Link Analysis Using Multilinear Algebra. Fifth IEEE International Conference on Data Mining* (*ICDM'*05). IEEE.

KRANTZ, D. H. (1965). The Scaling of Small and Large Color Differences Ph.D. thesis, University of Pennsylvania.

KUNISKY, D., WEIN, A. S. and BANDEIRA, A. S. (2019). Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. arXiv preprint. Available at arXiv:1907.11636.

KYNG, R., RAO, A. and SACHDEVA, S. (2015). Fast, provable algorithms for isotonic regression in all ℓ_p-norms. In *Advances in Neural Information Processing Systems* 2719–2727.

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

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

LIU, A. and MOITRA, A. (2020). Better algorithms for estimating non-parametric models in crowd-sourcing and rank aggregation. In *Proceedings of* 33*rd Conference on Learning Theory* (J. Abernethy and S. Agarwal, eds.). *Proceedings of Machine Learning Research*, *PMLR* **125** 2780–2829.

LUCE, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. Wiley, New York. MR0108411

LUO, Y. and ZHANG, A. R. (2020). Tensor clustering with planted structures: Statistical optimality and computational limits. arXiv preprint. Available at arXiv:2005.10743v2.

MA, R., CAI, T. T. and LI, H. (2021+). Optimal estimation of bacterial growth rates based on a permuted monotone matrix. *Biometrika*. to appear.

MA, Z. and WU, Y. (2015). Computational barriers in minimax submatrix detection. *Ann. Statist.* **43** 1089–1116. MR3346698 https://doi.org/10.1214/14-AOS1300

MAO, C., PANANJADY, A. and WAINWRIGHT, M. J. (2018). Breaking the $1/\sqrt{n}$ barrier: Faster rates for permutation-based models in polynomial time. In *Proceedings of the* 31*st Conference on Learning Theory* (S. Bubeck, V. Perchet and P. Rigollet, eds.). *Proceedings of Machine Learning Research*, *PMLR* **75** 2037–2042.

MAO, C., PANANJADY, A. and WAINWRIGHT, M. J. (2020). Towards optimal estimation of bivariate isotonic matrices with unknown permutations. *Ann. Statist.* **48** 3183–3205. MR4185805 https://doi.org/10.1214/19-AOS1925

MARSCHAK, J. and DAVIDSON, D. (1957). Experimental tests of stochastic decision theory Technical report, Cowles Foundation for Research in Economics, Yale University.

MCFADDEN, D. (1974). Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics* (P. Zarembka, ed.) 105–142. Academic Press, San Diego.

MCFADDEN, D. (1981). Econometric models of probabilistic choice. In *Structural Analysis of Discrete Data with Econometric Applications* (C. Manski and D. McFadden, eds.) 198–272. MIT press, Cambridge.

MCLAUGHLIN, D. H. and LUCE, R. D. (1965). Stochastic transitivity and cancellation of preferences between bitter-sweet solutions. *Psychon. Sci.* **2** 89–90.

MIRSKY, L. (1971). A dual of Dilworth's decomposition theorem. *Amer. Math. Monthly* **78** 876–877. MR0288054 https://doi.org/10.2307/2316481

MÖCKS, J. (1988). Decomposing event-related potentials: A new topographic components model. *Biol. Psychol.* **26** 199–215. https://doi.org/10.1016/0301-0511(88)90020-8

NEGAHBAN, S., OH, S., THEKUMPARAMPIL, K. K. and XU, J. (2018). Learning from comparisons and choices. *J. Mach. Learn. Res.* **19** 40. MR3862447

NEMIROVSKIĬ, A. S., POLYAK, B. T. and TSYBAKOV, A. B. (1985). The rate of convergence of nonparametric estimates of maximum likelihood type. *Problemy Peredachi Informatsii* **21** 17–33. MR0820705

NGUYEN, A., PIECH, C., HUANG, J. and GUIBAS, L. (2014). Codewebs: Scalable homework search for massive open online programming courses. In *Proceedings of the* 23*rd International Conference on World Wide Web* 491–502.

PANANJADY, A. and SAMWORTH, R. J (2022). Supplement to "Isotonic regression with unknown permutations: Statistics, computation and adaptation." https://doi.org/10.1214/21-AOS2107SUPP

PANANJADY, A., WAINWRIGHT, M. J. and COURTADE, T. A. (2017). Denoising linear models with permuted data. In 2017 *IEEE International Symposium on Information Theory* (*ISIT*) 446–450. IEEE.

PANANJADY, A., WAINWRIGHT, M. J. and COURTADE, T. A. (2018). Linear regression with shuffled data: Statistical and computational limits of permutation recovery. *IEEE Trans. Inf. Theory* **64** 3286–3300. MR3798377 https://doi.org/10.1109/TIT.2017.2776217

PANANJADY, A., MAO, C., MUTHUKUMAR, V., WAINWRIGHT, M. J. and COURTADE, T. A. (2020). Worst-case versus average-case design for estimation from partial pairwise comparisons. *Ann. Statist.* **48** 1072–1097. MR4102688 https://doi.org/10.1214/19-AOS1838

PITSOULIS, L. and PARDALOS, P. M. (2001). Quadratic assignment problem. In *Encyclopedia of Optimization* (C. A. Floudas and P. M. Pardalos, eds.) 2075–2107. Springer US, Boston, MA.

PLACKETT, R. L. (1975). The analysis of permutations. *J. Roy. Statist. Soc. Ser. C* **24** 193–202. MR0391338 https://doi.org/10.2307/2346567

RAKHLIN, A., SRIDHARAN, K. and TSYBAKOV, A. B. (2017). Empirical entropy, minimax regret and minimax risk. *Bernoulli* **23** 789–824. MR3606751 https://doi.org/10.3150/14-BEJ679

RIGOLLET, P. and WEED, J. (2019). Uncoupled isotonic regression via minimum Wasserstein deconvolution. *Inf. Inference* **8** 691–717. MR4045481 https://doi.org/10.1093/imaiai/iaz006

ROBERTSON, T. and WRIGHT, F. T. (1975). Consistency in generalized isotonic regression. *Ann. Statist.* **3** 350–362. MR0365871

SHAH, N. B., BALAKRISHNAN, S. and WAINWRIGHT, M. J. (2019a). Feeling the bern: Adaptive estimators for Bernoulli probabilities of pairwise comparisons. *IEEE Trans. Inf. Theory* **65** 4854–5874. MR3988527 https://doi.org/10.1109/TIT.2019.2903249

SHAH, N. B., BALAKRISHNAN, S. and WAINWRIGHT, M. J. (2019b). Low permutation-rank matrices: Structural properties and noisy completion. *J. Mach. Learn. Res.* **20** 101. MR3990455

SHAH, N. B., BALAKRISHNAN, S. and WAINWRIGHT, M. J. (2021). A permutation-based model for crowd labeling: Optimal estimation and robustness. *IEEE Trans. Inf. Theory* **67** 4162–4184.

SHAH, N. B., BALAKRISHNAN, S., GUNTUBOYINA, A. and WAINWRIGHT, M. J. (2017). Stochastically transitive models for pairwise comparisons: Statistical and computational issues. *IEEE Trans. Inf. Theory* **63** 934–959. MR3604649 https://doi.org/10.1109/TIT.2016.2634418

STOUT, Q. F. (2015). Isotonic regression for multiple independent variables. *Algorithmica* **71** 450–470. MR3331888 https://doi.org/10.1007/s00453-013-9814-z

TVERSKY, A. (1972). Elimination by aspects: A theory of choice. *Psychol. Rev.* **79** 281.

UNNIKRISHNAN, J., HAGHIGHATSHOAR, S. and VETTERLI, M. (2018). Unlabeled sensing with random linear measurements. *IEEE Trans. Inf. Theory* **64** 3237–3253. MR3798374 https://doi.org/10.1109/TIT.2018.2809002

VAN SCHUUR, W. H. (2003). Mokken scale analysis: Between the Guttman scale and parametric item response theory. *Polit. Anal.* **11** 139–163.

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

WANG, T., BERTHET, Q. and SAMWORTH, R. J. (2016). Statistical and computational trade-offs in estimation of sparse principal components. *Ann. Statist.* **44** 1896–1930. MR3546438 https://doi.org/10.1214/15-AOS1369

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

ZHANG, A. and XIA, D. (2018). Tensor SVD: Statistical and computational limits. *IEEE Trans. Inf. Theory* **64** 7311–7338. MR3876445 https://doi.org/10.1109/TIT.2018.2841377

ZHOU, D., HUANG, J. and SCHÖLKOPF, B. (2007). Learning with hypergraphs: Clustering, classification, and embedding. In *Advances in Neural Information Processing Systems* 1601–1608.

ZHOU, J., BHATTACHARYA, A., HERRING, A. H. and DUNSON, D. B. (2015). Bayesian factorizations of big sparse tensors. *J. Amer. Statist. Assoc.* **110** 1562–1576. MR3449055 https://doi.org/10.1080/01621459.2014.983233

The Annals of Statistics
2022, Vol. 50, No. 1, 351–375
https://doi.org/10.1214/21-AOS2109

ADMISSIBLE WAYS OF MERGING P-VALUES UNDER ARBITRARY DEPENDENCE

BY VLADIMIR VOVK[1], BIN WANG[2] AND RUODU WANG[3]

[1]*Department of Computer Science, Royal Holloway, University of London, v.vovk@rhul.ac.uk*
[2]*RCSDS, NCMIS, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, wangbin@amss.ac.cn*
[3]*Department of Statistics and Actuarial Science, University of Waterloo, wang@uwaterloo.ca*

Methods of merging several p-values into a single p-value are important in their own right and widely used in multiple hypothesis testing. This paper is the first to systematically study the admissibility (in Wald's sense) of p-merging functions and their domination structure, without any information on the dependence structure of the input p-values. As a technical tool, we use the notion of e-values, which are alternatives to p-values recently promoted by several authors. We obtain several results on the representation of admissible p-merging functions via e-values and on (in)admissibility of existing p-merging functions. By introducing new admissible p-merging functions, we show that some classic merging methods can be strictly improved to enhance power without compromising validity under arbitrary dependence.

REFERENCES

CHEN, Y., LIU, P., TAN, K. S. and WANG, R. (2020). Trade-off between validity and efficiency of merging p-values under arbitrary dependence. Preprint. Available at arXiv:2007.12366 [math.ST].

EFRON, B. (2010). *Large-Scale Inference*: *Empirical Bayes Methods for Estimation*, *Testing*, *and Prediction*. *Institute of Mathematical Statistics* (*IMS*) *Monographs* **1**. Cambridge Univ. Press, Cambridge. MR2724758 https://doi.org/10.1017/CBO9780511761362

FÖLLMER, H. and SCHIED, A. (2011). *Stochastic Finance*: *An Introduction in Discrete Time*, 3rd ed. de Gruyter, Berlin. MR2779313 https://doi.org/10.1515/9783110218053

GENOVESE, C. and WASSERMAN, L. (2004). A stochastic process approach to false discovery control. *Ann*. *Statist*. **32** 1035–1061. MR2065197 https://doi.org/10.1214/009053604000000283

GOEMAN, J. J., HEMERIK, J. and SOLARI, A. (2021). Only closed testing procedures are admissible for controlling false discovery proportions. *Ann*. *Statist*. **49** 1218–1238. MR4255125 https://doi.org/10.1214/20-aos1999

GOEMAN, J. J. and SOLARI, A. (2011). Multiple testing for exploratory research. *Statist*. *Sci*. **26** 584–597. MR2951390 https://doi.org/10.1214/11-STS356

GOOD, I. J. (1958). Significance tests in parallel and in series. *J*. *Amer*. *Statist*. *Assoc*. **53** 799–813. MR0103560

GRÜNWALD, P., DE HEIDE, R. and KOOLEN, W. M. (2020). Safe testing. Preprint. Available at arXiv:1906.07801 [math.ST].

HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. (1952). *Inequalities*, 2nd ed. Cambridge University Press, Cambridge. MR0046395

HOMMEL, G. (1983). Tests of the overall hypothesis for arbitrary dependence structures. *Biom*. *J*. **25** 423–430. MR0735888

JAKOBSONS, E., HAN, X. and WANG, R. (2016). General convex order on risk aggregation. *Scand*. *Actuar*. *J*. **8** 713–740. MR3507191 https://doi.org/10.1080/03461238.2015.1012223

LIU, Y. and XIE, J. (2020). Cauchy combination test: A powerful test with analytic p-value calculation under arbitrary dependency structures. *J*. *Amer*. *Statist*. *Assoc*. **115** 393–402. MR4078471 https://doi.org/10.1080/01621459.2018.1554485

RÜGER, B. (1978). Das maximale Signifikanzniveau des Tests: "Lehne H_0 ab, wenn k unter n gegebenen Tests zur Ablehnung führen". *Metrika* **25** 171–178. MR0526476 https://doi.org/10.1007/BF02204362

RÜSCHENDORF, L. (2009). On the distributional transform, Sklar's theorem, and the empirical copula process. *J*. *Statist*. *Plann*. *Inference* **139** 3921–3927. MR2553778 https://doi.org/10.1016/j.jspi.2009.05.030

MSC2020 subject classifications. Primary 62G10, 62C15; secondary 62F03.
Key words and phrases. Multiple hypothesis testing, global null, duality, admissible decisions, e-values.

RÜSCHENDORF, L. (2013). *Mathematical Risk Analysis*. *Springer Series in Operations Research and Financial Engineering*. Springer, Heidelberg. MR3051756 https://doi.org/10.1007/978-3-642-33590-7

SARKAR, S. K. (1998). Some probability inequalities for ordered MTP_2 random variables: A proof of the Simes conjecture. *Ann*. *Statist*. **26** 494–504. MR1626047 https://doi.org/10.1214/aos/1028144846

SHAFER, G. (2021). Testing by betting: A strategy for statistical and scientific communication. *J*. *Roy*. *Statist*. *Soc*. *Ser*. *A* **184** 407–478. MR4255905 https://doi.org/10.1111/rssa.12647

SIMES, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73** 751–754. MR0897872 https://doi.org/10.1093/biomet/73.3.751

VOVK, V. and WANG, R. (2020a). A class of ie-merging functions. Preprint. Available at arXiv:2007.06382 [math.ST].

VOVK, V. and WANG, R. (2020b). Combining p-values via averaging. *Biometrika* **107** 791–808. MR4186488 https://doi.org/10.1093/biomet/asaa027

VOVK, V. and WANG, R. (2021). E-values: Calibration, combination and applications. *Ann*. *Statist*. **49** 1736–1754. MR4298879 https://doi.org/10.1214/20-AOS2020

VOVK, V., WANG, B. and WANG, R. (2022). Supplement to "Admissible ways of merging p-values under arbitrary dependence." https://doi.org/10.1214/21-AOS2109SUPP

WANG, J. and OWEN, A. B. (2019). Admissibility in partial conjunction testing. *J*. *Amer*. *Statist*. *Assoc*. **114** 158–168. MR3941245 https://doi.org/10.1080/01621459.2017.1385465

WANG, B. and WANG, R. (2011). The complete mixability and convex minimization problems with monotone marginal densities. *J*. *Multivariate Anal*. **102** 1344–1360. MR2819953 https://doi.org/10.1016/j.jmva.2011.05.002

WANG, B. and WANG, R. (2016). Joint mixability. *Math*. *Oper*. *Res*. **41** 808–826. MR3520751 https://doi.org/10.1287/moor.2015.0755

WILSON, D. J. (2019). The harmonic mean p-value for combining dependent tests. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **116** 1195–1200. MR3904688 https://doi.org/10.1073/pnas.1814092116

WILSON, D. J. (2020). Generalized mean p-values for combining dependent tests: Comparison of generalized central limit theorem and robust risk analysis. *Wellcome Open Res*. **5** 1–21.

The Annals of Statistics
2022, Vol. 50, No. 1, 376–406
https://doi.org/10.1214/21-AOS2111

HIGH-DIMENSIONAL ASYMPTOTICS OF LIKELIHOOD RATIO TESTS IN THE GAUSSIAN SEQUENCE MODEL UNDER CONVEX CONSTRAINTS

BY QIYANG HAN[1], BODHISATTVA SEN[2] AND YANDI SHEN[3]

[1]*Department of Statistics, Rutgers University, qh85@stat.rutgers.edu*
[2]*Department of Statistics, Columbia University, bodhi@stat.columbia.edu*
[3]*Department of Statistics, University of Washington, ydshen@uw.edu*

In the Gaussian sequence model $Y = \mu + \xi$, we study the likelihood ratio test (LRT) for testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \in K$, where $\mu_0 \in K$, and K is a closed convex set in $\mathbb{R}^n$. In particular, we show that under the null hypothesis, normal approximation holds for the log-likelihood ratio statistic for a general pair (μ_0, K), in the high-dimensional regime where the estimation error of the associated least squares estimator diverges in an appropriate sense. The normal approximation further leads to a precise characterization of the power behavior of the LRT in the high-dimensional regime. These characterizations show that the power behavior of the LRT is in general nonuniform with respect to the Euclidean metric, and illustrate the conservative nature of existing minimax optimality and suboptimality results for the LRT. A variety of examples, including testing in the orthant/circular cone, isotonic regression, Lasso and testing parametric assumptions versus shape-constrained alternatives, are worked out to demonstrate the versatility of the developed theory.

REFERENCES

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

[2] ARIAS-CASTRO, E., CANDÈS, E. J. and PLAN, Y. (2011). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *Ann. Statist.* **39** 2533–2556. MR2906877 https://doi.org/10.1214/11-AOS910

[3] AZZALINI, A. and BOWMAN, A. (1993). On the use of nonparametric regression for checking linear relationships. *J. Roy. Statist. Soc. Ser. B* **55** 549–557. MR1224417

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

[5] BARAUD, Y., HUET, S. and LAURENT, B. (2005). Testing convex hypotheses on the mean of a Gaussian vector. Application to testing qualitative hypotheses on a regression function. *Ann. Statist.* **33** 214–257. MR2157802 https://doi.org/10.1214/009053604000000896

[6] BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions. The Theory and Application of Isotonic Regression. Wiley Series in Probability and Mathematical Statistics*. Wiley, London-Sydney. MR0326887

[7] BARTHOLOMEW, D. J. (1959). A test of homogeneity for ordered alternatives. *Biometrika* **46** 36–48. MR0104312 https://doi.org/10.1093/biomet/46.1-2.36

[8] BARTHOLOMEW, D. J. (1959). A test of homogeneity for ordered alternatives. II. *Biometrika* **46** 328–335. MR0112204 https://doi.org/10.1093/biomet/46.3-4.328

[9] BARTHOLOMEW, D. J. (1961). Ordered tests in the analysis of variance. *Biometrika* **48** 325–332. MR0132660 https://doi.org/10.1093/biomet/48.3-4.325

[10] BARTHOLOMEW, D. J. (1961). A test of homogeneity of means under restricted alternatives. *J. Roy. Statist. Soc. Ser. B* **23** 239–281. MR0143288

MSC2020 subject classifications. 60F17, 62E17.
Key words and phrases. Central limit theorem, isotonic regression, Lasso, normal approximation, power analysis, projection onto a closed convex set, second-order Poincaré inequalities, shape constraint.

[11] BELLEC, P. C. and ZHANG, C.-H. (2021). Second order Stein: Sure for sure and other applications in high-dimensional inference. *Ann. Statist.* (to appear). Available at arXiv:1811.04121.

[12] BESSON, O. (2006). Adaptive detection of a signal whose signature belongs to a cone. In *Fourth IEEE Workshop on Sensor Array and Multichannel Processing*, 2006 409–413. IEEE, Los Alamitos.

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

[14] BOUSQUET, O. (2003). Concentration inequalities for sub-additive functions using the entropy method. In *Stochastic Inequalities and Applications*. *Progress in Probability* **56** 213–247. Birkhäuser, Basel. MR2073435

[15] CARPENTIER, A. (2015). Testing the regularity of a smooth signal. *Bernoulli* **21** 465–488. MR3322327 https://doi.org/10.3150/13-BEJ575

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

[17] CARPENTIER, A., COLLIER, O., COMMINGES, L., TSYBAKOV, A. B. and WANG, Y. (2020). Estimation of the l_2-norm and testing in sparse linear regression with unknown variance. arXiv preprint. Available at arXiv:2010.13679.

[18] CARPENTIER, A. and VERZELEN, N. (2021). Optimal sparsity testing in linear regression model. *Bernoulli* **27** 727–750. MR4255213 https://doi.org/10.3150/20-bej1224

[19] CHATTERJEE, A. and LAHIRI, S. N. (2011). Bootstrapping lasso estimators. *J. Amer. Statist. Assoc.* **106** 608–625. MR2847974 https://doi.org/10.1198/jasa.2011.tm10159

[20] CHATTERJEE, S. (2009). Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields* **143** 1–40. MR2449121 https://doi.org/10.1007/s00440-007-0118-6

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

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

[23] CHERNOFF, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Stat.* **25** 573–578. MR0065087 https://doi.org/10.1214/aoms/1177728725

[24] CHRISTENSEN, R. and SUN, S. K. (2010). Alternative goodness-of-fit tests for linear models. *J. Amer. Statist. Assoc.* **105** 291–301. MR2656053 https://doi.org/10.1198/jasa.2009.tm08697

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

[26] COMMINGES, L. and DALALYAN, A. S. (2013). Minimax testing of a composite null hypothesis defined via a quadratic functional in the model of regression. *Electron. J. Stat.* **7** 146–190. MR3020417 https://doi.org/10.1214/13-EJS766

[27] COX, D., KOH, E., WAHBA, G. and YANDELL, B. S. (1988). Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *Ann. Statist.* **16** 113–119. MR0924859 https://doi.org/10.1214/aos/1176350693

[28] DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. MR2065195 https://doi.org/10.1214/009053604000000265

[29] DUROT, C. (2007). On the $\mathbb{L}_p$-error of monotonicity constrained estimators. *Ann. Statist.* **35** 1080–1104. MR2341699 https://doi.org/10.1214/009053606000001497

[30] DUROT, C. and TOCQUET, A.-S. (2001). Goodness of fit test for isotonic regression. *ESAIM Probab. Stat.* **5** 119–140. MR1875667 https://doi.org/10.1051/ps:2001105

[31] DYKSTRA, R. (1991). Asymptotic normality for chi-bar-square distributions. *Canad. J. Statist.* **19** 297–306. MR1144147 https://doi.org/10.2307/3315395

[32] EUBANK, R. L. and SPIEGELMAN, C. H. (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques. *J. Amer. Statist. Assoc.* **85** 387–392. MR1141739

[33] FAN, J. and HUANG, L.-S. (2001). Goodness-of-fit tests for parametric regression models. *J. Amer. Statist. Assoc.* **96** 640–652. MR1946431 https://doi.org/10.1198/016214501753168316

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

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

[36] GRECO, M., GINI, F. and FARINA, A. (2008). Radar detection and classification of jamming signals belonging to a cone class. *IEEE Trans*. *Signal Process*. **56** 1984–1993. MR2518695 https://doi.org/10.1109/TSP.2007.909326

[37] GROENEBOOM, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, *Vol*. *II* (*Berkeley*, *Calif*., 1983). *Wadsworth Statist*./*Probab*. *Ser*. 539–555. Wadsworth, Belmont, CA. MR0822052

[38] GROENEBOOM, P., HOOGHIEMSTRA, G. and LOPUHAÄ, H. P. (1999). Asymptotic normality of the L_1 error of the Grenander estimator. *Ann*. *Statist*. **27** 1316–1347. MR1740109 https://doi.org/10.1214/aos/1017938928

[39] GUERRE, E. and LAVERGNE, P. (2005). Data-driven rate-optimal specification testing in regression models. *Ann*. *Statist*. **33** 840–870. MR2163161 https://doi.org/10.1214/009053604000001200

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

[41] HAN, Q., JIANG, T. and SHEN, Y. (2021). A general method for power analysis in testing high dimensional covariance matrices. arXiv preprint. Available at arXiv:2101.11086.

[42] HAN, Q., SEN, B. and SHEN, Y. (2022). Supplement to "High dimensional asymptotics of likelihood ratio tests in the Gaussian sequence model under convex constraints." https://doi.org/10.1214/21-AOS2111SUPP

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

[44] HÄRDLE, W. and MAMMEN, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann*. *Statist*. **21** 1926–1947. MR1245774 https://doi.org/10.1214/aos/1176349403

[45] INGSTER, Y. I. and SUSLINA, I. A. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*. *Lecture Notes in Statistics* **169**. Springer, New York. MR1991446 https://doi.org/10.1007/978-0-387-21580-8

[46] INGSTER, Y. I., TSYBAKOV, A. B. and VERZELEN, N. (2010). Detection boundary in sparse regression. *Electron*. *J*. *Stat*. **4** 1476–1526. MR2747131 https://doi.org/10.1214/10-EJS589

[47] JUDITSKY, A. and NEMIROVSKI, A. (2002). On nonparametric tests of positivity/monotonicity/convexity. *Ann*. *Statist*. **30** 498–527. MR1902897 https://doi.org/10.1214/aos/1021379863

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

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

[50] KUDÔ, A. and CHOI, J. R. (1975). A generalized multivariate analogue of the one sided test. *Mem*. *Fac*. *Sci*., *Kyushu Univ*., *Ser*. *A* **29** 303–328. MR0386154 https://doi.org/10.2206/kyushumfs.29.303

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

[52] MCCOY, M. B. and TROPP, J. A. (2014). From Steiner formulas for cones to concentration of intrinsic volumes. *Discrete Comput*. *Geom*. **51** 926–963. MR3216671 https://doi.org/10.1007/s00454-014-9595-4

[53] MENÉNDEZ, J. A., RUEDA, C. and SALVADOR, B. (1992). Dominance of likelihood ratio tests under cone constraints. *Ann*. *Statist*. **20** 2087–2099. MR1193327 https://doi.org/10.1214/aos/1176348904

[54] MENÉNDEZ, J. A., RUEDA, C. and SALVADOR, B. (1992). Testing nonoblique hypotheses. *Comm*. *Statist*. *Theory Methods* **21** 471–484. MR1158571 https://doi.org/10.1080/03610929208830789

[55] MENÉNDEZ, J. A. and SALVADOR, B. (1991). Anomalies of the likelihood ratio tests for testing restricted hypotheses. *Ann*. *Statist*. **19** 889–898. MR1105850 https://doi.org/10.1214/aos/1176348126

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

[57] MEYER, M. C. (2003). A test for linear versus convex regression function using shape-restricted regression. *Biometrika* **90** 223–232. MR1966562 https://doi.org/10.1093/biomet/90.1.223

[58] MUKHERJEE, R. and SEN, S. (2020). On minimax exponents of sparse testing. arXiv preprint. Available at arXiv:2003.00570.

[59] NEUMEYER, N. and VAN KEILEGOM, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *J*. *Multivariate Anal*. **101** 1067–1078. MR2595293 https://doi.org/10.1016/j.jmva.2010.01.007

[60] NICKL, R. and VAN DE GEER, S. (2013). Confidence sets in sparse regression. *Ann*. *Statist*. **41** 2852–2876. MR3161450 https://doi.org/10.1214/13-AOS1170

[61] NOURDIN, I. and PECCATI, G. (2012). *Normal Approximations with Malliavin Calculus*. *Cambridge Tracts in Mathematics* **192**. Cambridge Univ. Press, Cambridge. From Stein's method to universality. MR2962301 https://doi.org/10.1017/CBO9781139084659

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

[63] ROBERTSON, T. and WEGMAN, E. J. (1978). Likelihood ratio tests for order restrictions in exponential families. *Ann. Statist.* **6** 485–505. MR0471147

[64] ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference. Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, Chichester. MR0961262

[65] ROCKAFELLAR, R. T. (1997). *Convex Analysis. Princeton Landmarks in Mathematics*. Princeton Univ. Press, Princeton, NJ. Reprint of the 1970 original, Princeton Paperbacks. MR1451876

[66] SEN, B. and MEYER, M. (2017). Testing against a linear regression model using ideas from shape-restricted estimation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 423–448. MR3611753 https://doi.org/10.1111/rssb.12178

[67] SHAH, R. D. and BÜHLMANN, P. (2018). Goodness-of-fit tests for high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 113–135. MR3744714 https://doi.org/10.1111/rssb.12234

[68] SHAPIRO, A. (1985). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika* **72** 133–144. MR0790208 https://doi.org/10.1093/biomet/72.1.133

[69] SHAPIRO, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *Int. Stat. Rev.* **56** 49–62. MR0963140 https://doi.org/10.2307/1403361

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

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

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

[73] VAN DER VAART, A. (2002). Semiparametric statistics. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1999). *Lecture Notes in Math.* **1781** 331–457. Springer, Berlin. MR1915446

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

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

[76] VERZELEN, N. and VILLERS, F. (2010). Goodness-of-fit tests for high-dimensional Gaussian linear models. *Ann. Statist.* **38** 704–752. MR2604699 https://doi.org/10.1214/08-AOS629

[77] WARRACK, G. and ROBERTSON, T. (1984). A likelihood ratio test regarding two nested but oblique order-restricted hypotheses. *J. Amer. Statist. Assoc.* **79** 881–886. MR0770282

[78] WEI, Y., WAINWRIGHT, M. J. and GUNTUBOYINA, A. (2019). The geometry of hypothesis testing over convex cones: Generalized likelihood ratio tests and minimax radii. *Ann. Statist.* **47** 994–1024. MR3909958 https://doi.org/10.1214/18-AOS1701

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

The Annals of Statistics
2022, Vol. 50, No. 1, 407–429
https://doi.org/10.1214/21-AOS2113

TESTING EQUIVALENCE OF CLUSTERING

BY CHAO GAO[1] AND ZONGMING MA[2]

[1]*Department of Statistics, University of Chicago, chaogao@galton.uchicago.edu*
[2]*Department of Statistics and Data Science, University of Pennsylvania, zongming@wharton.upenn.edu*

In this paper, we test whether two data sets measured on the same set of subjects share a common clustering structure. As a leading example, we focus on comparing clustering structures in two independent random samples from two deterministic two-component mixtures of multivariate Gaussian distributions. Mean parameters of these Gaussian distributions are treated as potentially unknown nuisance parameters and are allowed to differ. Assuming knowledge of mean parameters, we first determine the phase diagram of the testing problem over the entire range of signal-to-noise ratios by providing both lower bounds and tests that achieve them. When nuisance parameters are unknown, we propose tests that achieve the detection boundary adaptively as long as ambient dimensions of the data sets grow at a sublinear rate with the sample size.

REFERENCES

[1] ARIAS-CASTRO, E., HUANG, R. and VERZELEN, N. (2020). Detection of sparse positive dependence. *Electron. J. Stat.* **14** 702–730. MR4057605 https://doi.org/10.1214/19-EJS1675

[2] 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.

[3] BINKIEWICZ, N., VOGELSTEIN, J. T. and ROHE, K. (2017). Covariate-assisted spectral clustering. *Biometrika* **104** 361–377. MR3698259 https://doi.org/10.1093/biomet/asx008

[4] BIRNBAUM, A., JOHNSTONE, I. M., NADLER, B. and PAUL, D. (2013). Minimax bounds for sparse PCA with noisy high-dimensional data. *Ann. Statist.* **41** 1055–1084. MR3113803 https://doi.org/10.1214/12-AOS1014

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

[6] CAI, T. T., MA, Z. and WU, Y. (2013). Sparse PCA: Optimal rates and adaptive estimation. *Ann. Statist.* **41** 3074–3110. MR3161458 https://doi.org/10.1214/13-AOS1178

[7] CAI, T. T., SUN, W. and WANG, W. (2019). Covariate-assisted ranking and screening for large-scale two-sample inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 187–234. MR3928141

[8] CAI, T. T. and WU, Y. (2014). Optimal detection of sparse mixtures against a given null distribution. *IEEE Trans. Inf. Theory* **60** 2217–2232. MR3181520 https://doi.org/10.1109/TIT.2014.2304295

[9] CURTIS, C., SHAH, S. P., CHIN, S.-F. et al. (2012). The genomic and transcriptomic architecture of 2000 breast tumours reveals novel subgroups. *Nature* **486** 346.

[10] DESHPANDE, Y., SEN, S., MONTANARI, A. and MOSSEL, E. (2018). Contextual stochastic block models. In *Advances in Neural Information Processing Systems* 8581–8593.

[11] DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. MR2065195 https://doi.org/10.1214/009053604000000265

[12] DONOHO, D. and JIN, J. (2015). Higher criticism for large-scale inference, especially for rare and weak effects. *Statist. Sci.* **30** 1–25. MR3317751 https://doi.org/10.1214/14-STS506

[13] GAO, C. and MA, Z. (2022). Supplement to "Testing equivalence of clustering." https://doi.org/10.1214/21-AOS2113SUPP

[14] GAO, L. L., BIEN, J. and WITTEN, D. (2020). Are clusterings of multiple data views independent? *Biostatistics* **21** 692–708. MR4164052 https://doi.org/10.1093/biostatistics/kxz001

MSC2020 subject classifications. Primary 62C20, 62H15; secondary 62H30.
Key words and phrases. Minimax testing error, sparse mixture, phase transition, discrete structure inference, high-dimensional statistics.

[15] GAO, L. L., WITTEN, D. and BIEN, J. (2019). Testing for association in multi-view network data. Preprint. Available at arXiv:1909.11640.

[16] INGSTER, Y. I. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods Statist.* **6** 47–69. MR1456646

[17] INGSTER, Y. I. and SUSLINA, I. A. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models. Lecture Notes in Statistics* **169**. Springer, New York. MR1991446 https://doi.org/10.1007/978-0-387-21580-8

[18] JIN, J. and KE, Z. T. (2016). Rare and weak effects in large-scale inference: Methods and phase diagrams. *Statist. Sinica* **26** 1–34. MR3468343

[19] JIN, J., KE, Z. T. and WANG, W. (2017). Phase transitions for high dimensional clustering and related problems. *Ann. Statist.* **45** 2151–2189. MR3718165 https://doi.org/10.1214/16-AOS1522

[20] JIN, J. and WANG, W. (2016). Influential features PCA for high dimensional clustering. *Ann. Statist.* **44** 2323–2359. MR3576543 https://doi.org/10.1214/15-AOS1423

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

[22] LU, Y. and ZHOU, H. H. (2016). Statistical and computational guarantees of lloyd's algorithm and its variants. Preprint. Available at arXiv:1612.02099.

[23] MA, H., YANG, H., LYU, M. R. and KING, I. (2008). Sorec: Social recommendation using probabilistic matrix factorization. In *Proceedings of the* 17*th ACM Conference on Information and Knowledge Management* 931–940. ACM, New York.

[24] MA, H., ZHOU, D., LIU, C., LYU, M. R. and KING, I. (2011). Recommender systems with social regularization. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining* 287–296. ACM, New York.

[25] MA, Z. (2013). Sparse principal component analysis and iterative thresholding. *Ann. Statist.* **41** 772–801. MR3099121 https://doi.org/10.1214/13-AOS1097

[26] MO, Q., WANG, S., SESHAN, V. E., OLSHEN, A. B., SCHULTZ, N., SANDER, C., POWERS, R. S., LADANYI, M. and SHEN, R. (2013). Pattern discovery and cancer gene identification in integrated cancer genomic data. *Proc. Natl. Acad. Sci. USA* **110** 4245–4250.

[27] ROQUAIN, E. and VERZELEN, N. (2019). On using empirical null distribution in Benjamini–Hochberg procedure. Preprint. Available at arXiv:1912.03109.

[28] SHEN, R., OLSHEN, A. B. and LADANYI, M. (2009). Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics* **25** 2906–2912.

[29] VERZELEN, N. and ARIAS-CASTRO, E. (2017). Detection and feature selection in sparse mixture models. *Ann. Statist.* **45** 1920–1950. MR3718157 https://doi.org/10.1214/16-AOS1513

[30] VU, V. Q. and LEI, J. (2013). Minimax sparse principal subspace estimation in high dimensions. *Ann. Statist.* **41** 2905–2947. MR3161452 https://doi.org/10.1214/13-AOS1151

[31] WANG, H., TERROVITIS, M. and MAMOULIS, N. (2013). Location recommendation in location-based social networks using user check-in data. In *Proceedings of the* 21*st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* 374–383. ACM, New York.

[32] ZHAO, S. D., CAI, T. T. and LI, H. (2017). Optimal detection of weak positive latent dependence between two sequences of multiple tests. *J. Multivariate Anal.* **160** 169–184. MR3688697 https://doi.org/10.1016/j.jmva.2017.06.009

[33] ZHAO, S. D. and NGUYEN, Y. T. (2020). Nonparametric false discovery rate control for identifying simultaneous signals. *Electron. J. Stat.* **14** 110–142. MR4047996 https://doi.org/10.1214/19-EJS1663

The Annals of Statistics
2022, Vol. 50, No. 1, 430–459
https://doi.org/10.1214/21-AOS2115

MINIMAX NONPARAMETRIC ESTIMATION OF PURE QUANTUM STATES

BY SAMRIDDHA LAHIRY[1] AND MICHAEL NUSSBAUM[2]

[1]*Department of Statistics and Data Science, Cornell University, sl2938@cornell.edu*
[2]*Department of Mathematics, Cornell University, nussbaum@math.cornell.edu*

In classical statistics, Pinsker's theorem provides an exact asymptotic minimax bound in nonparametric estimation, improving upon optimal rates of convergence results. We obtain a quantum version of the theorem by establishing asymptotic minimax results for estimation of the displacement vector in a quantum Gaussian white noise model, given by a sequence of shifted vacuum states. Analogous results are then obtained for estimation of a general pure state from an ensemble of identically prepared, independent quantum systems, using the recently established local asymptotic equivalence to the quantum Gaussian white noise model. Optimality holds with respect to the most fundamental distance measure for quantum states, that is, trace norm distance, and in a true quantum sense, allowing for all possible measurements. Adaptive estimators are also obtained for the above cases. As an application, we obtain asymptotic minimax adaptive estimators for Wigner functions of pure states.

REFERENCES

[1] ALQUIER, P., MEZIANI, K. and PEYRÉ, G. (2013). Adaptive estimation of the density matrix in quantum homodyne tomography with noisy data. *Inverse Probl.* **29** 075017, 20. MR3080478 https://doi.org/10.1088/0266-5611/29/7/075017

[2] ARTILES, L. M., GILL, R. D. and GUŢĂ, M. I. (2005). An invitation to quantum tomography. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 109–134. MR2136642 https://doi.org/10.1111/j.1467-9868.2005.00491.x

[3] AUBRY, J.-M., BUTUCEA, C. and MEZIANI, K. (2009). State estimation in quantum homodyne tomography with noisy data. *Inverse Probl.* **25** 015003, 22. MR2465335 https://doi.org/10.1088/0266-5611/25/1/015003

[4] AUDENAERT, K. M. R., CALSAMIGLIA, J., MUÑOZ TAPIA, R., BAGAN, E., MASANES, L., ACIN, A. and VERSTRAETE, F. (2007). Discriminating states: The quantum Chernoff bound. *Phys. Rev. Lett.* **98** 160501.

[5] AUDENAERT, K. M. R., NUSSBAUM, M., SZKOŁA, A. and VERSTRAETE, F. (2008). Asymptotic error rates in quantum hypothesis testing. *Comm. Math. Phys.* **279** 251–283. MR2377635 https://doi.org/10.1007/s00220-008-0417-5

[6] BARNDORFF-NIELSEN, O. E., GILL, R. D. and JUPP, P. E. (2003). On quantum statistical inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 775–816. With discussion and reply by the authors. MR2017871 https://doi.org/10.1111/1467-9868.00415

[7] BELITSER, E. N. and LEVIT, B. Y. (1995). On minimax filtering over ellipsoids. *Math. Methods Statist.* **4** 259–273. MR1355248

[8] BRAUNSTEIN, S. L. and CAVES, C. M. (1994). Statistical distance and the geometry of quantum states. *Phys. Rev. Lett.* **72** 3439–3443. MR1274631 https://doi.org/10.1103/PhysRevLett.72.3439

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

[10] BUTUCEA, C., GUŢĂ, M. and ARTILES, L. (2007). Minimax and adaptive estimation of the Wigner function in quantum homodyne tomography with noisy data. *Ann. Statist.* **35** 465–494. MR2336856 https://doi.org/10.1214/009053606000001488

MSC2020 subject classifications. Primary 62G05; secondary 62G20, 62C20, 81P50.

Key words and phrases. Pinsker's theorem, pure quantum state, Hermite–Sobolev class, trace norm distance, quantum Gaussian sequence model, quantum Bayes estimation, local asymptotic equivalence, adaptive estimation, block Stein method, Wigner function.

[11] Butucea, C., Guţă, M. and Nussbaum, M. (2018). Local asymptotic equivalence of pure states ensembles and quantum Gaussian white noise. *Ann. Statist.* **46** 3676–3706. MR3852665 https://doi.org/10.1214/17-AOS1672

[12] Butucea, C., Guţă, M. and Nussbaum, M. (2018). Supplementary material to "Local asymptotic equivalence of pure states ensembles and quantum Gaussian white noise". *Ann. Statist.* **46**. MR3852665 https://doi.org/10.1214/17-AOS1672SUPP

[13] Cai, T., Kim, D., Wang, Y., Yuan, M. and Zhou, H. H. (2016). Optimal large-scale quantum state tomography with Pauli measurements. *Ann. Statist.* **44** 682–712. MR3476614 https://doi.org/10.1214/15-AOS1382

[14] Gill, R. D. and Guţă, M. I. (2013). On asymptotic quantum statistical inference. In *From Probability to Statistics and Back: High-Dimensional Models and Processes. Inst. Math. Stat.* (*IMS*) *Collect.* **9** 105–127. IMS, Beachwood, OH. MR3186752 https://doi.org/10.1214/12-IMSCOLL909

[15] Guţă, M. and Jenčová, A. (2007). Local asymptotic normality in quantum statistics. *Comm. Math. Phys.* **276** 341–379. MR2346393 https://doi.org/10.1007/s00220-007-0340-1

[16] Guţă, M. and Kahn, J. (2006). Local asymptotic normality for qubit states. *Phys. Rev. A* (3) **73** 052108, 15. MR2229156 https://doi.org/10.1103/PhysRevA.73.052108

[17] Hayashi, M. (2017). *Quantum Information Theory: Mathematical Foundation*, 2nd ed. *Graduate Texts in Physics*. Springer, Berlin. MR3558531 https://doi.org/10.1007/978-3-662-49725-8

[18] Helstrom, C. W. (1976). *Quantum Detection and Estimation Theory*. Academic Press, New York.

[19] Holevo, A. (2011). *Probabilistic and Statistical Aspects of Quantum Theory*, 2nd ed. *Quaderni/Monographs* **1**. Edizioni della Normale, Pisa. With a foreword from the second Russian edition by K. A. Valiev. MR2797301 https://doi.org/10.1007/978-88-7642-378-9

[20] Holevo, A. S. (1973). Optimal quantum measurements. *Theoret. Math. Phys.* **17** 1172–1177. MR0468885 https://doi.org/10.1007/BF01037594

[21] Holevo, A. S. (1978). *Investigations in the General Theory of Statistical Decisions*. Amer. Math. Soc., Providence, RI. Translated from the Russian by Lisa Rosenblatt, Proc. Steklov Inst. Math., **1**978, no. 3. MR0652516

[22] Holevo, A. S. (2001). *Statistical Structure of Quantum Theory. Lecture Notes in Physics. Monographs* **67**. Springer, Berlin. MR1889193 https://doi.org/10.1007/3-540-44998-1

[23] Kahn, J. and Guţă, M. (2009). Local asymptotic normality for finite dimensional quantum systems. *Comm. Math. Phys.* **289** 597–652. MR2506764 https://doi.org/10.1007/s00220-009-0787-3

[24] Koltchinskii, V. (2011). Von Neumann entropy penalization and low-rank matrix estimation. *Ann. Statist.* **39** 2936–2973. MR3012397 https://doi.org/10.1214/11-AOS926

[25] Koltchinskii, V. and Xia, D. (2015). Optimal estimation of low rank density matrices. *J. Mach. Learn. Res.* **16** 1757–1792. MR3417798

[26] Lahiry, S. and Nussbaum, M. (2022). Supplement to "Minimax nonparametric estimation of pure quantum states." https://doi.org/10.1214/21-AOS2115SUPP

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

[28] Leonhardt, U. (1997). *Measuring the Quantum State of Light*. Cambridge Univ. Press.

[29] Li, K. (2016). Discriminating quantum states: The multiple Chernoff distance. *Ann. Statist.* **44** 1661–1679. MR3519936 https://doi.org/10.1214/16-AOS1436

[30] Lounici, K., Meziani, K. and Peyré, G. (2018). Adaptive sup-norm estimation of the Wigner function in noisy quantum homodyne tomography. *Ann. Statist.* **46** 1318–1351. MR3798005 https://doi.org/10.1214/17-AOS1586

[31] Méziani, K. (2007). Nonparametric estimation of the purity of a quantum state in quantum homodyne tomography with noisy data. *Math. Methods Statist.* **16** 354–368. MR2378280 https://doi.org/10.3103/S1066530707040047

[32] Meziani, K. (2008). Nonparametric goodness-of fit testing in quantum homodyne tomography with noisy data. *Electron. J. Stat.* **2** 1195–1223. MR2461899 https://doi.org/10.1214/08-EJS286

[33] Naulet, Z. and Barat, É. (2017). Bayesian nonparametric estimation for quantum homodyne tomography. *Electron. J. Stat.* **11** 3595–3632. MR3709864 https://doi.org/10.1214/17-EJS1322

[34] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24** 2399–2430. MR1425959 https://doi.org/10.1214/aos/1032181160

[35] Nussbaum, M. (1999). Minimax risk: Pinsker bound. In *Encyclopedia of Statistical Sciences, Update Volume* **3** 451–460. Wiley, New York. MR1679440 https://doi.org/10.1002/0471667196.ess1098.pub2

[36] Nussbaum, M. and Szkoła, A. (2009). The Chernoff lower bound for symmetric quantum hypothesis testing. *Ann. Statist.* **37** 1040–1057. MR2502660 https://doi.org/10.1214/08-AOS593

[37] PARTHASARATHY, K. R. (1992). *An Introduction to Quantum Stochastic Calculus*. *Modern Birkhäuser Classics*. Birkhäuser/Springer Basel AG, Basel. [2012 reprint of the 1992 original] [MR1164866]. MR3012668

[38] PINSKER, M. S.(1980). Optimal filtering of square integrable signals in Gaussian white noise. *Problems Inform. Transmission* **16** no. 2, 120-133; translated from *Problemy Peredachi Informatsii* **16** (1980), no. 2, 52-68 (Russian). MR0624591

[39] TSYBAKOV, A. B. (1997). Asymptotically efficient estimation of a signal in L_2 under general loss functions. *Problems Inform. Transmission* **33** no. 1, 78-88; translated from *Problemy Peredachi Informatsii* **33** (1997), no. 1, 94-106 (Russian). MR1441532

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

[41] WANG, Y. and SONG, X. (2020). Quantum science and quantum technology. *Statist. Sci.* **35** 51–74. MR4071358 https://doi.org/10.1214/19-STS745

[42] YAMAGATA, K., FUJIWARA, A. and GILL, R. D. (2013). Quantum local asymptotic normality based on a new quantum likelihood ratio. *Ann. Statist.* **41** 2197–2217. MR3127863 https://doi.org/10.1214/13-AOS1147

The Annals of Statistics
2022, Vol. 50, No. 1, 460–486
https://doi.org/10.1214/21-AOS2116

CANONICAL THRESHOLDING FOR NONSPARSE HIGH-DIMENSIONAL LINEAR REGRESSION

BY IGOR SILIN[*] AND JIANQING FAN[†]

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

We consider a high-dimensional linear regression problem. Unlike many papers on the topic, we do not require sparsity of the regression coefficients; instead, our main structural assumption is a decay of eigenvalues of the covariance matrix of the data. We propose a new family of estimators, called the canonical thresholding estimators, which pick largest regression coefficients in the canonical form. The estimators admit an explicit form and can be linked to LASSO and Principal Component Regression (PCR). A theoretical analysis for both fixed design and random design settings is provided. Obtained bounds on the mean squared error and the prediction error of a specific estimator from the family allow to clearly state sufficient conditions on the decay of eigenvalues to ensure convergence. In addition, we promote the use of the relative errors, strongly linked with the out-of-sample R^2. The study of these relative errors leads to a new concept of joint effective dimension, which incorporates the covariance of the data and the regression coefficients simultaneously, and describes the complexity of a linear regression problem. Some minimax lower bounds are established to showcase the optimality of our procedure. Numerical simulations confirm good performance of the proposed estimators compared to the previously developed methods.

REFERENCES

BAIR, E., HASTIE, T., PAUL, D. and TIBSHIRANI, R. (2006). Prediction by supervised principal components. *J. Amer. Statist. Assoc.* **101** 119–137. MR2252436 https://doi.org/10.1198/016214505000000628

BARTLETT, P. L., LONG, P. M., LUGOSI, G. and TSIGLER, A. (2020). Benign overfitting in linear regression. *Proc. Natl. Acad. Sci. USA* **117** 30063–30070. MR4263288 https://doi.org/10.1073/pnas.1907378117

BELKIN, M. (2018). Approximation beats concentration? An approximation view on inference with smooth radial kernels. *Proc. Mach. Learn. Res.* **75** 1–18.

BELKIN, M., HSU, D. and XU, J. (2020). Two models of double descent for weak features. *SIAM J. Math. Data Sci.* **2** 1167–1180. MR4186534 https://doi.org/10.1137/20M1336072

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

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

BIETTI, A. and MAIRAL, J. (2019). On the inductive bias of neural tangent kernels. *Adv. Neural Inf. Process. Syst.* 12893–12904.

BOGDAN, M., VAN DEN BERG, E., SABATTI, C., SU, W. and CANDÈS, E. J. (2015). SLOPE—adaptive variable selection via convex optimization. *Ann. Appl. Stat.* **9** 1103–1140. MR3418717 https://doi.org/10.1214/15-AOAS842

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

CHINOT, G. and LERASLE, M. (2020). Benign overfitting in the large deviation regime. Available at arXiv:2003.05838.

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

MSC2020 subject classifications. Primary 62J05; secondary 62H12, 62H25.

Key words and phrases. High-dimensional linear regression, covariance eigenvalues decay, thresholding, relative errors, principal component regression.

DOBRIBAN, E. and LIU, S. (2019). Asymptotics for sketching in least squares regression. *Adv. Neural Inf. Process. Syst.*

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

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

DONOHO, D. L. (1995). De-noising by soft-thresholding. *IEEE Trans. Inf. Theory* **41** 613–627. MR1331258 https://doi.org/10.1109/18.382009

DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455. MR1311089 https://doi.org/10.1093/biomet/81.3.425

DONOHO, D. L. and JOHNSTONE, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90** 1200–1224. MR1379464

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

DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1995). Wavelet shrinkage: Asymptopia? *J. Roy. Statist. Soc. Ser. B* **57** 301–369. With discussion and a reply by the authors. MR1323344

DUDOIT, S. and VAN DER LAAN, M. J. (2005). Asymptotics of cross-validated risk estimation in estimator selection and performance assessment. *Stat. Methodol.* **2** 131–154. MR2161394 https://doi.org/10.1016/j.stamet.2005.02.003

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

FAN, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *J. Amer. Statist. Assoc.* **91** 674–688. MR1395735 https://doi.org/10.2307/2291663

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

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

FAN, J., XUE, L. and YAO, J. (2017). Sufficient forecasting using factor models. *J. Econometrics* **201** 292–306. MR3717565 https://doi.org/10.1016/j.jeconom.2017.08.009

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

GREENSHTEIN, E. and RITOV, Y. (2004). Persistence in high-dimensional linear predictor selection and the virtue of overparametrization. *Bernoulli* **10** 971–988. MR2108039 https://doi.org/10.3150/bj/1106314846

GYÖFRI, L., KOHLER, M., KRZYŻAK, A. and WALK, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. Springer, New York.

HASTIE, T., MONTANARI, A., ROSSET, S. and TIBSHIRANI, R. (2019). Surprises in high-dimensional ridgeless least squares interpolation. Available at arXiv:1903.08560.

JIRAK, M. and WAHL, M. (2018). Relative perturbation bounds with applications to empirical covariance operators. Available at arXiv:1802.02869.

JOLLIFFE, I. T. (1995). Rotation of principal components: Choice of normalization constraints. *J. Appl. Stat.* **22** 29–35. MR1342651 https://doi.org/10.1080/757584395

KUCHIBHOTLA, A. K. and CHAKRABORTTY, A. (2018). Moving beyond sub-Gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression. Available at arXiv:1804.02605.

LIANG, T. and RAKHLIN, A. (2020). Just interpolate: Kernel "ridgeless" regression can generalize. *Ann. Statist.* **48** 1329–1347. MR4124325 https://doi.org/10.1214/19-AOS1849

MA, S. and BELKIN, M. (2017). Diving into the shallows: A computational perspective on large-scale shallow learning. *Adv. Neural Inf. Process. Syst.* 3781–3790.

PAUL, D., BAIR, E., HASTIE, T. and TIBSHIRANI, R. (2008). "Preconditioning" for feature selection and regression in high-dimensional problems. *Ann. Statist.* **36** 1595–1618. MR2435449 https://doi.org/10.1214/009053607000000578

PEARSON, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science* **2** 559–572.

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

SILIN, I. and FAN, J. (2022). Supplement to "Canonical thresholding for nonsparse high-dimensional linear regression." https://doi.org/10.1214/21-AOS2116SUPP

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

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

ZUMBACH, G. (2009). Empirical properties of large covariance matrices. *RiskMetrics Journal* **9** 1.

The Annals of Statistics
2022, Vol. 50, No. 1, 487–510
https://doi.org/10.1214/21-AOS2117

SEMIPARAMETRIC LATENT-CLASS MODELS FOR MULTIVARIATE LONGITUDINAL AND SURVIVAL DATA

BY KIN YAU WONG[1], DONGLIN ZENG[2,*] AND D. Y. LIN[2,†]

[1]*Department of Applied Mathematics, The Hong Kong Polytechnic University, kin-yau.wong@polyu.edu.hk*
[2]*Department of Biostatistics, University of North Carolina at Chapel Hill, [*]dzeng@email.unc.edu; [†]lin@bios.unc.edu*

In long-term follow-up studies, data are often collected on repeated measures of multivariate response variables as well as on time to the occurrence of a certain event. To jointly analyze such longitudinal data and survival time, we propose a general class of semiparametric latent-class models that accommodates a heterogeneous study population with flexible dependence structures between the longitudinal and survival outcomes. We combine nonparametric maximum likelihood estimation with sieve estimation and devise an efficient EM algorithm to implement the proposed approach. We establish the asymptotic properties of the proposed estimators through novel use of modern empirical process theory, sieve estimation theory and semiparametric efficiency theory. Finally, we demonstrate the advantages of the proposed methods through extensive simulation studies and provide an application to the Atherosclerosis Risk in Communities study.

REFERENCES

[1] BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins Univ. Press, Baltimore, MD. MR1245941

[2] DE BOOR, C. (1976). Splines as linear combinations of B-splines. A survey. In *Approximation Theory*, *II* (*Proc*. *Internat*. *Sympos*., *Univ*. *Texas*, *Austin*, *Tex*., 1976) 1–47. MR0467092

[3] DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **39** 1–22. MR0501537 https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

[4] HENDERSON, R., DIGGLE, P. and DOBSON, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1** 465–480. https://doi.org/10.1093/biostatistics/1.4.465

[5] LIN, H., TURNBULL, B. W., MCCULLOCH, C. E. and SLATE, E. H. (2002). Latent class models for joint analysis of longitudinal biomarker and event process data: Application to longitudinal prostate-specific antigen readings and prostate cancer. *J. Amer. Statist. Assoc.* **97** 53–65. MR1947272 https://doi.org/10.1198/016214502753479220

[6] LIU, L., MA, J. Z. and O'QUIGLEY, J. (2008). Joint analysis of multi-level repeated measures data and survival: An application to the end stage renal disease (ESRD) data. *Stat. Med.* **27** 5679–5691. MR2573776 https://doi.org/10.1002/sim.3392

[7] LIU, Q. and PIERCE, D. A. (1994). A note on Gauss–Hermite quadrature. *Biometrika* **81** 624–629. MR1311107 https://doi.org/10.1093/biomet/81.3.624

[8] LIU, Y., LIN, Y., ZHOU, J. and LIU, L. (2020). A semi-parametric joint latent class model with longitudinal and survival data. *Stat. Interface* **13** 411–422. MR4091806 https://doi.org/10.4310/SII.2020.v13.n3.a10

[9] LIU, Y., LIU, L. and ZHOU, J. (2015). Joint latent class model of survival and longitudinal data: An application to CPCRA study. *Comput. Statist. Data Anal.* **91** 40–50. MR3368004 https://doi.org/10.1016/j.csda.2015.05.007

[10] LOUIS, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **44** 226–233. MR0676213 https://doi.org/10.1111/j.2517-6161.1982.tb01203.x

[11] MA, Y. and WANG, Y. (2012). Efficient distribution estimation for data with unobserved sub-population identifiers. *Electron. J. Stat.* **6** 710–737. MR2988426 https://doi.org/10.1214/12-EJS690

MSC2020 subject classifications. Primary 62N02; secondary 62G05, 62H30.
Key words and phrases. Censored data, joint analysis, mixture models, nonparametric estimation, sieve estimation.

[12] MURPHY, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.* **22** 712–731. MR1292537 https://doi.org/10.1214/aos/1176325492

[13] PROUST-LIMA, C., SÉNE, M., TAYLOR, J. M. G. and JACQMIN-GADDA, H. (2014). Joint latent class models for longitudinal and time-to-event data: A review. *Stat. Methods Med. Res.* **23** 74–90. MR3190688 https://doi.org/10.1177/0962280212445839

[14] SCHUMAKER, L. L. (2007). *Spline Functions: Basic Theory*, 3rd ed. Cambridge Univ. Press, Cambridge. MR2348176 https://doi.org/10.1017/CBO9780511618994

[15] SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464. MR0468014 https://doi.org/10.1214/aos/1176344136

[16] SHEN, R., OLSHEN, A. B. and LADANYI, M. (2009). Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics* **25** 2906–2912. https://doi.org/10.1093/bioinformatics/btp543

[17] THE ARIC INVESTIGATORS (1989). The atherosclerosis risk in communities (ARIC) study: Design and objectives. *Am. J. Epidemiol.* **129** 687–702. https://doi.org/10.1093/oxfordjournals.aje.a115184

[18] TSIATIS, A. A. and DAVIDIAN, M. (2004). Joint modeling of longitudinal and time-to-event data: An overview. *Statist. Sinica* **14** 809–834. MR2087974

[19] VARADHAN, R. and ROLAND, C. (2008). Simple and globally convergent methods for accelerating the convergence of any EM algorithm. *Scand. J. Stat.* **35** 335–353. MR2418745 https://doi.org/10.1111/j.1467-9469.2007.00585.x

[20] WANG, Y., GARCIA, T. P. and MA, Y. (2012). Nonparametric estimation for censored mixture data with application to the Cooperative Huntington's Observational Research Trial. *J. Amer. Statist. Assoc.* **107** 1324–1338. MR3036398 https://doi.org/10.1080/01621459.2012.699353

[21] WONG, K. Y., ZENG, D. and LIN, D. Y. (2022). Supplement to "Semiparametric latent-class models for multivariate longitudinal and survival data." https://doi.org/10.1214/21-AOS2117SUPP

[22] XU, C., BAINES, P. D. and WANG, J.-L. (2014). Standard error estimation using the EM algorithm for the joint modeling of survival and longitudinal data. *Biostatistics* **15** 731–744. https://doi.org/10.1093/biostatistics/kxu015

[23] XU, J. and ZEGER, S. L. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **50** 375–387. MR1856332 https://doi.org/10.1111/1467-9876.00241

[24] ZENG, D. and CAI, J. (2005). Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. *Ann. Statist.* **33** 2132–2163. MR2211082 https://doi.org/10.1214/009053605000000480

[25] ZENG, D. and LIN, D. Y. (2007). Maximum likelihood estimation in semiparametric regression models with censored data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 507–564. MR2370068 https://doi.org/10.1111/j.1369-7412.2007.00606.x

[26] ZENG, D. and LIN, D. Y. (2010). A general asymptotic theory for maximum likelihood estimation in semiparametric regression models with censored data. *Statist. Sinica* **20** 871–910. MR2682647

[27] ZENG, D., LIN, D. Y. and LIN, X. (2008). Semiparametric transformation models with random effects for clustered failure time data. *Statist. Sinica* **18** 355–377. MR2384992

[28] ZENG, D., MAO, L. and LIN, D. Y. (2016). Maximum likelihood estimation for semiparametric transformation models with interval-censored data. *Biometrika* **103** 253–271. MR3509885 https://doi.org/10.1093/biomet/asw013

The Annals of Statistics
2022, Vol. 50, No. 1, 511–536
https://doi.org/10.1214/21-AOS2118

ROBUST SUB-GAUSSIAN ESTIMATION OF A MEAN VECTOR IN NEARLY LINEAR TIME

BY JULES DEPERSIN[*] AND GUILLAUME LECUÉ[†]

CREST, ENSAE, IPParis, [*]*jules.depersin@ensae.fr;* [†]*guillaume.lecue@ensae.fr*

We construct an algorithm for estimating the mean of a heavy-tailed random variable when given an adversarial corrupted sample of N independent observations. The only assumption we make on the distribution of the non-corrupted (or *informative*) data is the existence of a covariance matrix Σ, unknown to the statistician. Our algorithm outputs $\hat{\mu}$, which is robust to the presence of $|\mathcal{O}|$ adversarial outliers and satisfies

$$\|\hat{\mu}-\mu\|_2 \lesssim \sqrt{\frac{\mathrm{Tr}(\Sigma)}{N}}+\sqrt{\frac{\|\Sigma\|_{\mathrm{op}}K}{N}} \tag{1}$$

with probability at least $1-\exp(-c_0K)-\exp(-c_1u)$, and runtime $\tilde{\mathcal{O}}(Nd+uKd)$ where $K \in \{600|\mathcal{O}|, \dots, N\}$ and $u \in \mathbb{N}^*$ are two parameters of the algorithm. The algorithm is fully data-dependent and does not use (1) in its construction, which combines recently developed tools for median-of-means estimators and covering semidefinite programming. We also show that this algorithm can automatically adapt to the number of outliers (adaptive choice of K) and that it satisfies the same bound in expectation.

REFERENCES

[1] ALLEN-ZHU, Z., GELASHVILI, R. and RAZENSHTEYN, I. (2016). Restricted isometry property for general p-norms. *IEEE Trans. Inf. Theory* **62** 5839–5854. MR3552427 https://doi.org/10.1109/TIT.2016.2598296

[2] ALLEN-ZHU, Z., LEE, Y. T. and ORECCHIA, L. (2016). Using optimization to obtain a width-independent, parallel, simpler, and faster positive SDP solver. In *Proceedings of the Twenty-Seventh Annual ACM–SIAM Symposium on Discrete Algorithms* 1824–1831. ACM, New York. MR3478503 https://doi.org/10.1137/1.9781611974331.ch127

[3] ALON, N., MATIAS, Y. and SZEGEDY, M. (1999). The space complexity of approximating the frequency moments. *J. Comput. System Sci.* **58** 137–147. Twenty-eighth Annual ACM Symposium on the Theory of Computing (Philadelphia, PA, 1996). MR1688610 https://doi.org/10.1006/jcss.1997.1545

[4] BIRGÉ, L. (1984). Stabilité et instabilité du risque minimax pour des variables indépendantes équidistribuées. *Ann. Inst. Henri Poincaré Probab. Stat.* **20** 201–223. MR0762855

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

[6] BUBECK, S., CESA-BIANCHI, N. and LUGOSI, G. (2013). Bandits with heavy tail. *IEEE Trans. Inf. Theory* **59** 7711–7717. MR3124669 https://doi.org/10.1109/TIT.2013.2277869

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

[8] CATONI, O. and GIULINI, I. (2017). Dimension-free PAC-Bayesian bounds for matrices, vectors, and linear least squares regression. Technical Report, CNRS and LSPM.

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

[10] CHENG, Y., DIAKONIKOLAS, I. and GE, R. (2019). High-dimensional robust mean estimation in nearly-linear time. In *Proceedings of the Thirtieth Annual ACM–SIAM Symposium on Discrete Algorithms* 2755–2771. SIAM, Philadelphia, PA. MR3909640 https://doi.org/10.1137/1.9781611975482.171

MSC2020 subject classifications. Primary 62F35, 62G08; secondary 62C20, 62G05, 62G20.
Key words and phrases. Empirical processes, robust statistics, algorithms, heavy-tailed data.

[11] CHERAPANAMJERI, Y., FLAMMARION, N. and BARTLETT, P. L. (2019). Fast mean estimation with sub-Gaussian rates.

[12] CHERAPANAMJERI, Y., HOPKINS, S. B., KATHURIA, T., RAGHAVENDRA, P. and TRIPURANENI, N. (2020). Algorithms for heavy-tailed statistics: Regression, covariance estimation, and beyond. In *STOC '*20*—Proceedings of the* 52*nd Annual ACM SIGACT Symposium on Theory of Computing* 601–609. ACM, New York. MR4141785

[13] DALALYAN, A. S. and MINASYAN, A. (2020). All-in-one robust estimator of the Gaussian mean. arXiv preprint arXiv:2002.01432.

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

[15] DEPERSIN, J. (2020). A spectral algorithm for robust regression with subgaussian rates. Technical Report, CREST–ENSAE.

[16] DEPERSIN, J. and LECUÉ, G. (2019). Fast algorithms for robust estimation of a mean vector.

[17] DEVROYE, L., LERASLE, M., LUGOSI, G. and OLIVEIRA, R. I. (2016). Sub-Gaussian mean estimators. *Ann*. *Statist*. **44** 2695–2725. MR3576558 https://doi.org/10.1214/16-AOS1440

[18] DIAKONIKOLAS, I., KAMATH, G., KANE, D., LI, J., MOITRA, A. and STEWART, A. (2019). Robust estimators in high-dimensions without the computational intractability. *SIAM J*. *Comput*. **48** 742–864. MR3945261 https://doi.org/10.1137/17M1126680

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

[20] DIAKONIKOLAS, I., KAMATH, G., KANE, D. M., LI, J., MOITRA, A. and STEWART, A. (2017). Being robust (in high dimensions) can be practical. arXiv preprint arXiv:1703.00893.

[21] DIAKONIKOLAS, I., KAMATH, G., KANE, D. M., LI, J., MOITRA, A. and STEWART, A. (2018). Robustly learning a Gaussian: Getting optimal error, efficiently. In *Proceedings of the Twenty-Ninth Annual ACM–SIAM Symposium on Discrete Algorithms* 2683–2702. SIAM, Philadelphia, PA. MR3775959 https://doi.org/10.1137/1.9781611975031.171

[22] DIAKONIKOLAS, I. and KANE, D. M. (2019). Recent advances in algorithmic high-dimensional robust statistics. arXiv preprint arXiv:1911.05911.

[23] DIAKONIKOLAS, I., KANE, D. M. and STEWART, A. (2018). List-decodable robust mean estimation and learning mixtures of spherical Gaussians. In *STOC'*18*—Proceedings of the* 50*th Annual ACM SIGACT Symposium on Theory of Computing* 1047–1060. ACM, New York. MR3826316 https://doi.org/10.1145/3188745.3188758

[24] DIAKONIKOLAS, I., KONG, W. and STEWART, A. (2019). Efficient algorithms and lower bounds for robust linear regression. In *Proceedings of the Thirtieth Annual ACM–SIAM Symposium on Discrete Algorithms* 2745–2754. SIAM, Philadelphia, PA. MR3909639 https://doi.org/10.1137/1.9781611975482.170

[25] DONOHO, D. L. and GASKO, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann*. *Statist*. **20** 1803–1827. MR1193313 https://doi.org/10.1214/aos/1176348890

[26] GOEMANS, M. X. and WILLIAMSON, D. P. (1995). Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J*. *Assoc*. *Comput*. *Mach*. **42** 1115–1145. MR1412228 https://doi.org/10.1145/227683.227684

[27] GROTHENDIECK, A. (1953). Résumé de la théorie métrique des produits tensoriels topologiques. *Bol*. *Soc*. *Mat*. *São Paulo* **8** 1–79. MR0094682

[28] HAMPEL, F. R. (1971). A general qualitative definition of robustness. *Ann*. *Math*. *Stat*. **42** 1887–1896. MR0301858 https://doi.org/10.1214/aoms/1177693054

[29] HAMPEL, F. R. (1973). Robust estimation: A condensed partial survey. *Z*. *Wahrsch*. *Verw*. *Gebiete* **27** 87–104. MR0359096 https://doi.org/10.1007/BF00536619

[30] HOPKINS, S. B. (2020). Mean estimation with sub-Gaussian rates in polynomial time. *Ann*. *Statist*. **48** 1193–1213. MR4102693 https://doi.org/10.1214/19-AOS1843

[31] HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann*. *Math*. *Stat*. **35** 73–101. MR0161415 https://doi.org/10.1214/aoms/1177703732

[32] HUBER, P. J. and RONCHETTI, E. M. (2009). *Robust Statistics*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2488795 https://doi.org/10.1002/9780470434697

[33] JERRUM, M. R., VALIANT, L. G. and VAZIRANI, V. V. (1986). Random generation of combinatorial structures from a uniform distribution. *Theoret*. *Comput*. *Sci*. **43** 169–188. MR0855970 https://doi.org/10.1016/0304-3975(86)90174-X

[34] KOLTCHINSKII, V., PANCHENKO, D. and LOZANO, F. (2003). Bounding the generalization error of convex combinations of classifiers: Balancing the dimensionality and the margins. *Ann. Appl. Probab.* **13** 213–252. MR1951998 https://doi.org/10.1214/aoap/1042765667
[35] LECUÉ, G. and LERASLE, M. (2019). Learning from MOM's principles: Le Cam's approach. *Stochastic Process. Appl.* **129** 4385–4410. MR4013866 https://doi.org/10.1016/j.spa.2018.11.024
[36] LECUÉ, G. and LERASLE, M. (2020). Robust machine learning by median-of-means: Theory and practice. *Ann. Statist.* **48** 906–931. MR4102681 https://doi.org/10.1214/19-AOS1828
[37] LEDOUX, M. (2001). *The Concentration of Measure Phenomenon. Mathematical Surveys and Monographs* **89**. Amer. Math. Soc., Providence, RI. MR1849347 https://doi.org/10.1090/surv/089
[38] LEDOUX, M. and TALAGRAND, M. (2011). *Probability in Banach Spaces: Isoperimetry and Processes. Classics in Mathematics*. Springer, Berlin. Reprint of the 1991 edition. MR2814399
[39] LEI, Z., LUH, K., VENKAT, P. and ZHANG, F. (2020). A fast spectral algorithm for mean estimation with sub-Gaussian rates. In *Conference on Learning Theory* 2598–2612.
[40] LEPSKIĬ, O. V. (1990). A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatn. Primen.* **35** 459–470. MR1091202 https://doi.org/10.1137/1135065
[41] LEPSKIĬ, O. V. (1991). Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Teor. Veroyatn. Primen.* **36** 645–659. MR1147167 https://doi.org/10.1137/1136085
[42] LERASLE, M. and OLIVEIRA, R. (2011). Robust empirical mean estimators. Technical Report, IMPA and CNRS.
[43] LERASLE, M., SZABO, Z., MATHIEU, T. and LECUÉ, G. (2017). MONK—outliers-robust mean embedding estimation by median-of-means. Technical Report, CNRS, Univ. Paris 11, Ecole Polytechnique and CREST.
[44] LUGOSI, G. and MENDELSON, S. (2019). Sub-Gaussian estimators of the mean of a random vector. *Ann. Statist.* **47** 783–794. MR3909950 https://doi.org/10.1214/17-AOS1639
[45] MINSKER, S. (2015). Geometric median and robust estimation in Banach spaces. *Bernoulli* **21** 2308–2335. MR3378468 https://doi.org/10.3150/14-BEJ645
[46] MINSKER, S. (2018). Sub-Gaussian estimators of the mean of a random matrix with heavy-tailed entries. *Ann. Statist.* **46** 2871–2903. MR3851758 https://doi.org/10.1214/17-AOS1642
[47] MINSKER, S. (2019). Distributed statistical estimation and rates of convergence in normal approximation. *Electron. J. Stat.* **13** 5213–5252. MR4043072 https://doi.org/10.1214/19-EJS1647
[48] NEMIROVSKY, A. S. and YUDIN, D. B. (1983). *Problem Complexity and Method Efficiency in Optimization. Wiley-Interscience Series in Discrete Mathematics*. Wiley, New York. Translated from the Russian and with a preface by E. R. Dawson. MR0702836
[49] NESTEROV, YU. (1998). Semidefinite relaxation and nonconvex quadratic optimization. *Optim. Methods Softw.* **9** 141–160. MR1618100 https://doi.org/10.1080/10556789808805690
[50] PENG, R., TANGWONGSAN, K. and ZHANG, P. (2012). Faster and simpler width-independent parallel algorithms for positive semidefinite programming.
[51] PISIER, G. (2012). Grothendieck's theorem, past and present. *Bull. Amer. Math. Soc.* (*N.S.*) **49** 237–323. MR2888168 https://doi.org/10.1090/S0273-0979-2011-01348-9
[52] SMALL, C. G. (1990). A survey of multidimensional medians. *Int. Stat. Rev.* 263–277.
[53] TUKEY, J. W. (1960). A survey of sampling from contaminated distributions. In *Contributions to Probability and Statistics* 448–485. Stanford Univ. Press, Stanford, CA. MR0120720
[54] TUKEY, J. W. (1962). The future of data analysis. *Ann. Math. Stat.* **33** 1–67. MR0133937 https://doi.org/10.1214/aoms/1177704711

The Annals of Statistics
2022, Vol. 50, No. 1, 537–559
https://doi.org/10.1214/21-AOS2119

ON AN EXTENSION OF THE PROMOTION TIME CURE MODEL

BY JAD BEYHUM[1,*], ANOUAR EL GHOUCH[2], FRANÇOIS PORTIER[3,4] AND INGRID VAN KEILEGOM[1,†]

[1]*ORSTAT, KU Leuven,* [*]*jad.beyhum@kuleuven.be;* [†]*ingrid.vankeilegom@kuleuven.be*
[2]*ISBA, UCLouvain, anouar.elghouch@uclouvain.be*
[3]*LTCI, Télécom Paris, Institut Polytechnique de Paris, francois.portier@gmail.com*
[4]*CREST, Ensai, Univ Rennes*

We consider the problem of estimating the distribution of time-to-event data that is subject to censoring and for which the event of interest might never occur, that is, some subjects are cured. To model this kind of data in the presence of covariates, one of the leading semiparametric models is the promotion time cure model (*Stochastic Models of Tumor Latency and Their Biostatistical Applications* (1996) World Scientific), which adapts the Cox model to the presence of cured subjects. Estimating the conditional distribution results in a complicated constrained optimization problem, and inference is difficult as no closed-formula for the variance is available. We propose a new model, inspired by the Cox model, that leads to a simple estimation procedure and that presents a closed formula for the variance. In this paper, we show (i) that the new model contains as a special case the promotion time cure model with an exponential link, and hence we have a simpler way to estimate the latter model than what is done so far in the literature; (ii) that in the latter special case, both estimators are equal to the partial likelihood estimator under the usual Cox model; (iii) that the estimators under the new model have certain asymptotic properties when the model is correct and when it is misspecified; (iv) that the error of LASSO type estimators is of order $\sqrt{\log(nd)/n}$ in the case of high-dimensional covariates with dimension d and sample size n. We also study the practical behaviour of our estimation procedure by means of simulations, and we apply our model and estimation method to a breast cancer data set.

REFERENCES

[1] ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10** 1100–1120. MR0673646

[2] BERTRAND, A., LEGRAND, C. and VAN KEILEGOM, I. (2016). miCoPTCM: Promotion time cure model with mis-measured covariates. R package version 1.0.

[3] BEYHUM, J., EL GHOUCH, A., PORTIER, F. and VAN KEILEGOM, I. (2022). Supplement to "On an extension of the promotion time cure model." https://doi.org/10.1214/21-AOS2119SUPP

[4] BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. *Johns Hopkins Series in the Mathematical Sciences*. Johns Hopkins Univ. Press, Baltimore, MD. MR1245941

[5] BRADIC, J., FAN, J. and JIANG, J. (2011). Regularization for Cox's proportional hazards model with NP-dimensionality. *Ann. Statist.* **39** 3092–3120. MR3012402 https://doi.org/10.1214/11-AOS911

[6] BRESLOW, N. (1972). Contribution to the discussion of the paper by D. R. Cox. *J. Roy. Statist. Soc. Ser. B* **34** 187–220.

[7] CHEN, M.-H., IBRAHIM, J. G. and SINHA, D. (1999). A new Bayesian model for survival data with a surviving fraction. *J. Amer. Statist. Assoc.* **94** 909–919. MR1723307 https://doi.org/10.2307/2670006

[8] CHEN, X., LINTON, O. and VAN KEILEGOM, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* **71** 1591–1608. MR2000259 https://doi.org/10.1111/1468-0262.00461

MSC2020 subject classifications. Primary 62N01, 62N02; secondary 62G05, 62G20.
Key words and phrases. Cox model, semiparametric efficiency, nonparametric MLE, variance estimation, misspecification, high-dimensional data, nonconvex penalties.

[9] COLLETT, D. (2015). *Modelling Survival Data in Medical Research*. CRC Press, Boca Raton.

[10] COX, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34** 187–220. MR0341758

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

[12] FLEMING, T. R. and HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Wiley, New York. MR1100924

[13] HUANG, J., SUN, T., YING, Z., YU, Y. and ZHANG, C.-H. (2013). Oracle inequalities for the LASSO in the Cox model. *Ann. Statist.* **41** 1142–1165. MR3113806 https://doi.org/10.1214/13-AOS1098

[14] IBRAHIM, J. G., CHEN, M.-H. and SINHA, D. (2001). Bayesian semiparametric models for survival data with a cure fraction. *Biometrics* **57** 383–388. MR1854229 https://doi.org/10.1111/j.0006-341X.2001.00383.x

[15] KLEIN, J. P. and MOESCHBERGER, M. L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer Science, New York.

[16] KONG, S. and NAN, B. (2014). Non-asymptotic oracle inequalities for the high-dimensional Cox regression via lasso. *Statist. Sinica* **24** 25–42. MR3184591

[17] KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference. Springer Series in Statistics*. Springer, New York. MR2724368 https://doi.org/10.1007/978-0-387-74978-5

[18] LIN, W. and KULASEKERA, K. B. (2007). Identifiability of single-index models and additive-index models. *Biometrika* **94** 496–501. MR2380574 https://doi.org/10.1093/biomet/asm029

[19] 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

[20] MA, Y. and YIN, G. (2008). Cure rate model with mismeasured covariates under transformation. *J. Amer. Statist. Assoc.* **103** 743–756. MR2524007 https://doi.org/10.1198/016214508000000319

[21] MEI, S., BAI, Y. and MONTANARI, A. (2018). The landscape of empirical risk for nonconvex losses. *Ann. Statist.* **46** 2747–2774. MR3851754 https://doi.org/10.1214/17-AOS1637

[22] MURPHY, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.* **22** 712–731. MR1292537 https://doi.org/10.1214/aos/1176325492

[23] MURPHY, S. A. and VAN DER VAART, A. W. (2000). On profile likelihood. *J. Amer. Statist. Assoc.* **95** 449–485. MR1803168 https://doi.org/10.2307/2669386

[24] NEWEY, W. K. and MCFADDEN, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics, Vol. IV. Handbooks in Econom.* **2** 2111–2245. North-Holland, Amsterdam. MR1315971

[25] PORTIER, F. and DELYON, B. (2013). Optimal transformation: A new approach for covering the central subspace. *J. Multivariate Anal.* **115** 84–107. MR3004547 https://doi.org/10.1016/j.jmva.2012.09.001

[26] PORTIER, F., EL GHOUCH, A. and VAN KEILEGOM, I. (2017). Efficiency and bootstrap in the promotion time cure model. *Bernoulli* **23** 3437–3468. MR3654812 https://doi.org/10.3150/16-BEJ852

[27] RITOV, Y. and WELLNER, J. A. (1988). Censoring, martingales, and the Cox model. In *Statistical Inference from Stochastic Processes (Ithaca, NY, 1987). Contemp. Math.* **80** 191–219. Amer. Math. Soc., Providence, RI. MR0999013 https://doi.org/10.1090/conm/080/999013

[28] SASIENI, P. (1992). Information bounds for the conditional hazard ratio in a nested family of regression models. *J. Roy. Statist. Soc. Ser. B* **54** 617–635. MR1160487

[29] STÄDLER, N., BÜHLMANN, P. and VAN DE GEER, S. (2010). ℓ_1-penalization for mixture regression models. *TEST* **19** 209–256. MR2677722 https://doi.org/10.1007/s11749-010-0197-z

[30] STRUTHERS, C. A. and KALBFLEISCH, J. D. (1986). Misspecified proportional hazard models. *Biometrika* **73** 363–369. MR0855896 https://doi.org/10.1093/biomet/73.2.363

[31] R CORE TEAM (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[32] THERNEAU, T. M. and GRAMBSCH, P. M. (2000). *Modeling Survival Data: Extending the Cox Model. Statistics for Biology and Health*. Springer, New York. MR1774977 https://doi.org/10.1007/978-1-4757-3294-8

[33] TIBSHIRANI, R. (1997). The lasso method for variable selection in the Cox model. *Stat. Med.* **16** 385–395.

[34] TSODIKOV, A. (1998). Asymptotic efficiency of a proportional hazards model with cure. *Statist. Probab. Lett.* **39** 237–244. MR1646259 https://doi.org/10.1016/S0167-7152(98)00056-X

[35] TSODIKOV, A. (1998). A proportional hazards model taking account of long-term survivors. *Biometrics* **54** 1508–1516.

[36] TSODIKOV, A. (2001). Estimation of survival based on proportional hazards when cure is a possibility. *Math. Comput. Modelling* **33** 1227–1236. MR1837407 https://doi.org/10.1016/S0895-7177(00)00311-3

[37] TSODIKOV, A. D., IBRAHIM, J. G. and YAKOVLEV, A. Y. (2003). Estimating cure rates from survival data: An alternative to two-component mixture models. *J. Amer. Statist. Assoc.* **98** 1063–1078. MR2055496 https://doi.org/10.1198/01622145030000001007

[38] VAN DE GEER, S. A. (2008). High-dimensional generalized linear models and the lasso. *Ann. Statist.* **36** 614–645. MR2396809 https://doi.org/10.1214/009053607000000929

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

[40] WANG, Y., KLIJN, J. G. M., SIEUWERTS, A. M., LOOK, M. P., YANG, F., TALANTOV, D., TIMMERMANS, M., VAN GELDER, M. E. M. M., YU, J. et al. (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* **365** 671–679.

[41] WELLNER, J. A. and ZHAN, Y. (1996). Bootstrapping Z-estimators. Technical report 308, Dept. Statistics, Univ. Washington.

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

[43] YAKOVLEV, A. and TSODIKOV, A. (1996). *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. World Scientific, Singapore.

[44] YU, Y., BRADIC, J. and SAMWORTH, R. J. (2018). Confidence intervals for high-dimensional Cox models. Preprint. Available at arXiv:1803.01150.

[45] ZENG, D., YIN, G. and IBRAHIM, J. G. (2006). Semiparametric transformation models for survival data with a cure fraction. *J. Amer. Statist. Assoc.* **101** 670–684. MR2256182 https://doi.org/10.1198/016214505000001122

The Annals of Statistics
2022, Vol. 50, No. 1, 560–583
https://doi.org/10.1214/21-AOS2121

BACKFITTING FOR LARGE SCALE CROSSED RANDOM EFFECTS REGRESSIONS

BY SWARNADIP GHOSH[*], TREVOR HASTIE[†] AND ART B. OWEN[‡]

Department of Statistics, Stanford University, [*]raswa281@stanford.edu; [†]hastie@stanford.edu; [‡]owen@stanford.edu

Regression models with crossed random effect errors can be very expensive to compute. The cost of both generalized least squares and Gibbs sampling can easily grow as $N^{3/2}$ (or worse) for N observations. Papaspiliopoulos, Roberts and Zanella (*Biometrika* **107** (2020) 25–40) present a collapsed Gibbs sampler that costs $O(N)$, but under an extremely stringent sampling model. We propose a backfitting algorithm to compute a generalized least squares estimate and prove that it costs $O(N)$. A critical part of the proof is in ensuring that the number of iterations required is $O(1)$, which follows from keeping a certain matrix norm below $1 - \delta$ for some $\delta > 0$. Our conditions are greatly relaxed compared to those for the collapsed Gibbs sampler, though still strict. Empirically, the backfitting algorithm has a norm below $1 - \delta$ under conditions that are less strict than those in our assumptions. We illustrate the new algorithm on a ratings data set from Stitch Fix.

REFERENCES

BATES, D., MÄCHLER, M., BOLKER, B. and WALKER, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67** 1–48. https://doi.org/10.18637/jss.v067.i01

BUJA, A., HASTIE, T. and TIBSHIRANI, R. (1989). Linear smoothers and additive models. *Ann. Statist.* **17** 453–555. MR0994249 https://doi.org/10.1214/aos/1176347115

CAMERON, A. C., GELBACH, J. B. and MILLER, D. L. (2011). Robust inference with multiway clustering. *J. Bus. Econom. Statist.* **29** 238–249. MR2807878 https://doi.org/10.1198/jbes.2010.07136

GAO, K. (2017). Scalable estimation and inference for massive linear mixed models with crossed random effects. Ph.D. thesis, Stanford Univ. MR4257222

GAO, K. and OWEN, A. (2017). Efficient moment calculations for variance components in large unbalanced crossed random effects models. *Electron. J. Stat.* **11** 1235–1296. MR3635913 https://doi.org/10.1214/17-EJS1236

GAO, K. and OWEN, A. B. (2020). Estimation and inference for very large linear mixed effects models. *Statist. Sinica* **30** 1741–1771. MR4260743 https://doi.org/10.5705/ss.202018.0029

GELMAN, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.* **1** 515–533. MR2221284 https://doi.org/10.1214/06-BA117A

GELMAN, A. and HILL, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge Univ. Press, Cambridge.

GELMAN, A. and SHIRLEY, K. (2011). Inference from simulations and monitoring convergence. In *Handbook of Markov Chain Monte Carlo. Chapman & Hall/CRC Handb. Mod. Stat. Methods* 163–174. CRC Press, Boca Raton, FL. MR2858448 https://doi.org/10.1201/b10905

GOODMAN, J. and SOKAL, A. D. (1989). Multigrid Monte Carlo method. Conceptual foundations. *Phys. Rev. D, Part. Fields* **40** 2035–2071. https://doi.org/10.1103/physrevd.40.2035

HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models. Monographs on Statistics and Applied Probability* **43**. CRC Press, London. MR1082147

OWEN, A. B. (2007). The pigeonhole bootstrap. *Ann. Appl. Stat.* **1** 386–411. MR2415741 https://doi.org/10.1214/07-AOAS122

PAPASPILIOPOULOS, O., ROBERTS, G. O. and ZANELLA, G. (2020). Scalable inference for crossed random effects models. *Biometrika* **107** 25–40. MR4064138 https://doi.org/10.1093/biomet/asz058

R CORE TEAM (2015). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

MSC2020 subject classifications. Primary 62J05; secondary 65C60.
Key words and phrases. Collapsed Gibbs sampler, mixed effect models, generalized least squares.

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

ROBINSON, G. K. (1991). That BLUP is a good thing: The estimation of random effects. *Statist. Sci.* **6** 15–51. MR1108815

SEARLE, S. R., CASELLA, G. and MCCULLOCH, C. E. (1992). *Variance Components. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Wiley, New York. MR1190470 https://doi.org/10.1002/9780470316856

STRASSEN, V. (1969). Gaussian elimination is not optimal. *Numer. Math.* **13** 354–356. MR0248973 https://doi.org/10.1007/BF02165411

VATS, D., FLEGAL, J. M. and JONES, G. L. (2019). Multivariate output analysis for Markov chain Monte Carlo. *Biometrika* **106** 321–337. MR3949306 https://doi.org/10.1093/biomet/asz002

The Annals of Statistics
2022, Vol. 50, No. 1, 584–613
https://doi.org/10.1214/21-AOS2123

TENSOR CLUSTERING WITH PLANTED STRUCTURES: STATISTICAL OPTIMALITY AND COMPUTATIONAL LIMITS

BY YUETIAN LUO[1,*] AND ANRU R. ZHANG[1,2,†]

[1]*Department of Statistics, University of Wisconsin-Madison,* [*]*yluo86@wisc.edu*
[2]*Department of Biostatistics & Bioinformatics, Duke University,* [†]*anru.zhang@duke.edu*

This paper studies the statistical and computational limits of high-order clustering with planted structures. We focus on two clustering models, constant high-order clustering (CHC) and rank-one higher-order clustering (ROHC), and study the methods and theory for testing whether a cluster exists (detection) and identifying the support of cluster (recovery).

Specifically, we identify the sharp boundaries of signal-to-noise ratio for which CHC and ROHC detection/recovery are statistically possible. We also develop the tight computational thresholds: when the signal-to-noise ratio is below these thresholds, we prove that polynomial-time algorithms cannot solve these problems under the computational hardness conjectures of hypergraphic planted clique (HPC) detection and hypergraphic planted dense subgraph (HPDS) recovery. We also propose polynomial-time tensor algorithms that achieve reliable detection and recovery when the signal-to-noise ratio is above these thresholds. Both sparsity and tensor structures yield the computational barriers in high-order tensor clustering. The interplay between them results in significant differences between high-order tensor clustering and matrix clustering in literature in aspects of statistical and computational phase transition diagrams, algorithmic approaches, hardness conjecture, and proof techniques. To our best knowledge, we are the first to give a thorough characterization of the statistical and computational trade-off for such a double computational-barrier problem. Finally, we provide evidence for the computational hardness conjectures of HPC detection (via low-degree polynomial and Metropolis methods) and HPDS recovery (via low-degree polynomial method).

REFERENCES

ALON, N., KRIVELEVICH, M. and SUDAKOV, B. (1998). Finding a large hidden clique in a random graph. *Random Structures Algorithms* **13** 457–466.

ALON, N., ANDONI, A., KAUFMAN, T., MATULEF, K., RUBINFELD, R. and XIE, N. (2007). Testing k-wise and almost k-wise independence. In *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing* 496–505. ACM, New York. MR2402475 https://doi.org/10.1145/1250790.1250863

AMAR, D., YEKUTIELI, D., MARON-KATZ, A., HENDLER, T. and SHAMIR, R. (2015). A hierarchical Bayesian model for flexible module discovery in three-way time-series data. *Bioinformatics* **31** i17–i26.

AMES, B. P. W. and VAVASIS, S. A. (2011). Nuclear norm minimization for the planted clique and biclique problems. *Math. Program.* **129** 69–89. MR2831403 https://doi.org/10.1007/s10107-011-0459-x

ANANDKUMAR, A., GE, R. and JANZAMIN, M. (2014). Guaranteed non-orthogonal tensor decomposition via alternating rank-1 updates. ArXiv preprint. Available at arXiv:1402.5180.

ARIAS-CASTRO, E. and VERZELEN, N. (2014). Community detection in dense random networks. *Ann. Statist.* **42** 940–969. MR3210992 https://doi.org/10.1214/14-AOS1208

ARORA, S., BARAK, B., BRUNNERMEIER, M. and GE, R. (2011). Computational complexity and information asymmetry in financial products. *Commun. ACM* **54** 101–107.

AWASTHI, P., CHARIKAR, M., LAI, K. A. and RISTESKI, A. (2015). Label optimal regret bounds for online local learning. In *Conference on Learning Theory* 150–166.

MSC2020 subject classifications. Primary 62H15; secondary 62C20.

Key words and phrases. Average-case complexity, high-order clustering, hypergraphic planted clique, hypergraphic planted dense subgraph, statistical-computational phase transition.

BALAKRISHNAN, S., KOLAR, M., RINALDO, A., SINGH, A. and WASSERMAN, L. (2011). Statistical and computational tradeoffs in biclustering. In *NIPS* 2011 *Workshop on Computational Trade-Offs in Statistical Learning* **4**.

BANDEIRA, A. S., KUNISKY, D. and WEIN, A. S. (2020). Computational hardness of certifying bounds on constrained PCA problems. *Innov. Theor. Comp. Sci.*

BARAK, B. and MOITRA, A. (2016). Noisy tensor completion via the sum-of-squares hierarchy. In *Conference on Learning Theory* 417–445.

BARAK, B., HOPKINS, S., KELNER, J., KOTHARI, P. K., MOITRA, A. and POTECHIN, A. (2019). A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM J. Comput.* **48** 687–735. MR3945259 https://doi.org/10.1137/17M1138236

BERTHET, Q. and BALDIN, N. (2020). Statistical and computational rates in graph logistic regression. In *International Conference on Artificial Intelligence and Statistics* 2719–2730.

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

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

BOLLOBÁS, B. and ERDŐS, P. (1976). Cliques in random graphs. *Math. Proc. Cambridge Philos. Soc.* **80** 419–427. MR0498256 https://doi.org/10.1017/S0305004100053056

BRENNAN, M. and BRESLER, G. (2019a). Optimal average-case reductions to sparse pca: From weak assumptions to strong hardness. In *Conference on Learning Theory* 469–470. PMLR.

BRENNAN, M. and BRESLER, G. (2019b). Average-case lower bounds for learning sparse mixtures, robust estimation and semirandom adversaries. ArXiv preprint. Available at arXiv:1908.06130.

BRENNAN, M. and BRESLER, G. (2020). Reducibility and statistical-computational gaps from secret leakage. In *Conference on Learning Theory* 648–847. PMLR.

BRENNAN, M., BRESLER, G. and HULEIHEL, W. (2018). Reducibility and computational lower bounds for problems with planted sparse structure. In *Conference on Learning Theory* 48–166. PMLR.

BRENNAN, M., BRESLER, G. and HULEIHEL, W. (2019). Universality of computational lower bounds for submatrix detection. In *Conference on Learning Theory* 417–468. PMLR.

BUSYGIN, S., PROKOPYEV, O. and PARDALOS, P. M. (2008). Biclustering in data mining. *Comput. Oper. Res.* **35** 2964–2987. MR2586410 https://doi.org/10.1016/j.cor.2007.01.005

BUTUCEA, C. and INGSTER, Y. I. (2013). Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli* **19** 2652–2688. MR3160567 https://doi.org/10.3150/12-BEJ470

BUTUCEA, C., INGSTER, Y. I. and SUSLINA, I. A. (2015). Sharp variable selection of a sparse submatrix in a high-dimensional noisy matrix. *ESAIM Probab. Stat.* **19** 115–134. MR3374872 https://doi.org/10.1051/ps/2014017

CAI, T. T., LIANG, T. and RAKHLIN, A. (2017). Computational and statistical boundaries for submatrix localization in a large noisy matrix. *Ann. Statist.* **45** 1403–1430. MR3670183 https://doi.org/10.1214/16-AOS1488

CAI, T. T. and WU, Y. (2020). Statistical and computational limits for sparse matrix detection. *Ann. Statist.* **48** 1593–1614. MR4124336 https://doi.org/10.1214/19-AOS1860

CANDOGAN, U. O. and CHANDRASEKARAN, V. (2018). Finding planted subgraphs with few eigenvalues using the Schur–Horn relaxation. *SIAM J. Optim.* **28** 735–759. MR3775143 https://doi.org/10.1137/16M1075144

CHANDRASEKARAN, V. and JORDAN, M. I. (2013). Computational and statistical tradeoffs via convex relaxation. *Proc. Natl. Acad. Sci. USA* **110** E1181–E1190. MR3047651 https://doi.org/10.1073/pnas.1302293110

CHARIKAR, M., NAAMAD, Y. and WU, J. (2018). On finding dense common subgraphs. ArXiv preprint. Available at arXiv:1802.06361.

CHEN, Y. (2015). Incoherence-optimal matrix completion. *IEEE Trans. Inf. Theory* **61** 2909–2923. MR3342311 https://doi.org/10.1109/TIT.2015.2415195

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.

CHI, E. C., ALLEN, G. I. and BARANIUK, R. G. (2017). Convex biclustering. *Biometrics* **73** 10–19. MR3632347 https://doi.org/10.1111/biom.12540

CHLAMTÁČ, E., DINITZ, M. and KRAUTHGAMER, R. (2012). Everywhere-sparse spanners via dense subgraphs. In 2012 *IEEE* 53*rd Annual Symposium on Foundations of Computer Science—FOCS* 2012 758–767. IEEE Computer Soc., Los Alamitos, CA. MR3186664

CHLAMTÁČ, E., DINITZ, M. and MAKARYCHEV, Y. (2017). Minimizing the union: Tight approximations for small set bipartite vertex expansion. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms* 881–899. SIAM, Philadelphia, PA. MR3627785 https://doi.org/10.1137/1.9781611974782.56

CHLAMTÁC, E. and MANURANGSI, P. (2018). Sherali-Adams integrality gaps matching the log-density threshold. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques* (*APPROX/RANDOM* 2018) Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

CICHOCKI, A., MANDIC, D., DE LATHAUWER, L., ZHOU, G., ZHAO, Q., CAIAFA, C. and PHAN, H. A. (2015). Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE Signal Process. Mag.* **32** 145–163.

DE LATHAUWER, L., DE MOOR, B. and VANDEWALLE, J. (2000a). A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.* **21** 1253–1278. MR1780272 https://doi.org/10.1137/S0895479896305696

DE LATHAUWER, L., DE MOOR, B. and VANDEWALLE, J. (2000b). On the best rank-1 and rank-$(R_1, R_2, \dots, R_N)$ approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl.* **21** 1324–1342. MR1780276 https://doi.org/10.1137/S0895479898346995

DEKEL, Y., GUREL-GUREVICH, O. and PERES, Y. (2014). Finding hidden cliques in linear time with high probability. *Combin. Probab. Comput.* **23** 29–49. MR3197965 https://doi.org/10.1017/S096354831300045X

DESHPANDE, Y. and MONTANARI, A. (2015a). Finding hidden cliques of size $\sqrt{N/e}$ in nearly linear time. *Found. Comput. Math.* **15** 1069–1128. MR3371378 https://doi.org/10.1007/s10208-014-9215-y

DESHPANDE, Y. and MONTANARI, A. (2015b). Improved sum-of-squares lower bounds for hidden clique and hidden submatrix problems. In *Conference on Learning Theory* 523–562.

DIAKONIKOLAS, I., KANE, D. M. and STEWART, A. (2017). Statistical query lower bounds for robust estimation of high-dimensional Gaussians and Gaussian mixtures (extended abstract). In *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 73–84. IEEE Computer Soc., Los Alamitos, CA. MR3734219 https://doi.org/10.1109/FOCS.2017.16

DIAKONIKOLAS, I., KONG, W. and STEWART, A. (2019). Efficient algorithms and lower bounds for robust linear regression. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms* 2745–2754. SIAM, Philadelphia, PA. MR3909639 https://doi.org/10.1137/1.9781611975482.170

DING, Y., KUNISKY, D., WEIN, A. S. and BANDEIRA, A. S. (2019). Subexponential-time algorithms for sparse PCA. ArXiv preprint. Available at arXiv:1907.11635.

DING, Y., KUNISKY, D., WEIN, A. S. and BANDEIRA, A. S. (2020). The average-case time complexity of certifying the restricted isometry property. ArXiv preprint. Available at arXiv:2005.11270.

DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2009). Message-passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci. USA* **106** 18914–18919.

DUDEJA, R. and HSU, D. (2021). Statistical query lower bounds for tensor PCA. *J. Mach. Learn. Res.* **22** 83. MR4253776

FAN, J., LIU, H., WANG, Z. and YANG, Z. (2018). Curse of heterogeneity: Computational barriers in sparse mixture models and phase retrieval. ArXiv preprint. Available at arXiv:1808.06996.

FAUST, K., SATHIRAPONGSASUTI, J. F., IZARD, J., SEGATA, N., GEVERS, D., RAES, J. and HUTTENHOWER, C. (2012). Microbial co-occurrence relationships in the human microbiome. *PLoS Comput. Biol.* **8** e1002606.

FEIGE, U. and KRAUTHGAMER, R. (2000). Finding and certifying a large hidden clique in a semirandom graph. *Random Structures Algorithms* **16** 195–208. MR1742351 https://doi.org/10.1002/(SICI)1098-2418(200003)16:2<195::AID-RSA5>3.3.CO;2-1

FEIZI, S., JAVADI, H. and TSE, D. (2017). Tensor biclustering. In *Advances in Neural Information Processing Systems* 1311–1320.

FELDMAN, V., PERKINS, W. and VEMPALA, S. (2018). On the complexity of random satisfiability problems with planted solutions. *SIAM J. Comput.* **47** 1294–1338. MR3827195 https://doi.org/10.1137/16M1078471

FELDMAN, V., GRIGORESCU, E., REYZIN, L., VEMPALA, S. S. and XIAO, Y. (2017). Statistical algorithms and a lower bound for detecting planted cliques. *J. ACM* **64** 8. MR3664576 https://doi.org/10.1145/3046674

FLORES, G. E., CAPORASO, J. G., HENLEY, J. B., RIDEOUT, J. R., DOMOGALA, D., CHASE, J., LEFF, J. W., VÁZQUEZ-BAEZA, Y., GONZALEZ, A. et al. (2014). Temporal variability is a personalized feature of the human microbiome. *Genome Biol.* **15** 531. https://doi.org/10.1186/s13059-014-0531-y

GAMARNIK, D. and SUDAN, M. (2014). Limits of local algorithms over sparse random graphs [extended abstract]. In *ITCS'14—Proceedings of the* 2014 *Conference on Innovations in Theoretical Computer Science* 369–375. ACM, New York. MR3359490

GAMARNIK, D. and ZADIK, I. (2019). The landscape of the planted clique problem: Dense subgraphs and the overlap gap property. ArXiv preprint. Available at arXiv:1904.07174.

GAO, C., MA, Z. and ZHOU, H. H. (2017). Sparse CCA: Adaptive estimation and computational barriers. *Ann. Statist.* **45** 2074–2101. MR3718162 https://doi.org/10.1214/16-AOS1519

GHOSHDASTIDAR, D. and DUKKIPATI, A. (2017). Consistency of spectral hypergraph partitioning under planted partition model. *Ann. Statist.* **45** 289–315. MR3611493 https://doi.org/10.1214/16-AOS1453

HAJEK, B., WU, Y. and XU, J. (2015). Computational lower bounds for community detection on random graphs. In *Conference on Learning Theory* 899–928.

HAJEK, B., WU, Y. and XU, J. (2016). Information limits for recovering a hidden community. In 2016 *IEEE International Symposium on Information Theory* (*ISIT*) 1894–1898. IEEE, Los Alamitos.

HAN, R., WILLETT, R. and ZHANG, A. (2020). An optimal statistical and computational framework for generalized tensor estimation. ArXiv preprint. Available at arXiv:2002.11255.

HENRIQUES, R. and MADEIRA, S. C. (2019). Triclustering algorithms for three-dimensional data analysis: A comprehensive survey. *ACM Comput*. *Surv*. **51** 95.

HILLAR, C. J. and LIM, L.-H. (2013). Most tensor problems are NP-hard. *J*. *ACM* **60** 45. MR3144915 https://doi.org/10.1145/2512329

HOPKINS, S. B. K. (2018). Statistical inference and the sum of squares method.

HOPKINS, S. B., SHI, J. and STEURER, D. (2015). Tensor principal component analysis via sum-of-square proofs. In *Conference on Learning Theory* 956–1006.

HOPKINS, S. B. and STEURER, D. (2017). Efficient Bayesian estimation from few samples: Community detection and related problems. In 58*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 379–390. IEEE Computer Soc., Los Alamitos, CA. MR3734245 https://doi.org/10.1109/FOCS.2017.42

HOPKINS, S. B., KOTHARI, P. K., POTECHIN, A., RAGHAVENDRA, P., SCHRAMM, T. and STEURER, D. (2017). The power of sum-of-squares for detecting hidden structures. In 58*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 720–731. IEEE Computer Soc., Los Alamitos, CA. MR3734275 https://doi.org/10.1109/FOCS.2017.72

JERRUM, M. (1992). Large cliques elude the Metropolis process. *Random Structures Algorithms* **3** 347–359. MR1179827 https://doi.org/10.1002/rsa.3240030402

JI, L., TAN, K.-L. and TUNG, A. K. (2006). Mining frequent closed cubes in 3D datasets. In *Proceedings of the* 32*nd International Conference on Very Large Data Bases* 811–822. VLDB Endowment.

JIANG, D., PEI, J., RAMANATHAN, M., TANG, C. and ZHANG, A. (2004). Mining coherent gene clusters from gene-sample-time microarray data. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 430–439. ACM, New York.

KANNAN, R. and VEMPALA, S. (2017). The hidden hubs problem. In *Conference on Learning Theory* 1190–1213.

KIM, C., BANDEIRA, A. S. and GOEMANS, M. X. (2017). Community detection in hypergraphs, spiked tensor models, and sum-of-squares. In 2017 *International Conference on Sampling Theory and Applications* (*SampTA*) 124–128. IEEE, Los Alamitos.

KOIRAN, P. and ZOUZIAS, A. (2014). Hidden cliques and the certification of the restricted isometry property. *IEEE Trans*. *Inf*. *Theory* **60** 4999–5006. MR3245368 https://doi.org/10.1109/TIT.2014.2331341

KOLAR, M., BALAKRISHNAN, S., RINALDO, A. and SINGH, A. (2011). Minimax localization of structural information in large noisy matrices. In *Advances in Neural Information Processing Systems* 909–917.

KOLDA, T. G. and BADER, B. W. (2009). Tensor decompositions and applications. *SIAM Rev*. **51** 455–500. MR2535056 https://doi.org/10.1137/07070111X

KUNISKY, D., WEIN, A. S. and BANDEIRA, A. S. (2019). Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. ArXiv preprint. Available at arXiv:1907.11636.

LESIEUR, T., MIOLANE, L., LELARGE, M., KRZAKALA, F. and ZDEBOROVÁ, L. (2017). Statistical and computational phase transitions in spiked tensor estimation. In 2017 *IEEE International Symposium on Information Theory* (*ISIT*) 511–515. IEEE, Los Alamitos.

LI, A. and TUCK, D. (2009). An effective tri-clustering algorithm combining expression data with gene regulation information. *Gene Regul*. *Syst*. *Biol*. **3** 49–64. https://doi.org/10.4137/grsb.s1150

LÖFFLER, M., WEIN, A. S. and BANDEIRA, A. S. (2020). Computationally efficient sparse clustering. ArXiv preprint. Available at arXiv:2005.10817.

LUO, Y. and ZHANG, A. R. (2020a). Open problem: Average-case hardness of hypergraphic planted clique detection. *Conf*. *Learn*. *Theory* **125** 3852–3856.

LUO, Y. and ZHANG, A. R. (2022b). Supplement to "Tensor clustering with planted structures: Statistical optimality and computational limits." https://doi.org/10.1214/21-AOS2123SUPP

MA, T. and WIGDERSON, A. (2015). Sum-of-squares lower bounds for sparse PCA. In *Advances in Neural Information Processing Systems* 1612–1620.

MA, Z. and WU, Y. (2015). Computational barriers in minimax submatrix detection. *Ann*. *Statist*. **43** 1089–1116. MR3346698 https://doi.org/10.1214/14-AOS1300

MADEIRA, S. C. and OLIVEIRA, A. L. (2004). Biclustering algorithms for biological data analysis: A survey. *IEEE/ACM Trans*. *Comput*. *Biol*. *Bioinform*. **1** 24–45.

MANKAD, S. and MICHAILIDIS, G. (2014). Biclustering three-dimensional data arrays with plaid models. *J*. *Comput*. *Graph*. *Statist*. **23** 943–965. MR3270705 https://doi.org/10.1080/10618600.2013.851608

MCSHERRY, F. (2001). Spectral partitioning of random graphs. In 42*nd IEEE Symposium on Foundations of Computer Science* (*Las Vegas*, *NV*, 2001) 529–537. IEEE Computer Soc., Los Alamitos, CA. MR1948742

MEKA, R., POTECHIN, A. and WIGDERSON, A. (2015). Sum-of-squares lower bounds for planted clique [extended abstract]. In *STOC'15—Proceedings of the* 2015 *ACM Symposium on Theory of Computing* 87–96. ACM, New York. MR3388186

MONTANARI, A. (2015). Finding one community in a sparse graph. *J. Stat. Phys.* **161** 273–299. MR3401018 https://doi.org/10.1007/s10955-015-1338-2

NILES-WEED, J. and ZADIK, I. (2020). The all-or-nothing phenomenon in sparse tensor PCA. *Adv. Neural Inf. Process. Syst.* **33**.

PERRY, A., WEIN, A. S. and BANDEIRA, A. S. (2020). Statistical limits of spiked tensor models. *Ann. Inst. Henri Poincaré Probab. Stat.* **56** 230–264. MR4058987 https://doi.org/10.1214/19-AIHP960

RAGHAVENDRA, P., SCHRAMM, T. and STEURER, D. (2018). High-dimensional estimation via sum-of-squares proofs. ArXiv preprint. Available at arXiv:1807.11419.

RICHARD, E. and MONTANARI, A. (2014). A statistical model for tensor PCA. In *Advances in Neural Information Processing Systems* 2897–2905.

RON, D. and FEIGE, U. (2010). Finding hidden cliques in linear time. *Discrete Math. Theor. Comput. Sci.*.

ROSSMAN, B. (2008). On the constant-depth complexity of k-clique. In *STOC'08* 721–730. ACM, New York. MR2582693 https://doi.org/10.1145/1374376.1374480

ROSSMAN, B. (2014). The monotone complexity of k-clique on random graphs. *SIAM J. Comput.* **43** 256–279. MR3166976 https://doi.org/10.1137/110839059

SCHRAMM, T. and WEIN, A. S. (2020). Computational barriers to estimation from low-degree polynomials. ArXiv preprint. Available at arXiv:2008.02269.

SIM, K., AUNG, Z. and GOPALKRISHNAN, V. (2010). Discovering correlated subspace clusters in 3D continuous-valued data. In 2010 *IEEE International Conference on Data Mining* 471–480. IEEE, Los Alamitos.

SUN, X. and NOBEL, A. B. (2013). On the maximal size of large-average and ANOVA-fit submatrices in a Gaussian random matrix. *Bernoulli* **19** 275–294. MR3019495 https://doi.org/10.3150/11-BEJ394

SUN, W. W., LU, J., LIU, H. and CHENG, G. (2017). Provable sparse tensor decomposition. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 899–916. MR3641413 https://doi.org/10.1111/rssb.12190

TANAY, A., SHARAN, R. and SHAMIR, R. (2002). Discovering statistically significant biclusters in gene expression data. *Bioinformatics* **18** S136–S144.

VERZELEN, N. and ARIAS-CASTRO, E. (2015). Community detection in sparse random networks. *Ann. Appl. Probab.* **25** 3465–3510. MR3404642 https://doi.org/10.1214/14-AAP1080

WANG, T., BERTHET, Q. and PLAN, Y. (2016). Average-case hardness of RIP certification. In *Advances in Neural Information Processing Systems* 3819–3827.

WANG, T., BERTHET, Q. and SAMWORTH, R. J. (2016). Statistical and computational trade-offs in estimation of sparse principal components. *Ann. Statist.* **44** 1896–1930. MR3546438 https://doi.org/10.1214/15-AOS1369

WANG, M., FISCHER, J. and SONG, Y. S. (2019). Three-way clustering of multi-tissue multi-individual gene expression data using semi-nonnegative tensor decomposition. *Ann. Appl. Stat.* **13** 1103–1127. MR3963564 https://doi.org/10.1214/18-AOAS1228

WANG, Z., GU, Q. and LIU, H. (2015). Sharp computational-statistical phase transitions via oracle computational model. ArXiv preprint. Available at arXiv:1512.08861.

WEIN, A. S., EL ALAOUI, A. and MOORE, C. (2019). The Kikuchi hierarchy and tensor PCA. In 2019 *IEEE* 60*th Annual Symposium on Foundations of Computer Science* 1446–1468. IEEE Comput. Soc. Press, Los Alamitos, CA. MR4228236

WU, Y. and XU, J. (2021). Statistical problems with planted structures: Information-theoretical and computational limits. *Inf. Theor. Methods Data Sci.* 383.

XIA, D. and ZHOU, F. (2019). The sup-norm perturbation of HOSVD and low rank tensor denoising. *J. Mach. Learn. Res.* **20** 61. MR3960915

ZHANG, T. and GOLUB, G. H. (2001). Rank-one approximation to high order tensors. *SIAM J. Matrix Anal. Appl.* **23** 534–550. MR1871328 https://doi.org/10.1137/S0895479899352045

ZHANG, A. and HAN, R. (2019). Optimal sparse singular value decomposition for high-dimensional high-order data. *J. Amer. Statist. Assoc.* **114** 1708–1725. MR4047294 https://doi.org/10.1080/01621459.2018.1527227

ZHANG, A. and XIA, D. (2018). Tensor SVD: Statistical and computational limits. *IEEE Trans. Inf. Theory* **64** 7311–7338. MR3876445 https://doi.org/10.1109/TIT.2018.2841377