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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

Erratum

Correction

Vol. 52, No. 1—February 2024

THE ANNALS OF STATISTICS Vol. 52, No. 1, pp. 1–414 February 2024

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 52, Number 1, February 2024. Published bimonthly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, OH 44094, USA. Periodicals postage paid at Cleveland, Ohio, and at additional mailing offices.

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

The Annals of Statistics
2024, Vol. 52, No. 1, 1–51
https://doi.org/10.1214/22-AOS2227

THE RIGHT COMPLEXITY MEASURE IN LOCALLY PRIVATE ESTIMATION: IT IS NOT THE FISHER INFORMATION

BY JOHN C. DUCHI[1,a] AND FENG RUAN[2,b]

[1]*Department of Statistics, Stanford University,* [a]*jduchi@stanford.edu*
[2]*Department of Statistics and Data Science, Northwestern University,* [b]*fengruan@northwestern.edu*

We identify fundamental tradeoffs between statistical utility and privacy under local models of privacy in which data is kept private even from the statistician, providing instance-specific bounds for private estimation and learning problems by developing the *local minimax risk*. In contrast to approaches based on worst-case (minimax) error, which are conservative, this allows us to evaluate the difficulty of individual problem instances and delineate the possibilities for adaptation in private estimation and inference. Our main results show that the local modulus of continuity of the estimand with respect to the variation distance—as opposed to the Hellinger distance central to classical statistics—characterizes rates of convergence under locally private estimation for many notions of privacy, including differential privacy and its relaxations. As consequences of these results, we identify an alternative to the Fisher information for private estimation, giving a more nuanced understanding of the challenges of adaptivity and optimality.

REFERENCES

[1] ABADI, M., CHU, A., GOODFELLOW, I., MCMAHAN, B., MIRONOV, I., TALWAR, K. and ZHANG, L. (2016). Deep learning with differential privacy. In 23*rd ACM Conference on Computer and Communications Security* (*ACM CCS*) 308–318.

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

[3] APPLE DIFFERENTIAL PRIVACY TEAM (2017). Learning with privacy at scale. Available at https://machinelearning.apple.com/2017/12/06/learning-with-privacy-at-scale.html.

[4] BALLE, B., BELL, J., GASCON, A. and NISSIM, K. (2019). The privacy blanket of the shuffle model. In 39*th Annual International Cryptology Conference* (*CRYPTO*).

[5] BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York. MR1623559

[6] BIRGÉ, L. (1983). Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsch. Verw. Gebiete* **65** 181–237. MR0722129 https://doi.org/10.1007/BF00532480

[7] BROWN, L. D. (1986). *Fundamentals of Statistical Exponential Families*. IMS, Hayward, CA.

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

[9] BUN, M. and STEINKE, T. (2016). Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography. Part I. Lecture Notes in Computer Science* **9985** 635–658. Springer, Berlin. MR3591832 https://doi.org/10.1007/978-3-662-53641-4_24

[10] CAI, T. T. and LOW, M. G. (2015). A framework for estimation of convex functions. *Statist. Sinica* **25** 423–456. MR3379081

[11] 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

[12] CASELLA, G. and STRAWDERMAN, W. E. (1981). Estimating a bounded normal mean. *Ann. Statist.* **9** 870–878. MR0619290

MSC2020 subject classifications. 62G05, 62G10, 68P27.
Key words and phrases. Locally private estimation, local minimax complexity, L^1 information, strong data processing inequalities.

[13] CHATTERJEE, S., DUCHI, J., LAFFERTY, J. and ZHU, Y. (2016). Local minimax complexity of stochastic convex optimization. In *Advances in Neural Information Processing Systems* **29**.

[14] COHEN, J. E., KEMPERMAN, J. H. B. and ZBĂGANU, G. (1998). *Comparisons of Stochastic Matrices*: *With Applications in Information Theory*, *Statistics*, *Economics*, *and Population Sciences*. Birkhäuser, Inc., Boston, MA. MR1638308

[15] COVER, T. M. and THOMAS, J. A. (2006). *Elements of Information Theory*, 2nd ed. Wiley Interscience, Hoboken, NJ. MR2239987

[16] DEL MORAL, P., LEDOUX, M. and MICLO, L. (2003). On contraction properties of Markov kernels. *Probab. Theory Related Fields* **126** 395–420. MR1992499 https://doi.org/10.1007/s00440-003-0270-6

[17] DOBRUSHIN, R. L. (1956). Central limit theorem for nonstationary Markov chains. I. *Theory Probab. Appl.* **1** 65–80.

[18] DONOHO, D. L. and LIU, R. C. (1987). Geometrizing rates of convergence I. Technical report 137, Univ. California, Berkeley, Dept. Statistics.

[19] DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence. II. *Ann. Statist.* **19** 633–667.

[20] DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence. III. *Ann. Statist.* **19** 668–701.

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

[22] DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2018). Minimax optimal procedures for locally private estimation. *J. Amer. Statist. Assoc.* **113** 182–201. MR3803452 https://doi.org/10.1080/01621459.2017.1389735

[23] DUCHI, J. C. and ROGERS, R. (2019). Lower bounds for locally private estimation via communication complexity. In *Proceedings of the Thirty Second Annual Conference on Computational Learning Theory*.

[24] DUCHI, J. C. and RUAN, F. (2018). A constrained risk inequality for general losses. Preprint. Available at arXiv:1804.08116 [stat.TH].

[25] DUCHI, J. C. and RUAN, F. (2021). Asymptotic optimality in stochastic optimization. *Ann. Statist.* **49** 21–48. MR4206668 https://doi.org/10.1214/19-AOS1831

[26] DWORK, C., MCSHERRY, F., NISSIM, K. and SMITH, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography*. *Lecture Notes in Computer Science* **3876** 265–284. Springer, Berlin. MR2241676 https://doi.org/10.1007/11681878_14

[27] DWORK, C. and ROTH, A. (2013). The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.* **9** 211–487. MR3254020 https://doi.org/10.1561/0400000042

[28] DWORK, C. and ROTHBLUM, G. (2016). Concentrated differential privacy. Preprint. Available at arXiv:1603.01887 [cs.DS].

[29] DWORK, C. and SMITH, A. (2009). Differential privacy for statistics: What we know and what we want to learn. *J. Priv. Confid.* **1** 135–154.

[30] ERLINGSSON, Ú., FELDMAN, V., MIRONOV, I., RAGHUNATHAN, A., TALWAR, K. and THAKURTA, A. (2019). Amplification by shuffling: From local to central differential privacy via anonymity. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms* 2468–2479. SIAM, Philadelphia, PA. MR3909620 https://doi.org/10.1137/1.9781611975482.151

[31] ERLINGSSON, U., PIHUR, V. and KOROLOVA, A. (2014). RAPPOR: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the* 21*st ACM Conference on Computer and Communications Security* (*CCS*).

[32] EUROPEAN UNION (2018). 2018 reform of EU data protection rules. Available at https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en. Accessed May 2018.

[33] EVFIMIEVSKI, A. V., GEHRKE, J. and SRIKANT, R. (2003). Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the Twenty-Second Symposium on Principles of Database Systems* 211–222.

[34] HARDT, M. and TALWAR, K. (2010). On the geometry of differential privacy. In *STOC'*10*—Proceedings of the* 2010 *ACM International Symposium on Theory of Computing* 705–714. ACM, New York. MR2743320

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

[36] HIRIART-URRUTY, J.-B. and LEMARÉCHAL, C. (1993). *Convex Analysis and Minimization Algorithms*. *I*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **305**. Springer, Berlin. Fundamentals. MR1261420

[37] HOMER, N., SZELINGER, S., REDMAN, M., DUGGAN, D., TEMBE, W., MUEHLING, J., PEARSON, J. V., STEPHAN, D. A., NELSON, S. F. et al. (2008). Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genet*. **4** e1000167.

[38] 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

[39] LE CAM, L. and YANG, G. L. (2000). *Asymptotics in Statistics*: *Some Basic Concepts*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR1784901 https://doi.org/10.1007/978-1-4612-1166-2

[40] LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. *Springer Texts in Statistics*. Springer, New York. MR2135927

[41] LEYSIEFFER, F. W. and WARNER, S. L. (1976). Respondent jeopardy and optimal designs in randomized response models. *J. Amer. Statist. Assoc.* **71** 649–656. MR0426236

[42] LIESE, F. and VAJDA, I. (2006). On divergences and informations in statistics and information theory. *IEEE Trans. Inf. Theory* **52** 4394–4412. MR2300826 https://doi.org/10.1109/TIT.2006.881731

[43] MIRONOV, I. (2017). Rényi differential privacy. In 30*th IEEE Computer Security Foundations Symposium* (*CSF*) 263–275.

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

[45] POLLARD, D. (1997). Another look at differentiability in quadratic mean. In *Festschrift for Lucien Le Cam* 305–314. Springer, New York. MR1462953

[46] POLYAK, B. T. and JUDITSKY, A. B. (1992). Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.* **30** 838–855. MR1167814 https://doi.org/10.1137/0330046

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

[48] STEIN, C. (1956). Efficient nonparametric testing and estimation. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, *Vol. I* 187–195. Univ. California Press, Berkeley–Los Angeles, CA. MR0084921

[49] TSYBAKOV, A. B. (1998). Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Ann. Statist.* **26** 2420–2469. MR1700239 https://doi.org/10.1214/aos/1024691478

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

[51] VAJDA, I. (1972). On the f-divergence and singularity of probability measures. *Period. Math. Hungar.* **2** 223–234. MR0335163 https://doi.org/10.1007/BF02018663

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

[53] 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

[54] WAINWRIGHT, M. J. and JORDAN, M. I. (2008). Graphical models, exponential families, and variational inference. *Found. Trends Mach. Learn.* **1** 1–305.

[55] WARNER, S. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc.* **60** 63–69.

[56] WASSERMAN, L. and ZHOU, S. (2010). A statistical framework for differential privacy. *J. Amer. Statist. Assoc.* **105** 375–389. MR2656057 https://doi.org/10.1198/jasa.2009.tm08651

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

The Annals of Statistics
2024, Vol. 52, No. 1, 52–77
https://doi.org/10.1214/23-AOS2292

MAXIMUM LIKELIHOOD FOR HIGH-NOISE GROUP ORBIT ESTIMATION AND SINGLE-PARTICLE CRYO-EM

BY ZHOU FAN[1,a], ROY R. LEDERMAN[1,b], YI SUN[2,e], TIANHAO WANG[1,c] AND SHENG XU[1,d]

[1]*Department of Statistics and Data Science, Yale University,* [a]*zhou.fan@yale.edu,* [b]*roy.lederman@yale.edu,* [c]*tianhao.wang@yale.edu,* [d]*sheng.xu@yale.edu*

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

Motivated by applications to single-particle cryo-electron microscopy (cryo-EM), we study several problems of function estimation in a high noise regime, where samples are observed after random rotation and possible linear projection of the function domain. We describe a stratification of the Fisher information eigenvalues according to transcendence degrees of graded pieces of the algebra of group invariants, and we relate critical points of the log-likelihood landscape to a sequence of moment optimization problems, extending previous results for a discrete rotation group without projections.

We then compute the transcendence degrees and forms of these optimization problems for several examples of function estimation under $\mathsf{SO}(2)$ and $\mathsf{SO}(3)$ rotations, including a simplified model of cryo-EM as introduced by Bandeira, Blum-Smith, Kileel, Niles-Weed, Perry and Wein. We affirmatively resolve conjectures that third-order moments are sufficient to locally identify a generic signal up to its rotational orbit in these examples.

For low-dimensional approximations of the electric potential maps of two small protein molecules, we empirically verify that the noise scalings of the Fisher information eigenvalues conform with our theoretical predictions over a range of SNR, in a model of $\mathsf{SO}(3)$ rotations without projections.

REFERENCES

[1] ABBE, E., BENDORY, T., LEEB, W., PEREIRA, J. M., SHARON, N. and SINGER, A. (2019). Multireference alignment is easier with an aperiodic translation distribution. *IEEE Trans. Inf. Theory* **65** 3565–3584. MR3959006 https://doi.org/10.1109/TIT.2018.2889674

[2] ABBE, E., BENDORY, T., LEEB, W., PEREIRA, J. M., SHARON, N. and SINGER, A. (2019). Multireference alignment is easier with an aperiodic translation distribution. *IEEE Trans. Inf. Theory* **65** 3565–3584. MR3959006 https://doi.org/10.1109/TIT.2018.2889674

[3] ABBE, E., PEREIRA, J. M. and SINGER, A. (2018). Estimation in the group action channel. In 2018 *IEEE International Symposium on Information Theory* (*ISIT*) 561–565. IEEE, Vail, CO, USA.

[4] BANDEIRA, A. S., BLUM-SMITH, B., KILEEL, J., NILES-WEED, J., PERRY, A. and WEIN, A. S. (2023). Estimation under group actions: Recovering orbits from invariants. *Appl. Comput. Harmon. Anal.* **66** 236–319. MR4609472 https://doi.org/10.1016/j.acha.2023.06.001

[5] BANDEIRA, A. S., NILES-WEED, J. and RIGOLLET, P. (2019). Optimal rates of estimation for multi-reference alignment. *Math. Stat. Learn.* **2** 25–75. MR4073147 https://doi.org/10.4171/msl/11

[6] BENDORY, T., BARTESAGHI, A. and SINGER, A. (2020). Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities. *IEEE Signal Process. Mag.* **37** 58–76. https://doi.org/10.1109/msp.2019.2957822

[7] BENDORY, T., BOUMAL, N., MA, C., ZHAO, Z. and SINGER, A. (2018). Bispectrum inversion with application to multireference alignment. *IEEE Trans. Signal Process.* **66** 1037–1050. MR3771661 https://doi.org/10.1109/TSP.2017.2775591

[8] BENDORY, T., EDIDIN, D., LEEB, W. and SHARON, N. (2022). Dihedral multi-reference alignment. *IEEE Trans. Inf. Theory* **68** 3489–3499. MR4433234 https://doi.org/10.1109/tit.2022.3146488

MSC2020 subject classifications. 62G08.
Key words and phrases. Mixture models, nonparametric function estimation, group invariance.

[9] BENDORY, T., JAFFE, A., LEEB, W., SHARON, N. and SINGER, A. (2022). Super-resolution multi-reference alignment. *Inf. Inference* **11** 533–555. MR4440674 https://doi.org/10.1093/imaiai/iaab003

[10] BENDORY, T., JAFFE, A., LEEB, W., SHARON, N. and SINGER, A. (2022). Super-resolution multi-reference alignment. *Inf. Inference* **11** 533–555. MR4440674 https://doi.org/10.1093/imaiai/iaab003

[11] BENDORY, T., MICKELIN, O. and SINGER, A. (2022). Sparse multi-reference alignment: Sample complexity and computational hardness. In *ICASSP* 2022-2022 *IEEE International Conference on Acoustics*, *Speech and Signal Processing* (*ICASSP*) 8977–8981. IEEE, Singapore, Singapore.

[12] BRUNEL, V.-E. (2019). Learning rates for Gaussian mixtures under group action. In *Conference on Learning Theory* 471–491.

[13] DUBOCHET, J., ADRIAN, M., CHANG, J.-J., HOMO, J.-C., LEPAULT, J., MCDOWALL, A. W. and SCHULTZ, P. (1988). Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* **21** 129–228.

[14] FAN, Z., LEDERMAN, R. R, SUN, Y., WANG, T. and XU, S. (2024). Supplement to "Maximum likelihood for high-noise group orbit estimation and single-particle cryo-EM." https://doi.org/10.1214/23-AOS2292SUPP

[15] FAN, Z., SUN, Y., WANG, T. and WU, Y. (2023). Likelihood landscape and maximum likelihood estimation for the discrete orbit recovery model. *Comm. Pure Appl. Math.* **76** 1208–1302. MR4582297

[16] FRANK, J. (2006). *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*: *Visualization of Biological Molecules in Their Native State*. Oxford Univ. Press, London.

[17] GE, R., HUANG, F., JIN, C. and YUAN, Y. (2015). Escaping from saddle points—online stochastic gradient for tensor decomposition. In *Conference on Learning Theory* 797–842.

[18] GHOSH, S. and RIGOLLET, P. (2023). Sparse multi-reference alignment: Phase retrieval, uniform uncertainty principles and the beltway problem. *Found. Comput. Math.* **23** 1851–1898. MR4649436 https://doi.org/10.1007/s10208-022-09584-6

[19] GOODALL, C. (1991). Procrustes methods in the statistical analysis of shape. *J. Roy. Statist. Soc. Ser. B* **53** 285–339. MR1108330

[20] GOWER, J. C. (1975). Generalized Procrustes analysis. *Psychometrika* **40** 33–51. MR0405725 https://doi.org/10.1007/BF02291478

[21] HENDERSON, R., BALDWIN, J. M., CESKA, T. A., ZEMLIN, F., BECKMANN, E. A. and DOWNING, K. H. (1990). Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* **213** 899–929.

[22] IBRAGIMOV, I. and HAS'MINSKII, R. (1981). *Statistical Estimation*: *Asymptotic Theory*. Springer, Berlin.

[23] JIN, C., GE, R., NETRAPALLI, P., KAKADE, S. M. and JORDAN, M. I. (2017). How to escape saddle points efficiently. In *Proceedings of the* 34*th International Conference on Machine Learning-Volume* 70 1724–1732. JMLR.org.

[24] JOHNSTONE, I. M. (2017). *Gaussian Estimation*: *Sequence and Wavelet Models*. Unpublished draft.

[25] KAM, Z. (1980). The reconstruction of structure from electron micrographs of randomly oriented particles. *J. Theoret. Biol.* **82** 15–39. https://doi.org/10.1016/0022-5193(80)90088-0

[26] KATSEVICH, A. E. and BANDEIRA, A. S. (2023). Likelihood maximization and moment matching in low SNR Gaussian mixture models. *Comm. Pure Appl. Math.* **76** 788–842. MR4569606 https://doi.org/10.1002/cpa.22051

[27] KHOSHOUEI, M., RADJAINIA, M., BAUMEISTER, W. and DANEV, R. (2017). Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate. *Nat. Commun.* **8** 1–6.

[28] LEE, J. D., SIMCHOWITZ, M., JORDAN, M. I. and RECHT, B. (2016). Gradient descent only converges to minimizers. In *Conference on Learning Theory* 1246–1257.

[29] LIU, A. and MOITRA, A. (2021). Algorithms from invariants: Smoothed analysis of orbit recovery over SO(3). arXiv e-prints arXiv–2106.

[30] MATHIEU, M., PETITPAS, I., NAVAZA, J., LEPAULT, J., KOHLI, E., POTHIER, P., PRASAD, B. V., COHEN, J. and REY, F. A. (2001). Atomic structure of the major capsid protein of rotavirus: Implications for the architecture of the virion. *EMBO J.* **20** 1485–1497. https://doi.org/10.1093/emboj/20.7.1485

[31] 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

[32] MITYAGIN, B. S. (2020). The zero set of a real analytic function. *Mat. Zametki* **107** 473–475. MR4070868 https://doi.org/10.4213/mzm12620

[33] PERRY, A., WEED, J., BANDEIRA, A. S., RIGOLLET, P. and SINGER, A. (2019). The sample complexity of multireference alignment. *SIAM J. Math. Data Sci.* **1** 497–517. MR4002723 https://doi.org/10.1137/18M1214317

[34] PUMIR, T., SINGER, A. and BOUMAL, N. (2021). The generalized orthogonal Procrustes problem in the high noise regime. *Inf. Inference* **10** 921–954. MR4312088 https://doi.org/10.1093/imaiai/iaaa035

[35] PUNJANI, A., RUBINSTEIN, J. L., FLEET, D. J. and BRUBAKER, M. A. (2017). cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14** 290–296. https://doi.org/10.1038/nmeth.4169

[36] ROMANOV, E., BENDORY, T. and ORDENTLICH, O. (2021). Multi-reference alignment in high dimensions: Sample complexity and phase transition. *SIAM J. Math. Data Sci.* **3** 494–523. MR4245326 https://doi.org/10.1137/20M1354994

[37] SCHERES, S. H., GAO, H., VALLE, M., HERMAN, G. T., EGGERMONT, P. P., FRANK, J. and CARAZO, J.-M. (2007). Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nat. Methods* **4** 27–29.

[38] SCHERES, S. H. W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180** 519–530. https://doi.org/10.1016/j.jsb.2012.09.006

[39] SHARON, N., KILEEL, J., KHOO, Y., LANDA, B. and SINGER, A. (2020). Method of moments for 3D single particle *ab initio* modeling with non-uniform distribution of viewing angles. *Inverse Probl.* **36** 044003. MR4072349 https://doi.org/10.1088/1361-6420/ab6139

[40] SIGWORTH, F. J. (1998). A maximum-likelihood approach to single-particle image refinement. *J. Struct. Biol.* **122** 328–339. https://doi.org/10.1006/jsbi.1998.4014

[41] SINGER, A. and SIGWORTH, F. J. (2020). Computational methods for single-particle electron cryomicroscopy. *Annu. Rev. Biomed. Data Sci.* **3** 163–190. https://doi.org/10.1146/annurev-biodatasci-021020-093826

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

[43] WU, Y. and ZHOU, H. H. (2021). Randomly initialized EM algorithm for two-component Gaussian mixture achieves near optimality in $O(\sqrt{n})$ iterations. *Math. Stat. Learn.* **4** 143–220. MR4383733 https://doi.org/10.4171/msl/29

[44] XU, J., HSU, D. and MALEKI, A. (2016). Global analysis of expectation maximization for mixtures of two Gaussians. In *Proceedings of the* 30*th International Conference on Neural Information Processing Systems* 2684–2692.

[45] ZHANG, X., SETTEMBRE, E., XU, C., DORMITZER, P. R., BELLAMY, R., HARRISON, S. C. and GRIGORIEFF, N. (2008). Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. *Proc. Natl. Acad. Sci. USA* **105** 1867–1872.

The Annals of Statistics
2024, Vol. 52, No. 1, 78–102
https://doi.org/10.1214/23-AOS2296

STARTREK: COMBINATORIAL VARIABLE SELECTION WITH FALSE DISCOVERY RATE CONTROL

BY LU ZHANG[1,a] AND JUNWEI LU[2,b]

[1]*Department of Statistics, Harvard University,* [a]*luzhang.r2@gmail.com*
[2]*Department of Biostatistics, Harvard University,* [b]*junweilu@hsph.harvard.edu*

Variable selection on the large-scale networks has been extensively studied in the literature. While most of the existing methods are limited to the local functionals especially the graph edges, this paper focuses on selecting the discrete hub structures of the networks. Specifically, we propose an inferential method, called StarTrek filter, to select the hub nodes with degrees larger than a certain thresholding level in the high-dimensional graphical models and control the false discovery rate (FDR). Discovering hub nodes in the networks is challenging: there is no straightforward statistic for testing the degree of a node due to the combinatorial structures; complicated dependence in the multiple testing problem is hard to characterize and control. In methodology, the StarTrek filter overcomes this by constructing p-values based on the maximum test statistics via the Gaussian multiplier bootstrap. In theory, we show that the StarTrek filter can control the FDR by providing accurate bounds on the approximation errors of the quantile estimation and addressing the dependence structures among the maximal statistics.

To this end, we establish novel Cramér-type comparison bounds for the high-dimensional Gaussian random vectors. Compared to the Gaussian comparison bound via the Kolmogorov distance established by Chernozhukov, Chetverikov and Kato (*Ann. Statist.* **42** (2014) 1787–1818), our Cramér-type comparison bounds establish the relative difference between the distribution functions of two high-dimensional Gaussian random vectors, which is essential in the theoretical analysis of FDR control. Moreover, the StarTrek filter can be applied to general statistical models for FDR control of discovering discrete structures such as simultaneously testing the sparsity levels of multiple high-dimensional linear models. We illustrate the validity of the StarTrek filter in a series of numerical experiments and apply it to the genotype-tissue expression dataset to discover central regulator genes.

REFERENCES

[1] BAI, J., ZHANG, X., KANG, X., JIN, L., WANG, P. and WANG, Z. (2019). Screening of core genes and pathways in breast cancer development via comprehensive analysis of multi gene expression datasets. *Oncol. Lett.* **18** 5821–5830.
[2] 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
[3] BARBER, R. F. and CANDÈS, E. J. (2019). A knockoff filter for high-dimensional selective inference. *Ann. Statist.* **47** 2504–2537. MR3988764 https://doi.org/10.1214/18-AOS1755
[4] BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Rev. Econ. Stud.* **81** 608–650. MR3207983 https://doi.org/10.1093/restud/rdt044
[5] BENJAMINI, Y. (2010). Discovering the false discovery rate. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 405–416. MR2758522 https://doi.org/10.1111/j.1467-9868.2010.00746.x
[6] BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. MR1325392

MSC2020 subject classifications. Primary 62H15, 62H22; secondary 60F99.
Key words and phrases. Graphical models, multiple testing, false discovery rate control, combinatorial inference, Gaussian multiplier bootstrap, comparison bounds.

[7] BENJAMINI, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188. MR1869245 https://doi.org/10.1214/aos/1013699998

[8] BLEIN, S., BARJHOUX, L., GENESIS INVESTIGATORS, DAMIOLA, F., DONDON, M.-G., EON-MARCHAIS, S., MARCOU, M., CARON, O., LORTHOLARY, A. et al. (2015). Targeted sequencing of the mitochondrial genome of women at high risk of breast cancer without detectable mutations in BRCA1/2. *PLoS ONE* **10** e0136192.

[9] CAI, T., LIU, W. and LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.* **106** 594–607. MR2847973 https://doi.org/10.1198/jasa.2011.tm10155

[10] CAI, T., LIU, W. and XIA, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. Amer. Statist. Assoc.* **108** 265–277. MR3174618 https://doi.org/10.1080/01621459.2012.758041

[11] CAI, T. T. and MA, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** 2359–2388. MR3160557 https://doi.org/10.3150/12-BEJ455

[12] CAI, T. T. and ZHANG, A. (2016). Inference for high-dimensional differential correlation matrices. *J. Multivariate Anal.* **143** 107–126. MR3431422 https://doi.org/10.1016/j.jmva.2015.08.019

[13] 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

[14] CHANG, J., SHAO, Q.-M. and ZHOU, W.-X. (2016). Cramér-type moderate deviations for Studentized two-sample U-statistics with applications. *Ann. Statist.* **44** 1931–1956. MR3546439 https://doi.org/10.1214/15-AOS1375

[15] CHEN, W.-C., WANG, C.-Y., HUNG, Y.-H., WENG, T.-Y., YEN, M.-C. and LAI, M.-D. (2016). Systematic analysis of gene expression alterations and clinical outcomes for long-chain acyl-coenzyme A synthetase family in cancer. *PLoS ONE* **11** e0155660.

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

[17] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Anti-concentration and honest, adaptive confidence bands. *Ann. Statist.* **42** 1787–1818. MR3262468 https://doi.org/10.1214/14-AOS1235

[18] 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

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

[20] DAI, C., LIN, B., XING, X. and LIU, J. S. (2022). False discovery rate control via data splitting. *J. Amer. Statist. Assoc.* 1–18. https://doi.org/10.1080/01621459.2022.2060113

[21] DAI, C., LIN, B., XING, X. and LIU, J. S. (2023). A scale-free approach for false discovery rate control in generalized linear models. *J. Amer. Statist. Assoc.* **118** 1551–1565. MR4646583 https://doi.org/10.1080/01621459.2023.2165930

[22] DAI, R. and BARBER, R. (2016). The knockoff filter for FDR control in group-sparse and multitask regression. In *International Conference on Machine Learning* 1851–1859. PMLR.

[23] DENG, H. and ZHANG, C.-H. (2020). Beyond Gaussian approximation: Bootstrap for maxima of sums of independent random vectors. *Ann. Statist.* **48** 3643–3671. MR4185823 https://doi.org/10.1214/20-AOS1946

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

[25] EISENACH, C., BUNEA, F., NING, Y. and DINICU, C. (2020). High-dimensional inference for cluster-based graphical models. *J. Mach. Learn. Res.* **21** Paper No. 53, 55. MR4095332

[26] FENG, H. and NING, Y. (2019). High-dimensional mixed graphical model with ordinal data: Parameter estimation and statistical inference. In *The 22nd International Conference on Artificial Intelligence and Statistics* 654–663. PMLR.

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

[28] GRONWALL, T. H. (1919). Note on the derivatives with respect to a parameter of the solutions of a system of differential equations. *Ann. of Math.* (2) **20** 292–296. MR1502565 https://doi.org/10.2307/1967124

[29] GU, Q., CAO, Y., NING, Y. and LIU, H. (2015). Local and global inference for high dimensional nonparanormal graphical models. Preprint. Available at arXiv:1502.02347.

[30] HELLWIG, B., MADJAR, K., EDLUND, K., MARCHAN, R., CADENAS, C., HEIMES, A.-S., ALMSTEDT, K., LEBRECHT, A., SICKING, I. et al. (2016). Epsin family member 3 and ribosome-related genes are associated with late metastasis in estrogen receptor-positive breast cancer and long-term survival in non-small cell lung cancer using a genome-wide identification and validation strategy. *PLoS ONE* **11** 1–18. https://doi.org/10.1371/journal.pone.0167585

[31] ILYAS, M. U., SHAFIQ, M. Z., LIU, A. X. and RADHA, H. (2011). A distributed and privacy preserving algorithm for identifying information hubs in social networks. In 2011 *Proceedings IEEE INFOCOM* 561–565. IEEE.

[32] JACCARD, P. (1901). Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. *Bull. Soc. Vaud. Sci. Nat.* **37** 241–272.

[33] JANKOVÁ, J. and VAN DE GEER, S. (2017). Honest confidence regions and optimality in high-dimensional precision matrix estimation. *TEST* **26** 143–162. MR3613609 https://doi.org/10.1007/s11749-016-0503-5

[34] JAVANMARD, A. and JAVADI, H. (2019). False discovery rate control via debiased lasso. *Electron. J. Stat.* **13** 1212–1253. MR3935848 https://doi.org/10.1214/19-ejs1554

[35] JAVANMARD, A. and MONTANARI, A. (2013). Nearly optimal sample size in hypothesis testing for high-dimensional regression. In 2013 51*st Annual Allerton Conference on Communication, Control, and Computing (Allerton)* 1427–1434. IEEE.

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

[37] JAVANMARD, A. and MONTANARI, A. (2014). Hypothesis testing in high-dimensional regression under the Gaussian random design model: Asymptotic theory. *IEEE Trans. Inf. Theory* **60** 6522–6554. MR3265038 https://doi.org/10.1109/TIT.2014.2343629

[38] JIN, J., KE, Z. T., LUO, S. and WANG, M. (2023). Optimal Estimation of the Number of Network Communities. *J. Amer. Statist. Assoc.* **118** 2101–2116.

[39] KE, Z. T., MA, Y. and LIN, X. (2023). Estimation of the number of spiked eigenvalues in a covariance matrix by bulk eigenvalue matching analysis. *J. Amer. Statist. Assoc.* **118** 374–392.

[40] KUCHIBHOTLA, A. K., MUKHERJEE, S. and BANERJEE, D. (2021). High-dimensional CLT: Improvements, non-uniform extensions and large deviations. *Bernoulli* **27** 192–217. MR4177366 https://doi.org/10.3150/20-BEJ1233

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

[42] LEE, R. K.-W., HOANG, T.-A. and LIM, E.-P. (2019). Discovering hidden topical hubs and authorities across multiple online social networks. *IEEE Trans. Knowl. Data Eng.* **33** 70–84.

[43] LI, J. and MAATHUIS, M. H. (2021). GGM knockoff filter: False discovery rate control for Gaussian graphical models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 534–558. MR4294543 https://doi.org/10.1111/rssb.12430

[44] LI, Y., GIORGI, E. E., BECKMAN, K. B., CABERTO, C., KAZMA, R., LUM-JONES, A., HAIMAN, C. A., MARCHAND, L. L., STRAM, D. O. et al. (2019). Association between mitochondrial genetic variation and breast cancer risk: The multiethnic cohort. *PLoS ONE* **14** 1–14. https://doi.org/10.1371/journal.pone.0222284

[45] LIU, M., XIA, Y., CHO, K. and CAI, T. (2021). Integrative high dimensional multiple testing with heterogeneity under data sharing constraints. *J. Mach. Learn. Res.* **22** Paper No. 126, 26. MR4279777

[46] LIU, W. (2013). Gaussian graphical model estimation with false discovery rate control. *Ann. Statist.* **41** 2948–2978. MR3161453 https://doi.org/10.1214/13-AOS1169

[47] LIU, W. and LUO, S. (2014). Hypothesis testing for high-dimensional regression models.

[48] LIU, W. and SHAO, Q.-M. (2010). Cramér-type moderate deviation for the maximum of the periodogram with application to simultaneous tests in gene expression time series. *Ann. Statist.* **38** 1913–1935. MR2662363 https://doi.org/10.1214/09-AOS774

[49] LIU, W. and SHAO, Q.-M. (2014). Phase transition and regularized bootstrap in large-scale t-tests with false discovery rate control. *Ann. Statist.* **42** 2003–2025. MR3262475 https://doi.org/10.1214/14-AOS1249

[50] LIU, Y., GU, H.-Y., ZHU, J., NIU, Y.-M., ZHANG, C. and GUO, G.-L. (2019). Identification of hub genes and key pathways associated with bipolar disorder based on weighted gene co-expression network analysis. *Front. Physiol.* **10** 1081.

[51] LIU, Y., YI, Y., WU, W., WU, K. and ZHANG, W. (2019). Bioinformatics prediction and analysis of hub genes and pathways of three types of gynecological cancer. *Oncol. Lett.* **18** 617–628.

[52] LONSDALE, J., THOMAS, J., SALVATORE, M., PHILLIPS, R., LO, E., SHAD, S., HASZ, R., WALTERS, G., GARCIA, F. et al. (2013). The genotype-tissue expression (GTEx) project. *Nat. Genet.* **45** 580–585.

[53] LOU, W., DING, B., WANG, S. and FU, P. (2020). Overexpression of GPX3, a potential biomarker for diagnosis and prognosis of breast cancer, inhibits progression of breast cancer cells in vitro. *Cancer Cell Int.* **20** 1–15.

[54] LU, J., NEYKOV, M. and LIU, H. (2017). Adaptive inferential method for monotone graph invariants. Preprint. Available at arXiv:1707.09114.

[55] LUSCOMBE, N. M., BABU, M. M., YU, H., SNYDER, M., TEICHMANN, S. A. and GERSTEIN, M. (2004). Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **431** 308–312.

[56] MALVIA, S., BAGADI, S. A. R., PRADHAN, D., CHINTAMANI, C., BHATNAGAR, A., ARORA, D., SARIN, R. and SAXENA, S. (2019). Study of gene expression profiles of breast cancers in Indian women. *Sci. Rep.* **9** 1–15.

[57] MARINO, N., GERMAN, R., RAO, X., SIMPSON, E., LIU, S., WAN, J., LIU, Y., SANDUSKY, G., JACOBSEN, M. et al. (2020). Upregulation of lipid metabolism genes in the breast prior to cancer diagnosis. *npj Breast Cancer* **6** 1–13.

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

[59] MOHAMED, M. M., SABET, S., PENG, D.-F., NOUH, M. A., EL-SHINAWI, M. and EL-RIFAI, W. (2014). Promoter hypermethylation and suppression of glutathione peroxidase 3 are associated with inflammatory breast carcinogenesis. *Oxid. Med. Cell. Longev.* **2014**.

[60] NEWMAN, M. E., WATTS, D. J. and STROGATZ, S. H. (2002). Random graph models of social networks. *Proc. Natl. Acad. Sci. USA* **99** 2566–2572.

[61] NEYKOV, M., LU, J. and LIU, H. (2019). Combinatorial inference for graphical models. *Ann. Statist.* **47** 795–827. MR3909951 https://doi.org/10.1214/17-AOS1650

[62] OBOZINSKI, G., TASKAR, B. and JORDAN, M. (2006). Multi-task feature selection. Statistics Department, UC Berkeley, Tech. Rep. **2** 2.

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

[64] POWER, J. D., SCHLAGGAR, B. L., LESSOV-SCHLAGGAR, C. N. and PETERSEN, S. E. (2013). Evidence for hubs in human functional brain networks. *Neuron* **79** 798–813. https://doi.org/10.1016/j.neuron.2013.07.035

[65] RAVIKUMAR, P., WAINWRIGHT, M. J., RASKUTTI, G. and YU, B. (2011). High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electron. J. Stat.* **5** 935–980. MR2836766 https://doi.org/10.1214/11-EJS631

[66] REN, Z., SUN, T., ZHANG, C.-H. and ZHOU, H. H. (2015). Asymptotic normality and optimalities in estimation of large Gaussian graphical models. *Ann. Statist.* **43** 991–1026. MR3346695 https://doi.org/10.1214/14-AOS1286

[67] ROTHMAN, A. J., BICKEL, P. J., LEVINA, E. and ZHU, J. (2008). Sparse permutation invariant covariance estimation. *Electron. J. Stat.* **2** 494–515. MR2417391 https://doi.org/10.1214/08-EJS176

[68] RUBINOV, M. and SPORNS, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage* **52** 1059–1069. https://doi.org/10.1016/j.neuroimage.2009.10.003

[69] SHAW, P., KABANI, N. J., LERCH, J. P., ECKSTRAND, K., LENROOT, R., GOGTAY, N., GREENSTEIN, D., CLASEN, L., EVANS, A. et al. (2008). Neurodevelopmental trajectories of the human cerebral cortex. *J. Neurosci.* **28** 3586–3594. https://doi.org/10.1523/JNEUROSCI.5309-07.2008

[70] SHEN, S. and LU, J. (2023). Combinatorial-probabilistic trade-off: P-values of community property test in the stochastic block models. *IEEE Trans. Inf. Theory* **69** 6605–6618. MR4650324

[71] SHEN, X., PAN, W. and ZHU, Y. (2012). Likelihood-based selection and sharp parameter estimation. *J. Amer. Statist. Assoc.* **107** 223–232. MR2949354 https://doi.org/10.1080/01621459.2011.645783

[72] SIROIS, I., AGUILAR-MAHECHA, A., LAFLEUR, J., FOWLER, E., VU, V., SCRIVER, M., BUCHANAN, M., CHABOT, C., RAMANATHAN, A. et al. (2019). A unique morphological phenotype in chemoresistant triple-negative breast cancer reveals metabolic reprogramming and PLIN4 expression as a molecular vulnerability. *Mol. Cancer Res.* **17** 2492–2507. https://doi.org/10.1158/1541-7786.MCR-19-0264

[73] SLEPIAN, D. (1962). The one-sided barrier problem for Gaussian noise. *Bell Syst. Tech. J.* **41** 463–501. MR0133183 https://doi.org/10.1002/j.1538-7305.1962.tb02419.x

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

[75] 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

[76] VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. and DEZEURE, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42** 1166–1202. MR3224285 https://doi.org/10.1214/14-AOS1221

[77] VAN DEN HEUVEL, M. P. and SPORNS, O. (2013). Network hubs in the human brain. *Trends Cogn. Sci.* **17** 683–696. https://doi.org/10.1016/j.tics.2013.09.012

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

[79] XIA, Y., CAI, T. and CAI, T. T. (2018). Multiple testing of submatrices of a precision matrix with applications to identification of between pathway interactions. *J. Amer. Statist. Assoc.* **113** 328–339. MR3803468 https://doi.org/10.1080/01621459.2016.1251930

[80] XING, X., ZHAO, Z. and LIU, J. S. (2023). Controlling false discovery rate using Gaussian mirrors. *J. Amer. Statist. Assoc.* **118** 222–241. MR4571118 https://doi.org/10.1080/01621459.2021.1923510

[81] YANG, Z., NING, Y. and LIU, H. (2018). On semiparametric exponential family graphical models. *J. Mach. Learn. Res.* **19** Paper No. 57, 59. MR3899759

[82] YU, M., GUPTA, V. and KOLAR, M. (2020). Simultaneous inference for pairwise graphical models with generalized score matching. *J. Mach. Learn. Res.* **21** Paper No. 91, 51. MR4119159

[83] YUAN, L., CHEN, L., QIAN, K., QIAN, G., WU, C.-L., WANG, X. and XIAO, Y. (2017). Co-expression network analysis identified six hub genes in association with progression and prognosis in human clear cell renal cell carcinoma (ccRCC). *Genomics Data* **14** 132–140.

[84] YUAN, M. and LIN, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94** 19–35. MR2367824 https://doi.org/10.1093/biomet/asm018

[85] ZHANG, L. and LU, J. (2024). Supplement to "StarTrek: Combinatorial variable selection with false discovery rate control." https://doi.org/10.1214/23-AOS2296SUPP

[86] ZHAO, Q., SUR, P. and CANDÈS, E. J. (2022). The asymptotic distribution of the MLE in high-dimensional logistic models: Arbitrary covariance. *Bernoulli* **28** 1835–1861. MR4411513 https://doi.org/10.3150/21-bej1401

[87] ZHAO, T., LIU, H., ROEDER, K., LAFFERTY, J. and WASSERMAN, L. (2012). The `huge` package for high-dimensional undirected graph estimation in R. *J. Mach. Learn. Res.* **13** 1059–1062. MR2930633

The Annals of Statistics
2024, Vol. 52, No. 1, 103–130
https://doi.org/10.1214/23-AOS2325

CHARACTERIZATION OF CAUSAL ANCESTRAL GRAPHS FOR TIME SERIES WITH LATENT CONFOUNDERS

BY ANDREAS GERHARDUS[a]

Institute of Data Science, German Aerospace Center, [a]*andreas.gerhardus@dlr.de*

In this paper, we introduce a novel class of graphical models for representing time-lag specific causal relationships and independencies of multivariate time series with unobserved confounders. We completely characterize these graphs and show that they constitute proper subsets of the currently employed model classes. As we show, from the novel graphs one can thus draw stronger causal inferences—without additional assumptions. We further introduce a graphical representation of Markov equivalence classes of the novel graphs. This graphical representation contains more causal knowledge than what current state-of-the-art causal discovery algorithms learn.

REFERENCES

ALI, R. A., RICHARDSON, T. S. and SPIRTES, P. (2009). Markov equivalence for ancestral graphs. *Ann. Statist.* **37** 2808–2837. MR2541448 https://doi.org/10.1214/08-AOS626

ANDREWS, B., SPIRTES, P. and COOPER, G. F. (2020). On the completeness of causal discovery in the presence of latent confounding with tiered background knowledge. 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** 4002–4011. PMLR.

ASSAAD, C. K., DEVIJVER, E. and GAUSSIER, E. (2022). Discovery of extended summary graphs in time series. In *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence* (J. Cussens and K. Zhang, eds.). *Proceedings of Machine Learning Research* **180** 96–106. PMLR.

BOLLEN, K. A. (1989). *Structural Equations with Latent Variables*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. Wiley, New York. MR0996025 https://doi.org/10.1002/9781118619179

BONGERS, S., BLOM, T. and MOOIJ, J. M. (2018). Causal modeling of dynamical systems. Preprint. Available at arXiv:1803.08784.

CHU, T. and GLYMOUR, C. (2008). Search for additive nonlinear time series causal models. *J. Mach. Learn. Res.* **9** 967–991. MR2417260

DIDELEZ, V. (2008). Graphical models for marked point processes based on local independence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 245–264. MR2412641 https://doi.org/10.1111/j.1467-9868.2007.00634.x

EICHLER, M. (2010). Graphical Gaussian modelling of multivariate time series with latent variables. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (Y. W. Teh and M. Titterington, eds.). *Proceedings of Machine Learning Research* **9** 193–200. PMLR.

EICHLER, M. and DIDELEZ, V. (2007). Causal reasoning in graphical time series models. In *Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence* (*UAI*2007) (R. Parr and L. van der Gaag, eds.) 109–116. AUAI Press.

EICHLER, M. and DIDELEZ, V. (2010). On Granger causality and the effect of interventions in time series. *Lifetime Data Anal.* **16** 3–32. MR2575937 https://doi.org/10.1007/s10985-009-9143-3

ENTNER, D. and HOYER, P. O. (2010). On causal discovery from time series data using FCI. In *Proceedings of the 5th European Workshop on Probabilistic Graphical Models* (P. Myllymäki, T. Roos and T. Jaakkola, eds.) 121–128. Helsinki Institute for Information Technology HIIT, Helsinki, Finland.

GAO, W. and TIAN, Z. (2010). Latent ancestral graph of structure vector autoregressive models. *J. Syst. Eng. Electron.* **21** 233–238.

GEIGER, D., VERMA, T. and PEARL, J. (1990). Identifying independence in Bayesian networks *Networks* **20** 507–534. MR1064736 https://doi.org/10.1002/net.3230200504

MSC2020 subject classifications. Primary 62A09, 62D20, 62M10; secondary 68T30, 68T37.

Key words and phrases. Causal graph, ancestral graph, time series, latent variable, causal inference, causal discovery.

GERHARDUS, A. (2024). Supplement to "Characterization of causal ancestral graphs for time series with latent confounders." https://doi.org/10.1214/23-AOS2325SUPP

GERHARDUS, A. and RUNGE, J. (2020). High-recall causal discovery for autocorrelated time series with latent confounders. In *Advances in Neural Information Processing Systems* (H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan and H. Lin, eds.) **33** 12615–12625. Curran Associates, Red Hook, NY.

GERHARDUS, A., WAHL, J., FALTENBACHER, S., NINAD, U. and RUNGE, J. (2023). Projecting infinite time series graphs to finite marginal graphs using number theory. Preprint. Available at arXiv:2310.05526.

GRANGER, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37** 424–438.

HUCKINS, J. F., DASILVA, A. W., HEDLUND, E. L., MURPHY, E. I., ROGERS, C., WANG, W., OBUCHI, M., HOLTZHEIMER, P. E., WAGNER, D. D. et al. (2020). Causal factors of anxiety and depression in college students: Longitudinal ecological momentary assessment and causal analysis using Peter and Clark momentary conditional independence. *JMIR Mental Health* **7** e16684.

HYVÄRINEN, A., ZHANG, K., SHIMIZU, S. and HOYER, P. O. (2010). Estimation of a structural vector autoregression model using non-Gaussianity. *J. Mach. Learn. Res.* **11** 1709–1731. MR2653353

KOLLER, D. and FRIEDMAN, N. (2009). *Probabilistic Graphical Models*: *Principles and Techniques*. MIT Press, Cambridge, MA.

KRETSCHMER, M., COUMOU, D., DONGES, J. F. and RUNGE, J. (2016). Using causal effect networks to analyze different Arctic drivers of midlatitude winter circulation. *J. Climate* **29** 4069–4081.

MAATHUIS, M. H. and COLOMBO, D. (2015). A generalized back-door criterion. *Ann. Statist.* **43** 1060–1088. MR3346697 https://doi.org/10.1214/14-AOS1295

MALINSKY, D. and SPIRTES, P. (2016). Estimating causal effects with ancestral graph Markov models. In *Proceedings of the Eighth International Conference on Probabilistic Graphical Models* (A. Antonucci, G. Corani and C. P. Campos, eds.). *Proceedings of Machine Learning Research* **52** 299–309. PMLR.

MALINSKY, D. and SPIRTES, P. (2018). Causal structure learning from multivariate time series in settings with unmeasured confounding. In *Proceedings of* 2018 *ACM SIGKDD Workshop on Causal Discovery* (T. D. Le, K. Zhang, E. Kıcıman, A. Hyvärinen and L. Liu, eds.). *Proceedings of Machine Learning Research* **92** 23–47. PMLR.

MOGENSEN, S. W. and HANSEN, N. R. (2020). Markov equivalence of marginalized local independence graphs. *Ann. Statist.* **48** 539–559. MR4065173 https://doi.org/10.1214/19-AOS1821

MOOIJ, J. M. and CLAASSEN, T. (2020). Constraint-based causal discovery using partial ancestral graphs in the presence of cycles. In *Proceedings of the* 36*th Conference on Uncertainty in Artificial Intelligence* (*UAI*) (J. Peters and D. Sontag, eds.). *Proceedings of Machine Learning Research* **124** 1159–1168. PMLR.

PAMFIL, R., SRIWATTANAWORACHAI, N., DESAI, S., PILGERSTORFER, P., GEORGATZIS, K., BEAUMONT, P. and ARAGAM, B. (2020). DYNOTEARS: Structure learning from time-series data. 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** 1595–1605. PMLR.

PEARL, J. (1988). *Probabilistic Reasoning in Intelligent Systems*: *Networks of Plausible Inference*. *The Morgan Kaufmann Series in Representation and Reasoning*. Morgan Kaufmann, San Francisco, CA. MR0965765

PEARL, J. (1995). Causal diagrams for empirical research. *Biometrika* **82** 669–688. MR1380809 https://doi.org/10.1093/biomet/82.4.669

PEARL, J. (2000). *Causality*: *Models*, *Reasoning*, *and Inference*. Cambridge Univ. Press, Cambridge. MR1744773

PEARL, J. (2009). *Causality*: *Models*, *Reasoning*, *and Inference*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2548166 https://doi.org/10.1017/CBO9780511803161

PERKOVIĆ, E., TEXTOR, J., KALISCH, M. and MAATHUIS, M. H. (2017). Complete graphical characterization and construction of adjustment sets in Markov equivalence classes of ancestral graphs. *J. Mach. Learn. Res.* **18** 1–62. MR3827108

PETERS, J., JANZING, D. and SCHÖLKOPF, B. (2013). Causal inference on time series using restricted structural equation models. In *Advances in Neural Information Processing Systems* (C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani and K. Q. Weinberger, eds.) **26**. Curran Associates, Red Hook, NY.

PETERS, J., JANZING, D. and SCHÖLKOPF, B. (2017). *Elements of Causal Inference*: *Foundations and Learning Algorithms*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3822088

RICHARDSON, T. and SPIRTES, P. (2002). Ancestral graph Markov models. *Ann. Statist.* **30** 962–1030. MR1926166 https://doi.org/10.1214/aos/1031689015

RUNGE, J. (2020). Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. In *Proceedings of the* 36*th Conference on Uncertainty in Artificial Intelligence* (*UAI*) (J. Peters and D. Sontag, eds.). *Proceedings of Machine Learning Research* **124** 1388–1397. PMLR.

RUNGE, J., HEITZIG, J., PETOUKHOV, V. and KURTHS, J. (2012). Escaping the curse of dimensionality in estimating multivariate transfer entropy. *Phys. Rev. Lett.* **108** 258701. https://doi.org/10.1103/PhysRevLett.108.258701

SAETIA, S., YOSHIMURA, N. and KOIKE, Y. (2021). Constructing brain connectivity model using causal network reconstruction approach. *Front*. *Neuroinform*. **15** 619557. https://doi.org/10.3389/fninf.2021.619557

SPIRTES, P., GLYMOUR, C. and SCHEINES, R. (1993). *Causation*, *Prediction*, *and Search*. *Lecture Notes in Statistics* **81**. Springer, New York. MR1227558 https://doi.org/10.1007/978-1-4612-2748-9

SPIRTES, P., GLYMOUR, C. and SCHEINES, R. (2000). *Causation*, *Prediction*, *and Search*, 2nd ed. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. (With additional material by David Heckerman, Christopher Meek, Gregory F. Cooper, and Thomas Richardson.) MR1815675

SPIRTES, P., MEEK, C. and RICHARDSON, T. (1995). Causal inference in the presence of latent variables and selection bias. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence* (*UAI*1995) (P. Besnard and S. Hanks, eds.) 499–506. Morgan Kaufmann, San Francisco, CA.

VERMA, T. and PEARL, J. (1990). Causal Networks: Semantics and Expressiveness. In *Uncertainty in Artificial Intelligence* (R. D. Shachter, T. S. Levitt, L. N. Kanal and J. F. Lemmer, eds.). *Machine Intelligence and Pattern Recognition* **9** 69-76. North-Holland. https://doi.org/10.1016/B978-0-444-88650-7.50011-1

ZHANG, J. (2006). Causal inference and reasoning in causally insufficient systems, Ph.D. thesis, Department of Philosophy, Carnegie Mellon University.

ZHANG, J. (2008a). Causal reasoning with ancestral graphs. *J*. *Mach*. *Learn*. *Res*. **9** 1437–1474. MR2426048

ZHANG, J. (2008b). On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artif*. *Intell*. **172** 1873–1896. MR2459793 https://doi.org/10.1016/j.artint.2008.08.001

The Annals of Statistics
2024, Vol. 52, No. 1, 131–156
https://doi.org/10.1214/23-AOS2331

STATISTICAL-COMPUTATIONAL TRADE-OFFS IN TENSOR PCA AND RELATED PROBLEMS VIA COMMUNICATION COMPLEXITY

BY RISHABH DUDEJA[1,a] AND DANIEL HSU[2,b]

[1]*Department of Statistics, University of Wisconsin–Madison,* [a]*rdudeja@wisc.edu*
[2]*Department of Computer Science, Columbia University,* [b]*djhsu@cs.columbia.edu*

Tensor PCA is a stylized statistical inference problem introduced by Montanari and Richard to study the computational difficulty of estimating an unknown parameter from higher-order moment tensors. Unlike its matrix counterpart, Tensor PCA exhibits a statistical-computational gap, that is, a sample size regime where the problem is information-theoretically solvable but conjectured to be computationally hard. This paper derives computational lower bounds on the run-time of memory bounded algorithms for Tensor PCA using communication complexity. These lower bounds specify a trade-off among the number of passes through the data sample, the sample size and the memory required by any algorithm that successfully solves Tensor PCA. While the lower bounds do not rule out polynomial-time algorithms, they do imply that many commonly-used algorithms, such as gradient descent and power method, must have a higher iteration count when the sample size is not large enough. Similar lower bounds are obtained for non-Gaussian component analysis, a family of statistical estimation problems in which low-order moment tensors carry no information about the unknown parameter. Finally, stronger lower bounds are obtained for an asymmetric variant of Tensor PCA and related statistical estimation problems. These results explain why many estimators for these problems use a memory state that is significantly larger than the effective dimensionality of the parameter of interest.

REFERENCES

[1] ACHARYA, J., CANONNE, C. L., SUN, Z. and TYAGI, H. (2020). Unified lower bounds for interactive high-dimensional estimation under information constraints. Preprint. Available at arXiv:2010.06562.

[2] ALON, N., MATIAS, Y. and SZEGEDY, M. (1999). The space complexity of approximating the frequency moments. *J. Comput. System Sci.* **58** 137–147. MR1688610 https://doi.org/10.1006/jcss.1997.1545

[3] ANANDKUMAR, A., DENG, Y., GE, R. and MOBAHI, H. (2017). Homotopy analysis for tensor PCA. In *Conference on Learning Theory* 79–104.

[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] BARNES, L. P., HAN, Y. and ÖZGÜR, A. (2020). Lower bounds for learning distributions under communication constraints via Fisher information. *J. Mach. Learn. Res.* **21** Paper No. 236, 30 pp. MR4209522

[6] BEAME, P., GHARAN, S. O. and YANG, X. (2018). Time-space tradeoffs for learning finite functions from random evaluations, with applications to polynomials. In *Conference on Learning Theory* 843–856. PMLR.

[7] BEN AROUS, G., GHEISSARI, R. and JAGANNATH, A. (2020). Algorithmic thresholds for tensor PCA. *Ann. Probab.* **48** 2052–2087. MR4124533 https://doi.org/10.1214/19-AOP1415

[8] BEN AROUS, G., MEI, S., MONTANARI, A. and NICA, M. (2019). The landscape of the spiked tensor model. *Comm. Pure Appl. Math.* **72** 2282–2330. MR4011861 https://doi.org/10.1002/cpa.21861

[9] BHATTIPROLU, V., GURUSWAMI, V. and LEE, E. (2017). Sum-of-squares certificates for maxima of random tensors on the sphere. In *Approximation*, *Randomization*, *and Combinatorial Optimization*. *Algorithms and Techniques*. *LIPIcs*. *Leibniz Int*. *Proc*. *Inform*. **81** Art. No. 31, 20 pp. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR3695598

MSC2020 subject classifications. 62F30, 68Q17.
Key words and phrases. Statistical-computational trade-offs, tensors.

[10] BIROLI, G., CAMMAROTA, C. and RICCI-TERSENGHI, F. (2019). How to iron out rough landscapes and get optimal performances: Replicated gradient descent and its application to tensor PCA. Preprint. Available at arXiv:1905.12294.

[11] BLANCHARD, G., KAWANABE, M., SUGIYAMA, M., SPOKOINY, V. and MÜLLER, K.-R. (2006). In search of non-Gaussian components of a high-dimensional distribution. *J. Mach. Learn. Res.* **7** 247–282. MR2274368

[12] 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

[13] BRAVERMAN, M., GARG, A., MA, T., NGUYEN, H. L. and WOODRUFF, D. P. (2016). Communication lower bounds for statistical estimation problems via a distributed data processing inequality. In *STOC'*16*—Proceedings of the* 48*th Annual ACM SIGACT Symposium on Theory of Computing* 1011–1020. ACM, New York. MR3536632 https://doi.org/10.1145/2897518.2897582

[14] BRENNAN, M. and BRESLER, G. (2020). Reducibility and statistical-computational gaps from secret leakage. In *Conference on Learning Theory* 648–847. PMLR.

[15] BRENNAN, M., BRESLER, G., HOPKINS, S. B., LI, J. and SCHRAMM, T. (2020). Statistical query algorithms and low-degree tests are almost equivalent. Preprint. Available at arXiv:2009.06107.

[16] BRUNA, J., REGEV, O., SONG, M. J. and TANG, Y. (2021). Continuous LWE. In *STOC '*21*—Proceedings of the* 53*rd Annual ACM SIGACT Symposium on Theory of Computing* 694–707. ACM, New York. MR4398874 https://doi.org/10.1145/3406325.3451000

[17] BUBECK, S., LEE, Y. T., PRICE, E. and RAZENSHTEYN, I. (2019). Adversarial examples from computational constraints. In *International Conference on Machine Learning* 831–840. PMLR.

[18] CELENTANO, M., MONTANARI, A. and WU, Y. (2020). The estimation error of general first order methods. In *Conference on Learning Theory* 1078–1141. PMLR.

[19] CHEN, X., GUNTUBOYINA, A. and ZHANG, Y. (2016). On Bayes risk lower bounds. *J. Mach. Learn. Res.* **17** Paper No. 219, 58 pp. MR3595153

[20] DAGAN, Y. and SHAMIR, O. (2018). Detecting correlations with little memory and communication. In *Conference on Learning Theory* 1145–1198. PMLR.

[21] DAVIS, D., DIAZ, M. and WANG, K. (2021). Clustering a mixture of Gaussians with unknown covariance. Preprint. Available at arXiv:2110.01602.

[22] DIAKONIKOLAS, I., KANE, D. and ZARIFIS, N. (2020). Near-optimal sq lower bounds for agnostically learning halfspaces and relus under Gaussian marginals. *Adv. Neural Inf. Process. Syst.* **33** 13586–13596.

[23] DIAKONIKOLAS, I. and KANE, D. M. (2020). Hardness of learning halfspaces with Massart noise. Preprint. Available at arXiv:2012.09720.

[24] DIAKONIKOLAS, I. and KANE, D. M. (2021). Non-Gaussian component analysis via lattice basis reduction. Preprint. Available at arXiv:2112.09104.

[25] DIAKONIKOLAS, I., KANE, D. M., KONTONIS, V., TZAMOS, C. and ZARIFIS, N. (2022). Learning general halfspaces with general Massart noise under the Gaussian distribution. In *STOC '*22*—Proceedings of the* 54*th Annual ACM SIGACT Symposium on Theory of Computing* 874–885. ACM, New York. MR4490047

[26] DIAKONIKOLAS, I., KANE, D. M., KONTONIS, V. and ZARIFIS, N. (2020). Algorithms and sq lower bounds for pac learning one-hidden-layer relu networks. In *Conference on Learning Theory* 1514–1539. PMLR.

[27] DIAKONIKOLAS, I., KANE, D. M., PITTAS, T. and ZARIFIS, N. (2021). The optimality of polynomial regression for agnostic learning under Gaussian marginals in the SQ model. In *Conference on Learning Theory* 1552–1584. PMLR.

[28] 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 58*th 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

[29] 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

[30] DUDEJA, R. and HSU, D. (2021). Statistical query lower bounds for tensor PCA. *J. Mach. Learn. Res.* **22** Paper No. 83, 51 pp. MR4253776

[31] DUDEJA, R. and HSU, D. (2022). Statistical-computational trade-offs in tensor PCA and related problems via communication complexity. Preprint. Available at arXiv:2204.07526.

[32] DUDEJA, R. and HSU, D. (2024). Supplement to "Statistical-computational trade-offs in tensor PCA and related problems via communication complexity." https://doi.org/10.1214/23-AOS2331SUPP
[33] 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** Art. 8, 37 pp. MR3664576 https://doi.org/10.1145/3046674
[34] 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
[35] GARG, S., KOTHARI, P. K., LIU, P. and RAZ, R. (2021). Memory-sample lower bounds for learning parity with noise. In 24*th International Conference on Approximation Algorithms for Combinatorial Optimization Problems*, *APPROX* 2021 *and* 25*th International Conference on Randomization and Computation*, *RANDOM* 2021 60. Schloss Dagstuhl-Leibniz-Zentrum fur Informatik GmbH, Dagstuhl Publishing.
[36] GARG, S., RAZ, R. and TAL, A. (2018). Extractor-based time-space lower bounds for learning. In *STOC*'18—*Proceedings of the* 50*th Annual ACM SIGACT Symposium on Theory of Computing* 990–1002. ACM, New York. MR3826311
[37] GARG, S., RAZ, R. and TAL, A. (2019). Time-space lower bounds for two-pass learning. In 34*th Computational Complexity Conference*. *LIPIcs*. *Leibniz Int*. *Proc*. *Inform*. **137** Art. No. 22, 39 pp. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR3984627
[38] GHOSH, M., JERONIMO, F. G., JONES, C., POTECHIN, A. and RAJENDRAN, G. (2020). Sum-of-squares lower bounds for Sherrington–Kirkpatrick via planted affine planes. In 2020 *IEEE* 61*st Annual Symposium on Foundations of Computer Science* 954–965. IEEE Computer Soc., Los Alamitos, CA. MR4232101
[39] GOYAL, N. and SHETTY, A. (2019). Non-Gaussian component analysis using entropy methods. In *STOC*'19—*Proceedings of the* 51*st Annual ACM SIGACT Symposium on Theory of Computing* 840–851. ACM, New York. MR4003388
[40] HAN, Y., ÖZGÜR, A. and WEISSMAN, T. (2021). Geometric lower bounds for distributed parameter estimation under communication constraints. *IEEE Trans*. *Inf*. *Theory* **67** 8248–8263. MR4346086 https://doi.org/10.1109/TIT.2021.3108952
[41] 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
[42] HOPKINS, S. B., SHI, J., SCHRAMM, T. and STEURER, D. (2016). Fast spectral algorithms from sum-of-squares proofs: Tensor decomposition and planted sparse vectors. In *STOC*'16—*Proceedings of the* 48*th Annual ACM SIGACT Symposium on Theory of Computing* 178–191. ACM, New York. MR3536564 https://doi.org/10.1145/2897518.2897529
[43] 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.
[44] HOPKINS, S. B. K. (2018). Statistical inference and the sum of squares method. PhD thesis, Cornell University. MR3864930
[45] KEARNS, M. (1998). Efficient noise-tolerant learning from statistical queries. *J. ACM* **45** 983–1006. MR1678849 https://doi.org/10.1145/293347.293351
[46] KOL, G., RAZ, R. and TAL, A. (2017). Time-space hardness of learning sparse parities. In *STOC*'17—*Proceedings of the* 49*th Annual ACM SIGACT Symposium on Theory of Computing* 1067–1080. ACM, New York. MR3678252 https://doi.org/10.1145/3055399.3055430
[47] KUNISKY, D., WEIN, A. S. and BANDEIRA, A. S. (2022). Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *Mathematical Analysis*, *Its Applications and Computation*. *Springer Proc*. *Math*. *Stat*. **385** 1–50. Springer, Cham. MR4461037 https://doi.org/10.1007/978-3-030-97127-4_1
[48] MAO, C. and WEIN, A. S. (2021). Optimal spectral recovery of a planted vector in a subspace. Preprint. Available at arXiv:2105.15081.
[49] MICCIANCIO, D. and REGEV, O. (2009). Lattice-based cryptography. In *Post-Quantum Cryptography* 147–191. Springer, Berlin. MR2590647 https://doi.org/10.1007/978-3-540-88702-7_5
[50] MONTANARI, A. and RICHARD, E. (2014). A statistical model for tensor PCA. Preprint. Available at arXiv:1411.1076.
[51] MOSHKOVITZ, D. and MOSHKOVITZ, M. (2017). Mixing implies lower bounds for space bounded learning. In *Conference on Learning Theory* 1516–1566. PMLR.

[52] MOSHKOVITZ, D. and MOSHKOVITZ, M. (2018). Entropy samplers and strong generic lower bounds for space bounded learning. In 9*th Innovations in Theoretical Computer Science*. *LIPIcs*. *Leibniz Int*. *Proc*. *Inform*. **94** Art. No. 28, 20 pp. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR3761764

[53] MOSHKOVITZ, M. and TISHBY, N. (2017). Mixing complexity and its applications to neural networks. Preprint. Available at arXiv:1703.00729.

[54] RAGHAVENDRA, P., SCHRAMM, T. and STEURER, D. (2018). High dimensional estimation via sum-of-squares proofs. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro* 2018. *Vol*. *IV*. *Invited Lectures* 3389–3423. World Sci. Publ., Hackensack, NJ. MR3966537

[55] RAZ, R. (2019). Fast learning requires good memory: A time-space lower bound for parity learning. *J. ACM* **66** Art. 3, 18 pp. MR3892562 https://doi.org/10.1145/3186563

[56] ROS, V., AROUS, G. B., BIROLI, G. and CAMMAROTA, C. (2019). Complex energy landscapes in spiked-tensor and simple glassy models: Ruggedness, arrangements of local minima, and phase transitions. *Phys*. *Rev*. *X* **9** 011003.

[57] SCHRAMM, T. and WEIN, A. S. (2022). Computational barriers to estimation from low-degree polynomials. *Ann*. *Statist*. **50** 1833–1858. MR4441142 https://doi.org/10.1214/22-aos2179

[58] SHAMIR, O. (2014). Fundamental limits of online and distributed algorithms for statistical learning and estimation. *Adv. Neural Inf. Process*. *Syst*. **27**.

[59] SHARAN, V., SIDFORD, A. and VALIANT, G. (2019). Memory-sample tradeoffs for linear regression with small error. In *STOC*'19*—Proceedings of the* 51*st Annual ACM SIGACT Symposium on Theory of Computing* 890–901. ACM, New York. MR4003393 https://doi.org/10.1145/3313276.3316403

[60] STEINHARDT, J., VALIANT, G. and WAGER, S. (2016). Memory, communication, and statistical queries. In *Conference on Learning Theory* 1490–1516. PMLR.

[61] TAN, Y. S. and VERSHYNIN, R. (2018). Polynomial time and sample complexity for non-Gaussian component analysis: Spectral methods. In *Conference on Learning Theory* 498–534. PMLR.

[62] VEMPALA, S. S. and XIAO, Y. (2011). Structure from local optima: Learning subspace juntas via higher order PCA. Preprint. Available at arXiv:1108.3329.

[63] WANG, K., YAN, Y. and DIAZ, M. (2020). Efficient clustering for stretched mixtures: Landscape and optimality. *Adv. Neural Inf. Process*. *Syst*. **33** 21309–21320.

[64] WANG, Z., GU, Q. and LIU, H. (2015). Sharp computational-statistical phase transitions via oracle computational model. Preprint. Available at arXiv:1512.08861.

[65] 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

[66] ZADIK, I., SONG, M. J., WEIN, A. S. and BRUNA, J. (2021). Lattice-based methods surpass sum-of-squares in clustering. Preprint. Available at arXiv:2112.03898.

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

[68] 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

[69] ZHENG, Q. and TOMIOKA, R. (2015). Interpolating convex and non-convex tensor decompositions via the subspace norm. In *Advances in Neural Information Processing Systems* 3106–3113.

The Annals of Statistics
2024, Vol. 52, No. 1, 157–183
https://doi.org/10.1214/23-AOS2338

ADAPTIVE NOVELTY DETECTION WITH FALSE DISCOVERY RATE GUARANTEE

BY ARIANE MARANDON[1,a], LIHUA LEI[2,c], DAVID MARY[3,d] AND ETIENNE ROQUAIN[1,b]

[1]*LPSM, Sorbonne Université, Université de Paris & CNRS,* [a]*ariane.marandon-carlhian@sorbonne-universite.fr,* [b]*etienne.roquain@sorbonne-universite.fr*

[2]*Graduate School of Business, Stanford University,* [c]*lihualei@stanford.edu*

[3]*Laboratoire Lagrange, Observatoire de la Côte d'Azur, Université Côte d'Azur & CNRS,* [d]*david.mary@unice.fr*

This paper studies the semisupervised novelty detection problem where a set of "typical" measurements is available to the researcher. Motivated by recent advances in multiple testing and conformal inference, we propose AdaDetect, a flexible method that is able to wrap around any probabilistic classification algorithm and control the false discovery rate (FDR) on detected novelties in finite samples without any distributional assumption other than exchangeability. In contrast to classical FDR-controlling procedures that are often committed to a pre-specified p-value function, AdaDetect learns the transformation in a data-adaptive manner to focus the power on the directions that distinguish between inliers and outliers. Inspired by the multiple testing literature, we further propose variants of AdaDetect that are adaptive to the proportion of nulls while maintaining the finite-sample FDR control. The methods are illustrated on synthetic datasets and real-world datasets, including an application in astrophysics.

REFERENCES

ANGELOPOULOS, A. N. and BATES, S. (2021). A gentle introduction to conformal prediction and distribution-free uncertainty quantification. ArXiv preprint. Available at arXiv:2107.07511.

AUDIBERT, J.-Y. and TSYBAKOV, A. B. (2007). Fast learning rates for plug-in classifiers. *Ann. Statist.* **35** 608–633. MR2336861 https://doi.org/10.1214/009053606000001217

BALASUBRAMANIAN, V., HO, S.-S. and VOVK, V. (2014). *Conformal Prediction for Reliable Machine Learning: Theory, Adaptations and Applications*. Newnes.

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

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

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

BARBER, R. F. and RAMDAS, A. (2017). The p-filter: Multilayer false discovery rate control for grouped hypotheses. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1247–1268. MR3689317 https://doi.org/10.1111/rssb.12218

BARTLETT, P. L., HARVEY, N., LIAW, C. and MEHRABIAN, A. (2019). Nearly-tight VC-dimension and pseudodimension bounds for piecewise linear neural networks. *J. Mach. Learn. Res.* **20** Paper No. 63. MR3960917

BASHARI, M., EPSTEIN, A., ROMANO, Y. and SESIA, M. (2023). Derandomized novelty detection with fdr control via conformal e-values.

BATES, S., CANDÈS, E., LEI, L., ROMANO, Y. and SESIA, M. (2023). Testing for outliers with conformal p-values. *Ann. Statist.* **51** 149–178. MR4564852 https://doi.org/10.1214/22-aos2244

BEKKER, J. and DAVIS, J. (2020). Learning from positive and unlabeled data: A survey. *Mach. Learn.* **109** 719–760. MR4094532 https://doi.org/10.1007/s10994-020-05877-5

BENJAMINI, Y. (2010). Discovering the false discovery rate. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 405–416. MR2758522 https://doi.org/10.1111/j.1467-9868.2010.00746.x

BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. MR1325392

MSC2020 subject classifications. Primary 62J15, 62G10; secondary 62H30.

Key words and phrases. Adaptive multiple testing, novelty detection, false discovery rate, conformal p-values, machine learning, classification, neural network, knockoff.

BENJAMINI, Y., KRIEGER, A. M. and YEKUTIELI, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93** 491–507. MR2261438 https://doi.org/10.1093/biomet/93.3.491

BENJAMINI, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188. MR1869245 https://doi.org/10.1214/aos/1013699998

BLANCHARD, G., LEE, G. and SCOTT, C. (2010). Semi-supervised novelty detection. *J. Mach. Learn. Res.* **11** 2973–3009. MR2746544

BLANCHARD, G. and ROQUAIN, É. (2009). Adaptive false discovery rate control under independence and dependence. *J. Mach. Learn. Res.* **10** 2837–2871. MR2579914

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

BOS, T. and SCHMIDT-HIEBER, J. (2022). Convergence rates of deep ReLU networks for multiclass classification. *Electron. J. Stat.* **16** 2724–2773. MR4406243 https://doi.org/10.1214/22-ejs2011

CAI, T. T. and SUN, W. (2009). Simultaneous testing of grouped hypotheses: Finding needles in multiple haystacks. *J. Amer. Statist. Assoc.* **104** 1467–1481. MR2597000 https://doi.org/10.1198/jasa.2009.tm08415

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

CALVO, B., LARRANAGA, P. and LOZANO, J. A. (2007). Learning Bayesian classifiers from positive and unlabeled examples. *Pattern Recogn. Lett.* **28** 2375–2384.

CANNON, A., HOWSE, J., HUSH, D. and SCOVEL, C. (2002). Learning with the neyman-pearson and min-max criteria. Los Alamos National Laboratory, Tech. Rep. LA-UR, pages 02–2951.

DU PLESSIS, M. C., NIU, G. and SUGIYAMA, M. (2014). Analysis of learning from positive and unlabeled data. *Adv. Neural Inf. Process. Syst.* **27** 703–711.

EFRON, B. (2004). Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. *J. Amer. Statist. Assoc.* **99** 96–104. MR2054289 https://doi.org/10.1198/016214504000000089

EFRON, B. (2007). Doing thousands of hypothesis tests at the same time. *Metron* **LXV** 3–21.

EFRON, B. (2008). Microarrays, empirical Bayes and the two-groups model. *Statist. Sci.* **23** 1–22. MR2431866 https://doi.org/10.1214/07-STS236

EFRON, B. (2009). Empirical Bayes estimates for large-scale prediction problems. *J. Amer. Statist. Assoc.* **104** 1015–1028. MR2562003 https://doi.org/10.1198/jasa.2009.tm08523

EFRON, B., TIBSHIRANI, R., STOREY, J. D. and TUSHER, V. (2001). Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.* **96** 1151–1160. MR1946571 https://doi.org/10.1198/016214501753382129

FERREIRA, J. A. and ZWINDERMAN, A. H. (2006). On the Benjamini–Hochberg method. *Ann. Statist.* **34** 1827–1849. MR2283719 https://doi.org/10.1214/009053606000000425

FISHER, A. (2021). Saffron and lord ensure online control of the false discovery rate under positive dependence. ArXiv preprint. Available at arXiv:2110.08161.

FITHIAN, W. and LEI, L. (2022). Conditional calibration for false discovery rate control under dependence. *Ann. Statist.* **50** 3091–3118. MR4524490 https://doi.org/10.1214/21-aos2137

FRIEDMAN, J. H. (2003). On multivariate goodness–of–fit and two–sample testing. *Stat. Probl. Part. Phys. Astrophys. Cosmol.* **1** 311.

GIRAUD, C. (2022). *Introduction to High-Dimensional Statistics. Monographs on Statistics and Applied Probability* **168**. CRC Press, Boca Raton, FL. MR4436193 https://doi.org/10.1201/9781003158745

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

GOODFELLOW, I., BENGIO, Y. and COURVILLE, A. (2016). *Deep Learning. Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3617773

GUO, T., XU, C., HUANG, J., WANG, Y., SHI, B., XU, C. and TAO, D. (2020). On positive-unlabeled classification in gan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*).

IVANOV, D. (2020). Dedpul: Difference-of-estimated-densities-based positive-unlabeled learning. In 2020 19*th IEEE International Conference on Machine Learning and Applications* (*ICMLA*) 782–790. IEEE, New York.

IVEZIĆ, Ž., CONNOLLY, A. J., VANDERPLAS, J. T. and GRAY, A. (2019). *Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data*. Princeton Univ. Press, Princeton.

IVEZIĆ, Ž., VIVAS, A. K., LUPTON, R. H. and ZINN, R. (2005). The selection of RR lyrae stars using single-epoch data. *Astron. J.* **129** 1096.

JAVANMARD, A. and JAVADI, H. (2019). False discovery rate control via debiased lasso. *Electron. J. Stat.* **13** 1212–1253. MR3935848 https://doi.org/10.1214/19-ejs1554

KORN, E. L., TROENDLE, J. F., MCSHANE, L. M. and SIMON, R. (2004). Controlling the number of false discoveries: Application to high-dimensional genomic data. *J. Statist. Plann. Inference* **124** 379–398. MR2080371 https://doi.org/10.1016/S0378-3758(03)00211-8

LECUN, Y. and CORTES, C. (2010). MNIST handwritten digit database.

LEI, L., D'AMOUR, A., DING, P., FELLER, A. and SEKHON, J. (2021). Distribution-free assessment of population overlap in observational studies Technical report.

LEI, L. and FITHIAN, W. (2018). AdaPT: An interactive procedure for multiple testing with side information. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 649–679. MR3849338 https://doi.org/10.1111/rssb.12253

LIANG, Z., SESIA, M. and SUN, W. (2022). Integrative conformal p-values for powerful out-of-distribution testing with labeled outliers.

LOPER, J. H., LEI, L., FITHIAN, W. and TANSEY, W. (2022). Smoothed nested testing on directed acyclic graphs. *Biometrika* **109** 457–471. MR4430968 https://doi.org/10.1093/biomet/asab041

MA, R., CAI, T. T. and LI, H. (2021). Global and simultaneous hypothesis testing for high-dimensional logistic regression models. *J. Amer. Statist. Assoc.* **116** 984–998. MR4270038 https://doi.org/10.1080/01621459.2019.1699421

MARANDON, A., LEI, L., MARY, D. and ROQUAIN, E. (2024). Supplement to "Adaptive Novelty Detection with false discovery rate guarantee." https://doi.org/10.1214/23-AOS2338SUPP

MARY, D. and ROQUAIN, E. (2022). Semi-supervised multiple testing. *Electron. J. Stat.* **16** 4926–4981. MR4490412 https://doi.org/10.1214/22-ejs2050

RAMDAS, A., CHEN, J., WAINWRIGHT, M. J. and JORDAN, M. I. (2019a). A sequential algorithm for false discovery rate control on directed acyclic graphs. *Biometrika* **106** 69–86. MR3912384 https://doi.org/10.1093/biomet/asy066

RAMDAS, A. K., BARBER, R. F., WAINWRIGHT, M. J. and JORDAN, M. I. (2019b). A unified treatment of multiple testing with prior knowledge using the p-filter. *Ann. Statist.* **47** 2790–2821. MR3988773 https://doi.org/10.1214/18-AOS1765

RAVA, B., SUN, W., JAMES, G. M. and TONG, X. (2021). A burden shared is a burden halved: A fairness-adjusted approach to classification. ArXiv preprint. Available at arXiv:2110.05720.

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

ROQUAIN, E. and VILLERS, F. (2011). Exact calculations for false discovery proportion with application to least favorable configurations. *Ann. Statist.* **39** 584–612. MR2797857 https://doi.org/10.1214/10-AOS847

ROSSET, S., HELLER, R., PAINSKY, A. and AHARONI, E. (2022). Optimal and maximin procedures for multiple testing problems. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 1105–1128. MR4494154 https://doi.org/10.1111/rssb.12507

RUFF, L., VANDERMEULEN, R. A., GÖRNITZ, N., DEECKE, L., SIDDIQUI, S. A., BINDER, A., MÜLLER, E. and KLOFT, M. (2018). Deep one-class classification. In *Proceedings of the* 35*th International Conference on Machine Learning* **80** 4393–4402.

SARKAR, S. K. (2008). On methods controlling the false discovery rate. *Sankhyā* **70** 135–168. MR2551809

SCHÖLKOPF, B., PLATT, J. C., SHAWE-TAYLOR, J., SMOLA, A. J. and WILLIAMSON, R. C. (2001). Estimating the support of a high-dimensional distribution. *Neural Comput.* **13** 1443–1471.

SCOTT, C. and NOWAK, R. (2005). A Neyman–Pearson approach to statistical learning. *IEEE Trans. Inf. Theory* **51** 3806–3819. MR2239000 https://doi.org/10.1109/TIT.2005.856955

STOREY, J. D., TAYLOR, J. E. and SIEGMUND, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: A unified approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 187–205. MR2035766 https://doi.org/10.1111/j.1467-9868.2004.00439.x

SUGIYAMA, M., SUZUKI, T. and KANAMORI, T. (2012). *Density Ratio Estimation in Machine Learning*. Cambridge Univ. Press, Cambridge. MR2895762 https://doi.org/10.1017/CBO9781139035613

SUN, W. and CAI, T. T. (2007). Oracle and adaptive compound decision rules for false discovery rate control. *J. Amer. Statist. Assoc.* **102** 901–912. MR2411657 https://doi.org/10.1198/016214507000000545

SUN, W. and CAI, T. T. (2009). Large-scale multiple testing under dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 393–424. MR2649603 https://doi.org/10.1111/j.1467-9868.2008.00694.x

VAPNIK, V. N. (1998). *Statistical Learning Theory*. *Adaptive and Learning Systems for Signal Processing*, *Communications*, *and Control*. Wiley, New York. MR1641250

VOVK, V. (2015). Cross-conformal predictors. *Ann. Math. Artif. Intell.* **74** 9–28. MR3353894 https://doi.org/10.1007/s10472-013-9368-4

VOVK, V., GAMMERMAN, A. and SHAFER, G. (2005). *Algorithmic Learning in a Random World*. Springer, New York. MR2161220

WANG, Y., KAJI, T. and ROCKOVA, V. (2022). Approximate Bayesian computation via classification. *J. Mach. Learn. Res.* **23** Paper No. 350. MR4577789

WEINSTEIN, A. (2021). On permutation invariant problems in large-scale inference. ArXiv preprint. Available at arXiv:2110.06250.

WEINSTEIN, A., BARBER, R. and CANDÈS, E. (2017). A power and prediction analysis for knockoffs with lasso statistics. Available at arXiv:1712.06465.

YANG, C.-Y., LEI, L., HO, N. and FITHIAN, W. (2021). Bonus: Multiple multivariate testing with a data-adaptive test statistic. Available at arXiv:2106.15743.

ZRNIC, T., RAMDAS, A. and JORDAN, M. I. (2021). Asynchronous online testing of multiple hypotheses. *J. Mach. Learn. Res.* **22** Paper No. 33. MR4253726 https://doi.org/10.1515/ijnsns-2019-0210

The Annals of Statistics
2024, Vol. 52, No. 1, 184–206
https://doi.org/10.1214/23-AOS2339

RANK-BASED INDICES FOR TESTING INDEPENDENCE BETWEEN TWO HIGH-DIMENSIONAL VECTORS

BY YEQING ZHOU[1,a], KAI XU[2,b], LIPING ZHU[3,c] AND RUNZE LI[4,d]

[1]*School of Mathematical Sciences and Key Laboratory of Intelligent Computing and Applications, Tongji University,* [a]*zhouyeqing@tongji.edu.cn*

[2]*School of Mathematics and Statistics, Anhui Normal University,* [b]*tjxxukai@ahnu.edu.cn*

[3]*Center for Applied Statistics, Institute of Statistics and Big Data, Renmin University of China Zhijiang Institute of Big Data and Statistics, School Statistics and Mathematics, Zhejiang Gongshang University,* [c]*zhu.liping@ruc.edu.cn*

[4]*Department of Statistics, The Pennsylvania State University at University Park,* [d]*rzli@psu.edu*

To test independence between two high-dimensional random vectors, we propose three tests based on the rank-based indices derived from Hoeffding's D, Blum–Kiefer–Rosenblatt's R and Bergsma–Dassios–Yanagimoto's τ^*. Under the null hypothesis of independence, we show that the distributions of the proposed test statistics converge to normal ones if the dimensions diverge arbitrarily with the sample size. We further derive an explicit rate of convergence. Thanks to the monotone transformation-invariant property, these distribution-free tests can be readily used to generally distributed random vectors including heavily-tailed ones. We further study the local power of the proposed tests and compare their relative efficiencies with two classic distance covariance/correlation based tests in high-dimensional settings. We establish explicit relationships between D, R, τ^* and Pearson's correlation for bivariate normal random variables. The relationships serve as a basis for power comparison. Our theoretical results show that under a Gaussian equicorrelation alternative: (i) the proposed tests are superior to the two classic distance covariance/correlation based tests if the components of random vectors have very different scales; (ii) the asymptotic efficiency of the proposed tests based on D, τ^* and R are sorted in a descending order.

REFERENCES

ALBERT, M., LAURENT, B., MARREL, A. and MEYNAOUI, A. (2022). Adaptive test of independence based on HSIC measures. *Ann. Statist.* **50** 858–879. MR4404921 https://doi.org/10.1214/21-aos2129

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley-Interscience, New York. MR1990662

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

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

BERGSMA, W. and DASSIOS, A. (2014). A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli* **20** 1006–1028. MR3178526 https://doi.org/10.3150/13-BEJ514

BERRETT, T. B. and SAMWORTH, R. J. (2019). Nonparametric independence testing via mutual information. *Biometrika* **106** 547–566. MR3992389 https://doi.org/10.1093/biomet/asz024

BLUM, J. R., KIEFER, J. and ROSENBLATT, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Stat.* **32** 485–498. MR0125690 https://doi.org/10.1214/aoms/1177705055

BODNAR, T., DETTE, H. and PAROLYA, N. (2019). Testing for independence of large dimensional vectors. *Ann. Statist.* **47** 2977–3008. MR3988779 https://doi.org/10.1214/18-AOS1771

CHAKRABORTY, S. and ZHANG, X. (2021). A new framework for distance and kernel-based metrics in high dimensions. *Electron. J. Stat.* **15** 5455–5522. MR4352549 https://doi.org/10.1214/21-ejs1889

MSC2020 subject classifications. Primary 62G10; secondary 62G20.

Key words and phrases. Bergsma–Dassios–Yanagimoto's τ^*, Blum–Kiefer–Rosenblatt's R, degenerate U-statistics, Hoeffding's D.

CHATTERJEE, S. (2021). A new coefficient of correlation. *J. Amer. Statist. Assoc.* **116** 2009–2022. MR4353729 https://doi.org/10.1080/01621459.2020.1758115

CHIANG, A. P., BECK, J. S., YEN, H., TAYEH, M. K., SCHEETZ, T. E. SWIDERSKI, R. E. NISHIMURA, D. Y. BRAUN, T. A., KIM, K.-Y. A. et al. (2006). Homozygosity mapping with SNP arrays identifies TRIM32, an E3 ubiquitin ligase, as a bardet–biedl syndrome gene (BBS11). *Proc. Natl. Acad. Sci. USA* **103** 6287–6292.

DEB, N. and SEN, B. (2023). Multivariate rank-based distribution-free nonparametric testing using measure transportation. *J. Amer. Statist. Assoc.* **118** 192–207. MR4571116 https://doi.org/10.1080/01621459.2021.1923508

DRTON, M., HAN, F. and SHI, H. (2020). High-dimensional consistent independence testing with maxima of rank correlations. *Ann. Statist.* **48** 3206–3227. MR4185806 https://doi.org/10.1214/19-AOS1926

GAO, L., FAN, Y., LV, J. and SHAO, Q.-M. (2021). Asymptotic distributions of high-dimensional distance correlation inference. *Ann. Statist.* **49** 1999–2020. MR4319239 https://doi.org/10.1214/20-aos2024

GORSKY, S. and MA, L. (2022). Multi-scale Fisher's independence test for multivariate dependence. *Biometrika* **109** 569–587. MR4472834 https://doi.org/10.1093/biomet/asac013

GRETTON, A., FUKUMIZU, K., TEO, C., SONG, L., SCHÖLKOPF, B. and SMOLA, A. (2008). A kernel statistical test of independence. In *Advances in Neural Information Processing Systems* 585–592.

HALL, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.* **14** 1–16. MR0734096 https://doi.org/10.1016/0047-259X(84)90044-7

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. *Probability and Mathematical Statistics*. Academic Press, New York. MR0624435

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

HELLER, R., HELLER, Y. and GORFINE, M. (2013). A consistent multivariate test of association based on ranks of distances. *Biometrika* **100** 503–510. MR3068450 https://doi.org/10.1093/biomet/ass070

HOEFFDING, W. (1948). A non-parametric test of independence. *Ann. Math. Stat.* **19** 546–557. MR0029139 https://doi.org/10.1214/aoms/1177730150

JIANG, T. and YANG, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Statist.* **41** 2029–2074. MR3127857 https://doi.org/10.1214/13-AOS1134

KENDALL, M. G. (1938). A new measure of rank correlation. *Biometrika* **30** 81–93.

LEE, D., ZHANG, K. and KOSOROK, M. R. (2023). The binary expansion randomized ensemble test. *Statist. Sinica* **33** 2381–2403. MR4647039

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

LI, R., XU, K., ZHOU, Y. and ZHU, L. (2023). Testing the effects of high-dimensional covariates via aggregating cumulative covariances. *J. Amer. Statist. Assoc.* **118** 2184–2194. MR4646635 https://doi.org/10.1080/01621459.2022.2044334

MOON, H. and CHEN, K. (2022). Interpoint-ranking sign covariance for the test of independence. *Biometrika* **109** 165–179. MR4374647 https://doi.org/10.1093/biomet/asab011

MUDHOLKAR, G. S. and WILDING, G. E. (2003). On the conventional wisdom regarding two consistent tests of bivariate independence. *Statistician* **52** 41–57. MR1973881 https://doi.org/10.1111/1467-9884.00340

PEARSON, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philos. Mag. Ser.* 5 **50** 157–175.

SCHEETZ, T. E., KIM, K.-Y. A., SWIDERSKI, R. E., PHILP, A. R., BRAUN, T. A., KNUDTSON, K. L., DORRANCE, A. M., DIBONA, G. F., HUANG, J. et al. (2006). Regulation of gene expression in the mammalian eye and its relevance to eye disease. *Proc. Natl. Acad. Sci. USA* **103** 14429–14434.

SHEN, C., PRIEBE, C. E. and VOGELSTEIN, J. T. (2020). From distance correlation to multiscale graph correlation. *J. Amer. Statist. Assoc.* **115** 280–291. MR4078463 https://doi.org/10.1080/01621459.2018.1543125

SHI, H., DRTON, M. and HAN, F. (2022a). On the power of Chatterjee's rank correlation. *Biometrika* **109** 317–333. MR4430960 https://doi.org/10.1093/biomet/asab028

SHI, H., DRTON, M. and HAN, F. (2022b). Distribution-free consistent independence tests via center-outward ranks and signs. *J. Amer. Statist. Assoc.* **117** 395–410. MR4399094 https://doi.org/10.1080/01621459.2020.1782223

SHI, H., HALLIN, M., DRTON, M. and HAN, F. (2022). On universally consistent and fully distribution-free rank tests of vector independence. *Ann. Statist.* **50** 1933–1959. MR4474478 https://doi.org/10.1214/21-aos2151

SPEARMAN, C. (1904). The proof and measurement of association between two things. *Amer. J. Psychol.* **15** 72–101.

SZÉKELY, G. J. and RIZZO, M. L. (2013). The distance correlation t-test of independence in high dimension. *J. Multivariate Anal.* **117** 193–213. MR3053543 https://doi.org/10.1016/j.jmva.2013.02.012

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

WEIHS, L., DRTON, M. and MEINSHAUSEN, N. (2018). Symmetric rank covariances: A generalized framework for nonparametric measures of dependence. *Biometrika* **105** 547–562. MR3842884 https://doi.org/10.1093/biomet/asy021

XU, K. and ZHU, L. (2022). Power analysis of projection-pursuit independence tests. *Statist. Sinica* **32** 417–433. MR4359639 https://doi.org/10.5705/ss.202019.0457

YAMADA, Y., HYODO, M. and NISHIYAMA, T. (2017). Testing block-diagonal covariance structure for high-dimensional data under non-normality. *J. Multivariate Anal.* **155** 305–316. MR3607897 https://doi.org/10.1016/j.jmva.2016.12.009

YANAGIMOTO, T. (1970). On measures of association and a related problem. *Ann. Inst. Statist. Math.* **22** 57–63.

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

YAO, S., ZHANG, X. and SHAO, X. (2018). Testing mutual independence in high dimension via distance covariance. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 455–480. MR3798874 https://doi.org/10.1111/rssb.12259

ZHANG, K. (2019). BET on independence. *J. Amer. Statist. Assoc.* **114** 1620–1637. MR4047288 https://doi.org/10.1080/01621459.2018.1537921

ZHOU, Y., XU, K., ZHU, L. and LI, R. (2024). Supplement to "Rank-based indices for testing independence between two high-dimensional vectors." https://doi.org/10.1214/23-AOS2339SUPP

ZHOU, Y. and ZHU, L. (2018). Model-free feature screening for ultrahigh dimensional datathrough a modified Blum–Kiefer–Rosenblatt correlation. *Statist. Sinica* **28** 1351–1370. MR3821008

ZHU, C., ZHANG, X., YAO, S. and SHAO, X. (2020). Distance-based and RKHS-based dependence metrics in high dimension. *Ann. Statist.* **48** 3366–3394. MR4185812 https://doi.org/10.1214/19-AOS1934

ZHU, L., XU, K., LI, R. and ZHONG, W. (2017). Projection correlation between two random vectors. *Biometrika* **104** 829–843. MR3737307 https://doi.org/10.1093/biomet/asx043

ZHU, L., ZHANG, Y. and XU, K. (2018). Measuring and testing for interval quantile dependence. *Ann. Statist.* **46** 2683–2710. MR3851752 https://doi.org/10.1214/17-AOS1635

The Annals of Statistics
2024, Vol. 52, No. 1, 207–232
https://doi.org/10.1214/23-AOS2341

TRANSFER LEARNING FOR CONTEXTUAL MULTI-ARMED BANDITS

BY CHANGXIAO CAI[1,a], T. TONY CAI[2,b] AND HONGZHE LI[3,c]

[1]*Department of Industrial and Operations Engineering, University of Michigan,* [a]*cxcai@umich.edu*

[2]*Department of Statistics and Data Science, University of Pennsylvania,* [b]*tcai@wharton.upenn.edu*

[3]*Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania,* [c]*hongzhe@pennmedicine.upenn.edu*

Motivated by a range of applications, we study in this paper the problem of transfer learning for nonparametric contextual multi-armed bandits under the covariate shift model, where we have data collected from source bandits before the start of the target bandit learning. The minimax rate of convergence for the cumulative regret is established and a novel transfer learning algorithm that attains the minimax regret is proposed. The results quantify the contribution of the data from the source domains for learning in the target domain in the context of nonparametric contextual multi-armed bandits.

In view of the general impossibility of adaptation to unknown smoothness, we develop a data-driven algorithm that achieves near-optimal statistical guarantees (up to a logarithmic factor) while automatically adapting to the unknown parameters over a large collection of parameter spaces under an additional self-similarity assumption. A simulation study is carried out to illustrate the benefits of utilizing the data from the source domains for learning in the target domain.

REFERENCES

ABE, N. and LONG, P. M. (1999). Associative reinforcement learning using linear probabilistic concepts. In *ICML* 3–11. Citeseer.

AGRAWAL, S., AVADHANULA, V., GOYAL, V. and ZEEVI, A. (2019). MNL-Bandit: A dynamic learning approach to assortment selection. *Oper. Res.* **67** 1453–1485. MR4014580 https://doi.org/10.1287/opre.2018.1832

AUDIBERT, J.-Y. and TSYBAKOV, A. B. (2007). Fast learning rates for plug-in classifiers. *Ann. Statist.* **35** 608–633. MR2336861 https://doi.org/10.1214/009053606000001217

AUER, P. (2002). Using confidence bounds for exploitation-exploration trade-offs. *J. Mach. Learn. Res.* **3** 397–422. MR1984023 https://doi.org/10.1162/153244303321897663

AUER, P., CESA-BIANCHI, N., FREUND, Y. and SCHAPIRE, R. E. (1995). Gambling in a rigged casino: The adversarial multi-armed bandit problem. In 36*th Annual Symposium on Foundations of Computer Science* (*Milwaukee*, *WI*, 1995) 322–331. IEEE Comput. Soc. Press, Los Alamitos, CA. MR1619094 https://doi.org/10.1109/SFCS.1995.492488

AUER, P. and ORTNER, R. (2010). UCB revisited: Improved regret bounds for the stochastic multi-armed bandit problem. *Period. Math. Hungar.* **61** 55–65. MR2728432 https://doi.org/10.1007/s10998-010-3055-6

BASTANI, H. and BAYATI, M. (2020). Online decision making with high-dimensional covariates. *Oper. Res.* **68** 276–294. MR4059503 https://doi.org/10.1287/opre.2019.1902

BASTANI, H., BAYATI, M. and KHOSRAVI, K. (2021). Mostly exploration-free algorithms for contextual bandits. *Manage. Sci.* **67** 1329–1349.

BEN-DAVID, S., BLITZER, J., CRAMMER, K. and PEREIRA, F. (2006). Analysis of representations for domain adaptation. *Adv. Neural Inf. Process. Syst.* **19**.

BLITZER, J., CRAMMER, K., KULESZA, A., PEREIRA, F. and WORTMAN, J. (2007). Learning bounds for domain adaptation. *Adv. Neural Inf. Process. Syst.* **20**.

BULL, A. D. (2012). Honest adaptive confidence bands and self-similar functions. *Electron. J. Stat.* **6** 1490–1516. MR2988456 https://doi.org/10.1214/12-EJS720

CAI, C., CAI, T. T. and LI, H. (2024). Supplement to "Transfer learning for contextual multi-armed bandits." https://doi.org/10.1214/23-AOS2341SUPP

MSC2020 subject classifications. Primary 62G08; secondary 62L12.

Key words and phrases. Contextual multi-armed bandit, transfer learning, covariate shift, minimax rate, regret bounds, adaptivity, self-similarity.

CAI, T. T. (2012). Minimax and adaptive inference in nonparametric function estimation. *Statist. Sci.* **27** 31–50. MR2953494 https://doi.org/10.1214/11-STS355

CAI, T. T. and LOW, M. G. (2004). An adaptation theory for nonparametric confidence intervals. *Ann. Statist.* **32** 1805–1840. MR2102494 https://doi.org/10.1214/009053604000000049

CAI, T. T., LOW, M. G. and XIA, Y. (2013). Adaptive confidence intervals for regression functions under shape constraints. *Ann. Statist.* **41** 722–750. MR3099119 https://doi.org/10.1214/12-AOS1068

CAI, T. T. and PU, H. (2022a). Transfer learning for nonparametric regression: Non-asymptotic minimax analysis and adaptive procedure. Preprint. Available at arXiv:2401.12272.

CAI, T. T. and PU, H. (2022b). Stochastic continuum-armed bandits with additive models: Minimax regrets and adaptive algorithm. *Ann. Statist.* **50** 2179–2204. MR4474487 https://doi.org/10.1214/22-aos2182

CAI, T. T. and WEI, H. (2021). Transfer learning for nonparametric classification: Minimax rate and adaptive classifier. *Ann. Statist.* **49** 100–128. MR4206671 https://doi.org/10.1214/20-AOS1949

CHEN, J. and JIANG, N. (2019). Information-theoretic considerations in batch reinforcement learning. In *International Conference on Machine Learning* 1042–1051. PMLR.

DEMIREL, I., CELIK, A. A. and TEKIN, C. (2022). Escada: Efficient safety and context aware dose allocation for precision medicine. *Adv. Neural Inf. Process. Syst.* **35** 27441–27454.

DING, K., LI, J. and LIU, H. (2019). Interactive anomaly detection on attributed networks. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining* 357–365.

DÜMBGEN, L. (1998). New goodness-of-fit tests and their application to nonparametric confidence sets. *Ann. Statist.* **26** 288–314. MR1611768 https://doi.org/10.1214/aos/1030563987

FARAHMAND, A.-M., SZEPESVÁRI, C. and MUNOS, R. (2010). Error propagation for approximate policy and value iteration. *Adv. Neural Inf. Process. Syst.* **23**.

GENOVESE, C. R. and WASSERMAN, L. (2005). Confidence sets for nonparametric wavelet regression. *Ann. Statist.* **33** 698–729. MR2163157 https://doi.org/10.1214/009053605000000011

GINÉ, E. and NICKL, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38** 1122–1170. MR2604707 https://doi.org/10.1214/09-AOS738

GOLDENSHLUGER, A. and ZEEVI, A. (2009). Woodroofe's one-armed bandit problem revisited. *Ann. Appl. Probab.* **19** 1603–1633. MR2538082 https://doi.org/10.1214/08-AAP589

GOLDENSHLUGER, A. and ZEEVI, A. (2013). A linear response bandit problem. *Stoch. Syst.* **3** 230–261. MR3353472 https://doi.org/10.1214/11-SSY032

GUR, Y., MOMENI, A. and WAGER, S. (2022). Smoothness-adaptive contextual bandits. *Oper. Res.* **70** 3198–3216. MR4538513 https://doi.org/10.1287/opre.2021.2215

HANNEKE, S. and KPOTUFE, S. (2019). On the value of target data in transfer learning. *Adv. Neural Inf. Process. Syst.* **32**.

HENGARTNER, N. W. and STARK, P. B. (1995). Finite-sample confidence envelopes for shape-restricted densities. *Ann. Statist.* **23** 525–550. MR1332580 https://doi.org/10.1214/aos/1176324534

HU, Y., KALLUS, N. and MAO, X. (2022). Smooth contextual bandits: Bridging the parametric and nondifferentiable regret regimes. *Oper. Res.* **70** 3261–3281. MR4538516

KALLUS, N. and UDELL, M. (2020). Dynamic assortment personalization in high dimensions. *Oper. Res.* **68** 1020–1037. MR4166283 https://doi.org/10.1287/opre.2019.1948

KLEINBERG, R. and LEIGHTON, T. (2003). The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In 44*th Annual IEEE Symposium on Foundations of Computer Science*, 2003. *Proceedings*. 594–605. IEEE, New York.

KPOTUFE, S. and MARTINET, G. (2021). Marginal singularity and the benefits of labels in covariate-shift. *Ann. Statist.* **49** 3299–3323. MR4352531 https://doi.org/10.1214/21-aos2084

KULIS, B., SAENKO, K. and DARRELL, T. (2011). What you saw is not what you get: Domain adaptation using asymmetric kernel transforms. In *CVPR* 2011 1785–1792. IEEE, New York.

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

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

LEPSKII, O. V. (1992). Asymptotically minimax adaptive estimation. I: Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.* **36** 682–697.

LEPSKII, O. V. (1993). Asymptotically minimax adaptive estimation. II. Schemes without optimal adaptation: Adaptive estimators. *Theory Probab. Appl.* **37** 433–448. MR1214353 https://doi.org/10.1137/1137095

LI, G., ZHAN, W., LEE, J. D., CHI, Y. and CHEN, Y. (2023). Reward-agnostic fine-tuning: Provable statistical benefits of hybrid reinforcement learning. Preprint. Available at arXiv:2305.10282.

LI, L., CHU, W., LANGFORD, J. and SCHAPIRE, R. E. (2010). A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the* 19*th International Conference on World Wide Web* 661–670.

LI, S., CAI, T. T. and LI, H. (2022). Transfer learning for high-dimensional linear regression: Prediction, estimation and minimax optimality. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 149–173. MR4400393

LI, S., CAI, T. T. and LI, H. (2023). Transfer learning in large-scale Gaussian graphical models with false discovery rate control. *J. Amer. Statist. Assoc.* **118** 2171–2183. MR4646634 https://doi.org/10.1080/01621459.2022.2044333

LI, W., DUAN, L., XU, D. and TSANG, I. W. (2013). Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.* **36** 1134–1148.

LOCATELLI, A. and CARPENTIER, A. (2018). Adaptivity to smoothness in X-armed bandits. In *Conference on Learning Theory* 1463–1492. PMLR.

LOW, M. G. (1997). On nonparametric confidence intervals. *Ann. Statist.* **25** 2547–2554. MR1604412 https://doi.org/10.1214/aos/1030741084

LUEDTKE, A. R. and VAN DER LAAN, M. J. (2016). Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Ann. Statist.* **44** 713–742. MR3476615 https://doi.org/10.1214/15-AOS1384

MA, C., PATHAK, R. and WAINWRIGHT, M. J. (2023). Optimally tackling covariate shift in RKHS-based nonparametric regression. *Ann. Statist.* **51** 738–761. MR4601000 https://doi.org/10.1214/23-aos2268

MAITY, S., SUN, Y. and BANERJEE, M. (2020). Minimax optimal approaches to the label shift problem. Preprint. Available at arXiv:2003.10443.

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

MANSOUR, Y., MOHRI, M. and ROSTAMIZADEH, A. (2009). Domain adaptation: Learning bounds and algorithms. Preprint. Available at arXiv:0902.3430.

MNIH, V., KAVUKCUOGLU, K., SILVER, D., RUSU, A. A., VENESS, J., BELLEMARE, M. G., GRAVES, A., RIEDMILLER, M., FIDJELAND, A. K. et al. (2015). Human-level control through deep reinforcement learning. *Nature* **518** 529–533.

MUNOS, R. (2007). Performance bounds in L_p-norm for approximate value iteration. *SIAM J. Control Optim.* **46** 541–561. MR2309039 https://doi.org/10.1137/040614384

NAKAMOTO, M., ZHAI, Y., SINGH, A., MARK, M. S., MA, Y., FINN, C., KUMAR, A. and LEVINE, S. (2023). Cal-QL: Calibrated offline RL pre-training for efficient online fine-tuning. Preprint. Available at arXiv:2303.05479.

NICKL, R. and SZABÓ, B. (2016). A sharp adaptive confidence ball for self-similar functions. *Stochastic Process. Appl.* **126** 3913–3934. MR3565485 https://doi.org/10.1016/j.spa.2016.04.017

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

PAN, S. J. and YANG, Q. (2009). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **22** 1345–1359.

PATHAK, R., MA, C. and WAINWRIGHT, M. (2022). A new similarity measure for covariate shift with applications to nonparametric regression. In *International Conference on Machine Learning* 17517–17530. PMLR.

PENG, M., LI, Y., WAMSLEY, B., WEI, Y. and ROEDER, K. (2021). Integration and transfer learning of single-cell transcriptomes via cFIT. *Proc. Natl. Acad. Sci. USA* **118** e2024383118.

PERCHET, V. and RIGOLLET, P. (2013). The multi-armed bandit problem with covariates. *Ann. Statist.* **41** 693–721. MR3099118 https://doi.org/10.1214/13-AOS1101

PICARD, D. and TRIBOULEY, K. (2000). Adaptive confidence interval for pointwise curve estimation. *Ann. Statist.* **28** 298–335. MR1762913 https://doi.org/10.1214/aos/1016120374

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

QIAN, W. and YANG, Y. (2016). Randomized allocation with arm elimination in a bandit problem with covariates. *Electron. J. Stat.* **10** 242–270. MR3466182 https://doi.org/10.1214/15-EJS1104

QUATTONI, A., COLLINS, M. and DARRELL, T. (2008). Transfer learning for image classification with sparse prototype representations. In 2008 *IEEE Conference on Computer Vision and Pattern Recognition* 1–8. IEEE, New York.

RABBI, M., AUNG, M. S., GAY, G., REID, M. C. and CHOUDHURY, T. (2018). Feasibility and acceptability of mobile phone-based auto-personalized physical activity recommendations for chronic pain self-management: Pilot study on adults. *J. Med. Internet Res.* **20** e10147. https://doi.org/10.2196/10147

RAGHU, M., ZHANG, C., KLEINBER, J. and BENGIO, S. (2019). Transfusion: Understanding transfer learning for medical imaging. In 33rd Conference on Neural Information Processing Systems (NeurIPS 2019).

RASHIDINEJAD, P., ZHU, B., MA, C., JIAO, J. and RUSSELL, S. (2022). Bridging offline reinforcement learning and imitation learning: A tale of pessimism. *IEEE Trans. Inf. Theory* **68** 8156–8196. MR4544936 https://doi.org/10.1109/tit.2022.3185139

REEVE, H. W. J., CANNINGS, T. I. and SAMWORTH, R. J. (2021). Adaptive transfer learning. *Ann. Statist.* **49** 3618–3649. MR4352543 https://doi.org/10.1214/21-aos2102

REEVE, H. W. J., MELLOR, J. and BROWN, G. (2018). The k-nearest neighbour UCB algorithm for multi-armed bandits with covariates. In *Algorithmic Learning Theory* 725–752. MR3857327

RIGOLLET, P. and ZEEVI, A. (2010). Nonparametric bandits with covariates. Preprint. Available at arXiv:1003.1630.

RINDTORFF, N. T., LU, M., PATEL, N. A., ZHENG, H. and D'AMOUR, A. (2019). A biologically plausible benchmark for contextual bandit algorithms in precision oncology using in vitro data. Preprint. Available at arXiv:1911.04389.

ROBBINS, H. (1952). Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.* **58** 527–535. MR0050246 https://doi.org/10.1090/S0002-9904-1952-09620-8

ROSS, S. and BAGNELL, J. A. (2012). Agnostic system identification for model-based reinforcement learning. In *Proceedings of the* 29*th International Conference on Machine Learning*.

ROTHSCHILD, M. (1974). A two-armed bandit theory of market pricing. *J. Econom. Theory* **9** 185–202. MR0496544 https://doi.org/10.1016/0022-0531(74)90066-0

SHI, C., LU, W. and SONG, R. (2020). Breaking the curse of nonregularity with subagging—Inference of the mean outcome under optimal treatment regimes. *J. Mach. Learn. Res.* **21** Paper No. 176, 67 pp. MR4209462

SHRESTHA, S. and JAIN, S. (2021). A Bayesian-bandit adaptive design for N-of-1 clinical trials. *Stat. Med.* **40** 1825–1844. MR4229804 https://doi.org/10.1002/sim.8873

SOEMERS, D., BRYS, T., DRIESSENS, K., WINANDS, M. and NOWÉ, A. (2018). Adapting to concept drift in credit card transaction data streams using contextual bandits and decision trees. In *Proceedings of the AAAI Conference on Artificial Intelligence* **32**.

SONG, Y., ZHOU, Y., SEKHARI, A., BAGNELL, J. A., KRISHNAMURTHY, A. and SUN, W. (2022). Hybrid RL: Using both offline and online data can make RL efficient. Preprint. Available at arXiv:2210.06718.

SUDLOW, C., GALLACHER, J., ALLEN, N., BERAL, V., BURTON, P., DANESH, J., DOWNEY, P., ELLIOTT, P., GREEN, J. et al. (2015). UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12** e1001779.

SUK, J. and KPOTUFE, S. (2021). Self-tuning bandits over unknown covariate-shifts. In *Algorithmic Learning Theory* 1114–1156. MR4227355

TEWARI, A. and MURPHY, S. A. (2017). From ads to interventions: Contextual bandits in mobile health. In *Mobile Health* 495–517. Springer, Berlin.

TSYBAKOV, A. B. (2004). Optimal aggregation of classifiers in statistical learning. *Ann. Statist.* **32** 135–166. MR2051002 https://doi.org/10.1214/aos/1079120131

WAGENMAKER, A. and PACCHIANO, A. (2023). Leveraging offline data in online reinforcement learning. In *International Conference on Machine Learning* 35300–35338. PMLR.

WANG, J., AGARWAL, D., HUANG, M., HU, G., ZHOU, Z., YE, C. and ZHANG, N. R. (2019). Data denoising with transfer learning in single-cell transcriptomics. *Nat. Methods* **16** 875–878.

WANG, Y., CHEN, B. and SIMCHI-LEVI, D. (2021). Multimodal dynamic pricing. *Manage. Sci.* **67** 6136–6152.

WEISS, K., KHOSHGOFTAAR, T. M. and WANG, D. (2016). A survey of transfer learning. *J. Big Data* **3** 1–40.

WOODROOFE, M. (1979). A one-armed bandit problem with a concomitant variable. *J. Amer. Statist. Assoc.* **74** 799–806. MR0556471

XIE, T. and JIANG, N. (2021). Batch value-function approximation with only realizability. In *International Conference on Machine Learning* 11404–11413. PMLR.

XIE, T., JIANG, N., WANG, H., XIONG, C. and BAI, Y. (2021). Policy finetuning: Bridging sample-efficient offline and online reinforcement learning. *Adv. Neural Inf. Process. Syst.* **34** 27395–27407.

YANG, Y. and ZHU, D. (2002). Randomized allocation with nonparametric estimation for a multi-armed bandit problem with covariates. *Ann. Statist.* **30** 100–121. MR1892657 https://doi.org/10.1214/aos/1015362186

YU, X., WANG, J., HONG, Q.-Q., TEKU, R., WANG, S.-H. and ZHANG, Y.-D. (2022). Transfer learning for medical images analyses: A survey. *Neurocomputing* **489** 230–254.

ZHOU, Z., WANG, Y., MAMANI, H. and COFFEY, D. G. (2019). How do tumor cytogenetics inform cancer treatments? Dynamic risk stratification and precision medicine using multi-armed bandits. *Dynamic Risk Stratification and Precision Medicine Using Multi-armed Bandits (June 17, 2019)*.

The Annals of Statistics
2024, Vol. 52, No. 1, 233–260
https://doi.org/10.1214/23-AOS2342

SETTLING THE SAMPLE COMPLEXITY OF MODEL-BASED OFFLINE REINFORCEMENT LEARNING

BY GEN LI[1,a], LAIXI SHI[2,b], YUXIN CHEN[3,c], YUEJIE CHI[4,e] AND YUTING WEI[3,d]

[1]*Department of Statistics, Chinese University of Hong Kong,* [a]*genli@cuhk.edu.hk*
[2]*Department of Computing Mathematical Sciences, California Institute of Technology,* [b]*laixis@caltech.edu*
[3]*Department of Statistics and Data Science, The Wharton School, University of Pennsylvania,* [c]*yuxinc@wharton.upenn.edu,* [d]*ytwei@wharton.upenn.edu*
[4]*Department of Electrical and Computer Engineering, Carnegie Mellon University,* [e]*yuejiechi@cmu.edu*

This paper is concerned with offline reinforcement learning (RL), which learns using precollected data without further exploration. Effective offline RL would be able to accommodate distribution shift and limited data coverage. However, prior results either suffer from suboptimal sample complexities or incur high burn-in cost to reach sample optimality, thus posing an impediment to efficient offline RL in sample-starved applications.

We demonstrate that the model-based (or "plug-in") approach achieves minimax-optimal sample complexity without any burn-in cost for tabular Markov decision processes (MDPs). Concretely, consider a γ-discounted infinite-horizon (resp., finite-horizon) MDP with S states and effective horizon $\frac{1}{1-\gamma}$ (resp., horizon H), and suppose the distribution shift of data is reflected by some single-policy clipped concentrability coefficient $C^{\star}_{\mathsf{clipped}}$. We prove that model-based offline RL yields ε-accuracy with a sample complexity of

$$\begin{cases} \dfrac{SC^{\star}_{\mathsf{clipped}}}{(1-\gamma)^3\varepsilon^2} & \text{(infinite-horizon MDPs)}, \\ \dfrac{H^4SC^{\star}_{\mathsf{clipped}}}{\varepsilon^2} & \text{(finite-horizon MDPs)}, \end{cases}$$

up to log factor, which is minimax optimal for the *entire ε-range*. The proposed algorithms are "pessimistic" variants of value iteration with Bernstein-style penalties, and do not require sophisticated variance reduction. Our analysis framework is established upon delicate leave-one-out decoupling arguments in conjunction with careful self-bounding techniques tailored to MDPs.

REFERENCES

AGARWAL, A., KAKADE, S. and YANG, L. F. (2020). Model-based reinforcement learning with a generative model is minimax optimal. In *Conference on Learning Theory* 67–83.

AUER, P. and ORTNER, R. (2006). Logarithmic online regret bounds for undiscounted reinforcement learning. *Adv. Neural Inf. Process. Syst.* **19**.

AZAR, M. G., OSBAND, I. and MUNOS, R. (2017). Minimax regret bounds for reinforcement learning. In *International Conference on Machine Learning* 263–272.

BAI, Y., XIE, T., JIANG, N. and WANG, Y.-X. (2019). Provably efficient Q-learning with low switching cost. In *Advances in Neural Information Processing Systems* 8002–8011.

BECK, C. L. and SRIKANT, R. (2012). Error bounds for constant step-size Q-learning. *Systems Control Lett.* **61** 1203–1208. MR2998204 https://doi.org/10.1016/j.sysconle.2012.08.014

BERTSEKAS, D. P. (2017). *Dynamic Programming and Optimal Control. Vol. I*, 4th ed. Athena Scientific, Belmont, MA. MR3644954

BOUREL, H., MAILLARD, O. and TALEBI, M. S. (2020). Tightening exploration in upper confidence reinforcement learning. In *International Conference on Machine Learning* 1056–1066.

MSC2020 subject classifications. 62C20.
Key words and phrases. Markov decision process, offline reinforcement learning, distribution shift, sample complexity, minimax optimality.

BUCKMAN, J., GELADA, C. and BELLEMARE, M. G. (2020). The importance of pessimism in fixed-dataset policy optimization. In *International Conference on Learning Representations*.

CHEN, J. and JIANG, N. (2019). Information-theoretic considerations in batch reinforcement learning. In *International Conference on Machine Learning* 1042–1051.

CHEN, M., LI, Y., WANG, E., YANG, Z., WANG, Z. and ZHAO, T. (2021). Pessimism meets invariance: Provably efficient offline mean-field multi-agent RL. *Adv. Neural Inf. Process. Syst.* **34**.

CHEN, Z., MAGULURI, S. T., SHAKKOTTAI, S. and SHANMUGAM, K. (2020). Finite-sample analysis of contractive stochastic approximation using smooth convex envelopes. *NeurIPS* **33** 8223–8234.

CHEN, Z., MAGULURI, S. T., SHAKKOTTAI, S. and SHANMUGAM, K. (2021). A Lyapunov theory for finite-sample guarantees of asynchronous Q-learning and TD-learning variants. arXiv preprint. Available at arXiv:2102.01567.

CUI, Q. and YANG, L. F. (2021). Minimax sample complexity for turn-based stochastic game. In *Uncertainty in Artificial Intelligence* 1496–1504.

DIEHL, C., SIEVERNICH, T., KRÜGER, M., HOFFMANN, F. and BERTRAN, T. (2021). Umbrella: Uncertainty-aware model-based offline reinforcement learning leveraging planning. arXiv preprint. Available at arXiv:2111.11097.

DONOHO, D. (2000). *High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality*. AMS Math Challenges Lecture.

DU, S. S., CHEN, J., LI, L., XIAO, L. and ZHOU, D. (2017). Stochastic variance reduction methods for policy evaluation. In *International Conference on Machine Learning* 1049–1058.

DUAN, Y., JIA, Z. and WANG, M. (2020). Minimax-optimal off-policy evaluation with linear function approximation. In *International Conference on Machine Learning* 2701–2709.

EBERT, F., FINN, C., DASARI, S., XIE, A., LEE, A. and LEVINE, S. (2018). Visual foresight: Model-based deep reinforcement learning for vision-based robotic control. arXiv preprint. Available at arXiv:1812.00568.

EVEN-DAR, E. and MANSOUR, Y. (2003). Learning rates for Q-learning. *J. Mach. Learn. Res.* **5** 1–25.

FILIPPI, S., CAPPÉ, O. and GARIVIER, A. (2010). Optimism in reinforcement learning and Kullback–Leibler divergence. In *Allerton Conference on Communication, Control, and Computing* 115–122. IEEE, Los Alamitos.

FRUIT, R., PIROTTA, M. and LAZARIC, A. (2020). Improved analysis of UCRL2 with empirical Bernstein inequality. arXiv preprint. Available at arXiv:2007.05456.

GHESHLAGHI AZAR, M., MUNOS, R. and KAPPEN, H. J. (2013). Minimax PAC bounds on the sample complexity of reinforcement learning with a generative model. *Mach. Learn.* **91** 325–349. MR3064431 https://doi.org/10.1007/s10994-013-5368-1

HE, J., ZHOU, D. and GU, Q. (2021). Nearly minimax optimal reinforcement learning for discounted MDPs. *Adv. Neural Inf. Process. Syst.* **34** 22288–22300.

JAKSCH, T., ORTNER, R. and AUER, P. (2010). Near-optimal regret bounds for reinforcement learning. *J. Mach. Learn. Res.* **11** 1563–1600. MR2645461

JIANG, N. and LI, L. (2016). Doubly robust off-policy value evaluation for reinforcement learning. In *International Conference on Machine Learning* 652–661.

JIN, C., ALLEN-ZHU, Z., BUBECK, S. and JORDAN, M. I. (2018). Is Q-learning provably efficient? In *Advances in Neural Information Processing Systems* 4863–4873.

JIN, C., YANG, Z., WANG, Z. and JORDAN, M. I. (2020). Provably efficient reinforcement learning with linear function approximation. In *Conference on Learning Theory* 2137–2143.

JIN, Y., YANG, Z. and WANG, Z. (2021). Is pessimism provably efficient for offline RL? In *International Conference on Machine Learning* 5084–5096.

JOHNSON, R. and ZHANG, T. (2013). Accelerating stochastic gradient descent using predictive variance reduction. In *Advances in Neural Information Processing Systems* 315–323.

KALLUS, N. and UEHARA, M. (2020). Double reinforcement learning for efficient off-policy evaluation in Markov decision processes. *J. Mach. Learn. Res.* **21** 167. MR4209453

KIDAMBI, R., RAJESWARAN, A., NETRAPALLI, P. and JOACHIMS, T. (2020). MOReL: Model-based offline reinforcement learning. *Adv. Neural Inf. Process. Syst.* **33** 21810–21823.

KUMAR, A., ZHOU, A., TUCKER, G. and LEVINE, S. (2020). Conservative Q-learning for offline reinforcement learning. In *Advances in Neural Information Processing Systems* **33** 1179–1191.

LAI, T. L. and ROBBINS, H. (1985). Asymptotically efficient adaptive allocation rules. *Adv. in Appl. Math.* **6** 4–22. MR0776826 https://doi.org/10.1016/0196-8858(85)90002-8

LANGE, S., GABEL, T. and RIEDMILLER, M. (2012). Batch reinforcement learning. In *Reinforcement Learning* 45–73. Springer, Berlin.

LATTIMORE, T. and SZEPESVÁRI, C. (2020). *Bandit Algorithms*. Cambridge Univ. Press, Cambridge.

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

LI, G., CAI, C., CHEN, Y., WEI, Y. and CHI, Y. (2023a). Is Q-learning minimax optimal? A tight sample complexity analysis. *Oper. Res.*

LI, G., SHI, L., CHEN, Y., CHI, Y. and WEI, Y. (2024). Supplement to "Settling the sample complexity of model-based offline reinforcement learning." https://doi.org/10.1214/23-AOS2342SUPP

LI, G., SHI, L., CHEN, Y., GU, Y. and CHI, Y. (2021). Breaking the sample complexity barrier to regret-optimal model-free reinforcement learning. *Adv. Neural Inf. Process. Syst.* **34**.

LI, G., WEI, Y., CHI, Y. and CHEN, Y. (2023c). Breaking the sample size barrier in model-based reinforcement learning with a generative model. *Oper. Res.*

LI, G., WEI, Y., CHI, Y., GU, Y. and CHEN, Y. (2020). Breaking the sample size barrier in model-based reinforcement learning with a generative model. In *Advances in Neural Information Processing Systems* **33**.

LI, G., WEI, Y., CHI, Y., GU, Y. and CHEN, Y. (2022). Sample complexity of asynchronous Q-learning: Sharper analysis and variance reduction. *IEEE Trans. Inf. Theory* **68** 448–473. MR4395423 https://doi.org/10.1109/tit.2021.3120096

LI, G., YAN, Y., CHEN, Y. and FAN, J. (2023a). Minimax-optimal reward-agnostic exploration in reinforcement learning. arXiv preprint. Available at arXiv:2304.07278.

LI, G., ZHAN, W., LEE, J. D., CHI, Y. and CHEN, Y. (2023b). Reward-agnostic fine-tuning: Provable statistical benefits of hybrid reinforcement learning. *Adv. Neural Inf. Process. Syst.*

LI, L., MUNOS, R. and SZEPESVÁRI, C. (2014). On minimax optimal offline policy evaluation. Available at arXiv:1409.3653.

MUNOS, R. (2007). Performance bounds in L_p-norm for approximate value iteration. *SIAM J. Control Optim.* **46** 541–561. MR2309039 https://doi.org/10.1137/040614384

NGUYEN-TANG, T., GUPTA, S. and VENKATESH, S. (2021). Sample complexity of offline reinforcement learning with deep ReLU networks. arXiv preprint. Available at arXiv:2103.06671.

PANAGANTI, K. and KALATHIL, D. (2022). Sample complexity of robust reinforcement learning with a generative model. In *International Conference on Artificial Intelligence and Statistics* 9582–9602.

PANANJADY, A. and WAINWRIGHT, M. J. (2021). Instance-dependent ℓ_∞-bounds for policy evaluation in tabular reinforcement learning. *IEEE Trans. Inf. Theory* **67** 566–585. MR4231973 https://doi.org/10.1109/TIT.2020.3027316

PRUDENCIO, R. F., MAXIMO, M. R. and COLOMBINI, E. L. (2022). A survey on offline reinforcement learning: Taxonomy, review, and open problems. arXiv preprint. Available at arXiv:2203.01387.

QU, G. and WIERMAN, A. (2020). Finite-time analysis of asynchronous stochastic approximation and Q-learning. In *Conference on Learning Theory* 3185–3205.

RASHIDINEJAD, P., ZHU, B., MA, C., JIAO, J. and RUSSELL, S. (2022). Bridging offline reinforcement learning and imitation learning: A tale of pessimism. *IEEE Trans. Inf. Theory* **68** 8156–8196. MR4544936 https://doi.org/10.1109/tit.2022.3185139

ROBBINS, H. and MONRO, S. (1951). A stochastic approximation method. *Ann. Math. Stat.* **22** 400–407. MR0042668 https://doi.org/10.1214/aoms/1177729586

SHI, C., LUO, S., LE, Y., ZHU, H. and SONG, R. (2022). Statistically efficient advantage learning for offline reinforcement learning in infinite horizons. *J. Amer. Statist. Assoc.* 1–14.

SHI, L. and CHI, Y. (2022). Distributionally robust model-based offline reinforcement learning with near-optimal sample complexity. arXiv preprint. Available at arXiv:2208.05767.

SHI, L., LI, G., WEI, Y., CHEN, Y. and CHI, Y. (2022). Pessimistic Q-learning for offline reinforcement learning: Towards optimal sample complexity. In *International Conference on Machine Learning* 19967–20025.

SHI, L., LI, G., WEI, Y., CHEN, Y., GEIST, M. and CHI, Y. (2023). The curious price of distributional robustness in reinforcement learning with a generative model. *Adv. Neural Inf. Process. Syst.*

SIDFORD, A., WANG, M., WU, X., YANG, L. and YE, Y. (2018). Near-optimal time and sample complexities for solving Markov decision processes with a generative model. In *NeurIPS* 5186–5196.

SILVER, D., SCHRITTWIESER, J., SIMONYAN, K., ANTONOGLOU, I., HUANG, A., GUEZ, A., HUBERT, T., BAKER, L., LAI, M. et al. (2017). Mastering the game of Go without human knowledge. *Nature* **550** 354–359.

SUTTON, R. S. and BARTO, A. G. (2018). *Reinforcement Learning: An Introduction*, 2nd ed. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3889951

SZEPESVÁRI, C. (1998). The asymptotic convergence-rate of Q-learning. In *Advances in Neural Information Processing Systems* 1064–1070.

TALEBI, M. S. and MAILLARD, O.-A. (2018). Variance-aware regret bounds for undiscounted reinforcement learning in MDPs. In *Algorithmic Learning Theory* 2018. *Proc. Mach. Learn. Res.* (*PMLR*) **83** 770–805. MR3857329

TANG, S. and WIENS, J. (2021). Model selection for offline reinforcement learning: Practical considerations for healthcare settings. In *Machine Learning for Healthcare Conference* 2–35.

UEHARA, M. and SUN, W. (2021). Pessimistic model-based offline reinforcement learning under partial coverage. arXiv preprint. Available at arXiv:2107.06226.

WAINWRIGHT, M. J. (2019a). *High-Dimensional Statistics: A Non-Asymptotic Viewpoint. Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge Univ. Press, Cambridge.

WAINWRIGHT, M. J. (2019b). Variance-reduced Q-learning is minimax optimal. Available at arXiv:1906.04697.

WANG, Y., DONG, K., CHEN, X. and WANG, L. (2019). Q-learning with UCB exploration is sample efficient for infinite-horizon MDP. In *International Conference on Learning Representations*.

WATKINS, C. J. and DAYAN, P. (1992). Q-learning. *Mach. Learn.* **8** 279–292.

XIE, T. and JIANG, N. (2021). Batch value-function approximation with only realizability. In *International Conference on Machine Learning* 11404–11413.

XIE, T., JIANG, N., WANG, H., XIONG, C. and BAI, Y. (2021). Policy finetuning: Bridging sample-efficient offline and online reinforcement learning. *Adv. Neural Inf. Process. Syst.* **34** 27395–27407.

XIONG, H., ZHAO, L., LIANG, Y. and ZHANG, W. (2020). Finite-time analysis for double Q-learning. *Adv. Neural Inf. Process. Syst.* **33**.

YAN, Y., LI, G., CHEN, Y. and FAN, J. (2022). Model-based reinforcement learning is minimax-optimal for offline zero-sum Markov games. arXiv preprint. Available at arXiv:2206.04044.

YAN, Y., LI, G., CHEN, Y. and FAN, J. (2023). The efficacy of pessimism in asynchronous Q-learning. *IEEE Trans. Inf. Theory* **69** 7185–7219. MR4660859

YIN, M., DUAN, Y., WANG, M. and WANG, Y.-X. (2021). Near-optimal offline reinforcement learning with linear representation: Leveraging variance information with pessimism. In *International Conference on Learning Representations*.

YIN, M. and WANG, Y.-X. (2021). Towards instance-optimal offline reinforcement learning with pessimism. *Adv. Neural Inf. Process. Syst.* **34**.

YU, T., THOMAS, G., YU, L., ERMON, S., ZOU, J. Y., LEVINE, S., FINN, C. and MA, T. (2020). MOPO: Model-based offline policy optimization. *Adv. Neural Inf. Process. Syst.* **33** 14129–14142.

ZANETTE, A., WAINWRIGHT, M. J. and BRUNSKILL, E. (2021). Provable benefits of actor-critic methods for offline reinforcement learning. *Adv. Neural Inf. Process. Syst.* **34**.

ZHAN, W., HUANG, B., HUANG, A., JIANG, N. and LEE, J. (2022). Offline reinforcement learning with realizability and single-policy concentrability. In *Conference on Learning Theory* 2730–2775.

ZHANG, Z., CHEN, Y., LEE, J. D. and DU, S. S. (2023). Settling the sample complexity of online reinforcement learning. arXiv preprint, 2307.13586.

ZHANG, Z., JI, X. and DU, S. (2021). Is reinforcement learning more difficult than bandits? A near-optimal algorithm escaping the curse of horizon. In *Conference on Learning Theory* 4528–4531.

ZHANG, Z., ZHOU, Y. and JI, X. (2020). Almost optimal model-free reinforcement learning via reference-advantage decomposition. *Adv. Neural Inf. Process. Syst.* **33**.

ZHANG, Z., ZHOU, Y. and JI, X. (2021). Model-free reinforcement learning: From clipped pseudo-regret to sample complexity. In *International Conference on Machine Learning* 12653–12662.

ZHOU, Z., ZHOU, Z., BAI, Q., QIU, L., BLANCHET, J. and GLYNN, P. (2021). Finite-sample regret bound for distributionally robust offline tabular reinforcement learning. In *AISTATS* 3331–3339.

The Annals of Statistics
2024, Vol. 52, No. 1, 261–284
https://doi.org/10.1214/23-AOS2346

RATES OF ESTIMATION FOR HIGH-DIMENSIONAL MULTIREFERENCE ALIGNMENT

BY ZEHAO DOU[a], ZHOU FAN[b] AND HARRISON H. ZHOU[c]

Department of Statistics and Data Science, Yale University, [a]*zehao.dou@yale.edu,* [b]*zhou.fan@yale.edu,* [c]*huibin.zhou@yale.edu*

We study the continuous multireference alignment model of estimating a periodic function on the circle from noisy and circularly-rotated observations. Motivated by analogous high-dimensional problems that arise in cryo-electron microscopy, we establish minimax rates for estimating generic signals that are explicit in the dimension K. In a high-noise regime with noise variance $\sigma^2 \gtrsim K$, for signals with Fourier coefficients of roughly uniform magnitude, the rate scales as σ^6 and has no further dependence on the dimension. This rate is achieved by a bispectrum inversion procedure, and our analyses provide new stability bounds for bispectrum inversion that may be of independent interest. In a low-noise regime where $\sigma^2 \lesssim K/\log K$, the rate scales instead as $K\sigma^2$, and we establish this rate by a sharp analysis of the maximum likelihood estimator that marginalizes over latent rotations. A complementary lower bound that interpolates between these two regimes is obtained using Assouad's hypercube lemma. We extend these analyses also to signals whose Fourier coefficients have a slow power law decay.

REFERENCES

[1] ABBE, E., BENDORY, T., LEEB, W., PEREIRA, J. M., SHARON, N. and SINGER, A. (2019). Multireference alignment is easier with an aperiodic translation distribution. *IEEE Trans. Inf. Theory* **65** 3565–3584. MR3959006 https://doi.org/10.1109/TIT.2018.2889674

[2] ABBE, E., PEREIRA, J. M. and SINGER, A. (2018). Estimation in the group action channel. In 2018 *IEEE International Symposium on Information Theory* (*ISIT*) 561–565. IEEE, New York.

[3] BANDEIRA, A. S., BLUM-SMITH, B., KILEEL, J., NILES-WEED, J., PERRY, A. and WEIN, A. S. (2023). Estimation under group actions: Recovering orbits from invariants. *Appl. Comput. Harmon. Anal.* **66** 236–319. MR4609472 https://doi.org/10.1016/j.acha.2023.06.001

[4] BANDEIRA, A. S., CHARIKAR, M., SINGER, A. and ZHU, A. (2014). Multireference alignment using semidefinite programming. In *ITCS'*14*—Proceedings of the* 2014 *Conference on Innovations in Theoretical Computer Science* 459–470. ACM, New York. MR3359498

[5] BANDEIRA, A. S., CHEN, Y. and SINGER, A. (2015). Non-unique games over compact groups and orientation estimation in cryo-EM. Preprint. Available at arXiv:1505.03840.

[6] BANDEIRA, A. S., NILES-WEED, J. and RIGOLLET, P. (2019). Optimal rates of estimation for multi-reference alignment. *Math. Stat. Learn.* **2** 25–75. MR4073147 https://doi.org/10.4171/msl/11

[7] BENDORY, T., BARTESAGHI, A. and SINGER, A. (2020). Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities. *IEEE Signal Process. Mag.* **37** 58–76. https://doi.org/10.1109/msp.2019.2957822

[8] BENDORY, T., BOUMAL, N., MA, C., ZHAO, Z. and SINGER, A. (2018). Bispectrum inversion with application to multireference alignment. *IEEE Trans. Signal Process.* **66** 1037–1050. MR3771661 https://doi.org/10.1109/TSP.2017.2775591

[9] BENDORY, T., EDIDIN, D., LEEB, W. and SHARON, N. (2022). Dihedral multi-reference alignment. *IEEE Trans. Inf. Theory* **68** 3489–3499. MR4433234 https://doi.org/10.1109/tit.2022.3146488

[10] BENDORY, T., JAFFE, A., LEEB, W., SHARON, N. and SINGER, A. (2022). Super-resolution multi-reference alignment. *Inf. Inference* **11** 533–555. MR4440674 https://doi.org/10.1093/imaiai/iaab003

[11] BOUMAL, N. (2016). Nonconvex phase synchronization. *SIAM J. Optim.* **26** 2355–2377. MR3566919 https://doi.org/10.1137/16M105808X

MSC2020 subject classifications. Primary 62G05; secondary 62C20.
Key words and phrases. Function estimation, group invariance, bispectrum inversion.

[12] BOUMAL, N., BENDORY, T., LEDERMAN, R. R. and SINGER, A. (2018). Heterogeneous multireference alignment: A single pass approach. In 2018 52*nd Annual Conference on Information Sciences and Systems* (*CISS*) 1–6. IEEE, New York.

[13] BROWN, L. G. (1992). A survey of image registration techniques. *ACM Comput*. *Surv*. **24** 325–376.

[14] BRUNEL, V.-E. (2019). Learning rates for Gaussian mixtures under group action. In *Conference on Learning Theory* 471–491. PMLR.

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

[16] DIAMOND, R. (1992). On the multiple simultaneous superposition of molecular structures by rigid body transformations. *Protein Sci*. **1** 1279–1287. https://doi.org/10.1002/pro.5560011006

[17] DOU, Z., FAN, Z. and ZHOU, H. (2024). Supplement to "Rates of estimation for high-dimensional multireference alignment." https://doi.org/10.1214/23-AOS2346SUPP

[18] FAN, Z., LEDERMAN, R. R., SUN, Y., WANG, T. and XU, S. (2021). Maximum likelihood for high-noise group orbit estimation and single-particle cryo-EM. Preprint. Available at arXiv:2107.01305.

[19] FAN, Z., SUN, Y., WANG, T. and WU, Y. (2023). Likelihood landscape and maximum likelihood estimation for the discrete orbit recovery model. *Comm*. *Pure Appl*. *Math*. **76** 1208–1302. MR4582297

[20] GHOSH, S. and RIGOLLET, P. (2021). Multi-reference alignment for sparse signals, uniform uncertainty principles and the beltway problem. Preprint. Available at arXiv:2106.12996.

[21] KAM, Z. (1980). The reconstruction of structure from electron micrographs of randomly oriented particles. *J*. *Theoret*. *Biol*. **82** 15–39. https://doi.org/10.1016/0022-5193(80)90088-0

[22] KATSEVICH, A. E. and BANDEIRA, A. S. (2023). Likelihood maximization and moment matching in low SNR Gaussian mixture models. *Comm*. *Pure Appl*. *Math*. **76** 788–842. MR4569606 https://doi.org/10.1002/cpa.22051

[23] LATAŁA, R. (2006). Estimates of moments and tails of Gaussian chaoses. *Ann*. *Probab*. **34** 2315–2331. MR2294983 https://doi.org/10.1214/009117906000000421

[24] LING, S. (2022). Near-optimal performance bounds for orthogonal and permutation group synchronization via spectral methods. *Appl*. *Comput*. *Harmon*. *Anal*. **60** 20–52. MR4387245 https://doi.org/10.1016/j.acha.2022.02.003

[25] PERRY, A., WEED, J., BANDEIRA, A. S., RIGOLLET, P. and SINGER, A. (2019). The sample complexity of multireference alignment. *SIAM J*. *Math*. *Data Sci*. **1** 497–517. MR4002723 https://doi.org/10.1137/18M1214317

[26] 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

[27] PUMIR, T., SINGER, A. and BOUMAL, N. (2021). The generalized orthogonal Procrustes problem in the high noise regime. *Inf*. *Inference* **10** 921–954. MR4312088 https://doi.org/10.1093/imaiai/iaaa035

[28] RITOV, Y. (1989). Estimating a signal with noisy nuisance parameters. *Biometrika* **76** 31–37. MR0991420 https://doi.org/10.1093/biomet/76.1.31

[29] ROMANOV, E., BENDORY, T. and ORDENTLICH, O. (2021). Multi-reference alignment in high dimensions: Sample complexity and phase transition. *SIAM J*. *Math*. *Data Sci*. **3** 494–523. MR4245326 https://doi.org/10.1137/20M1354994

[30] SADLER, B. M. and GIANNAKIS, G. B. (1992). Shift-and rotation-invariant object reconstruction using the bispectrum. *J*. *Opt*. *Soc*. *Amer*. *A* **9** 57–69.

[31] SCHERES, S. H. W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J*. *Struct*. *Biol*. **180** 519–530. https://doi.org/10.1016/j.jsb.2012.09.006

[32] SCHERES, S. H. W., VALLE, M., NUÑEZ, R., SORZANO, C. O. S., MARABINI, R., HERMAN, G. T. and CARAZO, J.-M. (2005). Maximum-likelihood multi-reference refinement for electron microscopy images. *J*. *Mol*. *Biol*. **348** 139–149. https://doi.org/10.1016/j.jmb.2005.02.031

[33] SHARON, N., KILEEL, J., KHOO, Y., LANDA, B. and SINGER, A. (2020). Method of moments for 3D single particle *ab initio* modeling with non-uniform distribution of viewing angles. *Inverse Probl*. **36** 044003, 40 pp. MR4072349 https://doi.org/10.1088/1361-6420/ab6139

[34] SIGWORTH, F. J. (1998). A maximum-likelihood approach to single-particle image refinement. *J*. *Struct*. *Biol*. **122** 328–339. https://doi.org/10.1006/jsbi.1998.4014

[35] SINGER, A. (2011). Angular synchronization by eigenvectors and semidefinite programming. *Appl*. *Comput*. *Harmon*. *Anal*. **30** 20–36. MR2737931 https://doi.org/10.1016/j.acha.2010.02.001

[36] SINGER, A., COIFMAN, R. R., SIGWORTH, F. J., CHESTER, D. W. and SHKOLNISKY, Y. (2010). Detecting consistent common lines in cryo-EM by voting. *J*. *Struct*. *Biol*. **169** 312–322. https://doi.org/10.1016/j.jsb.2009.11.003

[37] SINGER, A. and SHKOLNISKY, Y. (2011). Three-dimensional structure determination from common lines in cryo-EM by eigenvectors and semidefinite programming. *SIAM J. Imaging Sci.* **4** 543–572. MR2810897 https://doi.org/10.1137/090767777

[38] SINGER, A. and SIGWORTH, F. J. (2020). Computational methods for single-particle electron cryomicroscopy. *Annu. Rev. Biomed. Data Sci.* **3** 163–190. https://doi.org/10.1146/annurev-biodatasci-021020-093826

The Annals of Statistics
2024, Vol. 52, No. 1, 285–310
https://doi.org/10.1214/23-AOS2347

SUPERVISED HOMOGENEITY FUSION: A COMBINATORIAL APPROACH

BY WEN WANG[1,a], SHIHAO WU[2,c], ZIWEI ZHU[2,d], LING ZHOU[3,e] AND PETER X.-K. SONG[1,b]

[1]*Department of Biostatistics, University of Michigan, Ann Arbor,* [a]*wangwen@umich.edu,* [b]*pxsong@umich.edu*
[2]*Department of Statistics, University of Michigan, Ann Arbor,* [c]*wshihao@umich.edu,* [d]*zzw9348ustc@gmail.com*
[3]*Center of Statistical Research, Southwestern University of Finance and Economics,* [e]*lingzhou@swufe.edu.cn*

Fusing regression coefficients into homogeneous groups can unveil those coefficients that share a common value within each group. Such groupwise homogeneity reduces the intrinsic dimension of the parameter space and unleashes sharper statistical accuracy. We propose and investigate a new combinatorial grouping approach called L_0-Fusion that is amenable to mixed integer optimization (MIO). On the statistical aspect, we identify a fundamental quantity called *MSE grouping sensitivity* that underpins the difficulty of recovering the true groups. We show that L_0-Fusion achieves grouping consistency under the weakest possible requirement of the grouping sensitivity: if this requirement is violated, then the minimax risk of group misspecification will fail to converge to zero. Moreover, we show that in the high-dimensional regime, one can apply L_0-Fusion with a sure screening set of features without any essential loss of statistical efficiency, while reducing the computational cost substantially. On the algorithmic aspect, we provide an MIO formulation for L_0-Fusion along with a warm start strategy. Simulation and real data analysis demonstrate that L_0-Fusion exhibits superiority over its competitors in terms of grouping accuracy.

REFERENCES

ANA, L. F. and JAIN, A. K. (2003). Robust data clustering. In 2003 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2003. *Proceedings*. **2** II–II. IEEE, New York.

AUGER, I. E. and LAWRENCE, C. E. (1989). Algorithms for the optimal identification of segment neighborhoods. *Bull. Math. Biol.* **51** 39–54. MR0978902 https://doi.org/10.1016/S0092-8240(89)80047-3

BEALE, E. M. L., KENDALL, M. G. and MANN, D. W. (1967). The discarding of variables in multivariate analysis. *Biometrika* **54** 357–366. MR0221656 https://doi.org/10.1093/biomet/54.3-4.357

BEALE, E. M. L. and TOMLIN, J. A. (1970). Special facilities in a general mathematical programming system for non-convex problems using ordered sets of variables. *Oper. Res.* **69** 99.

BERTSIMAS, D., KING, A. and MAZUMDER, R. (2016). Best subset selection via a modern optimization lens. *Ann. Statist.* **44** 813–852. MR3476618 https://doi.org/10.1214/15-AOS1388

BERTSIMAS, D. and WEISMANTEL, R. (2005). *Optimization over Integers* **13**. Dynamic Ideas Belmont.

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

BRAUN, J. M., HOFFMAN, E., SCHWARTZ, J., SANCHEZ, B., SCHNAAS, L., MERCADO-GARCIA, A., SOLANO-GONZALEZ, M., BELLINGER, D. C., LANPHEAR, B. P. et al. (2012). Assessing windows of susceptibility to lead-induced cognitive deficits in Mexican children. *Neurotoxicolog* **33** 1040–1047.

COOK, W., LOVÁSZ, L. and SEYMOUR, P., eds. (1995) *Combinatorial Optimization*: *Papers from the DIMACS Special Year*. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* **20**. Amer. Math. Soc., Providence, RI. MR1338611 https://doi.org/10.1090/dimacs/020

DANTZIG, G. B. (1960). On the significance of solving linear programming problems with some integer variables. *Econometrica* **28** 30–44. MR0112748 https://doi.org/10.2307/1905292

MSC2020 subject classifications. Primary 60J05, 90C11; secondary 62P10.
Key words and phrases. Clustering, dimension reduction, mixed integer optimization, linear model, supervised learning.

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. and LV, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 849–911. MR2530322 https://doi.org/10.1111/j.1467-9868.2008.00674.x

GARSIDE, M. (1965). The best sub-set in multiple regression analysis. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **14** 196–200.

GUO, Y., ZHU, Z. and FAN, J. (2020). Best subset selection is robust against design dependence. Preprint. Available at arXiv:2007.01478.

GUROBI OPTIMIZATION, LLC (2021). Gurobi optimizer reference manual.

HOCKING, R. R. and LESLIE, R. N. (1967). Selection of the best subset in regression analysis. *Technometrics* **9** 531–540. MR0219192 https://doi.org/10.2307/1266192

JEON, J.-J., KWON, S. and CHOI, H. (2017). Homogeneity detection for the high-dimensional generalized linear model. *Comput. Statist. Data Anal.* **114** 61–74. MR3660839 https://doi.org/10.1016/j.csda.2017.04.001

JÜNGER, M. and REINELT, G. (2013). *Facets of Combinatorial Optimization*. Springer, Berlin.

KE, Y., LI, J. and ZHANG, W. (2016). Structure identification in panel data analysis. *Ann. Statist.* **44** 1193–1233. MR3485958 https://doi.org/10.1214/15-AOS1403

KE, Z. T., FAN, J. and WU, Y. (2015). Homogeneity pursuit. *J. Amer. Statist. Assoc.* **110** 175–194. MR3338495 https://doi.org/10.1080/01621459.2014.892882

KOBROSLY, R. W., PARLETT, L. E., STAHLHUT, R. W., BARRETT, E. S. and SWAN, S. H. (2012). Socioeconomic factors and phthalate metabolite concentrations among United States women of reproductive age. *Environ. Res.* **115** 11–17. https://doi.org/10.1016/j.envres.2012.03.008

LIAN, H., QIAO, X. and ZHANG, W. (2021). Homogeneity pursuit in single index models based panel data analysis. *J. Bus. Econom. Statist.* **39** 386–401. MR4235184 https://doi.org/10.1080/07350015.2019.1665531

MA, S. and HUANG, J. (2017). A concave pairwise fusion approach to subgroup analysis. *J. Amer. Statist. Assoc.* **112** 410–423. MR3646581 https://doi.org/10.1080/01621459.2016.1148039

MARIE, C., VENDITTELLI, F. and SAUVANT-ROCHAT, M. P. (2015). Obstetrical outcomes and biomarkers toassess exposure to phthalates: A review. *Environ. Int.* **83** 116–136.

MARKOWITZ, H. M. and MANNE, A. S. (1957). On the solution of discrete programming problems. *Econometrica* **25** 84–110. MR0085968 https://doi.org/10.2307/1907744

MARSEE, K., WOODRUFF, T. J., AXELRAD, D. A., CALAFAT, A. M. and SWAN, S. H. (2006). Estimated dailyphthalate exposures in a population of mothers of male infants exhibiting reduced anogenital distance. *Environ. Health Perspect.* **114** 805–809.

NEEDELL, D. and TROPP, J. A. (2009). CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.* **26** 301–321. MR2502366 https://doi.org/10.1016/j.acha.2008.07.002

NESTEROV, Y. (2004). *Introductory Lectures on Convex Optimization: A Basic Course*. *Applied Optimization* **87**. Kluwer Academic, Boston, MA. MR2142598 https://doi.org/10.1007/978-1-4419-8853-9

NESTEROV, Y. (2013). Gradient methods for minimizing composite functions. *Math. Program.* **140** 125–161. MR3071865 https://doi.org/10.1007/s10107-012-0629-5

SCHETTLER, T. (2006). Human exposure to phthalates via consumer products. *J. Androl.* **29** 134–139.

SHEN, J. and HE, X. (2015). Inference for subgroup analysis with a structured logistic-normal mixture model. *J. Amer. Statist. Assoc.* **110** 303–312. MR3338504 https://doi.org/10.1080/01621459.2014.894763

SHEN, X. and HUANG, H.-C. (2010). Grouping pursuit through a regularization solution surface. *J. Amer. Statist. Assoc.* **105** 727–739. MR2724856 https://doi.org/10.1198/jasa.2010.tm09380

SHEN, X., PAN, W., ZHU, Y. and ZHOU, H. (2013). On constrained and regularized high-dimensional regression. *Ann. Inst. Statist. Math.* **65** 807–832. MR3105798 https://doi.org/10.1007/s10463-012-0396-3

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

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

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

VIELMA, J. P. and NEMHAUSER, G. L. (2011). Modeling disjunctive constraints with a logarithmic number of binary variables and constraints. *Math. Program.* **128** 49–72. MR2810952 https://doi.org/10.1007/s10107-009-0295-4

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

WAINWRIGHT, M. J. (2019). *High-Dimensional Statistics*: *A Non-asymptotic Viewpoint*. *Cambridge Series in Statistical and Probabilistic Mathematics* **48**. Cambridge Univ. Press, Cambridge. MR3967104 https://doi.org/10.1017/9781108627771

WANG, W., WU, S., ZHU, Z., ZHOU, L. and SONG, P. X. (2024). Supplement to "Supervised homogeneity fusion: A combinatorial approach." https://doi.org/10.1214/23-AOS2347SUPP

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

ZHOU, L., SUN, S., FU, H. and SONG, P. X.-K. (2022). Subgroup-effects models for the analysis of personal treatment effects. *Ann*. *Appl*. *Stat*. **16** 80–103. MR4400504 https://doi.org/10.1214/21-aoas1503

ZHU, Y., SHEN, X. and PAN, W. (2013). Simultaneous grouping pursuit and feature selection over an undirected graph. *J*. *Amer*. *Statist*. *Assoc*. **108** 713–725. MR3174654 https://doi.org/10.1080/01621459.2013.770704

ZHU, Z. and WU, S. (2021). On the early solution path of best subset selection. Preprint. Available at arXiv:2107.06939.

The Annals of Statistics
2024, Vol. 52, No. 1, 311–334
https://doi.org/10.1214/23-AOS2348

TESTING FOR INDEPENDENCE IN HIGH DIMENSIONS BASED ON EMPIRICAL COPULAS

BY AXEL BÜCHER[1,a] AND CAMBYSE PAKZAD[2,b]

[1]*Fakultät für Mathematik, Ruhr-Universität Bochum,* [a]*axel.buecher@rub.de*
[2]*Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK,* [b]*cambyse.pakzad@inria.fr*

Testing for pairwise independence for the case where the number of variables may be of the same size or even larger than the sample size has received increasing attention in the recent years. We contribute to this branch of the literature by considering tests that allow to detect higher-order dependencies. The proposed methods are based on connecting the problem to copulas and making use of the Moebius transformation of the empirical copula process; an approach that is related to Lancaster interactions and that has already been used successfully for the case where the number of variables is fixed. Based on a martingale central limit theorem, it is shown that respective test statistics converge to the standard normal distribution, allowing for straightforward definition of critical values. The results are illustrated by a Monte Carlo simulation study.

REFERENCES

[1] BATTISTON, F., CENCETTI, G., IACOPINI, I., LATORA, V., LUCAS, M., PATANIA, A., YOUNG, J.-G. and PETRI, G. (2020). Networks beyond pairwise interactions: Structure and dynamics. *Phys. Rep.* **874** 1–92. MR4147650 https://doi.org/10.1016/j.physrep.2020.05.004

[2] BÜCHER, A. and PAKZAD, C. (2024). Supplement to "Testing for independence in high dimensions based on empirical copulas." https://doi.org/10.1214/23-AOS2348SUPP

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

[4] CAI, T. T. and MA, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** 2359–2388. MR3160557 https://doi.org/10.3150/12-BEJ455

[5] CHEN, S. X., ZHANG, L.-X. and ZHONG, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** 810–819. MR2724863 https://doi.org/10.1198/jasa.2010.tm09560

[6] CORMEN, T. H., LEISERSON, C. E., RIVEST, R. L. and STEIN, C. (2009). *Introduction to Algorithms*, 3rd ed. MIT Press, Cambridge, MA. MR2572804

[7] DEHEUVELS, P. (1979). La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. *Acad. Roy. Belg. Bull. Cl. Sci.* (5) **65** 274–292. MR0573609

[8] DEHEUVELS, P. (1981). An asymptotic decomposition for multivariate distribution-free tests of independence. *J. Multivariate Anal.* **11** 102–113. MR0612295 https://doi.org/10.1016/0047-259X(81)90136-6

[9] DEHEUVELS, P. (1981). A Kolmogorov–Smirnov type test for independence and multivariate samples. *Rev. Roumaine Math. Pures Appl.* **26** 213–226. MR0616038

[10] DRTON, M., HAN, F. and SHI, H. (2020). High-dimensional consistent independence testing with maxima of rank correlations. *Ann. Statist.* **48** 3206–3227. MR4185806 https://doi.org/10.1214/19-AOS1926

[11] FANG, H.-B., FANG, K.-T. and KOTZ, S. (2002). The meta-elliptical distributions with given marginals. *J. Multivariate Anal.* **82** 1–16. MR1918612 https://doi.org/10.1006/jmva.2001.2017

[12] GEISSER, S. and MANTEL, N. (1962). Pairwise independence of jointly dependent variables. *Ann. Math. Stat.* **33** 290–291. MR0137188 https://doi.org/10.1214/aoms/1177704732

[13] GENEST, C., NEŠLEHOVÁ, J. G., RÉMILLARD, B. and MURPHY, O. A. (2019). Testing for independence in arbitrary distributions. *Biometrika* **106** 47–68. MR3912383 https://doi.org/10.1093/biomet/asy059

MSC2020 subject classifications. Primary 62G10, 62H15; secondary 60F05.
Key words and phrases. Empirical copula process, high-dimensional statistics, higher-order dependence, Moebius transform, Lancaster interaction, rank based inference.

[14] GENEST, C., QUESSY, J.-F. and REMILLARD, B. (2007). Asymptotic local efficiency of Cramér–von Mises tests for multivariate independence. *Ann. Statist.* **35** 166–191. MR2332273 https://doi.org/10.1214/009053606000000984

[15] GENEST, C. and RÉMILLARD, B. (2004). Tests of independence and randomness based on the empirical copula process. *TEST* **13** 335–370. MR2154005 https://doi.org/10.1007/BF02595777

[16] GHOUDI, K., KULPERGER, R. J. and RÉMILLARD, B. (2001). A nonparametric test of serial independence for time series and residuals. *J. Multivariate Anal.* **79** 191–218. MR1868288 https://doi.org/10.1006/jmva.2000.1967

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

[18] HAN, Y. and WU, W. B. (2020). Test for high dimensional covariance matrices. *Ann. Statist.* **48** 3565–3588. MR4185820 https://doi.org/10.1214/20-AOS1943

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

[20] KOJADINOVIC, I. and HOLMES, M. (2009). Tests of independence among continuous random vectors based on Cramér–von Mises functionals of the empirical copula process. *J. Multivariate Anal.* **100** 1137–1154. MR2508377 https://doi.org/10.1016/j.jmva.2008.10.013

[21] LANCASTER, H. O. (1969). *The Chi-Squared Distribution*. Wiley, New York. MR0253452

[22] LEDOIT, O. and WOLF, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30** 1081–1102. MR1926169 https://doi.org/10.1214/aos/1031689018

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

[24] LIU, Z., PEACH, R. L., MEDIANO, P. A. M. and BARAHONA, M. (2023). Interaction measures, partition lattices and kernel tests for high-order interactions. ArXiv-preprint. Available at arXiv:2306.00904.

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

[26] RUBENSTEIN, P. K., CHWIALKOWSKI, K. P. and GRETTON, A. (2016). A kernel test for three-variable interactions with random processes. In *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence*. *UAI*'16 637–646. AUAI Press, Arlington, VA.

[27] SCHOTT, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika* **92** 951–956. MR2234197 https://doi.org/10.1093/biomet/92.4.951

[28] SEGERS, J. (2012). Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli* **18** 764–782. MR2948900 https://doi.org/10.3150/11-BEJ387

[29] SEJDINOVIC, D., GRETTON, A. and BERGSMA, W. (2013). A kernel test for three-variable interactions. In *Advances in Neural Information Processing Systems* (C. J. Burges, L. Bottou, M. Welling, Z. Ghahramani and K. Q. Weinberger, eds.) **26**. Curran Associates, Red Hook.

[30] SKLAR, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris* **8** 229–231. MR0125600

[31] STUTE, W. (1984). The oscillation behavior of empirical processes: The multivariate case. *Ann. Probab.* **12** 361–379. MR0735843

[32] YAO, S., ZHANG, X. and SHAO, X. (2018). Testing mutual independence in high dimension via distance covariance. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 455–480. MR3798874 https://doi.org/10.1111/rssb.12259

The Annals of Statistics
2024, Vol. 52, No. 1, 335–363
https://doi.org/10.1214/23-AOS2349

ADAPTIVE VARIATIONAL BAYES: OPTIMALITY, COMPUTATION AND APPLICATIONS

BY ILSANG OHN[1,a] AND LIZHEN LIN[2,b]

[1]*Department of Statistics, Inha University,* [a]*ilsang.ohn@inha.ac.kr*
[2]*Department of Mathematics, University of Maryland,* [b]*lizhen01@umd.edu*

In this paper, we explore adaptive inference based on variational Bayes. Although several studies have been conducted to analyze the contraction properties of variational posteriors, there is still a lack of a general and computationally tractable variational Bayes method that performs adaptive inference. To fill this gap, we propose a novel *adaptive variational Bayes* framework, which can operate on a collection of models. The proposed framework first computes a variational posterior over each individual model separately and then combines them with certain weights to produce a variational posterior over the entire model. It turns out that this combined variational posterior is the closest member to the posterior over the entire model in a predefined family of approximating distributions. We show that the adaptive variational Bayes attains optimal contraction rates adaptively under very general conditions. We also provide a methodology to maintain the tractability and adaptive optimality of the adaptive variational Bayes even in the presence of an enormous number of individual models, such as sparse models. We apply the general results to several examples, including deep learning and sparse factor models, and derive new and adaptive inference results. In addition, we characterize an implicit regularization effect of variational Bayes and show that the adaptive variational posterior can utilize this.

REFERENCES

ALQUIER, P. and RIDGWAY, J. (2020). Concentration of tempered posteriors and of their variational approximations. *Ann. Statist.* **48** 1475–1497. MR4124331 https://doi.org/10.1214/19-AOS1855

ALQUIER, P., RIDGWAY, J. and CHOPIN, N. (2016). On the properties of variational approximations of Gibbs posteriors. *J. Mach. Learn. Res.* **17** Paper No. 239, 41 pp. MR3595173

ARBEL, J., GAYRAUD, G. and ROUSSEAU, J. (2013). Bayesian optimal adaptive estimation using a sieve prior. *Scand. J. Stat.* **40** 549–570. MR3091697 https://doi.org/10.1002/sjos.12002

BABACAN, S. D., LUESSI, M., MOLINA, R. and KATSAGGELOS, A. K. (2012). Sparse Bayesian methods for low-rank matrix estimation. *IEEE Trans. Signal Process.* **60** 3964–3977. MR2960472 https://doi.org/10.1109/TSP.2012.2197748

BAI, J., SONG, Q. and CHENG, G. (2020). Efficient variational inference for sparse deep learning with theoretical guarantee. In *Proceedings of the* 34*th International Conference on Neural Information Processing Systems* **33** 466–476.

BELITSER, E. and GHOSAL, S. (2003). Adaptive Bayesian inference on the mean of an infinite-dimensional normal distribution. *Ann. Statist.* **31** 536–559. MR1983541 https://doi.org/10.1214/aos/1051027880

BHATTACHARYA, A., PATI, D. and YANG, Y. (2019). Bayesian fractional posteriors. *Ann. Statist.* **47** 39–66. MR3909926 https://doi.org/10.1214/18-AOS1712

BISHOP, C. M. (2006). *Pattern Recognition and Machine Learning. Information Science and Statistics.* Springer, New York. MR2247587 https://doi.org/10.1007/978-0-387-45528-0

CAI, T., MA, Z. and WU, Y. (2015). Optimal estimation and rank detection for sparse spiked covariance matrices. *Probab. Theory Related Fields* **161** 781–815. MR3334281 https://doi.org/10.1007/s00440-014-0562-z

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

MSC2020 subject classifications. Primary 62C10; secondary 62G20.
Key words and phrases. Variational Bayes, adaptive inference, posterior contraction rates, model selection consistency, deep neural networks, quasi-posteriors.

CHAE, M., LIN, L. and DUNSON, D. B. (2019). Bayesian sparse linear regression with unknown symmetric error. *Inf. Inference* **8** 621–653. MR3994400 https://doi.org/10.1093/imaiai/iay022

CHÉRIEF-ABDELLATIF, B.-E. (2019). Consistency of ELBO maximization for model selection. In *Symposium on Advances in Approximate Bayesian Inference*. *Proc. Mach. Learn. Res.* (*PMLR*) **96** 11–31. Proceedings of Machine Learning Research PMLR. MR3980812

CHÉRIEF-ABDELLATIF, B.-E. (2020). Convergence rates of variational inference in sparse deep learning. In *Proceedings of the* 37*th International Conference on Machine Learning* 1831–1842. PMLR.

CHÉRIEF-ABDELLATIF, B.-E. and ALQUIER, P. (2018). Consistency of variational Bayes inference for estimation and model selection in mixtures. *Electron. J. Stat.* **12** 2995–3035. MR3855643 https://doi.org/10.1214/18-EJS1475

FOONG, A., BURT, D., LI, Y. and TURNER, R. (2020). On the expressiveness of approximate inference in Bayesian neural networks. In *Proceedings of the* 34*th International Conference on Neural Information Processing Systems* **33** 15897–15908.

FORTUIN, V. (2022). Priors in Bayesian deep learning: A review. *Int. Stat. Rev.* **90** 563–591. MR4524825

GAO, C., VAN DER VAART, A. W. and ZHOU, H. H. (2020). A general framework for Bayes structured linear models. *Ann. Statist.* **48** 2848–2878. MR4152123 https://doi.org/10.1214/19-AOS1909

GENG, J., BHATTACHARYA, A. and PATI, D. (2019). Probabilistic community detection with unknown number of communities. *J. Amer. Statist. Assoc.* **114** 893–905. MR3963189 https://doi.org/10.1080/01621459.2018.1458618

GHOSAL, S., GHOSH, J. K. and VAN DER VAART, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531. MR1790007 https://doi.org/10.1214/aos/1016218228

GHOSAL, S., LEMBER, J. and VAN DER VAART, A. (2008). Nonparametric Bayesian model selection and averaging. *Electron. J. Stat.* **2** 63–89. MR2386086 https://doi.org/10.1214/07-EJS090

GHOSAL, S. and VAN DER VAART, A. (2007). Convergence rates of posterior distributions for non-i.i.d. observations. *Ann. Statist.* **35** 192–223. MR2332274 https://doi.org/10.1214/009053606000001172

HAN, Q. (2021). Oracle posterior contraction rates under hierarchical priors. *Electron. J. Stat.* **15** 1085–1153. MR4255298 https://doi.org/10.1214/21-ejs1811

HUANG, X., WANG, J. and LIANG, F. (2016). A variational algorithm for Bayesian variable selection. Preprint. Available at arXiv:1602.07640.

JIANG, S. and TOKDAR, S. T. (2021). Variable selection consistency of Gaussian process regression. *Ann. Statist.* **49** 2491–2505. MR4338372 https://doi.org/10.1214/20-aos2043

KOHLER, M. and LANGER, S. (2021). On the rate of convergence of fully connected deep neural network regression estimates. *Ann. Statist.* **49** 2231–2249. MR4319248 https://doi.org/10.1214/20-aos2034

KONG, I., YANG, D., LEE, J., OHN, I., BAEK, G. and KIM, Y. (2023). Masked Bayesian neural networks: Theoretical guarantee and its posterior inference. In *Proceedings of the* 40*th International Conference on Machine Learning* 17462–17491. PMLR.

LEE, K., CHAE, M. and LIN, L. (2021). Bayesian high-dimensional semi-parametric inference beyond sub-Gaussian errors. *J. Korean Statist. Soc.* **50** 511–527. MR4265000 https://doi.org/10.1007/s42952-020-00091-4

LEMBER, J. and VAN DER VAART, A. (2007). On universal Bayesian adaptation. *Statist. Decisions* **25** 127–152. MR2388859 https://doi.org/10.1524/stnd.2007.25.2.127

MILLER, J. W. and HARRISON, M. T. (2018). Mixture models with a prior on the number of components. *J. Amer. Statist. Assoc.* **113** 340–356. MR3803469 https://doi.org/10.1080/01621459.2016.1255636

NING, B. (2021). Spike and slab Bayesian sparse principal component analysis. Preprint. Available at arXiv:2102.00305.

OHN, I. and KIM, Y. (2022). Posterior consistency of factor dimensionality in high-dimensional sparse factor models. *Bayesian Anal.* **17** 491–514. MR4483228 https://doi.org/10.1214/21-ba1261

OHN, I. and LIN, L. (2023). Optimal Bayesian estimation of Gaussian mixtures with growing number of components. *Bernoulli* **29** 1195–1218. MR4550220 https://doi.org/10.3150/22-bej1495

OHN, I. and LIN, L. (2024). Supplement to "Adaptive variational Bayes: Optimality, computation and applications." https://doi.org/10.1214/23-AOS2349SUPP

PATI, D., BHATTACHARYA, A., PILLAI, N. S. and DUNSON, D. (2014). Posterior contraction in sparse Bayesian factor models for massive covariance matrices. *Ann. Statist.* **42** 1102–1130. MR3210997 https://doi.org/10.1214/14-AOS1215

PATI, D. BHATTACHARYA, A. and YANG, Y. (2018). On statistical optimality of variational Bayes. In *Proceedings of the* 21*st International Conference on Artificial Intelligence and Statistics* 1579–1588. PMLR.

RAY, K. and SZABÓ, B. (2022). Variational Bayes for high-dimensional linear regression with sparse priors. *J. Amer. Statist. Assoc.* **117** 1270–1281. MR4480711 https://doi.org/10.1080/01621459.2020.1847121

RAY, K., SZABÓ, B. and CLARA, G. (2020). Spike and slab variational Bayes for high dimensional logistic regression. In *Proceedings of the* 34*th International Conference on Neural Information Processing Systems* 14423–14434.

ROUSSEAU, J. and SZABO, B. (2017). Asymptotic behaviour of the empirical Bayes posteriors associated to maximum marginal likelihood estimator. *Ann. Statist.* **45** 833–865. MR3650402 https://doi.org/10.1214/16-AOS1469

SCHMIDT-HIEBER, J. (2020). Nonparametric regression using deep neural networks with ReLU activation function. *Ann. Statist.* **48** 1875–1897. MR4134774 https://doi.org/10.1214/19-AOS1875

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

SYRING, N. and MARTIN, R. (2023). Gibbs posterior concentration rates under sub-exponential type losses. *Bernoulli* **29** 1080–1108. MR4550216 https://doi.org/10.3150/22-bej1491

WALKER, S. and HJORT, N. L. (2001). On Bayesian consistency. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 811–821. MR1872068 https://doi.org/10.1111/1467-9868.00314

XIE, F., CAPE, J., PRIEBE, C. E. and XU, Y. (2022). Bayesian sparse spiked covariance model with a continuous matrix shrinkage prior. *Bayesian Anal.* **17** 1193–1217. MR4506026 https://doi.org/10.1214/21-ba1292

YANG, Y. and MARTIN, R. (2020). Variational approximations of empirical Bayes posteriors in high-dimensional linear models. Preprint. Available at arXiv:2007.15930.

YANG, Y. and PATI, D. (2017). Bayesian model selection consistency and oracle inequality with intractable marginal likelihood. Preprint. Available at arXiv:1701.00311.

YANG, Y., PATI, D. and BHATTACHARYA, A. (2020). α-variational inference with statistical guarantees. *Ann. Statist.* **48** 886–905. MR4102680 https://doi.org/10.1214/19-AOS1827

YANG, Y. and TOKDAR, S. T. (2015). Minimax-optimal nonparametric regression in high dimensions. *Ann. Statist.* **43** 652–674. MR3319139 https://doi.org/10.1214/14-AOS1289

YAO, D., XIE, F. and XU, Y. (2022). Bayesian sparse Gaussian mixture model in high dimensions. Preprint. Available at arXiv:2207.10301.

ZHANG, A. Y. and ZHOU, H. H. (2020). Theoretical and computational guarantees of mean field variational inference for community detection. *Ann. Statist.* **48** 2575–2598. MR4152113 https://doi.org/10.1214/19-AOS1898

ZHANG, F. and GAO, C. (2020). Convergence rates of variational posterior distributions. *Ann. Statist.* **48** 2180–2207. MR4134791 https://doi.org/10.1214/19-AOS1883

ZHANG, T. (2006). From ε-entropy to KL-entropy: Analysis of minimum information complexity density estimation. *Ann. Statist.* **34** 2180–2210. MR2291497 https://doi.org/10.1214/009053606000000704

The Annals of Statistics
2024, Vol. 52, No. 1, 364–391
https://doi.org/10.1214/23-AOS2350

RANK AND FACTOR LOADINGS ESTIMATION IN TIME SERIES TENSOR FACTOR MODEL BY PRE-AVERAGING

BY WEILIN CHEN[a] AND CLIFFORD LAM[b]

Department of Statistics, London School of Economics and Political Science, [a]*w.chen56@lse.ac.uk,* [b]*c.lam2@lse.ac.uk*

The idiosyncratic components of a tensor time series factor model can exhibit serial correlations, (e.g., finance or economic data), ruling out many state-of-the-art methods that assume white/independent idiosyncratic components. While the traditional higher order orthogonal iteration (HOOI) is proved to be convergent to a set of factor loading matrices, the closeness of them to the true underlying factor loading matrices are in general not established, or only under i.i.d. Gaussian noises. Under the presence of serial and cross-correlations in the idiosyncratic components and time series variables with only bounded fourth-order moments, for tensor time series data with tensor order two or above, we propose a pre-averaging procedure that can be considered a random projection method. The estimated directions corresponding to the strongest factors are then used for projecting the data for a potentially improved re-estimation of the factor loading spaces themselves, with theoretical guarantees and rate of convergence spelt out when not all factors are pervasive. We also propose a new rank estimation method, which utilizes correlation information from the projected data. Extensive simulations are performed and compared to other state-of-the-art or traditional alternatives. A set of tensor-valued NYC taxi data is also analyzed.

REFERENCES

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

BAI, J. and NG, S. (2023). Approximate factor models with weaker loadings. *J. Econometrics* **235** 1893–1916. MR4602937 https://doi.org/10.1016/j.jeconom.2023.01.027

BARIGOZZI, M., HE, Y., LI, L. and TRAPANI, L. (2023). Robust estimation of large factor models for tensor-valued time series. Available at arXiv:2303.18163.

CHEN, E. Y. and FAN, J. (2023). Statistical inference for high-dimensional matrix-variate factor models. *J. Amer. Statist. Assoc.* **118** 1038–1055. MR4595475 https://doi.org/10.1080/01621459.2021.1970569

CHEN, W. and LAM, C. (2024). Supplement to "Rank and factor loadings estimation in time series tensor factor model by pre-averaging." https://doi.org/10.1214/23-AOS2350SUPPA, https://doi.org/10.1214/23-AOS2350SUPPB

CHEN, E. Y., XIA, D., CAI, C. and FAN, J. (2024). Semi-parametric tensor factor analysis by iteratively projected singular value decomposition. *J. Roy. Stat. Soc. Ser. B Stat. Methodol.* https://doi.org/10.1093/jrsssb/qkae001

CHEN, E. Y., YUN, X., CHEN, R. and YAO, Q. (2020). Modeling multivariate spatial-temporal data with latent low-dimensional dynamics. Preprint. Available at arXiv:2002.01305.

CHEN, R., HAN, Y., LI, Z., XIAO, H., YANG, D. and YU, R. (2022). Analysis of tensor time series: TensorTS. Submitted.

CHEN, R., YANG, D. and ZHANG, C.-H. (2022). Factor models for high-dimensional tensor time series. *J. Amer. Statist. Assoc.* **117** 94–116. MR4399070 https://doi.org/10.1080/01621459.2021.1912757

CHENG, C., WEI, Y. and CHEN, Y. (2021). Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise. *IEEE Trans. Inf. Theory* **67** 7380–7419. MR4345128 https://doi.org/10.1109/TIT.2021.3111828

FAN, J., GUO, J. and ZHENG, S. (2022). Estimating number of factors by adjusted eigenvalues thresholding. *J. Amer. Statist. Assoc.* **117** 852–861. MR4436317 https://doi.org/10.1080/01621459.2020.1825448

MSC2020 subject classifications. 62H25, 62H12.

Key words and phrases. Core rank tensor, tensor fibres pre-averaging, strongest factors projection, iterative projection algorithm, bootstrap tensor fibres.

FREYALDENHOVEN, S. (2022). Factor models with local factors—determining the number of relevant factors. *J. Econometrics* **229** 80–102. MR4414013 https://doi.org/10.1016/j.jeconom.2021.04.006

HAN, Y., CHEN, R., YANG, D. and ZHANG, C.-H. (2020). Tensor factor model estimation by iterative projection. Methodology.

HAN, Y., CHEN, R. and ZHANG, C.-H. (2022). Rank determination in tensor factor model. *Electron. J. Stat.* **16** 1726–1803. MR4393792 https://doi.org/10.1214/22-ejs1991

HAN, Y. and ZHANG, C.-H. (2023). Tensor principal component analysis in high dimensional CP models. *IEEE Trans. Inf. Theory* **69** 1147–1167. MR4564648

HAN, Y., ZHANG, C.-H. and CHEN, R. (2022). CP factor model for dynamic tensors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* Under revision.

HARTIGAN, J. A. (2014). Bounding the maximum of dependent random variables. *Electron. J. Stat.* **8** 3126–3140. MR3301303 https://doi.org/10.1214/14-EJS974

HE, Y., KONG, X., YU, L., ZHANG, X. and ZHAO, C. (2024). Matrix factor analysis: From least squares to iterative projection. *J. Bus. Econom. Statist.* **42** 322–334. MR4684510 https://doi.org/10.1080/07350015.2023.2191676

HE, Y., KONG, X.-B., LIU, D. and ZHAO, R. (2023). Robust statistical inference for large-dimensional matrix-valued time series via iterative Huber regression. Available at arXiv:2306.03317.

HE, Y., LI, L. and TRAPANI, L. (2022). Statistical inference for large-dimensional tensor factor model by weighted/unweighted projection. Available at arXiv:2206.09800.

HE, Y., WANG, Y., YU, L., ZHOU, W. and ZHOU, W.-X. (2022). Matrix Kendall's tau in high-dimensions: A robust statistic for matrix factor model. Available at arXiv:2207.09633.

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

LAM, C., YAO, Q. and BATHIA, N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika* **98** 901–918. MR2860332 https://doi.org/10.1093/biomet/asr048

LIU, T., YUAN, M. and ZHAO, H. (2022). Characterizing spatiotemporal transcriptome of the human brain via low-rank tensor decomposition. *Stat. Biosci.*. https://doi.org/10.1007/s12561-021-09331-5

LIU, X. and CHEN, E. Y. (2022). Identification and estimation of threshold matrix-variate factor models. *Scand. J. Stat.* **49** 1383–1417. MR4471290

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

TAO, M., SU, J. and WANG, L. (2019). Land cover classification of PolSAR image using tensor representation and learning. *J. Appl. Remote Sens.* **13** 016516. https://doi.org/10.1117/1.JRS.13.016516

WANG, D., LIU, X. and CHEN, R. (2019). Factor models for matrix-valued high-dimensional time series. *J. Econometrics* **208** 231–248. MR3906969 https://doi.org/10.1016/j.jeconom.2018.09.013

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

YU, L., HE, Y., KONG, X. and ZHANG, X. (2022). Projected estimation for large-dimensional matrix factor models. *J. Econometrics* **229** 201–217. MR4414019 https://doi.org/10.1016/j.jeconom.2021.04.001

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

The Annals of Statistics
2024, Vol. 52, No. 1, 392–411
https://doi.org/10.1214/24-AOS2355

ESTIMATION AND INFERENCE FOR MINIMIZER AND MINIMUM OF CONVEX FUNCTIONS: OPTIMALITY, ADAPTIVITY AND UNCERTAINTY PRINCIPLES

BY T. TONY CAI[a], RAN CHEN[b] AND YUANCHENG ZHU[c]

Department of Statistics and Data Science, The Wharton School, University Of Pennsylvania, [a]*tcai@wharton.upenn.edu,* [b]*ran1chen@wharton.upenn.edu,* [c]*yuancheng.zhu@gmail.com*

Optimal estimation and inference for both the minimizer and minimum of a convex regression function under the white noise and nonparametric regression models are studied in a nonasymptotic local minimax framework, where the performance of a procedure is evaluated at individual functions. Fully adaptive and computationally efficient algorithms are proposed and sharp minimax lower bounds are given for both the estimation accuracy and expected length of confidence intervals for the minimizer and minimum.

The nonasymptotic local minimax framework brings out new phenomena in simultaneous estimation and inference for the minimizer and minimum. We establish a novel uncertainty principle that provides a fundamental limit on how well the minimizer and minimum can be estimated simultaneously for any convex regression function. A similar result holds for the expected length of the confidence intervals for the minimizer and minimum.

REFERENCES

AGARWAL, A., FOSTER, D. P., HSU, D. J., KAKADE, S. M. and RAKHLIN, A. (2011). Stochastic convex optimization with bandit feedback. In *Advances in Neural Information Processing Systems* (J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira and K. Q. Weinberger, eds.) **24**. Curran Associates, Red Hook.

AUER, P., ORTNER, R. and SZEPESVÁRI, C. (2007). Improved rates for the stochastic continuum-armed bandit problem. In *Learning Theory*. *Lecture Notes in Computer Science* **4539** 454–468. Springer, Berlin. MR2397605 https://doi.org/10.1007/978-3-540-72927-3_33

BELITSER, E., GHOSAL, S. and VAN ZANTEN, H. (2012). Optimal two-stage procedures for estimating location and size of the maximum of a multivariate regression function. *Ann. Statist.* **40** 2850–2876. MR3097962 https://doi.org/10.1214/12-AOS1053

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

BLUM, J. R. (1954). Multidimensional stochastic approximation methods. *Ann. Math. Stat.* **25** 737–744. MR0065092 https://doi.org/10.1214/aoms/1177728659

CAI, T. T., CHEN, R. and ZHU, Y. (2024). Supplement to "Estimation and Inference for Minimizer and Minimum of Convex Functions: Optimality, Adaptivity and Uncertainty Principles." https://doi.org/10.1214/24-AOS2355SUPP

CAI, T. T. and LOW, M. G. (2015). A framework for estimation of convex functions. *Statist. Sinica* **25** 423–456. MR3379081

CAI, T. T., LOW, M. G. and XIA, Y. (2013). Adaptive confidence intervals for regression functions under shape constraints. *Ann. Statist.* **41** 722–750. MR3099119 https://doi.org/10.1214/12-AOS1068

CHATTERJEE, S., DUCHI, J. C., LAFFERTY, J. and ZHU, Y. (2016). Local minimax complexity of stochastic convex optimization. In *Advances in Neural Information Processing Systems* (D. Lee, M. Sugiyama, U. Luxburg, I. Guyon and R. Garnett, eds.) **29**. Curran Associates, Red Hook.

CHEN, H. (1988). Lower rate of convergence for locating a maximum of a function. *Ann. Statist.* **16** 1330–1334. MR0959206 https://doi.org/10.1214/aos/1176350965

CHEN, H., HUANG, M.-N. L. and HUANG, W.-J. (1996). Estimation of the location of the maximum of a regression function using extreme order statistics. *J. Multivariate Anal.* **57** 191–214. MR1391168 https://doi.org/10.1006/jmva.1996.0029

MSC2020 subject classifications. Primary 62G08; secondary 62G99, 62G20.

Key words and phrases. Adaptivity, confidence interval, nonparametric regression, minimax optimality, modulus of continuity, uncertainty principle, white noise model.

DENG, H., HAN, Q. and SEN, B. (2023). Inference for local parameters in convexity constrained models. *J. Amer. Statist. Assoc.* **118** 2721–2735. MR4681616 https://doi.org/10.1080/01621459.2022.2071721

DIPPON, J. (2003). Accelerated randomized stochastic optimization. *Ann. Statist.* **31** 1260–1281. MR2001650 https://doi.org/10.1214/aos/1059655913

DÜMBGEN, L. (1998). New goodness-of-fit tests and their application to nonparametric confidence sets. *Ann. Statist.* **26** 288–314. MR1611768 https://doi.org/10.1214/aos/1030563987

FACER, M. R. and MÜLLER, H.-G. (2003). Nonparametric estimation of the location of a maximum in a response surface. *J. Multivariate Anal.* **87** 191–217. MR2007268 https://doi.org/10.1016/S0047-259X(03)00030-7

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

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

HENGARTNER, N. W. and STARK, P. B. (1995). Finite-sample confidence envelopes for shape-restricted densities. *Ann. Statist.* **23** 525–550. MR1332580 https://doi.org/10.1214/aos/1176324534

KIEFER, J. (1982). Optimum rates for non-parametric density and regression estimates under order restrictions. *Stat. Probab. Essays Honor CR Rao* **419** 428.

KIEFER, J. and WOLFOWITZ, J. (1952). Stochastic estimation of the maximum of a regression function. *Ann. Math. Stat.* **23** 462–466. MR0050243 https://doi.org/10.1214/aoms/1177729392

KLEINBERG, R. (2004). Nearly tight bounds for the continuum-armed bandit problem. In *Proceedings of the* 17*th International Conference on Neural Information Processing Systems*, *NIPS'*04 697–704. MIT Press, Cambridge, MA, USA.

KLEINBERG, R., SLIVKINS, A. and UPFAL, E. (2019). Bandits and experts in metric spaces. *J. ACM* **66** 30. MR3962354 https://doi.org/10.1145/3299873

MOKKADEM, A. and PELLETIER, M. (2007). A companion for the Kiefer–Wolfowitz–Blum stochastic approximation algorithm. *Ann. Statist.* **35** 1749–1772. MR2351104 https://doi.org/10.1214/009053606000001451

MÜLLER, H.-G. (1989). Adaptive nonparametric peak estimation. *Ann. Statist.* **17** 1053–1069. MR1015137 https://doi.org/10.1214/aos/1176347255

POLYAK, B. T. and TSYBAKOV, A. B. (1990). Optimal orders of accuracy for search algorithms of stochastic optimization. *Problemy Peredachi Informatsii* **26** 45–53. MR1074128

SHOUNG, J.-M. and ZHANG, C.-H. (2001). Least squares estimators of the mode of a unimodal regression function. *Ann. Statist.* **29** 648–665. MR1865335 https://doi.org/10.1214/aos/1009210684

The Annals of Statistics
2024, Vol. 52, No. 1, 412
https://doi.org/10.1214/23-AOS2344
Main article: https://doi.org/10.1214/12-AOS1067

ERRATUM: IMPROVED MULTIVARIATE NORMAL MEAN ESTIMATION WITH UNKNOWN COVARIANCE WHEN *p* IS GREATER THAN *n*

BY DIDIER CHÉTELAT[a]

Polytechnique Montréal, [a]*didier.chetelat@polymtl.ca*

This erratum offers a correction to Chételat and Wells ((2012) *Ann. Statist.* **40** 3137–3160), following the note of Foroushani and Nkurunziza ((2023) arXiv:2311.13140).

REFERENCES

CHÉTELAT, D. and WELLS, M. T. (2012). Improved multivariate normal mean estimation with unknown covariance when p is greater than n. *Ann. Statist.* **40** 3137–3160. MR3097972 https://doi.org/10.1214/12-AOS1067

FOROUSHANI, A. A. and NKURUNZIZA, S. (2023). A note on improved multivariate normal mean estimation with unknown covariance when p is greater than n. arXiv preprint. Available at arXiv:2311.13140.

MSC2020 subject classifications. Primary 62F10; secondary 62C20, 62H12.
Key words and phrases. Minimax estimation, James–Stein estimator, high-dimensional inference.

The Annals of Statistics
2024, Vol. 52, No. 1, 413–414
https://doi.org/10.1214/24-AOS2351
Main article: https://doi.org/10.1214/19-AOS1875

CORRECTION TO "NONPARAMETRIC REGRESSION USING DEEP NEURAL NETWORKS WITH RELU ACTIVATION FUNCTION"

BY JOHANNES SCHMIDT-HIEBER[1,a] AND DON VU[2,b]

[1]*Department of Applied Mathematics, University of Twente,* [a]*a.j.schmidt-hieber@utwente.nl*
[2]*Department of Mathematics, Vrije Universiteit Amsterdam,* [b]*don@abnormalstatistics.nl*

REFERENCES

[1] SCHMIDT-HIEBER, J. (2020). Nonparametric regression using deep neural networks with ReLU activation function. *Ann*. *Statist*. **48** 1875–1897. MR4134774 https://doi.org/10.1214/19-AOS1875