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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

Erratum

Vol. 54, No. 3—June 2026

THE ANNALS OF STATISTICS Vol. 54, No. 3, pp. 1101–1653 June 2026

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

The purpose of the Institute is to foster the development and dissemination of the theory and applications of statistics and probability.

IMS OFFICERS

IMS EDITORS

The Annals of Statistics **[ISSN 0090-5364 (print); ISSN 2168-8966 (online)],** Volume 54, Number 3, June 2026. 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
2026, Vol. 54, No. 3, 1101–1127
https://doi.org/10.1214/25-AOS2551

SPARSE PCA: A NEW SCALABLE ESTIMATOR BASED ON INTEGER PROGRAMMING

BY KAYHAN BEHDIN[1,a] AND RAHUL MAZUMDER[1,2,b]

[1]*Operations Research Center, Massachusetts Institute of Technology*
[2]*Sloan School of Management, Massachusetts Institute of Technology,* [a]*behdink@mit.edu,* [b]*rahulmaz@mit.edu*

We consider the sparse principal component analysis (SPCA) problem under the well-known spiked covariance model. Recent work has shown that the SPCA problem can be reformulated as a mixed integer program (MIP) and can be solved to global optimality, leading to estimators that are known to enjoy optimal statistical properties. However, prior MIP algorithms for SPCA appear to be limited in terms of scalability to up to a thousand features or so. In this paper we propose a new estimator for SPCA which can be formulated as a MIP. Different from earlier work, we make use of the underlying spiked covariance model and properties of the multivariate Gaussian distribution to arrive at our estimator. We establish statistical guarantees for our proposed estimator in terms of estimation error and support recovery. We derive guarantees under departures from the spiked covariance model and for approximate solutions to the optimization problem. We propose a custom algorithm to solve the MIP, which scales better than off-the-shelf solvers, and demonstrate that our approach can be much more computationally attractive compared to earlier exact MIP-based approaches for the SPCA problem. Our numerical experiments on synthetic and real datasets show that our algorithms can address problems with up to 20,000 features in minutes; and generally result in favorable statistical properties compared to existing popular approaches for SPCA.

REFERENCES

[1] AKTÜRK, M. S., ATAMTÜRK, A. and GÜREL, S. (2009). A strong conic quadratic reformulation for machine-job assignment with controllable processing times. *Oper. Res. Lett.* **37** 187–191. MR2528379 https://doi.org/10.1016/j.orl.2008.12.009

[2] AMINI, A. A. and WAINWRIGHT, M. J. (2009). High-dimensional analysis of semidefinite relaxations for sparse principal components. *Ann. Statist.* **37** 2877–2921. MR2541450 https://doi.org/10.1214/08-AOS664

[3] BEHDIN, K. and MAZUMDER, R. (2024). Sparse NMF with archetypal regularization: Computational and robustness properties. *J. Mach. Learn. Res.* **25** Paper No. [36], 62. MR4723886

[4] BEHDIN, K. and MAZUMDER, R. (2026). Supplement to "Sparse PCA: A new scalable estimator based on integer programming." https://doi.org/10.1214/25-AOS2551SUPP

[5] BERK, L. and BERTSIMAS, D. (2019). Certifiably optimal sparse principal component analysis. *Math. Program. Comput.* **11** 381–420. MR3989322 https://doi.org/10.1007/s12532-018-0153-6

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

[7] BERTSIMAS, D., CORY-WRIGHT, R. and PAUPHILET, J. (2022). Solving large-scale sparse PCA to certifiable (near) optimality. *J. Mach. Learn. Res.* **23** Paper No. [13], 35. MR4420738

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

[9] BERTSIMAS, D. and VAN PARYS, B. (2020). Sparse high-dimensional regression: Exact scalable algorithms and phase transitions. *Ann. Statist.* **48** 300–323. MR4065163 https://doi.org/10.1214/18-AOS1804

[10] BRESLER, G., PARK, S. M. and PERSU, M. (2018). Sparse PCA from sparse linear regression. In *Proceedings of the* 32*nd International Conference on Neural Information Processing Systems*. *NIPS'*18 10965–10975. Curran Associates, Red, Hook, NY, USA.

MSC2020 subject classifications. 62H25.
Key words and phrases. Sparse PCA, mixed integer programming, spiked covariance model.

[11] 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
[12] CORNUÉJOLS, G. (2008). Valid inequalities for mixed integer linear programs. *Math. Program.* **112** 3–44. MR2327000 https://doi.org/10.1007/s10107-006-0086-0
[13] D'ASPREMONT, A., BACH, F. and EL GHAOUI, L. (2008). Optimal solutions for sparse principal component analysis. *J. Mach. Learn. Res.* **9** 1269–1294. MR2426043
[14] D'ASPREMONT, A., EL GHAOUI, L., JORDAN, M. I. and LANCKRIET, G. R. G. (2007). A direct formulation for sparse PCA using semidefinite programming. *SIAM Rev.* **49** 434–448. MR2353806 https://doi.org/10.1137/050645506
[15] DESHPANDE, Y. and MONTANARI, A. (2016). Sparse PCA via covariance thresholding. *J. Mach. Learn. Res.* **17** Paper No. 141, 41. MR3555032
[16] DEY, S. S., MAZUMDER, R. and WANG, G. (2018). A convex integer programming approach for optimal sparse PCA. Preprint. Available at arXiv:1810.09062.
[17] DURAN, M. A. and GROSSMANN, I. E. (1986). An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Math. Program.* **36** 307–339. MR0866413 https://doi.org/10.1007/BF02592064
[18] FRANGIONI, A. and GENTILE, C. (2006). Perspective cuts for a class of convex 0-1 mixed integer programs. *Math. Program.* **106** 225–236. MR2208082 https://doi.org/10.1007/s10107-005-0594-3
[19] GATARIC, M., WANG, T. and SAMWORTH, R. J. (2020). Sparse principal component analysis via axis-aligned random projections. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 329–359. MR4084167 https://doi.org/10.1111/rssb.12360
[20] GÜNLÜK, O. and LINDEROTH, J. (2010). Perspective reformulations of mixed integer nonlinear programs with indicator variables. *Math. Program.* **124** 183–205. MR2679988 https://doi.org/10.1007/s10107-010-0360-z
[21] GUYON, I., GUNN, S., BEN-HUR, A. and DROR, G. (2005). Result analysis of the NIPS 2003 feature selection challenge. In *Advances in Neural Information Processing Systems* **17**. MIT Press, Cambridge.
[22] HASTIE, T., TIBSHIRANI, R. and WAINWRIGHT, M. (2015). *Statistical Learning with Sparsity*: *The Lasso and Generalizations*. *Monographs on Statistics and Applied Probability* **143**. CRC Press, Boca Raton, FL. MR3616141
[23] HAZIMEH, H. and MAZUMDER, R. (2020). Fast best subset selection: Coordinate descent and local combinatorial optimization algorithms. *Oper. Res.* **68** 1517–1537. MR4166310
[24] HAZIMEH, H., MAZUMDER, R. and SAAB, A. (2022). Sparse regression at scale: Branch-and-bound rooted in first-order optimization. *Math. Program.* **196** 347–388. MR4506808 https://doi.org/10.1007/s10107-021-01712-4
[25] HOTELLING, H. (1933). Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **24** 417.
[26] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961 https://doi.org/10.1214/aos/1009210544
[27] JOHNSTONE, I. M. and LU, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* **104** 682–693. MR2751448 https://doi.org/10.1198/jasa.2009.0121
[28] JOLLIFFE, I. T., TRENDAFILOV, N. T. and UDDIN, M. (2003). A modified principal component technique based on the LASSO. *J. Comput. Graph. Statist.* **12** 531–547. MR2002634 https://doi.org/10.1198/1061860032148
[29] KRAUTHGAMER, R., NADLER, B. and VILENCHIK, D. (2015). Do semidefinite relaxations solve sparse PCA up to the information limit? *Ann. Statist.* **43** 1300–1322. MR3346704 https://doi.org/10.1214/15-AOS1310
[30] LEI, J. and VU, V. Q. (2015). Sparsistency and agnostic inference in sparse PCA. *Ann. Statist.* **43** 299–322. MR3311861 https://doi.org/10.1214/14-AOS1273
[31] LI, Y. and XIE, W. (2025). Exact and approximation algorithms for sparse principal component analysis. *INFORMS J. Comput.* **37** 582–602. MR4923940
[32] LUSS, R. and TEBOULLE, M. (2013). Conditional gradient algorithms for rank-one matrix approximations with a sparsity constraint. *SIAM Rev.* **55** 65–98. MR3032839 https://doi.org/10.1137/110839072
[33] 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
[34] MOGHADDAM, B., WEISS, Y. and AVIDAN, S. (2006). Spectral bounds for sparse PCA: Exact and greedy algorithms. In *Advances in Neural Information Processing Systems* 915–922.
[35] PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17** 1617–1642. MR2399865

[36] RICHTÁRIK, P., JAHANI, M., AHIPAŞAOĞLU, S. D. and TAKÁČ, M. (2021). Alternating maximization: Unifying framework for 8 sparse PCA formulations and efficient parallel codes. *Optim. Eng.* **22** 1493–1519. MR4322709 https://doi.org/10.1007/s11081-020-09562-3

[37] STEWART, G. W. and SUN, J. G. (1990). *Matrix Perturbation Theory*. *Computer Science and Scientific Computing*. Academic Press, Boston, MA. MR1061154

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

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

[40] WITTEN, D. M., TIBSHIRANI, R., and HASTIE, T., (2009). A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* **10** 515–534.

[41] YUAN, X.-T. and ZHANG, T. (2013). Truncated power method for sparse eigenvalue problems. *J. Mach. Learn. Res.* **14** 899–925. MR3063614

[42] ZOU, H., HASTIE, T. and TIBSHIRANI, R. (2006). Sparse principal component analysis. *J. Comput. Graph. Statist.* **15** 265–286. MR2252527 https://doi.org/10.1198/106186006X113430

The Annals of Statistics
2026, Vol. 54, No. 3, 1128–1152
https://doi.org/10.1214/25-AOS2574

ADAPTIVE ROBUST CONFIDENCE INTERVALS

BY YUETIAN LUO[1,a] AND CHAO GAO[2,b]

[1]*Data Science Institute, University of Chicago,* [a]*yuetian@uchicago.edu*
[2]*Department of Statistics, University of Chicago,* [b]*chaogao@uchicago.edu*

This paper studies the construction of adaptive confidence intervals under Huber's contamination model when the contamination proportion is unknown. For the robust confidence interval of a Gaussian mean, we show that the optimal length of an adaptive interval must be exponentially wider than that of a nonadaptive one. An optimal construction is achieved through simultaneous uncertainty quantification of quantiles at all levels. The results are further extended beyond the Gaussian location model by addressing a general family of robust hypothesis testing. In contrast to adaptive robust estimation, our findings reveal that the optimal length of an adaptive robust confidence interval critically depends on the distribution's shape.

REFERENCES

BALAKRISHNAN, S., DU, S. S., LI, J. and SINGH, A. (2017). Computationally efficient robust sparse estimation in high dimensions. In *Conference on Learning Theory* 169–212. PMLR.

BALCAN, M.-F., BLUM, A. and VEMPALA, S. (2008). A discriminative framework for clustering via similarity functions. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing* 671–680. ACM, New York. MR2582690 https://doi.org/10.1145/1374376.1374474

BARAUD, Y. (2004). Confidence balls in Gaussian regression. *Ann. Statist.* **32** 528–551. MR2060168 https://doi.org/10.1214/009053604000000085

BATENI, A.-H. and DALALYAN, A. S. (2020). Confidence regions and minimax rates in outlier-robust estimation on the probability simplex. *Electron. J. Stat.* **14** 2653–2677. MR4124558 https://doi.org/10.1214/20-EJS1731

BERAN, R. (1977). Robust location estimates. *Ann. Statist.* **5** 431–444. MR0448699

BHATT, S., FANG, G., LI, P. and SAMORODNITSKY, G. (2022). Minimax m-estimation under adversarial contamination. In *International Conference on Machine Learning* 1906–1924. PMLR.

BICKEL, P. J. (1965). On some robust estimates of location. *Ann. Math. Statist.* **36** 847–858. MR0177484 https://doi.org/10.1214/aoms/1177700058

BIRGÉ, L. (1979). Sur un théoreme de minimax et son application aux tests. Univ. de Paris-Sud, Dép. de Mathématique.

BULL, A. D. and NICKL, R. (2013). Adaptive confidence sets in L^2. *Probab. Theory Related Fields* **156** 889–919. MR3078289 https://doi.org/10.1007/s00440-012-0446-z

CAI, T. T. and GUO, Z. (2017). Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *Ann. Statist.* **45** 615–646. MR3650395 https://doi.org/10.1214/16-AOS1461

CAI, T. T., LOW, M. and MA, Z. (2014). Adaptive confidence bands for nonparametric regression functions. *J. Amer. Statist. Assoc.* **109** 1054–1070. MR3265680 https://doi.org/10.1080/01621459.2013.879260

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. and LOW, M. G. (2006). Adaptive confidence balls. *Ann. Statist.* **34** 202–228. MR2275240 https://doi.org/10.1214/009053606000000146

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

CANONNE, C., HOPKINS, S. B., LI, J., LIU, A. and NARAYANAN, S. (2023). The full landscape of robust mean testing: Sharp separations between oblivious and adaptive contamination. In 2023 *IEEE* 64*th Annual Symposium on Foundations of Computer Science—FOCS* 2023 2159–2168. IEEE Computer Soc., Los Alamitos, CA. MR4720375 https://doi.org/10.1109/FOCS57990.2023.00133

MSC2020 subject classifications. Primary 62F35, 62F03; secondary 62C20, 62C25.

Key words and phrases. Adaptivity, confidence intervals, Huber's contamination model, location model, minimax hypothesis testing.

CARPENTIER, A. (2013). Honest and adaptive confidence sets in L_p. *Electron. J. Stat.* **7** 2875–2923. MR3148371 https://doi.org/10.1214/13-EJS867
CARPENTIER, A., KLOPP, O., LÖFFLER, M. and NICKL, R. (2018). Adaptive confidence sets for matrix completion. *Bernoulli* **24** 2429–2460. MR3779691 https://doi.org/10.3150/17-BEJ933
CARPENTIER, A. and NICKL, R. (2015). On signal detection and confidence sets for low rank inference problems. *Electron. J. Stat.* **9** 2675–2688. MR3432430 https://doi.org/10.1214/15-EJS1087
CATONI, O. (2012). Challenging the empirical mean and empirical variance: A deviation study. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 1148–1185. MR3052407 https://doi.org/10.1214/11-AIHP454
CHEN, M., GAO, C. and REN, Z. (2016). A general decision theory for Huber's ϵ-contamination model. *Electron. J. Stat.* **10** 3752–3774. MR3579675 https://doi.org/10.1214/16-EJS1216
CHEN, M., GAO, C. and REN, Z. (2018). Robust covariance and scatter matrix estimation under Huber's contamination model. *Ann. Statist.* **46** 1932–1960. MR3845006 https://doi.org/10.1214/17-AOS1607
CHENG, Y., DIAKONIKOLAS, I. and GE, R. (2019). High-dimensional robust mean estimation in nearly-linear time. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms* 2755–2771. SIAM, Philadelphia, PA. MR3909640 https://doi.org/10.1137/1.9781611975482.171
CHENG, Y., DIAKONIKOLAS, I., GE, R. and WOODRUFF, D. P. (2019). Faster algorithms for high-dimensional robust covariance estimation. In *Conference on Learning Theory* 727–757. PMLR.
CHERAPANAMJERI, Y., FLAMMARION, N. and BARTLETT, P. L. (2019). Fast mean estimation with sub-Gaussian rates. In *Conference on Learning Theory* 786–806. PMLR.
CHERAPANAMJERI, Y., HOPKINS, S. B., KATHURIA, T., RAGHAVENDRA, P. and TRIPURANENI, N. (2020). Algorithms for heavy-tailed statistics: Regression, covariance estimation, and beyond. In *STOC '20—Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing* 601–609. ACM, New York. MR4141785
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
COLLINS, J. R. (1976). Robust estimation of a location parameter in the presence of asymmetry. *Ann. Statist.* **4** 68–85. MR0400538
CROW, E. L. and SIDDIQUI, M. M. (1967). Robust estimation of location. *J. Amer. Statist. Assoc.* **62** 353–389. MR0212953
DALALYAN, A. S. and MINASYAN, A. (2022). All-in-one robust estimator of the Gaussian mean. *Ann. Statist.* **50** 1193–1219. MR4404933 https://doi.org/10.1214/21-aos2145
DEPERSIN, J. and LECUÉ, G. (2022). Robust sub-Gaussian estimation of a mean vector in nearly linear time. *Ann. Statist.* **50** 511–536. MR4382026 https://doi.org/10.1214/21-aos2118
DIAKONIKOLAS, I., HOPKINS, S. B., KANE, D. and KARMALKAR, S. (2020). Robustly learning any clusterable mixture of Gaussians. arXiv preprint. Available at arXiv:2005.06417.
DIAKONIKOLAS, I., KAMATH, G., KANE, D. M., LI, J., MOITRA, A. and STEWART, A. (2016). Robust estimators in high dimensions without the computational intractability. In *57th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2016 655–664. IEEE Computer Soc., Los Alamitos, CA. MR3631028 https://doi.org/10.1109/FOCS.2016.85
DIAKONIKOLAS, I., KANE, D., PENSIA, A. and PITTAS, T. (2024). Near-optimal algorithms for Gaussians with Huber contamination: Mean estimation and linear regression. *Adv. Neural Inf. Process. Syst.* **36** 43384–43422.
DIAKONIKOLAS, I. and KANE, D. M. (2021). The sample complexity of robust covariance testing. In *Conference on Learning Theory* 1511–1521. PMLR.
DIAKONIKOLAS, I. and KANE, D. M. (2023). *Algorithmic High-Dimensional Robust Statistics*. Cambridge Univ. Press, Cambridge.
DIAKONIKOLAS, I., KANE, D. M. and STEWART, A. (2017). Statistical query lower bounds for robust estimation of high-dimensional Gaussians and Gaussian mixtures (extended abstract). In *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 73–84. IEEE Computer Soc., Los Alamitos, CA. MR3734219 https://doi.org/10.1109/FOCS.2017.16
DIAKONIKOLAS, I., KONG, W. and STEWART, A. (2019). Efficient algorithms and lower bounds for robust linear regression. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms* 2745–2754. SIAM, Philadelphia, PA. MR3909639 https://doi.org/10.1137/1.9781611975482.170
DÜMBGEN, L. (2003). Optimal confidence bands for shape-restricted curves. *Bernoulli* **9** 423–449. MR1997491 https://doi.org/10.3150/bj/1065444812
EFRON, B. and TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap*. *Monographs on Statistics and Applied Probability* **57**. CRC Press, New York. MR1270903 https://doi.org/10.1007/978-1-4899-4541-9
EL KAROUI, N., BEAN, D., BICKEL, P. J., LIM, C. and YU, B. (2013). On robust regression with high-dimensional predictors. *Proc. Natl. Acad. Sci. USA* **110** 14557–14562.
GAO, C. (2020). Robust regression via mutivariate regression depth. *Bernoulli* **26** 1139–1170. MR4058363 https://doi.org/10.3150/19-BEJ1144

GENOVESE, C. and WASSERMAN, L. (2008). Adaptive confidence bands. *Ann. Statist.* **36** 875–905. MR2396818 https://doi.org/10.1214/07-AOS500
GINÉ, E. and NICKL, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38** 1122–1170. MR2604707 https://doi.org/10.1214/09-AOS738
GINÉ, E. and NICKL, R. (2015). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge Univ. Press, New York. MR3588285 https://doi.org/10.1017/CBO9781107337862
GUPTA, S., HOPKINS, S. and PRICE, E. (2024). Beyond Catoni: Sharper rates for heavy-tailed and robust mean estimation. In *The Thirty Seventh Annual Conference on Learning Theory* 2232–2269. PMLR.
HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. *Springer Series in Statistics*. Springer, New York. MR1145237 https://doi.org/10.1007/978-1-4612-4384-7
HAMPEL, F. R. (1968). Contributions to the theory of robust estimation. Univ. California, Berkeley. MR2617979
HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics*: *The Approach Based on Influence Functions*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York. MR0829458
HOFFMANN, M. and NICKL, R. (2011). On adaptive inference and confidence bands. *Ann. Statist.* **39** 2383–2409. MR2906872 https://doi.org/10.1214/11-AOS903
HOPKINS, S., LI, J. and ZHANG, F. (2020). Robust and heavy-tailed mean estimation made simple, via regret minimization. *Adv. Neural Inf. Process. Syst.* **33** 11902–11912.
HOPKINS, S. B. (2020). Mean estimation with sub-Gaussian rates in polynomial time. *Ann. Statist.* **48** 1193–1213. MR4102693 https://doi.org/10.1214/19-AOS1843
HOPKINS, S. B. and LI, J. (2018). Mixture models, robustness, and sum of squares proofs. In *STOC'*18*—Proceedings of the* 50*th Annual ACM SIGACT Symposium on Theory of Computing* 1021–1034. ACM, New York. MR3826314 https://doi.org/10.1145/3188745.3188748
HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101. MR0161415 https://doi.org/10.1214/aoms/1177703732
HUBER, P. J. (1965). A robust version of the probability ratio test. *Ann. Math. Statist.* **36** 1753–1758. MR0185747 https://doi.org/10.1214/aoms/1177699803
HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proc. Fifth Berkeley Sympos. Math. Statist. And Probability* (*Berkeley, Calif.*, 1965/66), *Vol. I*: *Statistics* 221–233. Univ. California Press, Berkeley, CA. MR0216620
HUBER, P. J. (1968). Robust confidence limits. *Z. Wahrsch. Verw. Gebiete* **10** 269–278. MR0242330 https://doi.org/10.1007/BF00531848
HUBER, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821. MR0356373
HUBER, P. J. and RONCHETTI, E. M. (2009). *Robust Statistics*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2488795 https://doi.org/10.1002/9780470434697
JAIN, A., ORLITSKY, A. and RAVINDRAKUMAR, V. (2022). Robust estimation algorithms don't need to know the corruption level. arXiv preprint. Available at arXiv:2202.05453.
JUDITSKY, A. and LAMBERT-LACROIX, S. (2003). Nonparametric confidence set estimation. *Math. Methods Statist.* **12** 410–428. MR2054156
KUCHIBHOTLA, A. K., BALAKRISHNAN, S. and WASSERMAN, L. (2024). The HulC: Confidence regions from convex hulls. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **86** 586–622. MR4859071 https://doi.org/10.1093/jrsssb/qkad134
LAI, K. A., RAO, A. B. and VEMPALA, S. (2016). Agnostic estimation of mean and covariance. In 57*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2016 665–674. IEEE Computer Soc., Los Alamitos, CA. MR3631029
LECAM, L. (1973). Convergence of estimates under dimensionality restrictions. *Ann. Statist.* **1** 38–53. MR0334381
LEE, J. C. H. and VALIANT, P. (2022). Optimal sub-Gaussian mean estimation in $\mathbb{R}$. In 2021 *IEEE* 62*nd Annual Symposium on Foundations of Computer Science—FOCS* 2021 672–683. IEEE Computer Soc., Los Alamitos, CA. MR4399724 https://doi.org/10.1109/FOCS52979.2021.00071
LEPSKIĬ, O. V. (1991). On a problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatn. Primen.* **35** 454–466. MR1091202 https://doi.org/10.1137/1135065
LEPSKIĬ, O. V. (1992). Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Teor. Veroyatn. Primen.* **36** 682–697. MR1147167 https://doi.org/10.1137/1136085
LEPSKIĬ, O. V. (1993). Asymptotically minimax adaptive estimation. II. Schemes without optimal adaptation. Adaptive estimates. *Teor. Veroyatn. Primen.* **37** 433–448. MR1214353 https://doi.org/10.1137/1137095
LOH, P.-L. (2017). Statistical consistency and asymptotic normality for high-dimensional robust M-estimators. *Ann. Statist.* **45** 866–896. MR3650403 https://doi.org/10.1214/16-AOS1471

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

LUGOSI, G. and MENDELSON, S. (2021). Robust multivariate mean estimation: The optimality of trimmed mean. *Ann. Statist.* **49** 393–410. MR4206683 https://doi.org/10.1214/20-AOS1961

LUO, Y. and GAO, C. (2026). Supplement to "Adaptive robust confidence intervals." https://doi.org/10.1214/25-AOS2574SUPP

MINASYAN, A. and ZHIVOTOVSKIY, N. (2025). Statistically optimal robust mean and covariance estimation for anisotropic Gaussians. *Math. Stat. Learn.* **8** 33–69. MR4948548 https://doi.org/10.4171/msl/48

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

MINSKER, S. and NDAOUD, M. (2021). Robust and efficient mean estimation: An approach based on the properties of self-normalized sums. *Electron. J. Stat.* **15** 6036–6070. MR4355703 https://doi.org/10.1214/21-ejs1925

NEYMAN, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philos. Trans. R. Soc. Lond. Ser. A, Math. Phys. Sci.* **236** 333–380.

NICKL, R. and SZABÓ, B. (2016). A sharp adaptive confidence ball for self-similar functions. *Stoch. 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

PARK, B., BALAKRISHNAN, S. and WASSERMAN, L. (2023). Robust universal inference for misspecified models. arXiv preprint. Available at arXiv:2307.04034.

POLITIS, D. N. and ROMANO, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22** 2031–2050. MR1329181 https://doi.org/10.1214/aos/1176325770

PORTNOY, S. (1985). Asymptotic behavior of M estimators of p regression parameters when p^2/n is large. II. Normal approximation. *Ann. Statist.* **13** 1403–1417. MR0811499 https://doi.org/10.1214/aos/1176349744

PRASAD, A., SUGGALA, A. S., BALAKRISHNAN, S. and RAVIKUMAR, P. (2020). Robust estimation via robust gradient estimation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 601–627. MR4112778 https://doi.org/10.1111/rssb.12364

ROBINS, J. and VAN DER VAART, A. (2006). Adaptive nonparametric confidence sets. *Ann. Statist.* **34** 229–253. MR2275241 https://doi.org/10.1214/009053605000000877

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

SZABÓ, B., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2015). Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Ann. Statist.* **43** 1391–1428. MR3357861 https://doi.org/10.1214/14-AOS1270

TUKEY, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians* (*Vancouver, B.C.*, 1974), *Vol.* 2 523–531. Canad. Math. Congr., Montreal, QC. MR0426989

VAN DER PAS, S., SZABÓ, B. and VAN DER VAART, A. (2017). Adaptive posterior contraction rates for the horseshoe. *Electron. J. Stat.* **11** 3196–3225. MR3705450 https://doi.org/10.1214/17-EJS1316

WAHBA, G. (1983). Bayesian "confidence intervals" for the cross-validated smoothing spline. *J. Roy. Statist. Soc. Ser. B, Methodol.* **45** 133–150. MR0701084

WANG, H. and RAMDAS, A. (2023). Huber-robust confidence sequences. In *International Conference on Artificial Intelligence and Statistics*. 9662–9679. PMLR.

WASSERMAN, L., RAMDAS, A. and BALAKRISHNAN, S. (2020). Universal inference. *Proc. Natl. Acad. Sci. USA* **117** 16880–16890. MR4242731 https://doi.org/10.1073/pnas.1922664117

YOHAI, V. J. and MARONNA, R. A. (1979). Asymptotic behavior of M-estimators for the linear model. *Ann. Statist.* **7** 258–268. MR0520237

ZHU, B., JIAO, J. and STEINHARDT, J. (2022). Robust estimation via generalized quasi-gradients. *Inf. Inference* **11** 581–636. MR4474345 https://doi.org/10.1093/imaiai/iaab018

The Annals of Statistics
2026, Vol. 54, No. 3, 1153–1177
https://doi.org/10.1214/25-AOS2588

ONLINE TENSOR LEARNING: COMPUTATIONAL AND STATISTICAL TRADE-OFFS, ADAPTIVITY AND OPTIMAL REGRET

BY JINGYANG LI[1,a], JIAN-FENG CAI[2,b], YANG CHEN[3,d] AND DONG XIA[2,c]

[1]*Department of Statistics and Data Science, School of Management, Fudan University,* [a]*jjyyli.acad@gmail.com*
[2]*Department of Mathematics, Hong Kong University of Science and Technology,* [b]*jfcai@ust.hk,* [c]*madxia@ust.hk*
[3]*Department of Statistics, University of Michigan,* [d]*ychenang@umich.edu*

Large tensor learning algorithms are typically computationally expensive and require storing a vast amount of data. In this paper we propose a unified online Riemannian gradient descent (oRGrad) algorithm for tensor learning, which is computationally efficient, consumes much less memory, and can handle sequentially arriving data while making timely predictions. The algorithm is applicable to both linear and generalized linear models. If the time horizon T is known, oRGrad achieves statistical optimality by choosing an appropriate fixed step size. We find that noisy tensor completion particularly benefits from online algorithms by avoiding the trimming procedure and ensuring sharp entrywise statistical error, which is often technically challenging for offline methods. The regret of oRGrad is analyzed, revealing a fascinating trilemma concerning the computational convergence rate, statistical error, and regret bound. By selecting an appropriate constant step size, oRGrad achieves an $O(T^{1/2})$ regret. We then introduce the adaptive-oRGrad algorithm, which can achieve the optimal $O(\log T)$ regret by adaptively selecting step sizes, regardless of whether the time horizon is known. The adaptive-oRGrad algorithm can attain a statistically optimal error rate without knowing the horizon. Comprehensive numerical simulations corroborate our theoretical findings. We show that oRGrad significantly outperforms its offline counterpart in predicting the solar F10.7 index with tensor predictors that monitor space weather impacts.

REFERENCES

[1] AROUS, G. B., 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

[2] ARROYO, J., ATHREYA, A., CAPE, J., CHEN, G., PRIEBE, C. E. and VOGELSTEIN, J. T. (2021). Inference for multiple heterogeneous networks with a common invariant subspace. *J. Mach. Learn. Res.* **22** Paper No. 142, 49. MR4318498

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

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

[5] CAI, C., LI, G., POOR, H. V. and CHEN, Y. (2019). Nonconvex low-rank tensor completion from noisy data. *Adv. Neural Inf. Process. Syst.* **32**.

[6] CAI, J.-F., LI, J. and XIA, D. (2022). Provable tensor-train format tensor completion by Riemannian optimization. *J. Mach. Learn. Res.* **23** 1–77. MR4577075

[7] CAI, J.-F., LI, J. and XIA, D. (2023). Generalized low-rank plus sparse tensor estimation by fast Riemannian optimization. *J. Amer. Statist. Assoc.* **118** 2588–2604. MR4681606 https://doi.org/10.1080/01621459.2022.2063131

[8] CAI, T. and ZHOU, W.-X. (2013). A max-norm constrained minimization approach to 1-bit matrix completion. *J. Mach. Learn. Res.* **14** 3619–3647. MR3159403

MSC2020 subject classifications. Primary 62J05; secondary 62F10.

Key words and phrases. Online learning, tensor completion, generalized linear model, statistical optimality, regret analysis.

[9] CHANG, S., ZHANG, Y., TANG, J., YIN, D., CHANG, Y., HASEGAWA-JOHNSON, M. A. and HUANG, T. S. (2017). Streaming recommender systems. In *Proceedings of the 26th International Conference on World Wide Web* 381–389.
[10] CHEN, E. Y., XIA, D., CAI, C. and FAN, J. (2020). Semiparametric Tensor Factor Analysis by iteratively projected svd. arXiv preprint. Available at arXiv:2007.02404.
[11] CHEN, H., RASKUTTI, G. and YUAN, M. (2019). Non-convex projected gradient descent for generalized low-rank tensor regression. *J. Mach. Learn. Res.* **20** 172–208. MR3911412
[12] CHEN, Y., FAN, J., MA, C. and YAN, Y. (2019). Inference and uncertainty quantification for noisy matrix completion. *Proc. Natl. Acad. Sci. USA* **116** 22931–22937. MR4036123 https://doi.org/10.1073/pnas.1910053116
[13] CHEN, Y., FAN, J., MA, C. and YAN, Y. (2021). Bridging convex and nonconvex optimization in robust PCA: Noise, outliers and missing data. *Ann. Statist.* **49** 2948–2971. MR4338899 https://doi.org/10.1214/21-aos2066
[14] CHEN, Y. and WAINWRIGHT, M. J. (2015). Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees. arXiv preprint. Available at arXiv:1509.03025.
[15] DAVENPORT, M. A., PLAN, Y., VAN DEN BERG, E. and WOOTTERS, M. (2014). 1-bit matrix completion. *Inf. Inference* **3** 189–223. MR3311452 https://doi.org/10.1093/imaiai/iau006
[16] DE SA, C., RE, C. and OLUKOTUN, K. (2015). Global convergence of stochastic gradient descent for some non-convex matrix problems. In *International Conference on Machine Learning* 2332–2341. PMLR.
[17] FAN, J., GONG, W., LI, C. J. and SUN, Q. (2018). Statistical sparse online regression: A diffusion approximation perspective. In *International Conference on Artificial Intelligence and Statistics* 1017–1026. PMLR.
[18] GANDY, S., RECHT, B. and YAMADA, I. (2011). Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Probl.* **27** 025010, 19. MR2765628 https://doi.org/10.1088/0266-5611/27/2/025010
[19] 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. PMLR.
[20] GROSS, D., LIU, Y.-K., FLAMMIA, S. T., BECKER, S. and EISERT, J. (2010). Quantum state tomography via compressed sensing. *Phys. Rev. Lett.* **105** 150401.
[21] GUO, W., KOTSIA, I. and PATRAS, I. (2012). Tensor learning for regression. *IEEE Trans. Image Process.* **21** 816–827. MR2932176 https://doi.org/10.1109/TIP.2011.2165291
[22] HAN, Q., SUN, W. W. and ZHANG, Y. (2022). Online statistical inference for matrix contextual bandit. arXiv preprint. Available at arXiv:2212.11385.
[23] HAN, R., WILLETT, R. and ZHANG, A. R. (2022). An optimal statistical and computational framework for generalized tensor estimation. *Ann. Statist.* **50** 1–29. MR4382094 https://doi.org/10.1214/21-AOS2061
[24] HAZAN, E., AGARWAL, A. and KALE, S. (2007). Logarithmic regret algorithms for online convex optimization. *Mach. Learn.* **69** 169–192.
[25] HU, W., SHEN, W., ZHOU, H. and KONG, D. (2020). Matrix linear discriminant analysis. *Technometrics* **62** 196–205. MR4095745 https://doi.org/10.1080/00401706.2019.1610069
[26] HUANG, F., NIRANJAN, U. N., HAKEEM, M. U. and ANANDKUMAR, A. (2015). Online tensor methods for learning latent variable models. *J. Mach. Learn. Res.* **16** 2797–2835. MR3450524
[27] HUANG, H.-H., YU, F., FAN, X. and ZHANG, T. (2022). Robust Regularized Low-Rank Matrix Models for Regression and Classification. arXiv preprint. Available at arXiv:2205.07106.
[28] JAIN, P., JIN, C., KAKADE, S. M., NETRAPALLI, P. and SIDFORD, A. (2016). Streaming pca: Matching matrix Bernstein and near-optimal finite sample guarantees for oja's algorithm. In *Conference on Learning Theory* 1147–1164. PMLR.
[29] JAIN, P. and OH, S. (2014). Provable tensor factorization with missing data. *Adv. Neural Inf. Process. Syst.* **27**.
[30] JIN, C., KAKADE, S. M. and NETRAPALLI, P. (2016). Provable efficient online matrix completion via non-convex stochastic gradient descent. *Adv. Neural Inf. Process. Syst.* **29**.
[31] JING, B.-Y., LI, T., LYU, Z. and XIA, D. (2021). Community detection on mixture multilayer networks via regularized tensor decomposition. *Ann. Statist.* **49** 3181–3205. MR4352527 https://doi.org/10.1214/21-aos2079
[32] KE, Z. T., SHI, F. and XIA, D. (2019). Community detection for hypergraph networks via regularized tensor power iteration. arXiv preprint. Available at arXiv:1909.06503.
[33] KOLDA, T. G. and BADER, B. W. (2009). Tensor decompositions and applications. *SIAM Rev.* **51** 455–500. MR2535056 https://doi.org/10.1137/07070111X
[34] KRESSNER, D., STEINLECHNER, M. and VANDEREYCKEN, B. (2014). Low-rank tensor completion by Riemannian optimization. *BIT* **54** 447–468. MR3223510 https://doi.org/10.1007/s10543-013-0455-z

[35] LANGFORD, J., LI, L. and ZHANG, T. (2009). Sparse online learning via truncated gradient. *J. Mach. Learn. Res.* **10** 777–801. MR2491758

[36] LI, J., CAI, J.-F., CHEN, Y. and XIA, D. (2026). Supplement to "Online tensor learning: Computational and statistical trade-offs, adaptivity and optimal regret." https://doi.org/10.1214/25-AOS2588SUPP

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

[38] LI, P., FENG, J., JIN, X., ZHANG, L., XU, X. and YAN, S. (2018). Online robust low-rank tensor modeling for streaming data analysis. *IEEE Trans. Neural Netw. Learn. Syst.* **30** 1061–1075. MR3943037 https://doi.org/10.1109/tnnls.2018.2860964

[39] LIU, T., YUAN, M. and ZHAO, H. (2022). Characterizing spatiotemporal transcriptome of the human brain via low-rank tensor decomposition. *Stat. Biosci.* 1–29.

[40] LUO, Y. and ZHANG, A. R. (2024). Tensor-on-tensor regression: Riemannian optimization, over-parameterization, statistical-computational gap and their interplay. *Ann. Statist.* **52** 2583–2612. MR4842819 https://doi.org/10.1214/24-aos2396

[41] LYU, Z. and XIA, D. (2026). Optimal clustering by Lloyd's algorithm for low-rank mixture model. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **88** 43–65. MR5030085 https://doi.org/10.1093/jrsssb/qkaf041

[42] LYU, Z., XIA, D. and ZHANG, Y. (2023). Latent space model for higher-order networks and generalized tensor decomposition. *J. Comput. Graph. Statist.* **32** 1320–1336. MR4669249 https://doi.org/10.1080/10618600.2022.2164289

[43] MAI, Q., ZHANG, X., PAN, Y. and DENG, K. (2022). A doubly enhanced EM algorithm for model-based tensor clustering. *J. Amer. Statist. Assoc.* **117** 2120–2134. MR4528493 https://doi.org/10.1080/01621459.2021.1904959

[44] MAIRAL, J., BACH, F., PONCE, J. and SAPIRO, G. (2010). Online learning for matrix factorization and sparse coding. *J. Mach. Learn. Res.* **11** 19–60. MR2591620

[45] MEKA, R., JAIN, P., CARAMANIS, C. and DHILLON, I. S. (2008). Rank minimization via online learning. In *Proceedings of the* 25*th International Conference on Machine Learning* 656–663.

[46] MOLINA, R. (1994). On the hierarchical Bayesian approach to image restoration: Applications to astronomical images. *IEEE Trans. Pattern Anal. Mach. Intell.* **16** 1122–1128.

[47] MU, C., HUANG, B., WRIGHT, J. and GOLDFARB, D. (2014). Square deal: Lower bounds and improved relaxations for tensor recovery. In *International Conference on Machine Learning* 73–81. PMLR.

[48] RIDEOUT, W. and COSTER, A. (2006). Automated GPS processing for global total electron content data. *GPS Solut.* **10** 219–228.

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

[50] SHALIT, U., WEINSHALL, D. and CHECHIK, G. (2012). Online learning in the embedded manifold of low-rank matrices. *J. Mach. Learn. Res.* **13** 429–458. MR2913705

[51] SHAN, L., LIN, L., SUN, C. and WANG, X. (2016). Predicting ad click-through rates via feature-based fully coupled interaction tensor factorization. *Electron. Commer. Res. Appl.* **16** 30–42.

[52] SHEN, Y., LI, J., CAI, J.-F. and XIA, D. (2025). Computationally efficient and statistically optimal robust high-dimensional linear regression. *Ann. Statist.* **53** 374–399. MR4865020 https://doi.org/10.1214/24-aos2473

[53] SHI, C., SHEN, C. and SIDIROPOULOS, N. D. (2023). On high-dimensional and low-rank tensor bandits. arXiv preprint. Available at arXiv:2305.03884.

[54] SUN, H., CHEN, Y., ZOU, S., REN, J., CHANG, Y., WANG, Z. and COSTER, A. (2023). Complete global total electron content map dataset based on a video imputation algorithm VISTA. *Sci. Data* **10** 236.

[55] SUN, H., HUA, Z., REN, J., ZOU, S., SUN, Y. and CHEN, Y. (2022). Matrix completion methods for the total electron content video reconstruction. *Ann. Appl. Stat.* **16** 1333–1358. MR4455883 https://doi.org/10.1214/21-aoas1541

[56] TONG, T., MA, C., PRATER-BENNETTE, A., TRIPP, E. and CHI, Y. (2022). Scaling and scalability: Provable nonconvex low-rank tensor estimation from incomplete measurements. *J. Mach. Learn. Res.* **23** Paper No. [163], 77. MR4577115

[57] WANG, H., CHEN, J. and WEI, K. (2021). Entrywise convergence of Riemannian gradient method for low rank tensor completion via Tucker decomposition. arXiv preprint. Available at arXiv:2108.07899.

[58] WANG, M. and LI, L. (2020). Learning from binary multiway data: Probabilistic tensor decomposition and its statistical optimality. *J. Mach. Learn. Res.* **21** Paper No. 154, 38. MR4209440

[59] WANG, Z., ZOU, S., SUN, H. and CHEN, Y. (2023). Forecast global ionospheric TEC: Apply modified U-net on VISTA TEC data set. *Space Weather* **21** e2023SW003494.

[60] XIA, D. (2021). Normal approximation and confidence region of singular subspaces. *Electron. J. Stat.* **15** 3798–3851. MR4298986 https://doi.org/10.1214/21-ejs1876

[61] XIA, D. and KOLTCHINSKII, V. (2016). Estimation of low rank density matrices: Bounds in Schatten norms and other distances. *Electron. J. Stat.* **10** 2717–2745. MR3546973 https://doi.org/10.1214/16-EJS1192

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

[63] XIA, D. and YUAN, M. (2021). Statistical inferences of linear forms for noisy matrix completion. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 58–77. MR4220984 https://doi.org/10.1111/rssb.12400

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

[65] YANG, W. and XU, H. (2015). Streaming sparse principal component analysis. In *International Conference on Machine Learning* 494–503. PMLR.

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

[67] 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

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

[69] ZHANG, T. (2004). Solving large scale linear prediction problems using stochastic gradient descent algorithms. In *Proceedings of the Twenty-First International Conference on Machine Learning* **116**.

[70] ZHANG, X. and NG, M. K. (2021). Low rank tensor completion with Poisson observations. *IEEE Trans. Pattern Anal. Mach. Intell.* **44** 4239–4251.

[71] ZHENG, S., HUANG, W., VANDEREYCKEN, B. and ZHANG, X. (2022). Riemannian optimization using three different metrics for Hermitian PSD fixed-rank constraints: an extended version. arXiv preprint. Available at arXiv:2204.07830.

[72] ZHOU, J., HAO, B., WEN, Z., ZHANG, J. and SUN, W. W. (2025). Stochastic low-rank tensor bandits for multi-dimensional online decision making. *J. Amer. Statist. Assoc.* **120** 198–211. MR4893551 https://doi.org/10.1080/01621459.2024.2311364

The Annals of Statistics
2026, Vol. 54, No. 3, 1178–1204
https://doi.org/10.1214/25-AOS2589

STATISTICAL INFERENCE FOR LOW-RANK TENSORS: HETEROSKEDASTICITY, SUBGAUSSIANITY, AND APPLICATIONS

BY JOSHUA AGTERBERG[1,a] AND ANRU R. ZHANG[2,b]

[1]*Department of Statistics, University of Illinois Urbana-Champaign,* [a]*jagt@illinois.edu*
[2]*Department of Biostatistics & Bioinformatics and Department of Computer Science, Duke University,* [b]*anru.zhang@duke.edu*

In this paper, we consider inference and uncertainty quantification for low Tucker rank tensors with additive noise in the high-dimensional regime. Focusing on the output of the *higher-order orthogonal iteration* (HOOI) algorithm, a commonly used algorithm for tensor singular value decomposition, we establish nonasymptotic distributional theory and study how to construct confidence regions and intervals for both the estimated singular vectors and the tensor entries in the presence of heteroskedastic subgaussian noise, which are further shown to be optimal for homoskedastic Gaussian noise. Furthermore, as a byproduct of our theoretical results, we establish the entrywise convergence of HOOI when initialized via diagonal deletion. To further illustrate the utility of our theoretical results, we then consider several concrete statistical inference tasks. First, in the tensor mixed-membership blockmodel, we consider a two-sample test for equality of membership profiles, and we propose a test statistic with consistency under local alternatives that exhibits a power improvement relative to the corresponding matrix test considered in several previous works. Next, we consider simultaneous inference for small collections of entries of the tensor, and we obtain consistent confidence regions. Finally, focusing on the particular case of testing whether entries of the tensor are equal, we propose a consistent test statistic that shows how index overlap results in different asymptotic standard deviations. All of our proposed procedures are fully data-driven, adaptive to noise distribution and signal strength, and do not rely on sample-splitting, and our main results highlight the effect of higher-order structures on estimation relative to the matrix setting. Our theoretical results are demonstrated through numerical simulations.

REFERENCES

[1] ABBE, E., FAN, J., WANG, K. and ZHONG, Y. (2020). Entrywise eigenvector analysis of random matrices with low expected rank. *Ann. Statist.* **48** 1452–1474. MR4124330 https://doi.org/10.1214/19-AOS1854

[2] AGTERBERG, J., LUBBERTS, Z. and PRIEBE, C. E. (2022). Entrywise estimation of singular vectors of low-rank matrices with heteroskedasticity and dependence. *IEEE Trans. Inf. Theory* **68** 4618–4650. MR4449064 https://doi.org/10.1109/tit.2022.3159085

[3] AGTERBERG, J. and ZHANG, A. R. (2025). Estimating higher-order mixed memberships via the $\ell_{2,\infty}$ tensor perturbation bound. *J. Amer. Statist. Assoc.* **120** 1214–1224. MR4917290 https://doi.org/10.1080/01621459.2024.2404265

[4] AGTERBERG, J. and ZHANG, A. R (2026). Supplement to "Statistical inference for low-rank tensors: Heteroskedasticity, subgaussianity, and applications." https://doi.org/10.1214/25-AOS2589SUPP

[5] AUDDY, A. and YUAN, M. (2022). On estimating rank-one spiked tensors in the presence of heavy tailed errors. *IEEE Trans. Inf. Theory* **68** 8053–8075. MR4544931 https://doi.org/10.1109/tit.2022.3191883

[6] AUDDY, A. and YUAN, M. (2023). Perturbation bounds for (nearly) orthogonally decomposable tensors with statistical applications. *Inf. Inference* **12** 1044–1072. MR4565757 https://doi.org/10.1093/imaiai/iaac033

[7] CAI, C., LI, G., CHI, Y., POOR, H. V. and CHEN, Y. (2021). Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees. *Ann. Statist.* **49** 944–967. MR4255114 https://doi.org/10.1214/20-aos1986

MSC2020 subject classifications. Primary 62G15; secondary 62E7.
Key words and phrases. Tensor data analysis, spectral methods, singular value decomposition.

[8] CAI, C., LI, G., POOR, H. V. and CHEN, Y. (2022). Nonconvex low-rank tensor completion from noisy data. *Oper. Res.* **70** 1219–1237. MR4409613

[9] CAI, C., POOR, H. V. and CHEN, Y. (2023). Uncertainty quantification for nonconvex tensor completion: Confidence intervals, heteroscedasticity and optimality. *IEEE Trans. Inf. Theory* **69** 407–452. MR4544966 https://doi.org/10.1109/tit.2022.3205781

[10] CAI, J.-F., LI, J. and XIA, D. (2022). Provable tensor-train format tensor completion by Riemannian optimization. *J. Mach. Learn. Res.* **23** Paper No. [123], 77. MR4577075

[11] CAI, T. T. and GUO, Z. (2017). Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *Ann. Statist.* **45** 615–646. MR3650395 https://doi.org/10.1214/16-AOS1461

[12] CHEN, Y., CHI, Y., FAN, J. and MA, C. (2021). Spectral methods for data science: A statistical perspective. *Found. Trends Mach. Learn.* **14** 566–806.

[13] CHI, E. C., GAINES, B. R., SUN, W. W., ZHOU, H. and YANG, J. (2020). Provable convex co-clustering of tensors. *J. Mach. Learn. Res.* **21** Paper No. 214, 58. MR4209500

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

[15] DU, X. and TANG, M. (2023). Hypothesis testing for equality of latent positions in random graphs. *Bernoulli* **29** 3221–3254. MR4632137 https://doi.org/10.3150/22-bej1581

[16] FAN, J., FAN, Y., HAN, X. and LV, J. (2022). SIMPLE: Statistical inference on membership profiles in large networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 630–653. MR4413000 https://doi.org/10.1111/rssb.12505

[17] HAN, R., LUO, Y., WANG, M. and ZHANG, A. R. (2022). Exact clustering in tensor block model: Statistical optimality and computational limit. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 1666–1698. MR4515554 https://doi.org/10.1111/rssb.12547

[18] HAN, R., SHI, P. and ZHANG, A. R. (2024). Guaranteed functional tensor singular value decomposition. *J. Amer. Statist. Assoc.* **119** 995–1007. MR4765964 https://doi.org/10.1080/01621459.2022.2153689

[19] HAN, R., WILLETT, R. and ZHANG, A. R. (2022). An optimal statistical and computational framework for generalized tensor estimation. *Ann. Statist.* **50** 1–29. MR4382094 https://doi.org/10.1214/21-AOS2061

[20] HAN, X., YANG, Q. and FAN, Y. (2023). Universal rank inference via residual subsampling with application to large networks. *Ann. Statist.* **51** 1109–1133. MR4630942 https://doi.org/10.1214/23-aos2282

[21] HAN, Y. and ZHANG, C.-H. (2023). Tensor principal component analysis in high dimensional CP models. *IEEE Trans. Inf. Theory* **69** 1147–1167. MR4564648 https://doi.org/10.1109/tit.2022.3203972

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

[23] HU, J. and WANG, M. (2023). Multiway spherical clustering via degree-corrected tensor block models. *IEEE Trans. Inf. Theory* **69** 3880–3919. MR4604474 https://doi.org/10.1109/tit.2023.3239521

[24] HUANG, J., HUANG, D. Z., YANG, Q. and CHENG, G. (2022). Power iteration for tensor PCA. *J. Mach. Learn. Res.* **23** Paper No. [128], 47. MR4577080

[25] 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. MR4646629 https://doi.org/10.1080/01621459.2022.2035736

[26] JING, B.-Y., LI, T., LYU, Z. and XIA, D. (2021). Community detection on mixture multilayer networks via regularized tensor decomposition. *Ann. Statist.* **49** 3181–3205. MR4352527 https://doi.org/10.1214/21-aos2079

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

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

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

[30] LUO, Y., RASKUTTI, G., YUAN, M. and ZHANG, A. R. (2021). A sharp blockwise tensor perturbation bound for orthogonal iteration. *J. Mach. Learn. Res.* **22** Paper No. 179, 48. MR4318535

[31] LUO, Y. and ZHANG, A. R. (2022). Tensor clustering with planted structures: Statistical optimality and computational limits. *Ann. Statist.* **50** 584–613. MR4382029 https://doi.org/10.1214/21-aos2123

[32] LYU, Z. and XIA, D. (2023). Optimal estimation and computational limit of low-rank Gaussian mixtures. *Ann. Statist.* **51** 646–667. MR4600996 https://doi.org/10.1214/23-aos2264

[33] LYU, Z. and XIA, D. (2026). Optimal clustering by Lloyd's algorithm for low-rank mixture model. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **88** 43–65. MR5030085 https://doi.org/10.1093/jrsssb/qkaf041

[34] LYU, Z., XIA, D. and ZHANG, Y. (2023). Latent space model for higher-order networks and generalized tensor decomposition. *J. Comput. Graph. Statist.* **32** 1320–1336. MR4669249 https://doi.org/10.1080/10618600.2022.2164289

[35] MARTINO, C., SHENHAV, L., MAROTZ, C. A., ARMSTRONG, G., MCDONALD, D., VÁZQUEZ-BAEZA, Y., MORTON, J. T., JIANG, L., GLORIA DOMINGUEZ-BELLO, M. et al. (2021). Context-aware dimensionality reduction deconvolutes gut microbial community dynamics. *Nat. Biotechnol.* **39** 165–168.

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

[37] 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

[38] WANG, H., CHEN, J. and WEI, K. (2023). Implicit regularization and entrywise convergence of Riemannian optimization for low Tucker-rank tensor completion. *J. Mach. Learn. Res.* **24** Paper No. [347], 84. MR4690296

[39] WANG, M. and ZENG, Y. (2019). Multiway clustering via tensor block models. In *Advances in Neural Information Processing Systems* **32**. Curran Associates, Red Hook.

[40] WU, T., BENSON, A. R. and GLEICH, D. F. (2016). General tensor spectral co-clustering for higher-order data. In *Advances in Neural Information Processing Systems* **29**. Curran Associates, Red Hook.

[41] XIA, D., ZHANG, A. R. and ZHOU, Y. (2022). Inference for low-rank tensors—no need to debias. *Ann. Statist.* **50** 1220–1245. MR4404934 https://doi.org/10.1214/21-aos2146

[42] YAN, Y., CHEN, Y. and FAN, J. (2024). Inference for heteroskedastic PCA with missing data. *Ann. Statist.* **52** 729–756. MR4744194 https://doi.org/10.1214/24-aos2366

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

[44] 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

[45] ZHANG, A. R., CAI, T. T. and WU, Y. (2022). Heteroskedastic PCA: Algorithm, optimality, and applications. *Ann. Statist.* **50** 53–80. MR4382008 https://doi.org/10.1214/21-aos2074

[46] ZHANG, C., HAN, R., ZHANG, A. R. and VOYLES, P. M. (2020). Denoising atomic resolution 4D scanning transmission electron microscopy data with tensor singular value decomposition. *Ultramicroscopy* **219** 113123.

[47] ZHANG, Z., ALLEN, G. I., ZHU, H. and DUNSON, D. (2019). Tensor network factorizations: Relationships between brain structural connectomes and traits. *NeuroImage* **197** 330–343.

[48] ZHOU, Y. and CHEN, Y. (2025). Deflated HeteroPCA: Overcoming the curse of ill-conditioning in heteroskedastic PCA. *Ann. Statist.* **53** 91–116. MR4865009 https://doi.org/10.1214/24-aos2456

[49] ZHOU, Y., ZHANG, A. R., ZHENG, L. and WANG, Y. (2022). Optimal high-order tensor SVD via tensor-train orthogonal iteration. *IEEE Trans. Inf. Theory* **68** 3991–4019. MR4433265 https://doi.org/10.1109/tit.2022.3152733

[50] ZHU, M. and GHODSI, A. (2006). Automatic dimensionality selection from the scree plot via the use of profile likelihood. *Comput. Statist. Data Anal.* **51** 918–930. MR2297497 https://doi.org/10.1016/j.csda.2005.09.010

The Annals of Statistics
2026, Vol. 54, No. 3, 1205–1231
https://doi.org/10.1214/25-AOS2594

LEARNING EXTREMAL GRAPHICAL STRUCTURES IN HIGH DIMENSIONS

BY SEBASTIAN ENGELKE[1,a], MICHAËL LALANCETTE[2,b] AND STANISLAV VOLGUSHEV[3,c]

[1]*Research Center for Statistics, University of Geneva,* [a]*sebastian.engelke@unige.ch*
[2]*Département de mathématiques, Université du Québec à Montréal,* [b]*lalancette.michael@uqam.ca*
[3]*Department of Statistical Sciences, University of Toronto,* [c]*stanislav.volgushev@utoronto.ca*

Extremal graphical models encode the conditional independence structure of multivariate extremes. Key statistics for learning extremal graphical structures are empirical extremal variograms, for which we prove nonasymptotic concentration bounds that hold under general domain of attraction conditions. For the popular class of Hüsler–Reiss models, we propose a majority voting algorithm for learning the underlying graph from data through L^1 regularized optimization. Our concentration bounds are used to derive explicit conditions that ensure the consistent recovery of any connected graph. The methodology is illustrated through a simulation study as well as the analysis of river discharge and currency exchange data.

REFERENCES

ALBERT, R. and BARABÁSI, A.-L. (2002). Statistical mechanics of complex networks. *Rev. Modern Phys*. **74** 47–97. MR1895096 https://doi.org/10.1103/RevModPhys.74.47

AMÉNDOLA, C., HOLLERING, B., SULLIVANT, S. and TRAN, N. (2021). Markov equivalence of max-linear Bayesian networks. In *Uncertainty in Artificial Intelligence* 1746–1755. PMLR.

AMÉNDOLA, C., KLÜPPELBERG, C., LAURITZEN, S. and TRAN, N. M. (2022). Conditional independence in max-linear Bayesian networks. *Ann. Appl. Probab.* **32** 1–45. MR4386833 https://doi.org/10.1214/21-AAP1670

ASADI, P., DAVISON, A. C. and ENGELKE, S. (2015). Extremes on river networks. *Ann. Appl. Stat*. **9** 2023–2050. MR3456363 https://doi.org/10.1214/15-AOAS863

ASENOVA, S., MAZO, G. and SEGERS, J. (2021). Inference on extremal dependence in the domain of attraction of a structured Hüsler–Reiss distribution motivated by a Markov tree with latent variables. *Extremes* **24** 461–500. MR4301085 https://doi.org/10.1007/s10687-021-00407-5

ASENOVA, S. and SEGERS, J. (2023). Extremes of Markov random fields on block graphs: Max-stable limits and structured Hüsler–Reiss distributions. *Extremes* **26** 433–468. MR4627351 https://doi.org/10.1007/s10687-023-00467-9

BEIRLANT, J., GOEGEBEUR, Y., TEUGELS, J. and SEGERS, J. (2004). *Statistics of Extremes*: *Theory and Applications*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR2108013 https://doi.org/10.1002/0470012382

CASEY, A. and PAPASTATHOPOULOS, I. (2023). Decomposable tail graphical models. arXiv preprint. Available at arXiv:2302.05182.

CHANG, A. and ALLEN, G. I. (2023). Subbotin graphical models for extreme value dependencies with applications to functional neuronal connectivity. *Ann. Appl. Stat*. **17** 2364–2386. MR4637671 https://doi.org/10.1214/22-aoas1723

CLÉMENÇON, S., JALALZAI, H., LHAUT, S., SABOURIN, A. and SEGERS, J. (2023). Concentration bounds for the empirical angular measure with statistical learning applications. *Bernoulli* **29** 2797–2827. MR4632121 https://doi.org/10.3150/22-bej1562

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

DE FONDEVILLE, R. and DAVISON, A. C. (2018). High-dimensional peaks-over-threshold inference. *Biometrika* **105** 575–592. MR3842886 https://doi.org/10.1093/biomet/asy026

MSC2020 subject classifications. Primary 60G70, 62H22; secondary 62G30.

Key words and phrases. Multivariate Pareto distribution, peaks over threshold modeling, Hüsler–Reiss distribution, graphical lasso, neighborhood selection.

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

DOMBRY, C., ENGELKE, S. and OESTING, M. (2016). Exact simulation of max-stable processes. *Biometrika* **103** 303–317. MR3509888 https://doi.org/10.1093/biomet/asw008

DREES, H., DE HAAN, L. and RESNICK, S. (2000). How to make a Hill plot. *Ann. Statist.* **28** 254–274. MR1762911 https://doi.org/10.1214/aos/1016120372

DRTON, M. and MAATHUIS, M. H. (2017). Structure learning in graphical modeling. *Annu. Rev. Stat. Appl.* **4** 365–393.

EINMAHL, J. H. J., KRAJINA, A. and SEGERS, J. (2008). A method of moments estimator of tail dependence. *Bernoulli* **14** 1003–1026. MR2543584 https://doi.org/10.3150/08-BEJ130

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

ENGELKE, S., DE FONDEVILLE, R. and OESTING, M. (2019). Extremal behaviour of aggregated data with an application to downscaling. *Biometrika* **106** 127–144. MR3912387 https://doi.org/10.1093/biomet/asy052

ENGELKE, S., GNECCO, N. and RÖTTGER, F. (2025). Extremes of structural causal models. arXiv preprint. Available at arXiv:2503.0653.

ENGELKE, S., HENTSCHEL, M., LALANCETTE, M. and RÖTTGER, F. (2024). Graphical models for multivariate extremes. arXiv preprint. Available at arXiv:2402.02187.

ENGELKE, S. and HITZ, A. S. (2020). Graphical models for extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 871–932. MR4136498

ENGELKE, S. and IVANOVS, J. (2021). Sparse structures for multivariate extremes. *Annu. Rev. Stat. Appl.* **8** 241–270. MR4243547 https://doi.org/10.1146/annurev-statistics-040620-041554

ENGELKE, S., IVANOVS, J. and STROKORB, K. (2025). Graphical models for infinite measures with applications to extremes. *Ann. Appl. Probab.* **35** 3490-3542. MR4975054 https://doi.org/10.1214/25-AAP2201

ENGELKE, S., LALANCETTE, M. and VOLGUSHEV, S. (2026a). Supplement to "Learning extremal graphical structures in high dimensions." https://doi.org/10.1214/25-AOS2594SUPPA

ENGELKE, S., LALANCETTE, M. and VOLGUSHEV, S. (2026b). Supplement to "Learning extremal graphical structures in high dimensions." https://doi.org/10.1214/25-AOS2594SUPPB

ENGELKE, S., MALINOWSKI, A., KABLUCHKO, Z. and SCHLATHER, M. (2015). Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 239–265. MR3299407 https://doi.org/10.1111/rssb.12074

ENGELKE, S. and TAEB, A. (2025). Extremal graphical modeling with latent variables via convex optimization. *J. Mach. Learn. Res.* **26** 1–68. MR4896091

ENGELKE, S. and VOLGUSHEV, S. (2022). Structure learning for extremal tree models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 2055–2087. MR4515566

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

FARRIS, J. S., KLUGE, A. G. and ECKARDT, M. J. (1970). A numerical approach to phylogenetic systematics. *Syst. Zool.* **19** 172–189.

FOUGÈRES, A.-L., DE HAAN, L. and MERCADIER, C. (2015). Bias correction in multivariate extremes. *Ann. Statist.* **43** 903–934. MR3325714 https://doi.org/10.1214/14-AOS1305

FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441.

GISSIBL, N. and KLÜPPELBERG, C. (2018). Max-linear models on directed acyclic graphs. *Bernoulli* **24** 2693–2720. MR3779699 https://doi.org/10.3150/17-BEJ941

GISSIBL, N., KLÜPPELBERG, C. and LAURITZEN, S. (2021). Identifiability and estimation of recursive max-linear models. *Scand. J. Stat.* **48** 188–211. MR4233170 https://doi.org/10.1111/sjos.12446

GNECCO, N., MEINSHAUSEN, N., PETERS, J. and ENGELKE, S. (2021). Causal discovery in heavy-tailed models. *Ann. Statist.* **49** 1755–1778. MR4298880 https://doi.org/10.1214/20-aos2021

GOIX, N., SABOURIN, A. and CLÉMENÇON, S. (2015). Learning the dependence structure of rare events: A non-asymptotic study. In *Conf. Learn. Theory* 843–860. PMLR.

GUDENDORF, G. and SEGERS, J. (2010). Extreme-value copulas. In *Copula Theory and Its Applications*. *Lect. Notes Stat. Proc.* **198** 127–145. Springer, Heidelberg. MR3051266 https://doi.org/10.1007/978-3-642-12465-5_6

HENTSCHEL, M., ENGELKE, S. and SEGERS, J. (2025). Statistical inference for Hüsler–Reiss graphical models through matrix completions. *J. Amer. Statist. Assoc.* **120** 909–921. MR4917266 https://doi.org/10.1080/01621459.2024.2371978

HU, S., PENG, Z. and SEGERS, J. (2024). Modeling multivariate extreme value distributions via Markov trees. *Scand. J. Stat.* **51** 760–800. MR4828108 https://doi.org/10.1111/sjos.12698

HÜSLER, J. and REISS, R.-D. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statist. Probab. Lett.* **7** 283–286. MR0980699 https://doi.org/10.1016/0167-7152(89)90106-5

KABLUCHKO, Z., SCHLATHER, M. and DE HAAN, L. (2009). Stationary max-stable fields associated to negative definite functions. *Ann. Probab.* **37** 2042–2065. MR2561440 https://doi.org/10.1214/09-AOP455

KEEF, C., TAWN, J. A. and LAMB, R. (2013). Estimating the probability of widespread flood events. *Environmetrics* **24** 13–21. MR3042270 https://doi.org/10.1002/env.2190

KLÜPPELBERG, C. and KRALI, M. (2021). Estimating an extreme Bayesian network via scalings. *J. Multivariate Anal.* **181** Paper No. 104672, 23. MR4157685 https://doi.org/10.1016/j.jmva.2020.104672

KLÜPPELBERG, C. and LAURITZEN, S. (2019). Bayesian networks for max-linear models. In *Network Science* 79–97. Springer, Cham. MR3971300

KRALI, M., DAVISON, A. C. and KLÜPPELBERG, C. (2023). Heavy-tailed max-linear structural equation models in networks with hidden nodes. arXiv preprint. Available at arXiv:2306.15356.

LALANCETTE, M. (2023). On pairwise interaction multivariate Pareto models. *Stat* **12** Paper No. e613, 10. MR4641234 https://doi.org/10.1002/sta4.613

LALANCETTE, M., ENGELKE, S. and VOLGUSHEV, S. (2021). Rank-based estimation under asymptotic dependence and independence, with applications to spatial extremes. *Ann. Statist.* **49** 2552–2576. MR4338375 https://doi.org/10.1214/20-aos2046

LAURITZEN, S. L. (1996). *Graph. Models. Oxford Statistical Science Series* **17**. Clarendon, Oxford. MR1419991

LEDERER, J. and OESTING, M. (2024). Extremes in high dimensions: Methods and scalable algorithms. arXiv preprint. Available at arXiv:2303.04258.

LHAUT, S., SABOURIN, A. and SEGERS, J. (2022). Uniform concentration bounds for frequencies of rare events. *Statist. Probab. Lett.* **189** Paper No. 109610, 7. MR4452775 https://doi.org/10.1016/j.spl.2022.109610

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

LIU, H., XU, M., GU, H., GUPTA, A., LAFFERTY, J. and WASSERMAN, L. (2011). Forest density estimation. *J. Mach. Learn. Res.* **12** 907–951. MR2786914

LOH, P.-L. and WAINWRIGHT, M. J. (2013). Structure estimation for discrete graphical models: Generalized covariance matrices and their inverses. *Ann. Statist.* **41** 3022–3049. MR3161456 https://doi.org/10.1214/13-AOS1162

LOH, P.-L. and WAINWRIGHT, M. J. (2017). Support recovery without incoherence: A case for nonconvex regularization. *Ann. Statist.* **45** 2455–2482. MR3737898 https://doi.org/10.1214/16-AOS1530

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

MHALLA, L., CHAVEZ-DEMOULIN, V. and DUPUIS, D. J. (2020). Causal mechanism of extreme river discharges in the upper Danube basin network. *J. R. Stat. Soc., Ser. C, Appl. Stat.* **69** 741–764. MR4133145 https://doi.org/10.1111/rssc.12415

PAPASTATHOPOULOS, I. and STROKORB, K. (2016). Conditional independence among max-stable laws. *Statist. Probab. Lett.* **108** 9–15. MR3420420 https://doi.org/10.1016/j.spl.2015.08.008

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

ROOTZÉN, H., SEGERS, J. and WADSWORTH, J. L. (2018). Multivariate peaks over thresholds models. *Extremes* **21** 115–145. MR3764613 https://doi.org/10.1007/s10687-017-0294-4

ROOTZÉN, H. and TAJVIDI, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli* **12** 917–930. MR2265668 https://doi.org/10.3150/bj/1161614952

RÖTTGER, F., COONS, J. I. and GROSDOS, A. (2023). Parametric and nonparametric symmetries in graphical models for extremes. arXiv preprint. Available at arXiv:2306.00703.

RÖTTGER, F., ENGELKE, S. and ZWIERNIK, P. (2023). Total positivity in multivariate extremes. *Ann. Statist.* **51** 962–1004. MR4630937 https://doi.org/10.1214/23-aos2272

RÖTTGER, F. and SCHMITZ, Q. (2022). On the local metric property in multivariate extremes. arXiv preprint. Available at arXiv:2212.10350.

SEGERS, J. (2020). One- versus multi-component regular variation and extremes of Markov trees. *Adv. Appl. Probab.* **52** 855–878. MR4153591 https://doi.org/10.1017/apr.2020.22

TAWN, J. A. (1990). Modelling multivariate extreme value distributions. *Biometrika* **77** 245–253.

THIBAUD, E., MUTZNER, R. and DAVISON, A. C. (2013). Threshold modeling of extreme spatial rainfall. *Water Resour. Res.* **49** 4633–4644.

TRAN, N. M., BUCK, J. and KLÜPPELBERG, C. (2024). Estimating a directed tree for extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **86** 771–792. MR4792085 https://doi.org/10.1093/jrsssb/qkad165

WAN, P. and ZHOU, C. (2025). Graphical lasso for extremes. arXiv preprint. Available at arXiv:2307.15004.
WANG, H., LI, B. and LENG, C. (2009). Shrinkage tuning parameter selection with a diverging number of parameters. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 671–683. MR2749913 https://doi.org/10.1111/j.1467-9868.2008.00693.x
WANG, M. and ALLEN, G. I. (2023). Thresholded graphical lasso adjusts for latent variables. *Biometrika* **110** 681–697. MR4627778 https://doi.org/10.1093/biomet/asac060
YING, J., CARDOSO, J. M. and PALOMAR, D. (2021). Minimax estimation of Laplacian constrained precision matrices. In *Int. Conf. Artif. Intell. Stat* 3736–3744. PMLR.
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
ZHAO, P. and YU, B. (2006). On model selection consistency of Lasso. *J. Mach. Learn. Res.* **7** 2541–2563. MR2274449
ZHOU, C. (2010). Are banks too big to fail? Measuring systemic importance of financial institutions. *Int. J. Cent. Bank.* **6** 205–250.
ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279469 https://doi.org/10.1198/016214506000000735

The Annals of Statistics
2026, Vol. 54, No. 3, 1232–1261
https://doi.org/10.1214/25-AOS2600

GRADIENT DESCENT INFERENCE IN EMPIRICAL RISK MINIMIZATION

BY QIYANG HAN[1,a] AND XIAOCONG XU[2,b]

[1]*Department of Statistics, Rutgers University,* [a]*qh85@stat.rutgers.edu*
[2]*Data Sciences and Operations Department, Marshall School of Business, University of Southern California,* [b]*xuxiaoco@marshall.usc.edu*

Gradient descent is one of the most widely used iterative algorithms in modern statistical learning. However, its precise algorithmic dynamics in high-dimensional settings remain only partially understood, which has limited its broader potential for statistical inference applications.

This paper provides a precise, nonasymptotic joint distributional characterization of gradient descent iterates and their debiased statistics in a broad class of empirical risk minimization problems, in the so-called mean-field regime where the sample size is proportional to the signal dimension. Our nonasymptotic state evolution theory holds for both general nonconvex loss functions and non-Gaussian data, and reveals the central role of two Onsager correction matrices that precisely characterize the nontrivial dependence among all gradient descent iterates in the mean-field regime.

Leveraging the joint state evolution characterization, we show that the gradient descent iterate retrieves approximate normality after a debiasing correction via a linear combination of observable loss derivative directions from all past iterates. Crucially, the debiasing coefficients are directly linked to the Onsager correction matrices, which can be estimated in a fully data-driven manner via the proposed *gradient descent inference algorithm*. This leads to a new algorithmic statistical inference framework based on debiased gradient descent, which (i) applies to a broad class of models with both convex and nonconvex losses, (ii) remains valid at each iteration without requiring algorithmic convergence and (iii) exhibits a certain robustness to possible model misspecification. As a by-product, our framework also provides algorithmic estimates of the generalization error at each iteration.

We demonstrate our theory and inference methods in the canonical single-index regression model and a generalized logistic regression model, where the natural loss functions may exhibit arbitrarily nonconvex landscapes. Our analysis further shows that, in linear regression with squared loss, the proposed debiased gradient descent iterate eventually coincides with the debiased convex regularized estimator in a mean-field distributional sense, and the quality of statistical inference for the unknown signal aligns exactly with the generalization error achieved along the algorithmic trajectory.

REFERENCES

[1] AGORITSAS, E., BIROLI, G., URBANI, P. and ZAMPONI, F. (2018). Out-of-equilibrium dynamical mean-field equations for the perceptron model. *J. Phys. A* **51** 085002, 36. MR3763797 https://doi.org/10.1088/1751-8121/aaa68d

[2] ALI, A., DOBRIBAN, E. and TIBSHIRANI, R. (2020). The implicit regularization of stochastic gradient flow for least squares. In *International Conference on Machine Learning* 233–244. PMLR.

[3] ALI, A., KOLTER, J. Z. and TIBSHIRANI, R. J. (2019). A continuous-time view of early stopping for least squares regression. In *The* 22*nd International Conference on Artificial Intelligence and Statistics* 1370–1378. PMLR.

[4] ALTIERI, A., BIROLI, G. and CAMMAROTA, C. (2020). Dynamical mean-field theory and aging dynamics. *J. Phys. A* **53** 375006, 34. MR4152046 https://doi.org/10.1088/1751-8121/aba3dd

MSC2020 subject classifications. Primary 60E15; secondary 60G15.
Key words and phrases. Debiased statistical inference, empirical risk minimization, gradient descent, linear regression, logistic regression, state evolution, universality.

[5] BALASUBRAMANIAN, K., GHOSAL, P. and HE, Y. (2025). High-dimensional scaling limits and fluctuations of online least-squares SGD with smooth covariance. *Ann. Appl. Probab.* **35** 2983–3045. MR4975041 https://doi.org/10.1214/24-aap2123

[6] BAO, Z., HAN, Q. and XU, X. (2025). A leave-one-out approach to approximate message passing. *Ann. Appl. Probab.* **35** 2716–2766. MR4945090 https://doi.org/10.1214/25-AAP2186

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

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

[9] BELLEC, P. C. (2025). Observable adjustments in single-index models for regularized M-estimators with bounded p/n. *Ann. Statist.* **53** 531–560. MR4900157 https://doi.org/10.1214/24-aos2464

[10] BELLEC, P. C. and TAN, K. (2024). Uncertainty quantification for iterative algorithms in linear models with application to early stopping. arXiv preprint. Available at arXiv:2404.17856.

[11] BELLEC, P. C. and ZHANG, C.-H. (2023). Debiasing convex regularized estimators and interval estimation in linear models. *Ann. Statist.* **51** 391–436. MR4600987 https://doi.org/10.1214/22-aos2243

[12] BEN AROUS, G., GHEISSARI, R. and JAGANNATH, A. (2021). Online stochastic gradient descent on non-convex losses from high-dimensional inference. *J. Mach. Learn. Res.* **22** Paper No. 106, 51. MR4279757

[13] BEN AROUS, G., GHEISSARI, R. and JAGANNATH, A. (2024). High-dimensional limit theorems for SGD: Effective dynamics and critical scaling. *Comm. Pure Appl. Math.* **77** 2030–2080. MR4692886 https://doi.org/10.1002/cpa.22169

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

[15] CELENTANO, M., CHENG, C. and MONTANARI, A. (2021). The high-dimensional asymptotics of first order methods with random data. arXiv preprint. Available at arXiv:2112.07572.

[16] CELENTANO, M., MONTANARI, A. and WEI, Y. (2023). The Lasso with general Gaussian designs with applications to hypothesis testing. *Ann. Statist.* **51** 2194–2220. MR4678801 https://doi.org/10.1214/23-aos2327

[17] CHEN, X., LEE, J. D., TONG, X. T. and ZHANG, Y. (2020). Statistical inference for model parameters in stochastic gradient descent. *Ann. Statist.* **48** 251–273. MR4065161 https://doi.org/10.1214/18-AOS1801

[18] COLLINS-WOODFIN, E., PAQUETTE, C., PAQUETTE, E. and SEROUSSI, I. (2024). Hitting the high-dimensional notes: An ODE for SGD learning dynamics on GLMs and multi-index models. *Inf. Inference* **13** Paper No. iaae028, 107. MR4811680 https://doi.org/10.1093/imaiai/iaae028

[19] CUGLIANDOLO, L. F. and KURCHAN, J. (1993). Analytical solution of the off-equilibrium dynamics of a long-range spin-glass model. *Phys. Rev. Lett.* **71** 173.

[20] DANDI, Y., TROIANI, E., ARNABOLDI, L., PESCE, L., ZDEBOROVÁ, L. and KRZAKALA, F. (2024). The benefits of reusing batches for gradient descent in two-layer networks: Breaking the curse of information and leap exponents. arXiv preprint. Available at arXiv:2402.03220.

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

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

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

[24] FAN, Z. (2022). Approximate message passing algorithms for rotationally invariant matrices. *Ann. Statist.* **50** 197–224. MR4382014 https://doi.org/10.1214/21-aos2101

[25] FANG, Y., XU, J. and YANG, L. (2018). Online bootstrap confidence intervals for the stochastic gradient descent estimator. *J. Mach. Learn. Res.* **19** Paper No. 78, 21. MR3899780

[26] GERBELOT, C., TROIANI, E., MIGNACCO, F., KRZAKALA, F. and ZDEBOROVÁ, L. (2024). Rigorous dynamical mean-field theory for stochastic gradient descent methods. *SIAM J. Math. Data Sci.* **6** 400–427. MR4741502 https://doi.org/10.1137/23M1594388

[27] HAN, Q. (2023). Noisy linear inverse problems under convex constraints: Exact risk asymptotics in high dimensions. *Ann. Statist.* **51** 1611–1638. MR4658570 https://doi.org/10.1214/23-aos2301

[28] HAN, Q. (2025). Entrywise dynamics and universality of general first order methods. *Ann. Statist.* **53** 1783–1807. MR4959811 https://doi.org/10.1214/25-AOS2544
[29] HAN, Q. (2025). Long-time dynamics and universality of nonconvex gradient descent. arXiv preprint. Available at arXiv:2509.11426.
[30] HAN, Q. and SHEN, Y. (2023). Universality of regularized regression estimators in high dimensions. *Ann. Statist.* **51** 1799–1823. MR4658577 https://doi.org/10.1214/23-aos2309
[31] HAN, Q. and XU, X. (2023). The distribution of ridgeless least squares interpolators. arXiv preprint. Available at arXiv:2307.02044.
[32] HAN, Q. and XU, X. (2026). Supplement to "Gradient descent inference in empirical risk minimization." https://doi.org/10.1214/25-AOS2600SUPP
[33] HUANG, J., JIAO, Y., LU, X. and ZHU, L. (2018). Robust decoding from 1-bit compressive sampling with ordinary and regularized least squares. *SIAM J. Sci. Comput.* **40** A2062–A2086. MR3820388 https://doi.org/10.1137/17M1154102
[34] JAVANMARD, A. and MONTANARI, A. (2013). State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Inf. Inference* **2** 115–144. MR3311445 https://doi.org/10.1093/imaiai/iat004
[35] JAVANMARD, A. and MONTANARI, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909. MR3277152
[36] 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
[37] JONES, C. and PESENTI, L. (2025). Fourier analysis of iterative algorithms. In *52nd International Colloquium on Automata, Languages, and Programming. LIPIcs. Leibniz Int. Proc. Inform.* **334** Art. No. 102, 21. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR4934254 https://doi.org/10.4230/lipics.icalp.2025.102
[38] LEE, J. D., OKO, K., SUZUKI, T. and WU, D. (2024). Neural network learns low-dimensional polynomials with SGD near the information-theoretic limit. *Adv. Neural Inf. Process. Syst.* **37** 58716–58756.
[39] LI, G. and WEI, Y. (2022). A non-asymptotic framework for approximate message passing in spiked models. arXiv preprint. Available at arXiv:2208.03313.
[40] MIGNACCO, F., KRZAKALA, F., URBANI, P. and ZDEBOROVÁ, L. (2020). Dynamical mean-field theory for stochastic gradient descent in Gaussian mixture classification. *Adv. Neural Inf. Process. Syst.* **33** 9540–9550.
[41] MIGNACCO, F. and URBANI, P. (2022). The effective noise of stochastic gradient descent. *J. Stat. Mech. Theory Exp.* **8** Paper No. 083405, 23. MR4504824 https://doi.org/10.1088/1742-5468/ac841d
[42] MIOLANE, L. and MONTANARI, A. (2021). The distribution of the Lasso: Uniform control over sparse balls and adaptive parameter tuning. *Ann. Statist.* **49** 2313–2335. MR4319252 https://doi.org/10.1214/20-aos2038
[43] MONTANARI, A. (2018). Mean field asymptotics in high-dimensional statistics: From exact results to efficient algorithms. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018, Vol. IV. Invited Lectures* 2973–2994. World Scientific, Hackensack, NJ. MR3966519
[44] PAQUETTE, C., LEE, K., PEDREGOSA, F. and PAQUETTE, E. (2021). SGD in the large: Average-case analysis, asymptotics, and stepsize criticality. In *Conference on Learning Theory* 3548–3626. PMLR.
[45] PAQUETTE, C. and PAQUETTE, E. (2021). Dynamics of stochastic momentum methods on large-scale, quadratic models. *Adv. Neural Inf. Process. Syst.* **34** 9229–9240.
[46] PATIL, P., WU, Y. and TIBSHIRANI, R. (2024). Failures and successes of cross-validation for early-stopped gradient descent. In *International Conference on Artificial Intelligence and Statistics* 2260–2268. PMLR.
[47] 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
[48] RUPPERT, D. (1988). Efficient estimations from a slowly convergent Robbins-Monro process. Tech. Report, Cornell Univ. Operations Research and Industrial Engineering.
[49] RUSH, C. and VENKATARAMANAN, R. (2018). Finite sample analysis of approximate message passing algorithms. *IEEE Trans. Inf. Theory* **64** 7264–7286. MR3876443 https://doi.org/10.1109/TIT.2018.2816681
[50] STOJNIC, M. (2013). A framework to characterize performance of lasso algorithms. arXiv preprint. Available at arXiv:1303.7291.
[51] 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

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

[53] TAN, K. and BELLEC, P. C. (2024). Estimating generalization performance along the trajectory of proximal SGD in robust regression. *Adv. Neural Inf. Process. Syst.*

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

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

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

[57] YADLOWSKY, S., YUN, T., MCLEAN, C. Y. and D'AMOUR, A. (2021). SLOE: A faster method for statistical inference in high-dimensional logistic regression. *Adv. Neural Inf. Process. Syst.* **34** 29517–29528.

[58] ZHU, W., CHEN, X. and WU, W. B. (2023). Online covariance matrix estimation in stochastic gradient descent. *J. Amer. Statist. Assoc.* **118** 393–404. MR4571129 https://doi.org/10.1080/01621459.2021.1933498

The Annals of Statistics
2026, Vol. 54, No. 3, 1262–1288
https://doi.org/10.1214/25-AOS2603

STATISTICAL-COMPUTATIONAL TRADE-OFFS FOR RECURSIVE ADAPTIVE PARTITIONING ESTIMATORS

BY YAN SHUO TAN[1,a], JASON M. KLUSOWSKI[2,b] AND KRISHNAKUMAR BALASUBRAMANIAN[3,c]

[1]*Department of Statistics and Data Science, National University of Singapore,* [a]*yanshuo@nus.edu.sg*
[2]*Department of Operations Research and Financial Engineering, Princeton University,* [b]*jason.klusowski@princeton.edu*
[3]*Department of Statistics, University of California, Davis,* [c]*kbala@ucdavis.edu*

Models based on recursive adaptive partitioning such as decision trees and their ensembles are popular for high-dimensional regression as they can potentially avoid the curse of dimensionality. Because empirical risk minimization (ERM) is computationally infeasible, these models are typically trained using greedy algorithms. Although effective in many cases, these algorithms have been empirically observed to get stuck at local optima. We explore this phenomenon in the context of learning sparse regression functions over d binary features, showing that when the true regression function f^* does not satisfy Abbe et al. (In *Conference on Learning Theory* (2022) 4782–4887 PMLR)'s Merged Staircase Property (MSP), a form of heredity restriction similar to that used in classical ANOVA modeling, greedy training requires $\exp(\Omega(d))$ samples to achieve low estimation error. Conversely, when f^* does satisfy MSP, greedy training can attain small estimation error with only $O(\log d)$ samples. This dichotomy mirrors that of two-layer neural networks trained with stochastic gradient descent (SGD) in the mean-field regime, thereby establishing a head-to-head comparison between SGD-trained neural networks and greedy recursive partitioning estimators. Furthermore, ERM-trained recursive partitioning estimators achieve low estimation error with $O(\log d)$ samples irrespective of whether f^* satisfies MSP, thereby demonstrating a statistical-computational trade-off for greedy training. Our proofs are based on a novel interpretation of greedy recursive partitioning using stochastic process theory and a coupling technique that may be of independent interest.

REFERENCES

[1] ABBE, E., ADSERA, E. B. and MISIAKIEWICZ, T. (2022). The merged-staircase property: A necessary and nearly sufficient condition for SGD learning of sparse functions on two-layer neural networks. In *Conference on Learning Theory* 4782–4887. PMLR.

[2] ABBE, E., ADSERA, E. B. and MISIAKIEWICZ, T. (2023). SGD learning on neural networks: Leap complexity and saddle-to-saddle dynamics. In *The Thirty Sixth Annual Conference on Learning Theory* 2552–2623. PMLR.

[3] ABBE, E. and BOIX-ADSERÀ, E. (2022). On the non-universality of deep learning: Quantifying the cost of symmetry. In *Advances in Neural Information Processing Systems* (A. H. Oh, A. Agarwal, D. Belgrave and K. Cho, eds.).

[4] ABBE, E., BOIX-ADSERA, E., BRENNAN, M., BRESLER, G. and NAGARAJ, D. (2021). The staircase property: How hierarchical structure can guide deep learning.

[5] ABBE, E. and SANDON, C. (2023). Polynomial-time universality and limitations of deep learning. *Comm. Pure Appl. Math.* **76** 3493–3549. MR4642823 https://doi.org/10.1002/cpa.22121

[6] AGARWAL, A., TAN, Y. S., RONEN, O., SINGH, C. and YU, B. (2022). Hierarchical shrinkage: Improving the accuracy and interpretability of tree-based models. In *International Conference on Machine Learning* 111–135. PMLR.

MSC2020 subject classifications. 62G08, 68Q32.

Key words and phrases. CART, random forest, greedy, tree, recursive partitioning, statistical-computational trade-offs.

[7] AGHAEI, S., GÓMEZ, A. and VAYANOS, P. (2025). Strong optimal classification trees. *Oper. Res.* **73** 2223–2241. MR4958675

[8] AGLIN, G., NIJSSEN, S. and SCHAUS, P. (2020). Learning optimal decision trees using caching branch-and-bound search. *Proc. AAAI Conf. Artif. Intell.* **34** 3146–3153.

[9] ATHEY, S. and IMBENS, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proc. Natl. Acad. Sci. USA* **113** 7353–7360. MR3531135 https://doi.org/10.1073/pnas.1510489113

[10] BARAK, B., EDELMAN, B., GOEL, S., KAKADE, S., MALACH, E. and ZHANG, C. (2022). Hidden progress in deep learning: SGD learns parities near the computational limit. *Adv. Neural Inf. Process. Syst.* **35** 21750–21764.

[11] BERTSIMAS, D. and DUNN, J. (2017). Optimal classification trees. *Mach. Learn.* **106** 1039–1082. MR3665788 https://doi.org/10.1007/s10994-017-5633-9

[12] BREIMAN, L. (2001). Random forests. *Mach. Learn.* **45** 5–32.

[13] BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. and STONE, C. J. (1984). *Classification and Regression Trees*. *Wadsworth Statistics/Probability Series*. Wadsworth Advanced Books and Software, Belmont, CA. MR0726392

[14] CARUANA, R., KARAMPATZIAKIS, N. and YESSENALINA, A. (2008). An empirical evaluation of supervised learning in high dimensions. In *Proceedings of the* 25*th International Conference on Machine Learning* 96–103.

[15] CATTANEO, M. D., CHANDAK, R. and KLUSOWSKI, J. M. (2024). Convergence rates of oblique regression trees for flexible function libraries. *Ann. Statist.* **52** 466–490. MR4744184 https://doi.org/10.1214/24-aos2354

[16] CATTANEO, M. D., KLUSOWSKI, J. M. and TIAN, P. M. (2022). On the pointwise behavior of recursive partitioning and its implications for heterogeneous causal effect estimation. Preprint. Available at arXiv:2211.10805.

[17] CHEN, T. and GUESTRIN, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the* 22*nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794.

[18] CHI, C.-M., VOSSLER, P., FAN, Y. and LV, J. (2022). Asymptotic properties of high-dimensional random forests. *Ann. Statist.* **50** 3415–3438. MR4524502 https://doi.org/10.1214/22-aos2234

[19] CHIPMAN, H. A., GEORGE, E. I. and MCCULLOCH, R. E. (1998). Bayesian CART model search. *J. Amer. Statist. Assoc.* **93** 935–948.

[20] CHIPMAN, H. A., GEORGE, E. I. and MCCULLOCH, R. E. (2010). BART: Bayesian additive regression trees. *Ann. Appl. Stat.* **4** 266–298. MR2758172 https://doi.org/10.1214/09-AOAS285

[21] DEMIROVIĆ, E., LUKINA, A., HEBRARD, E., CHAN, J., BAILEY, J., LECKIE, C., RAMAMOHANARAO, K. and STUCKEY, P. J. (2022). MurTree: Optimal decision trees via dynamic programming and search. *J. Mach. Learn. Res.* **23** Paper No. [26], 47. MR4420751

[22] FERNÁNDEZ-DELGADO, M., CERNADAS, E., BARRO, S. and AMORIM, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *J. Mach. Learn. Res.* **15** 3133–3181. MR3277155

[23] FRIEDMAN, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Statist.* **29** 1189–1232. MR1873328 https://doi.org/10.1214/aos/1013203451

[24] GLASGOW, M. R. (2024). SGD finds then tunes features in two-layer neural networks with near-optimal sample complexity: A case study in the XOR problem. In *The Twelfth International Conference on Learning Representations*.

[25] GRINSZTAJN, L., OYALLON, E. and VAROQUAUX, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? *Adv. Neural Inf. Process. Syst.* **35** 507–520.

[26] HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2001). *The Elements of Statistical Learning*: *Data Mining*, *Inference*, *and Prediction*. *Springer Series in Statistics*. Springer, New York. MR1851606 https://doi.org/10.1007/978-0-387-21606-5

[27] HU, X., RUDIN, C. and SELTZER, M. (2019). Optimal sparse decision trees. *Adv. Neural Inf. Process. Syst.*.

[28] ISHWARAN, H., KOGALUR, U. B., BLACKSTONE, E. H. and LAUER, M. S. (2008). Random survival forests. *Ann. Appl. Stat.* **2** 841–860. MR2516796 https://doi.org/10.1214/08-AOAS169

[29] JOSHI, N., MISIAKIEWICZ, T. and SREBRO, N. (2024). On the complexity of learning sparse functions with statistical and gradient queries. In *The Thirty-Eighth Annual Conference on Neural Information Processing Systems*.

[30] KE, G., MENG, Q., FINLEY, T., WANG, T., CHEN, W., MA, W., YE, Q. and LIU, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **30**.

[31] KLUSOWSKI, J. M. (2020). Sparse learning with CART. In *Advances in Neural Information Processing Systems* (H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan and H. Lin, eds.) **33** 11612–11622. Curran Associates, Red Hook.

[32] KLUSOWSKI, J. M. and TIAN, P. (2021). Nonparametric variable screening with optimal decision stumps. In *Proceedings of the* 24*th International Conference on Artificial Intelligence and Statistics* (A. Banerjee and K. Fukumizu, eds.). *Proceedings of Machine Learning Research* **130** 748–756. PMLR.
[33] KLUSOWSKI, J. M. and TIAN, P. M. (2024). Large scale prediction with decision trees. *J. Amer. Statist. Assoc.* **119** 525–537. MR4713911 https://doi.org/10.1080/01621459.2022.2126782
[34] KOU, Y., CHEN, Z., GU, Q. and KAKADE, S. M. (2024). Matching the statistical query lower bound for k-sparse parity problems with stochastic gradient descent. Preprint. Available at arXiv:2404.12376.
[35] LIN, J., ZHONG, C., HU, D., RUDIN, C. and SELTZER, M. (2020). Generalized and scalable optimal sparse decision trees. In *International Conference on Machine Learning* 6150–6160. PMLR.
[36] LOH, W.-Y. (2002). Regression trees with unbiased variable selection and interaction detection. *Statist. Sinica* **12** 361–386. MR1902715
[37] MAZUMDER, R. and WANG, H. (2024). On the convergence of CART under sufficient impurity decrease condition. *Adv. Neural Inf. Process. Syst.* **36**.
[38] MEI, S., MONTANARI, A. and NGUYEN, P.-M. (2018). A mean field view of the landscape of two-layer neural networks. *Proc. Natl. Acad. Sci. USA* **115** E7665–E7671. MR3845070 https://doi.org/10.1073/pnas.1806579115
[39] MENDELSON, S. (2018). Learning without concentration for general loss functions. *Probab. Theory Related Fields* **171** 459–502. MR3800838 https://doi.org/10.1007/s00440-017-0784-y
[40] MURDOCH, W. J., SINGH, C., KUMBIER, K., ABBASI-ASL, R. and YU, B. (2019). Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci. USA* **116** 22071–22080. MR4030584 https://doi.org/10.1073/pnas.1900654116
[41] OLSON, R. S., CAVA, W. L., MUSTAHSAN, Z., VARIK, A. and MOORE, J. H. (2018). Data-driven advice for applying machine learning to bioinformatics problems. In *Biocomputing* 2018: *Proceedings of the Pacific Symposium* 192–203. World Scientific, Singapore.
[42] PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A. et al. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12** 2825–2830. MR2854348
[43] RUDIN, C., CHEN, C., CHEN, Z., HUANG, H., SEMENOVA, L. and ZHONG, C. (2022). Interpretable machine learning: Fundamental principles and 10 grand challenges. *Stat. Surv.* **16** 1–85. MR4361744 https://doi.org/10.1214/21-ss133
[44] STEIN, E. M. and SHAKARCHI, R. (2003). *Fourier Analysis*: *An Introduction*. *Princeton Lectures in Analysis* **1**. Princeton Univ. Press, Princeton, NJ. MR1970295
[45] SYRGKANIS, V. and ZAMPETAKIS, M. (2020). Estimation and inference with trees and forests in high dimensions. In *Conference on Learning Theory* 3453–3454. PMLR.
[46] TAN, Y. S., AGARWAL, A. and YU, B. (2022). A cautionary tale on fitting decision trees to data from additive models: Generalization lower bounds. In *International Conference on Artificial Intelligence and Statistics* 9663-9685. PMLR.
[47] TAN, Y. S, KLUSOWSKI, J. M and BALASUBRAMANIAN, K. (2026). Supplement to "Statistical-computational trade-offs for recursive adaptive partitioning estimators." https://doi.org/10.1214/25-AOS2603SUPP
[48] TAN, Y. S., RONEN, O., SAARINEN, T. and YU, B. (2024). The computational curse of big data for Bayesian Additive Regression Trees: a hitting time analysis. Preprint. Available at arXiv:2406.19958.
[49] TAN, Y. S. and VERSHYNIN, R. (2023). Online stochastic gradient descent with arbitrary initialization solves non-smooth, non-convex phase retrieval. *J. Mach. Learn. Res.* **24** Paper No. [58], 47. MR4582480
[50] TANG, C., GARREAU, D. and VON LUXBURG, U. (2018). When do random forests fail? In *Advances in Neural Information Processing Systems* **31** (S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi and R. Garnett, eds.). Curran Associates, Red Hook.
[51] UHLER, C., RASKUTTI, G., BÜHLMANN, P. and YU, B. (2013). Geometry of the faithfulness assumption in causal inference. *Ann. Statist.* **41** 436–463. MR3099109 https://doi.org/10.1214/12-AOS1080
[52] VERWER, S. and ZHANG, Y. (2019). Learning optimal classification trees using a binary linear program formulation. *Proc. AAAI Conf. Artif. Intell.* **33** 1625–1632.
[53] WAHBA, G. (1990). *Spline Models for Observational Data*. *CBMS-NSF Regional Conference Series in Applied Mathematics* **59**. SIAM, Philadelphia, PA. MR1045442 https://doi.org/10.1137/1.9781611970128
[54] 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
[55] WASSERMAN, L. and LAFFERTY, J. (2005). Rodeo: Sparse nonparametric regression in high dimensions. *Adv. Neural Inf. Process. Syst.* **18**.
[56] WRIGHT, M. N., ZIEGLER, A. et al. (2017). ranger: A fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Softw.* **77**.

[57] WU, C. F. J. and HAMADA, M. S. (2009). *Experiments*: *Planning*, *Analysis*, *and Optimization*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2583259

[58] ZHANG, R., XIN, R., SELTZER, M. and RUDIN, C. (2023). Optimal sparse regression trees. *Proc. AAAI Conf. Artif. Intell.* **37** 11270–11279.

[59] ZHU, H., MURALI, P., PHAN, D., NGUYEN, L. and KALAGNANAM, J. (2020). A scalable MIP-based method for learning optimal multivariate decision trees. *Adv. Neural Inf. Process. Syst.* **33** 1771–1781.

The Annals of Statistics
2026, Vol. 54, No. 3, 1289–1314
https://doi.org/10.1214/25-AOS2604

VECCHIA GAUSSIAN PROCESSES: ON PROBABILISTIC AND STATISTICAL PROPERTIES

BY BOTOND SZABO[1,a] AND YICHEN ZHU[2,b]

[1]*Bocconi Institute for Data Science and Analytics and Department of Decision Sciences, Bocconi University,* [a]*botond.szabo@unibocconi.it*

[2]*Department of Statistics and Actuarial Science, University of Hong Kong,* [b]*yczhu@hku.hk*

Gaussian Processes (GPs) are widely used to model dependencies in spatial statistics and machine learning. However, exact inference is computationally intractable for GP regression, with a time complexity of $O(n^3)$. The Vecchia approximation scales up computation by introducing sparsity into the spatial dependency structure, represented by a directed acyclic graph (DAG). Despite its practical popularity, this approach lacks rigorous theoretical foundations, and the choice of DAG structure remains an open problem.

In this paper, we systematically study the Vecchia approximation of the popular, isotropic Matérn GP as standalone stochastic process and uncover key probabilistic and statistical properties. We propose selecting parent sets as norming sets with fixed cardinality in the Vecchia approximation. On the probabilistic side, we show that the conditional distributions of Matérn GPs, as well as their Vecchia approximations, can be characterized by polynomial interpolations. This enables us to establish several results on small ball probabilities and the Reproducing Kernel Hilbert Spaces (RKHSs) of Vecchia GPs. Building on these probabilistic results, we prove that in the nonparametric regression model, the corresponding posterior contracts around the truth at the optimal minimax rate, both under oracle rescaling and hierarchical tuning of the prior.

We illustrate the theoretical findings through numerical experiments on synthetic datasets. Our core algorithms are implemented in C++ with an R interface.

REFERENCES

[1] AURZADA, F., IBRAGIMOV, I. A., LIFSHITS, M. A. and VAN ZANTEN, J. (2009). Small deviations of smooth stationary Gaussian processes. *Teor. Veroyatn. Primen.* **53** 697–707. MR2766145 https://doi.org/10.1137/S0040585X97983912

[2] BAI, Y., SONG, P. X.-K. and RAGHUNATHAN, T. E. (2012). Joint composite estimating functions in spatiotemporal models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 799–824. MR2988907 https://doi.org/10.1111/j.1467-9868.2012.01035.x

[3] BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2015). *Hierarchical Modeling and Analysis for Spatial Data*, 2nd ed. *Monographs on Statistics and Applied Probability* **135**. CRC Press, Boca Raton, FL. MR3362184

[4] BANERJEE, S., GELFAND, A. E., FINLEY, A. O. and SANG, H. (2008). Gaussian predictive process models for large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 825–848. MR2523906 https://doi.org/10.1111/j.1467-9868.2008.00663.x

[5] BEVILACQUA, M. and GAETAN, C. (2015). Comparing composite likelihood methods based on pairs for spatial Gaussian random fields. *Stat. Comput.* **25** 877–892. MR3375623 https://doi.org/10.1007/s11222-014-9460-6

[6] BHATT, S., WEISS, D., CAMERON, E., BISANZIO, D., MAPPIN, B., DALRYMPLE, U., BATTLE, K., MOYES, C., HENRY, A. et al. (2015). The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature* **526** 207–211.

[7] BOS, L. (2018). Fekete points as norming sets. *Dolomites Res. Not. Approx.* **11** 26–34. MR3895933

MSC2020 subject classifications. Primary 62G08; secondary 60G15, 62H11.
Key words and phrases. Gaussian process, minimax rate, norming set, Vecchia approximation.

[8] Cao, J., Guinness, J., Genton, M. G. and Katzfuss, M. (2022). Scalable Gaussian-process regression and variable selection using Vecchia approximations. *J. Mach. Learn. Res.* **23** Paper No. [348], 30. MR4577787

[9] Cressie, N. and Johannesson, G. (2008). Fixed rank Kriging for very large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 209–226. MR2412639 https://doi.org/10.1111/j.1467-9868.2007.00633.x

[10] Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2848400

[11] Datta, A., Banerjee, S., Finley, A. O. and Gelfand, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *J. Amer. Statist. Assoc.* **111** 800–812. MR3538706 https://doi.org/10.1080/01621459.2015.1044091

[12] Daxberger, E. A. and Low, B. K. H. (2017). Distributed batch Gaussian process optimization. In *International Conference on Machine Learning* 951–960. PMLR.

[13] Deisenroth, M. and Ng, J. W. (2015). Distributed Gaussian processes. In *International Conference on Machine Learning* 1481–1490. PMLR.

[14] Driscoll, T. A. and Fornberg, B. (2002). Interpolation in the limit of increasingly flat radial basis functions. *Comput. Math. Appl.* **43** 413–422. MR1883576 https://doi.org/10.1016/S0898-1221(01)00295-4

[15] Fang, X. and Bhadra, A. (2025). Posterior concentration for Gaussian process priors under rescaled and hierarchical Matérn and confluent hypergeometric covariance functions. *Electron. J. Stat.* **19** 3333–3369. MR4947572 https://doi.org/10.1214/25-ejs2417

[16] Furrer, R., Genton, M. G. and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *J. Comput. Graph. Statist.* **15** 502–523. MR2291261 https://doi.org/10.1198/106186006X132178

[17] Gelfand, A. E., Diggle, P. J., Fuentes, M. and Guttorp, P. (2010). *Handbook of Spatial Statistics*. *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. CRC Press, Boca Raton, FL. MR2761512 https://doi.org/10.1201/9781420072884

[18] Gelfand, A. E. and Schliep, E. M. (2016). Spatial statistics and Gaussian processes: A beautiful marriage. *Spat. Stat.* **18** 86–104. MR3573271 https://doi.org/10.1016/j.spasta.2016.03.006

[19] Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. *Cambridge Series in Statistical and Probabilistic Mathematics* **44**. Cambridge Univ. Press, Cambridge. MR3587782 https://doi.org/10.1017/9781139029834

[20] Guinness, J. (2018). Permutation and grouping methods for sharpening Gaussian process approximations. *Technometrics* **60** 415–429. MR3878098 https://doi.org/10.1080/00401706.2018.1437476

[21] Jetter, K., Stöckler, J. and Ward, J. D. (1999). Error estimates for scattered data interpolation on spheres. *Math. Comp.* **68** 733–747. MR1642746 https://doi.org/10.1090/S0025-5718-99-01080-7

[22] Kang, M., Schäfer, F., Guinness, J. and Katzfuss, M. (2024). Asymptotic properties of Vecchia approximation for Gaussian processes. arXiv preprint. Available at arXiv:2401.15813.

[23] Katzfuss, M. and Guinness, J. (2021). A general framework for Vecchia approximations of Gaussian processes. *Statist. Sci.* **36** 124–141. MR4194207 https://doi.org/10.1214/19-STS755

[24] Katzfuss, M., Guinness, J., Gong, W. and Zilber, D. (2020). Vecchia approximations of Gaussian-process predictions. *J. Agric. Biol. Environ. Stat.* **25** 383–414. MR4139037 https://doi.org/10.1007/s13253-020-00401-7

[25] Kaufman, C. G., Schervish, M. J. and Nychka, D. W. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *J. Amer. Statist. Assoc.* **103** 1545–1555. MR2504203 https://doi.org/10.1198/016214508000000959

[26] Kuelbs, J., Li, W. V. and Linde, W. (1994). The Gaussian measure of shifted balls. *Probab. Theory Related Fields* **98** 143–162. MR1258983 https://doi.org/10.1007/BF01192511

[27] Larsson, E. and Fornberg, B. (2005). Theoretical and computational aspects of multivariate interpolation with increasingly flat radial basis functions. *Comput. Math. Appl.* **49** 103–130. MR2123189 https://doi.org/10.1016/j.camwa.2005.01.010

[28] Ledoux, M. (2006). Isoperimetry and Gaussian analysis. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1994). *Lecture Notes in Math.* **1648** 165–294. Springer, Berlin. MR1600888 https://doi.org/10.1007/BFb0095676

[29] Lee, J., Bahri, Y., Novak, R., Schoenholz, S. S., Pennington, J. and Sohl-Dickstein, J. (2017). Deep neural networks as Gaussian processes. arXiv preprint. Available at arXiv:1711.00165.

[30] Lee, Y. J., Micchelli, C. A. and Yoon, J. (2015). A study on multivariate interpolation by increasingly flat kernel functions. *J. Math. Anal. Appl.* **427** 74–87. MR3318187 https://doi.org/10.1016/j.jmaa.2015.02.006

[31] Li, W. V. and Linde, W. (1999). Approximation, metric entropy and small ball estimates for Gaussian measures. *Ann. Probab.* **27** 1556–1578. MR1733160 https://doi.org/10.1214/aop/1022677459

[32] LI, W. V. and SHAO, Q.-M. (2001). Gaussian processes: Inequalities, small ball probabilities and applications. In *Stochastic Processes*: *Theory and Methods*. *Handbook of Statist*. **19** 533–597. North-Holland, Amsterdam. MR1861734 https://doi.org/10.1016/S0169-7161(01)19019-X

[33] LIU, H., ONG, Y.-S., SHEN, X. and CAI, J. (2020). When Gaussian process meets big data: A review of scalable GPs. *IEEE Trans. Neural Netw. Learn. Syst.* **31** 4405–4423. MR4169962 https://doi.org/10.1109/tnnls.2019.2957109

[34] NEIDINGER, R. D. (2019). Multivariate polynomial interpolation in Newton forms. *SIAM Rev.* **61** 361–381. MR3947286 https://doi.org/10.1137/17M1124188

[35] PERUZZI, M., BANERJEE, S. and FINLEY, A. O. (2022). Highly scalable Bayesian geostatistical modeling via meshed Gaussian processes on partitioned domains. *J. Amer. Statist. Assoc.* **117** 969–982. MR4436326 https://doi.org/10.1080/01621459.2020.1833889

[36] QUIÑONERO-CANDELA, J. and RASMUSSEN, C. E. (2005). A unifying view of sparse approximate Gaussian process regression. *J. Mach. Learn. Res.* **6** 1939–1959. MR2249877

[37] RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR2514435

[38] 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

[39] ROUSSEAU, J. and SZABO, B. (2020). Asymptotic frequentist coverage properties of Bayesian credible sets for sieve priors. *Ann. Statist.* **48** 2155–2179. MR4134790 https://doi.org/10.1214/19-AOS1881

[40] SCHÄFER, F., SULLIVAN, T. J. and OWHADI, H. (2021). Compression, inversion, and approximate PCA of dense kernel matrices at near-linear computational complexity. *Multiscale Model. Simul.* **19** 688–730. MR4243658 https://doi.org/10.1137/19M129526X

[41] SEEGER, M. (2004). Gaussian processes for machine learning. *Int. J. Neural Syst.* **14** 69–106.

[42] SONG, G., RIDDLE, J., FASSHAUER, G. E. and HICKERNELL, F. J. (2012). Multivariate interpolation with increasingly flat radial basis functions of finite smoothness. *Adv. Comput. Math.* **36** 485–501. MR2893476 https://doi.org/10.1007/s10444-011-9192-5

[43] STEIN, M. L. (2013). Statistical properties of covariance tapers. *J. Comput. Graph. Statist.* **22** 866–885. MR3173747 https://doi.org/10.1080/10618600.2012.719844

[44] SZABO, B., HADJI, A. and VAN DER VAART, A. (2025). Adaptation using spatially distributed Gaussian processes. *J. Amer. Statist. Assoc.* **120** 2784–2795. MR5006881 https://doi.org/10.1080/01621459.2025.2501717

[45] SZABÓ, B. and VAN ZANTEN, H. (2019). An asymptotic analysis of distributed nonparametric methods. *J. Mach. Learn. Res.* **20** Paper No. 87, 30. MR3960941

[46] SZABÓ, B., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2015). Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Ann. Statist.* **43** 1391–1428. MR3357861 https://doi.org/10.1214/14-AOS1270

[47] SZABO, B. and ZHU, Y. (2026). Supplement to "Vecchia Gaussian Processes: on probabilistic and statistical properties." https://doi.org/10.1214/25-AOS2604SUPPA, https://doi.org/10.1214/25-AOS2604SUPPB

[48] SZABÓ, B. T., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2013). Empirical Bayes scaling of Gaussian priors in the white noise model. *Electron. J. Stat.* **7** 991–1018. MR3044507 https://doi.org/10.1214/13-EJS798

[49] TITSIAS, M. (2009). Variational learning of inducing variables in sparse Gaussian processes. In *Proceedings of the* 12*th International Conference on Artificial. Intelligence and Statistics* **5** 567–574. PMLR.

[50] VAN DER VAART, A. and VAN ZANTEN, H. (2011). Information rates of nonparametric Gaussian process methods. *J. Mach. Learn. Res.* **12** 2095–2119. MR2819028

[51] VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2008). Rates of contraction of posterior distributions based on Gaussian process priors. *Ann. Statist.* **36** 1435–1463. MR2418663 https://doi.org/10.1214/009053607000000613

[52] VECCHIA, A. V. (1988). Estimation and model identification for continuous spatial processes. *J. Roy. Statist. Soc. Ser. B, Methodol.* **50** 297–312. MR0964183

[53] WENDLAND, H. (2004). *Scattered Data Approximation*. *Cambridge Monographs on Applied and Computational Mathematics* **17**. Cambridge Univ. Press, Cambridge. MR2131724

[54] ZHANG, L., TANG, W. and BANERJEE, S. (2024). Fixed-domain asymptotics under Vecchia's approximation of spatial process likelihoods. *Statist. Sinica* **34** 1863–1881. MR4805552

[55] ZHU, Y., PERUZZI, M., LI, C. and DUNSON, D. B. (2024). Radial neighbours for provably accurate scalable approximations of Gaussian processes. *Biometrika* **111** 1151–1167. MR4830051 https://doi.org/10.1093/biomet/asae029

The Annals of Statistics
2026, Vol. 54, No. 3, 1315–1344
https://doi.org/10.1214/25-AOS2606

UNCERTAINTY QUANTIFICATION FOR ITERATIVE ALGORITHMS IN LINEAR MODELS WITH APPLICATION TO EARLY STOPPING

BY PIERRE C. BELLEC[1,a] AND KAI TAN[2,b]

[1]*Department of Statistics, Rutgers University,* [a]*pierre.bellec@rutgers.edu*
[2]*Department of Statistics, The Ohio State University,* [b]*tan.1457@osu.edu*

This paper investigates the iterates $\widehat{\boldsymbol{b}}^1, \ldots, \widehat{\boldsymbol{b}}^T$ obtained from iterative algorithms in high-dimensional linear regression problems, in the regime where the feature dimension p is comparable with the sample size n, that is, $p \asymp n$. The analysis and proposed estimators are applicable to Gradient Descent (GD), proximal GD and their accelerated variants such as Fast Iterative Soft-Thresholding (FISTA). The paper proposes novel estimators for the generalization error of the iterate $\widehat{\boldsymbol{b}}^t$ for any fixed iteration t along the trajectory. These estimators are proved to be $\sqrt{n}$-consistent under Gaussian designs. Applications to early-stopping are provided: when the generalization error of the iterates is a U-shape function of the iteration t, the estimates allow to select from the data an iteration $\hat{t}$ that achieves the smallest generalization error along the trajectory. Additionally, we provide a technique for developing debiasing corrections and valid confidence intervals for the components of the true coefficient vector from the iterate $\widehat{\boldsymbol{b}}^t$ at any finite iteration t. Extensive simulations on synthetic data illustrate the theoretical results.

REFERENCES

AGARWAL, A., NEGAHBAN, S. and WAINWRIGHT, M. J. (2012). Fast global convergence of gradient methods for high-dimensional statistical recovery. *Ann. Statist.* **40** 2452–2482. MR3097609 https://doi.org/10.1214/12-AOS1032

ALI, A., DOBRIBAN, E. and TIBSHIRANI, R. (2020). The implicit regularization of stochastic gradient flow for least squares. In *International Conference on Machine Learning* 233–244. PMLR.

ALI, A., KOLTER, J. Z. and TIBSHIRANI, R. J. (2019). A continuous-time view of early stopping for least squares regression. In *The* 22*nd International Conference on Artificial Intelligence and Statistics* 1370–1378. PMLR.

AMBROSIO, L. and DAL MASO, G. (1990). A general chain rule for distributional derivatives. *Proc. Amer. Math. Soc.* **108** 691–702. MR0969514 https://doi.org/10.2307/2047789

AUDDY, A., ZOU, H., RAHNAMARAD, K. and MALEKI, A. (2024). Approximate leave-one-out cross validation for regression with ℓ_1 regularizers. In *Proceedings of the* 27*th International Conference on Artificial Intelligence and Statistics* 2377–2385. PMLR.

BANDEIRA, A. S., DOBRIBAN, E., MIXON, D. G. and SAWIN, W. F. (2013). Certifying the restricted isometry property is hard. *IEEE Trans. Inf. Theory* **59** 3448–3450. MR3061257 https://doi.org/10.1109/TIT.2013.2248414

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

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

BECK, A. and TEBOULLE, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2** 183–202. MR2486527 https://doi.org/10.1137/080716542

BELLEC, P. C. (2023). Out-of-sample error estimation for M-estimators with convex penalty. *Inf. Inference* **12** Paper No. iaad031, 36. MR4660702 https://doi.org/10.1093/imaiai/iaad031

BELLEC, P. C. (2025). Observable adjustments in single-index models for regularized M-estimators with bounded p/n. *Ann. Statist.* **53** 531–560. MR4900157 https://doi.org/10.1214/24-aos2464

MSC2020 subject classifications. 62J07.
Key words and phrases. High-dimensional models, risk estimate, Lasso, generalization error.

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

BELLEC, P. C. and ROMON, G. (2021). Chi-square and normal inference in high-dimensional multi-task regression. arXiv preprint. Available at arXiv:2107.07828.

BELLEC, P. C. and SHEN, Y. (2022). Derivatives and residual distribution of regularized m-estimators with application to adaptive tuning. In *Conference on Learning Theory* 1912–1947. PMLR.

BELLEC, P. C. and TAN, K. (2026). Supplement to "Uncertainty quantification for iterative algorithms in linear models with application to early stopping." https://doi.org/10.1214/25-AOS2606SUPPA, https://doi.org/10.1214/25-AOS2606SUPPB

BELLEC, P. C. and ZHANG, C.-H. (2022). De-biasing the Lasso with degrees-of-freedom adjustment. *Bernoulli* **28** 713–743. MR4389062 https://doi.org/10.3150/21-BEJ1348

BELLEC, P. C. and ZHANG, C.-H. (2023). Debiasing convex regularized estimators and interval estimation in linear models. *Ann. Statist.* **51** 391–436. MR4600987 https://doi.org/10.1214/22-aos2243

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

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

BLUMENSATH, T. and DAVIES, M. E. (2009). Iterative hard thresholding for compressed sensing. *Appl. Comput. Harmon. Anal.* **27** 265–274. MR2559726 https://doi.org/10.1016/j.acha.2009.04.002

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

BOTTOU, L. (2010). Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010* 177–186. Physica-Verlag, Heidelberg. MR3362066

BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data*: *Methods*, *Theory and Applications*. *Springer Series in Statistics*. Springer, Heidelberg. MR2807761 https://doi.org/10.1007/978-3-642-20192-9

BÜHLMANN, P. and YU, B. (2003). Boosting with the L_2 loss: Regression and classification. *J. Amer. Statist. Assoc.* **98** 324–339. MR1995709 https://doi.org/10.1198/016214503000125

CAI, T. T. and GUO, Z. (2017). Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *Ann. Statist.* **45** 615–646. MR3650395 https://doi.org/10.1214/16-AOS1461

CARPENTIER, A. and KIM, A. K. H. (2018). An iterative hard thresholding estimator for low rank matrix recovery with explicit limiting distribution. *Statist. Sinica* **28** 1371–1393. MR3821009

CELENTANO, M., CHENG, C. and MONTANARI, A. (2021). The high-dimensional asymptotics of first order methods with random data. arXiv preprint. Available at arXiv:2112.07572.

CELENTANO, M. and MONTANARI, A. (2022). Fundamental barriers to high-dimensional regression with convex penalties. *Ann. Statist.* **50** 170–196. MR4382013 https://doi.org/10.1214/21-aos2100

CELENTANO, M., MONTANARI, A. and WEI, Y. (2023). The Lasso with general Gaussian designs with applications to hypothesis testing. *Ann. Statist.* **51** 2194–2220. MR4678801 https://doi.org/10.1214/23-aos2327

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

CHANDRASEKHER, K. A., LOU, M. and PANANJADY, A. (2024). Alternating minimization for generalized rank-1 matrix sensing: Sharp predictions from a random initialization. *Inf. Inference* **13** Paper No. iaae025, 94. MR4798949 https://doi.org/10.1093/imaiai/iaae025

CHANDRASEKHER, K. A., PANANJADY, A. and THRAMPOULIDIS, C. (2023). Sharp global convergence guarantees for iterative nonconvex optimization with random data. *Ann. Statist.* **51** 179–210. MR4564853 https://doi.org/10.1214/22-aos2246

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

DAUBECHIES, I., DEFRISE, M. and DE MOL, C. (2004). An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.* **57** 1413–1457. MR2077704 https://doi.org/10.1002/cpa.20042

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

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

DUDEJA, R., LU, Y. M. and SEN, S. (2023). Universality of approximate message passing with semirandom matrices. *Ann. Probab.* **51** 1616–1683. MR4642220 https://doi.org/10.1214/23-aop1628

El Alaoui, A., Montanari, A. and Sellke, M. (2022). Sampling from the Sherrington-Kirkpatrick Gibbs measure via algorithmic stochastic localization. In 2022 *IEEE 63rd Annual Symposium on Foundations of Computer Science—FOCS* 2022 323–334. IEEE Comput. Soc., Los Alamitos, CA. MR4537214 https://doi.org/10.1109/FOCS54457.2022.00038

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

El Karoui, N., Bean, D., Bickel, P. J., Lim, C. and Yu, B. (2013). On robust regression with high-dimensional predictors. *Proc. Natl. Acad. Sci. USA* **110** 14557–14562.

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

Feng, L. and Zhang, C.-H. (2019). Sorted concave penalized regression. *Ann. Statist.* **47** 3069–3098. MR4025735 https://doi.org/10.1214/18-AOS1759

Friedman, J. H., Hastie, T. and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33** 1–22.

Gerbelot, C., Troiani, E., Mignacco, F., Krzakala, F. and Zdeborová, L. (2024). Rigorous dynamical mean-field theory for stochastic gradient descent methods. *SIAM J. Math. Data Sci.* **6** 400–427. MR4741502 https://doi.org/10.1137/23M1594388

Han, Q. and Shen, Y. (2023). Universality of regularized regression estimators in high dimensions. *Ann. Statist.* **51** 1799–1823. MR4658577 https://doi.org/10.1214/23-aos2309

Hastie, T., Montanari, A., Rosset, S. and Tibshirani, R. J. (2022). Surprises in high-dimensional ridgeless least squares interpolation. *Ann. Statist.* **50** 949–986. MR4404925 https://doi.org/10.1214/21-aos2133

Hoppe, F., Mayrink Verdun, C., Laus, H., Krahmer, F. and Rauhut, H. (2023). Uncertainty quantification for learned ista. In 2023 *IEEE 33rd International Workshop on Machine Learning for Signal Processing* (*MLSP*) 1–6. IEEE.

Hu, H. and Lu, Y. M. (2023). Universality laws for high-dimensional learning with random features. *IEEE Trans. Inf. Theory* **69** 1932–1964. MR4564688 https://doi.org/10.1109/tit.2022.3217698

Hutchinson, M. F. (1990). A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *Comm. Statist. Simulation Comput.* **19** 433–450. MR1075456 https://doi.org/10.1080/03610919008812864

Javanmard, A. and Montanari, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909. MR3277152

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

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

Lei, L., Bickel, P. J. and El Karoui, N. (2018). Asymptotics for high dimensional regression M-estimates: Fixed design results. *Probab. Theory Related Fields* **172** 983–1079. MR3877551 https://doi.org/10.1007/s00440-017-0824-7

Lou, M. (2024). Kabir Aladin Verchand, and Ashwin Pananjady. Hyperparameter tuning via trajectory predictions: Stochastic prox-linear methods in matrix sensing. arXiv preprint. Available at arXiv:2402.01599.

Loureiro, B., Gerbelot, C., Cui, H., Goldt, S., Krzakala, F., Mézard, M. and Zdeborová, L. (2022). Learning curves of generic features maps for realistic datasets with a teacher-student model. *J. Stat. Mech. Theory Exp.* **11** Paper No. 114001, 78. MR4535572 https://doi.org/10.1088/1742-5468/ac9825

Luo, Y., Ren, Z. and Barber, R. (2023). Iterative approximate cross-validation. In *International Conference on Machine Learning* 23083–23102. PMLR.

Mignacco, F., Krzakala, F., Urbani, P. and Zdeborová, L. (2021). Dynamical mean-field theory for stochastic gradient descent in Gaussian mixture classification. *J. Stat. Mech. Theory Exp.* **12** Paper No. 124008, 23. MR4412836 https://doi.org/10.1088/1742-5468/ac3a80

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

Montanari, A. and Saeed, B. N. (2022). Universality of empirical risk minimization. In *Conference on Learning Theory* 4310–4312. PMLR.

Montanari, A. and Venkataramanan, R. (2021). Estimation of low-rank matrices via approximate message passing. *Ann. Statist.* **49** 321–345. MR4206680 https://doi.org/10.1214/20-AOS1958

Montanari, A. and Wu, Y. (2024). Statistically optimal firstorder algorithms: A proof via orthogonalization. *Inf. Inference* **13** Paper No. iaae027, 24. MR4806510 https://doi.org/10.1093/imaiai/iaae027

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. E. (1983). A method for solving the convex programming problem with convergence rate $O(1/k^2)$. *Dokl. Akad. Nauk SSSR* **269** 543–547. MR0701288

PATIL, P., WU, Y. and TIBSHIRANI, R. J. (2024). Failures and successes of cross-validation for early-stopped gradient descent. arXiv preprint. Available at arXiv:2402.16793.

PESCE, L., KRZAKALA, F., LOUREIRO, B. and STEPHAN, L. (2023). Are Gaussian data all you need? The extents and limits of universality in high-dimensional generalized linear estimation. In *International Conference on Machine Learning* 27680–27708. PMLR.

RAHNAMA RAD, K. and MALEKI, A. (2020). A scalable estimate of the out-of-sample prediction error via approximate leave-one-out cross-validation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 965–996. MR4136500

RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2014). Early stopping and non-parametric regression: An optimal data-dependent stopping rule. *J. Mach. Learn. Res.* **15** 335–366. MR3190843

ROSSET, S., ZHU, J. and HASTIE, T. (2004). Boosting as a regularized path to a maximum margin classifier. *J. Mach. Learn. Res.* **5** 941–973. MR2248005

SHENG, Y. and ALI, A. (2022). Accelerated gradient flow: Risk, stability, and implicit regularization. arXiv preprint. Available at arXiv:2201.08311.

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

TAN, K., ROMON, G. and BELLEC, P. C. (2024). Noise covariance estimation in multi-task high-dimensional linear models. *Bernoulli* **30** 1695–1722. MR4746046 https://doi.org/10.3150/23-BEJ1644

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

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

VAN DE GEER, S. (2016). *Estimation and Testing Under Sparsity*: *École d'Été de Probabilités de Saint-Flour XLV*–2015. *Lecture Notes in Math.* **2159**. Springer, Cham. MR3526202 https://doi.org/10.1007/978-3-319-32774-7

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

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

WAINWRIGHT, M. J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso). *IEEE Trans. Inf. Theory* **55** 2183–2202. MR2729873 https://doi.org/10.1109/TIT.2009.2016018

WANG, S., WENG, H. and MALEKI, A. (2020). Which bridge estimator is the best for variable selection? *Ann. Statist.* **48** 2791–2823. MR4152121 https://doi.org/10.1214/19-AOS1906

YE, F. and ZHANG, C.-H. (2010). Rate minimaxity of the Lasso and Dantzig selector for the ℓ_q loss in ℓ_r balls. *J. Mach. Learn. Res.* **11** 3519–3540. MR2756192

YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 49–67. MR2212574 https://doi.org/10.1111/j.1467-9868.2005.00532.x

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

ZHANG, C.-H. and ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 217–242. MR3153940 https://doi.org/10.1111/rssb.12026

ZIEMER, W. P. (1989). *Weakly Differentiable Functions*: *Sobolev Spaces and Functions of Bounded Variation*. *Graduate Texts in Mathematics* **120**. Springer, New York. MR1014685 https://doi.org/10.1007/978-1-4612-1015-3

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

The Annals of Statistics
2026, Vol. 54, No. 3, 1345–1371
https://doi.org/10.1214/25-AOS2607

MARKOV STICK-BREAKING PROCESSES

BY MARÍA F. GIL-LEYVA[1,a], ANTONIO LIJOI[2,c], RAMSÉS H. MENA[1,b] AND IGOR PRÜNSTER[2,d]

[1]*Department of Probability and Statistics, IIMAS, UNAM,* [a]*marifer@sigma.iimas.unam.mx,* [b]*ramses@sigma.iimas.unam.mx*
[2]*Bocconi Institute for Data Science and Analytics, Bocconi University,* [c]*antonio.lijoi@unibocconi.it,* [d]*igor@unibocconi.it*

Stick-breaking has a long history and is one of the most popular procedures for constructing random discrete distributions in statistics and machine learning. In particular, due to their intuitive construction and computational tractability they are ubiquitous in modern Bayesian nonparametric inference. Most widely used models, such as the Dirichlet and the Pitman–Yor processes, rely on i.i.d. or independent length variables. Here, we pursue a completely unexplored research direction by considering Markov length variables and investigate the corresponding general class of stick-breaking processes, which we term *Markov stick-breaking processes*. We establish conditions under which the associated species sampling process is proper and the distribution of a Markov stick-breaking process has full topological support, two fundamental *desiderata* for Bayesian nonparametric models. We also analyze the stochastic ordering of the weights and provide a new characterization of the Pitman–Yor process as the only stick-breaking process invariant under size-biased permutations, under mild conditions. Moreover, we identify two notable subclasses of Markov stick-breaking processes that enjoy appealing properties and include Dirichlet, Pitman–Yor and geometric priors as special cases. Our findings include distributional results enabling posterior inference algorithms and methodological insights.

REFERENCES

ARRATIA, R., BARBOUR, A. D. and TAVARÉ, S. (2003). *Logarithmic Combinatorial Structures*: *A Probabilistic Approach*. *EMS Monographs in Mathematics*. European Mathematical Society (EMS), Zürich. MR2032426 https://doi.org/10.4171/000

BISSIRI, P. G. and ONGARO, A. (2014). On the topological support of species sampling priors. *Electron*. *J*. *Stat*. **8** 861–882. MR3229100 https://doi.org/10.1214/14-EJS912

BLACKWELL, D. and MACQUEEN, J. B. (1973). Ferguson distributions via Pólya urn schemes. *Ann*. *Statist*. **1** 353–355. MR0362614

CAMERLENGHI, F., CORRADIN, R. and ONGARO, A. (2024). Contaminated Gibbs-type priors. *Bayesian Anal*. **19** 347–376. MR4729435 https://doi.org/10.1214/22-ba1358

CARON, F., DAVY, M. and DOUCET, A. (2007). Generalized Polya Urn for time-varying Dirichlet process mixtures. In 23*rd Conference on Uncertainty in Artificial Intelligence* (*UAI'*2007).

CARON, F., NEISWANGER, W., WOOD, F., DOUCET, A. and DAVY, M. (2017). Generalized Pólya urn for time-varying Pitman–Yor processes. *J*. *Mach*. *Learn*. *Res*. **18** Paper No. 27, 32. MR3634894

DE BLASI, P., FAVARO, S., LIJOI, A., MENA, R. H., PRÜNSTER, I. and RUGGIERO, M. (2015). Are Gibbs-type priors the most natural generalization of the Dirichlet process? *IEEE Trans*. *Pattern Anal*. *Mach*. *Intell*. **37** 212–229.

DE BLASI, P. and GIL-LEYVA, M. F. (2023). Gibbs sampling for mixtures in order of appearance: The ordered allocation sampler. *J*. *Comput*. *Graph*. *Statist*. **32** 1416–1424. MR4669257 https://doi.org/10.1080/10618600.2023.2177298

DE BLASI, P., MENA, R. H. and PRÜNSTER, I. (2022). Asymptotic behavior of the number of distinct values in a sample from the geometric stick-breaking process. *Ann*. *Inst*. *Statist*. *Math*. **74** 143–165. MR4358167 https://doi.org/10.1007/s10463-021-00791-6

MSC2020 subject classifications. 62F15, 62C10, 60G57.

Key words and phrases. Bayesian nonparametrics, Dirichlet process, Markovian dependence, species sampling model, size-biased permutation, stick-breaking construction, Pitman–Yor process.

ESCOBAR, M. D. and WEST, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90** 577–588. MR1340510

FAVARO, S., LIJOI, A., NAVA, C., NIPOTI, B., PRÜNSTER, I. and TEH, Y. W. (2016). On the stick-breaking representation for homogeneous NRMIs. *Bayesian Anal.* **11** 697–724. MR3498043 https://doi.org/10.1214/15-BA964

FAVARO, S., LIJOI, A. and PRÜNSTER, I. (2012). On the stick-breaking representation of normalized inverse Gaussian priors. *Biometrika* **99** 663–674. MR2966776 https://doi.org/10.1093/biomet/ass023

FAVARO, S., LIJOI, A. and PRÜNSTER, I. (2013). Conditional formulae for Gibbs-type exchangeable random partitions. *Ann. Appl. Probab.* **23** 1721–1754. MR3114915 https://doi.org/10.1214/12-AAP843

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230. MR0350949

FREEDMAN, D. A. (1963). On the asymptotic behavior of Bayes' estimates in the discrete case. *Ann. Math. Statist.* **34** 1386–1403. MR0158483 https://doi.org/10.1214/aoms/1177703871

FUENTES-GARCÍA, R., MENA, R. H. and WALKER, S. G. (2010). A new Bayesian nonparametric mixture model. *Comm. Statist. Simulation Comput.* **39** 669–682. MR2785657 https://doi.org/10.1080/03610910903580963

GHOSAL, S. and VAN DER VAART, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. *Cambridge Series in Statistical and Probabilistic Mathematics* **44**. Cambridge Univ. Press, Cambridge. MR3587782 https://doi.org/10.1017/9781139029834

GIL-LEYVA, M. F, LIJOI, A., MENA, R. H and PRÜNSTER, I. (2026). Supplement to "Markov stick-breaking processes." https://doi.org/10.1214/25-AOS2607SUPP

GIL-LEYVA, M. F. and MENA, R. H. (2023). Stick-breaking processes with exchangeable length variables. *J. Amer. Statist. Assoc.* **118** 537–550. MR4571140 https://doi.org/10.1080/01621459.2021.1941054

GIL-LEYVA, M. F., MENA, R. H. and NICOLERIS, T. (2020). Beta-binomial stick-breaking non-parametric prior. *Electron. J. Stat.* **14** 1479–1507. MR4082475 https://doi.org/10.1214/20-EJS1694

GIORDANO, R., LIU, R., JORDAN, M. I. and BRODERICK, T. (2023). Evaluating sensitivity to the stick-breaking prior in Bayesian nonparametrics (with discussion). *Bayesian Anal.* **18** 287–366. MR4557246 https://doi.org/10.1214/22-ba1309

GNEDIN, A. (2010). A species sampling model with finitely many types. *Electron. Commun. Probab.* **15** 79–88. MR2606505 https://doi.org/10.1214/ECP.v15-1532

GNEDIN, A. and PITMAN, J. (2006). Exchangeable Gibbs partitions and Stirling triangles. *J. Math. Sci.* **138** 5674–5685.

GRIFFIN, J. E. and STEEL, M. F. J. (2011). Stick-breaking autoregressive processes. *J. Econometrics* **162** 383–396. MR2795625 https://doi.org/10.1016/j.jeconom.2011.03.001

HALMOS, P. R. (1944). Random alms. *Ann. Math. Statist.* **15** 182–189. MR0010342 https://doi.org/10.1214/aoms/1177731283

HATJISPYROS, S. J., MERKATAS, C. and WALKER, S. G. (2023). Mixture models with decreasing weights. *Comput. Statist. Data Anal.* **179** Paper No. 107651, 13. MR4503390 https://doi.org/10.1016/j.csda.2022.107651

HJORT, N. L., HOLMES, C., MÜLLER, P. and WALKER, S. G., eds. (2010). *Bayesian Nonparametrics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **28**. Cambridge Univ. Press, Cambridge. MR2722987 https://doi.org/10.1017/CBO9780511802478

HORIGUCHI, A., CHAN, C. and MA, L. (2025). A tree perspective on stick-breaking models in covariate-dependent mixtures (with discussion). *Bayesian Anal.* **20** 1139–1230. MR4964231 https://doi.org/10.1214/24-ba1462

HU, Y. and ZHANG, J. (2022). Functional central limit theorems for stick-breaking priors. *Bayesian Anal.* **17** 1101–1120. MR4506023 https://doi.org/10.1214/21-ba1290

ISHWARAN, H. and JAMES, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.* **96** 161–173. MR1952729 https://doi.org/10.1198/016214501750332758

KALLI, M., GRIFFIN, J. E. and WALKER, S. G. (2011). Slice sampling mixture models. *Stat. Comput.* **21** 93–105. MR2746606 https://doi.org/10.1007/s11222-009-9150-y

KINGMAN, J. F. C. (1978). The representation of partition structures. *J. Lond. Math. Soc.* (2) **18** 374–380. MR0509954 https://doi.org/10.1112/jlms/s2-18.2.374

KINGMAN, J. F. C. (1982). The coalescent. *Stoch. Process. Appl.* **13** 235–248. MR0671034 https://doi.org/10.1016/0304-4149(82)90011-4

KINGMAN, J. F. C., TAYLOR, S. J., HAWKES, A. G., WALKER, A. M., COX, D. R., SMITH, A. F. M., HILL, B. M., BURVILLE, P. J. and LEONARD, T. (1975). Random discrete distribution. *J. Roy. Statist. Soc. Ser. B, Methodol.* **37** 1–22. MR0368264

LIJOI, A., MENA, R. H. and PRÜNSTER, I. (2007). Bayesian nonparametric estimation of the probability of discovering new species. *Biometrika* **94** 769–786. MR2416792 https://doi.org/10.1093/biomet/asm061

LIJOI, A., PRÜNSTER, I. and WALKER, S. G. (2008). Bayesian nonparametric estimators derived from conditional Gibbs structures. *Ann. Appl. Probab.* **18** 1519–1547. MR2434179 https://doi.org/10.1214/07-AAP495

LINERO, A. R. and DU, J. (2023). Gibbs priors for Bayesian nonparametric variable selection with weak learners. *J. Comput. Graph. Statist.* **32** 1046–1059. MR4641480 https://doi.org/10.1080/10618600.2022.2142594

LO, A. Y. (1984). On a class of Bayesian nonparametric estimates. I. Density estimates. *Ann. Statist.* **12** 351–357. MR0733519 https://doi.org/10.1214/aos/1176346412

MACEACHERN, S. N. (1994). Estimating normal means with a conjugate style Dirichlet process prior. *Comm. Statist. Simulation Comput.* **23** 727–741. MR1293996 https://doi.org/10.1080/03610919408813196

MACEACHERN, S. N. (1999). Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science*. Am. Statist. Assoc., Alexandria.

MACEACHERN, S. N. (2000). Dependent Dirichlet processes. Tech. Report, Department of Statistics, The Ohio State Univ.

MACEACHERN, S. N., KOTTAS, A. and GELFAND, A. E. (2001). Spatial nonparametric Bayesian models. In *American Statistical Association* 2001 *Proceedings of the Section on Bayesian Statistics*.

MCCLOSKEY, J. W. T. (1965). *A Model for the Distribution of Individuals by Species in an Environment*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—Michigan State University. MR2615013

MENA, R. H. and WALKER, S. G. (2015). On the Bayesian mixture model and identifiability. *J. Comput. Graph. Statist.* **24** 1155–1169. MR3432934 https://doi.org/10.1080/10618600.2014.950376

MÜLLER, P., QUINTANA, F. A., JARA, A. and HANSON, T. (2015). *Bayesian Nonparametric Data Analysis*. *Springer Series in Statistics*. Springer, Cham. MR3309338 https://doi.org/10.1007/978-3-319-18968-0

NIETO-BARAJAS, L. E., MÜLLER, P., JI, Y., LU, Y. and MILLS, G. B. (2012). A time-series DDP for functional proteomics profiles. *Biometrics* **68** 859–868. MR3055190 https://doi.org/10.1111/j.1541-0420.2011.01724.x

PAPASPILIOPOULOS, O. and ROBERTS, G. O. (2008). Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. *Biometrika* **95** 169–186. MR2409721 https://doi.org/10.1093/biomet/asm086

PERMAN, M., PITMAN, J. and YOR, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Probab. Theory Related Fields* **92** 21–39. MR1156448 https://doi.org/10.1007/BF01205234

PITMAN, J. (1995). Exchangeable and partially exchangeable random partitions. *Probab. Theory Related Fields* **102** 145–158. MR1337249 https://doi.org/10.1007/BF01213386

PITMAN, J. (1996a). Some developments of the Blackwell–MacQueen urn scheme. In *Statistics*, *Probability and Game Theory*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **30** 245–267. IMS, Hayward, CA. MR1481784 https://doi.org/10.1214/lnms/1215453576

PITMAN, J. (1996b). Random discrete distributions invariant under size-biased permutation. *Adv. Appl. Probab.* **28** 525–539. MR1387889 https://doi.org/10.2307/1428070

PITMAN, J. (2003). Poisson–Kingman partitions. In *Statistics and Science*: *A Festschrift for Terry Speed*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **40** 1–34. IMS, Beachwood, OH. MR2004330 https://doi.org/10.1214/lnms/1215091133

PITMAN, J. (2006). *Combinatorial Stochastic Processes*. *Lecture Notes in Math.* **1875**. Springer, Berlin. MR2245368

PITMAN, J. and YOR, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25** 855–900. MR1434129 https://doi.org/10.1214/aop/1024404422

QUINTANA, F. A., MÜLLER, P., JARA, A. and MACEACHERN, S. N. (2022). The dependent Dirichlet process and related models. *Statist. Sci.* **37** 24–41. MR4371095 https://doi.org/10.1214/20-sts819

REGAZZINI, E., LIJOI, A. and PRÜNSTER, I. (2003). Distributional results for means of normalized random measures with independent increments. *Ann. Statist.* **31** 560–585. MR1983542 https://doi.org/10.1214/aos/1051027881

RICHARDSON, S. and GREEN, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *J. Roy. Statist. Soc. Ser. B, Methodol.* **59** 731–792. MR1483213 https://doi.org/10.1111/1467-9868.00095

RUGGIERO, M. and SORDELLO, M. (2018). Clustering dynamics in a class of normalised generalised gamma dependent priors. *Ann. Inst. Statist. Math.* **70** 83–98. MR3742819 https://doi.org/10.1007/s10463-016-0583-8

SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* **4** 639–650. MR1309433

SORIANO, J. and MA, L. (2019). Mixture modeling on related samples by ψ-stick breaking and kernel perturbation. *Bayesian Anal.* **14** 161–180. MR3910042 https://doi.org/10.1214/18-BA1106

STEFANUCCI, M. and CANALE, A. (2021). Multiscale stick-breaking mixture models. *Stat. Comput.* **31** Paper No. 13, 13. MR4206653 https://doi.org/10.1007/s11222-020-09991-1

TEH, Y. W., JORDAN, M. I., BEAL, M. J. and BLEI, D. M. (2006). Hierarchical Dirichlet processes. *J. Amer. Statist. Assoc.* **101** 1566–1581. MR2279480 https://doi.org/10.1198/016214506000000302

WALKER, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Comm. Statist. Simulation Comput.* **36** 45–54. MR2370888 https://doi.org/10.1080/03610910601096262

The Annals of Statistics
2026, Vol. 54, No. 3, 1372–1397
https://doi.org/10.1214/25-AOS2608

IDENTIFICATION AND ESTIMATION FOR MATRIX TIME-SERIES CP-FACTOR MODELS

BY JINYUAN CHANG[1,2,3,a], YUE DU[1,b], GUANGLIN HUANG[1,c] AND QIWEI YAO[4,d]

[1]*Joint Laboratory of Data Science and Business Intelligence, Southwestern University of Finance and Economics,* [a]*changjinyuan@swufe.edu.cn,* [b]*122071400004@smail.swufe.edu.cn,* [c]*huanggl@swufe.edu.cn*

[2]*State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences*

[3]*China Center for Economic Research, Peking University*

[4]*Department of Statistics, London School of Economics and Political Science,* [d]*q.yao@lse.ac.uk*

We propose a new method for identifying and estimating the CP-factor models for matrix time series. Unlike the generalized eigenanalysis-based method (*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **85** (2023) 127–148) for which the convergence rates of the associated estimators may suffer from small eigengaps as the asymptotic theory is based on some matrix perturbation analysis, the proposed new method enjoys faster convergence rates which are free from any eigengaps. It achieves this by turning the problem into a joint diagonalization of several matrices whose elements are determined by a basis of a linear system, and by choosing the basis carefully to avoid near colinearity (see Proposition 5 and Section 4.3). Furthermore, unlike the generalized eigenanalysis-based method which requires the two factor loading matrices to be full-ranked, the proposed new method can handle rank-deficient factor loading matrices. Illustration with both simulated and real matrix time-series data shows the advantages of the proposed new method.

REFERENCES

AFSARI, B. (2008). Sensitivity analysis for the problem of matrix joint diagonalization. *SIAM J. Matrix Anal. Appl.* **30** 1148–1171. MR2447446 https://doi.org/10.1137/060655997

CHANG, J., DU, Y., HUANG, G. and YAO, Q. (2026). Supplement to "Identification and Estimation for Matrix Time-Series CP-factor Models." https://doi.org/10.1214/25-AOS2608SUPP

CHANG, J., GUO, B. and YAO, Q. (2015). High dimensional stochastic regression with latent factors, endogeneity and nonlinearity. *J. Econometrics* **189** 297–312. MR3414901 https://doi.org/10.1016/j.jeconom.2015.03.024

CHANG, J., GUO, B. and YAO, Q. (2018). Principal component analysis for second-order stationary vector time series. *Ann. Statist.* **46** 2094–2124. MR3845012 https://doi.org/10.1214/17-AOS1613

CHANG, J., HE, J., LIN, C. and YAO, Q. (2024). HDTSA: An R package for high-dimensional time series analysis. Available at arXiv:2412.17341.

CHANG, J., HE, J., YANG, L. and YAO, Q. (2023). Modelling matrix time series via a tensor CP-decomposition. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **85** 127–148. MR4726961 https://doi.org/10.1093/jrsssb/qkac011

CHEN, E. Y. and CHEN, R. (2022). Modeling dynamic transport network with matrix factor models: An application to international trade flow. *J. Data Sci.* **21** 490–507.

CHEN, E. Y., TSAY, R. S. and CHEN, R. (2020). Constrained factor models for high-dimensional matrix-variate time series. *J. Amer. Statist. Assoc.* **115** 775–793. MR4107679 https://doi.org/10.1080/01621459.2019.1584899

CHEN, R., XIAO, H. and YANG, D. (2021). Autoregressive models for matrix-valued time series. *J. Econometrics* **222** 539–560. MR4234832 https://doi.org/10.1016/j.jeconom.2020.07.015

DE LATHAUWER, L. (2006). A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization. *SIAM J. Matrix Anal. Appl.* **28** 642–666. MR2262974 https://doi.org/10.1137/040608830

FAMA, E. F. and MACBETH, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *J. Polit. Econ.* **81** 607–636.

MSC2020 subject classifications. Primary 62M10; secondary 62H25.

Key words and phrases. CP-decomposition, dimension reduction, matrix time series, nonorthogonal joint diagonalization.

HAN, Y., CHEN, R., YANG, D. and ZHANG, C.-H. (2024a). Tensor factor model estimation by iterative projection. *Ann. Statist.* **52** 2641–2667. MR4842821 https://doi.org/10.1214/24-aos2412

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., YANG, D., ZHANG, C.-H. and CHEN, R. (2024b). CP factor model for dynamic tensors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **86** 1383–1413. MR4896641 https://doi.org/10.1093/jrsssb/qkae036

HAN, Y. and ZHANG, C.-H. (2023). Tensor principal component analysis in high dimensional CP models. *IEEE Trans. Inf. Theory* **69** 1147–1167. MR4564648 https://doi.org/10.1109/tit.2022.3203972

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. and YAO, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *Ann. Statist.* **40** 694–726. MR2933663 https://doi.org/10.1214/12-AOS970

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

PHAM, D.-T. and CARDOSO, J.-F. (2001). Blind separation of instantaneous mixtures of nonstationary sources. *IEEE Trans. Signal Process.* **49** 1837–1848. MR1852133 https://doi.org/10.1109/78.942614

VOLLGRAF, R. and OBERMAYER, K. (2006). Quadratic optimization for simultaneous matrix diagonalization. *IEEE Trans. Signal Process.* **54** 3270–3278.

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

ZIEHE, A., LASKOV, P., NOLTE, G. and MÜLLER, K.-R. (2004). A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation. *J. Mach. Learn. Res.* **5** 777–800. MR2247999

The Annals of Statistics
2026, Vol. 54, No. 3, 1398–1424
https://doi.org/10.1214/25-AOS2609

TOWARDS A UNIFIED THEORY FOR SEMIPARAMETRIC DATA FUSION WITH INDIVIDUAL-LEVEL DATA

BY ELLEN GRAHAM[a], MARCO CARONE[b] AND ANDREA ROTNITZKY[c]

Department of Biostatistics, University of Washington, [a]*graham18@uw.edu,* [b]*mcarone@uw.edu,* [c]*arotnitz@uw.edu*

We consider inference about a finite-dimensional parameter integrating samples from independent sources. A recently developed theory considers scenarios where sources align with subsets of the conditional distributions of a single factorization of the joint target distribution. While this theory applies in many settings, it falls short in important data fusion problems, such as two-sample instrumental variable analysis, settings that integrate data from epidemiological studies with diverse designs, and studies with mismeasured variables supplemented by external validation studies. In this paper we derive a comprehensive theory that, in particular, covers these settings by allowing the integration of sources aligned with conditional distributions that do not correspond to a single factorization of the target distribution. We provide a universal characterization of the influence functions of regular and asymptotically linear estimators and the efficient influence function of a target parameter, irrespective of the parameter of interest or the statistical model for the target distribution, thus paving the way for a unified theory for machine-learning debiased, semiparametric efficient estimation.

REFERENCES

ANDERSON, T. W. and RUBIN, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Ann. Math. Stat.* **20** 46–63. MR0028546 https://doi.org/10.1214/aoms/1177730090

ANGRIST, J. D. and KRUEGER, A. B. (1992). The effect of age at school entry on educational attainment: An application of instrumental variables with moments from two samples. *J. Amer. Statist. Assoc.* **87** 328–336.

ARNOLD, B. C., CASTILLO, E. and SARABIA, J. M. (1996). Specification of distributions by combinations of marginal and conditional distributions. *Statist. Probab. Lett.* **26** 153–157. MR1381466 https://doi.org/10.1016/0167-7152(95)00005-4

BAREINBOIM, E. and PEARL, J. (2014). Transportability from multiple environments with limited experiments: Completeness results. In *Advances in Neural Information Processing Systems* **27**. Curran Associates, Red Hook.

BAREINBOIM, E. and PEARL, J. (2016). Causal inference and the data-fusion problem. *Proc. Natl. Acad. Sci. USA* **113** 7345–7352.

BAREINBOIM, E. and TIAN, J. (2015). Recovering causal effects from selection bias. *Proc. AAAI Conf. Artif. Intell.* **29**.

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

CHANG, J. T. and POLLARD, D. (1997). Conditioning as disintegration. *Stat. Neerl.* **51** 287–317. MR1484954 https://doi.org/10.1111/1467-9574.00056

CHATTERJEE, N., CHEN, Y.-H., MAAS, P. and CARROLL, R. J. (2016). Constrained maximum likelihood estimation for model calibration using summary-level information from external big data sources. *J. Amer. Statist. Assoc.* **111** 107–117. MR3494641 https://doi.org/10.1080/01621459.2015.1123157

CHEN, X., HONG, H. and TAROZZI, A. (2008). Semiparametric efficiency in GMM models of nonclassical measurement errors, missing data and treatment effects. Available at https://cowles.yale.edu/sites/default/files/2022-08/d1644.pdf.

COLE, S. R., EDWARDS, J. K., BRESKIN, A., ROSIN, S., ZIVICH, P. N., SHOOK-SA, B. E. and HUDGENS, M. G. (2023). Illustration of 2 fusion designs and estimators. *Amer. J. Epidemiol.* **192** 467–474.

MSC2020 subject classifications. 62G05.

Key words and phrases. Causal transportability, data integration, external validation studies, retrospective sampling, semiparametric efficiency.

D'HAULTFOEUILLE, X. (2010). A new instrumental method for dealing with endogenous selection. *J. Econometrics* **154** 1–15. MR2558947 https://doi.org/10.1016/j.jeconom.2009.06.005

DAHABREH, I. J. and HERNÁN, M. A. (2019). Extending inferences from a randomized trial to a target population. *Eur. J. Epidemiol.* **34** 719–722.

DAHABREH, I. J., PETITO, L. C., ROBERTSON, S. E., HERNÁN, M. A. and STEINGRIMSSON, J. A. (2020a). Towards causally interpretable meta-analysis: Transporting inferences from multiple randomized trials to a new target population. *Epidemiology* **31** 334–344.

DAHABREH, I. J., ROBERTSON, S. E., STEINGRIMSSON, J. A., STUART, E. A. and HERNÁN, M. A. (2020b). Extending inferences from a randomized trial to a new target population. *Stat. Med.* **39** 1999–2014. MR4105273 https://doi.org/10.1002/sim.8426

EVANS, R. J. and DIDELEZ, V. (2024). Parameterizing and simulating from causal models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **86** 535–568. MR4779257 https://doi.org/10.1093/jrsssb/qkad058

FLEMING, T. R. and POWERS, J. H. (2012). Biomarkers and surrogate endpoints in clinical trials. *Stat. Med.* **31** 2973–2984. MR2993599 https://doi.org/10.1002/sim.5403

GELMAN, A. and SPEED, T. P. (1993). Characterizing a joint probability distribution by conditionals. *J. Roy. Statist. Soc. Ser. B, Methodol.* **55** 185–188. MR1210430

GELMAN, A. and SPEED, T. P. (1999). Corrigendum: "Characterizing a joint probability distribution by conditionals" [J. Roy. Statist. Soc. Ser. B **55** (1993), no. 1, 185–188; MR1210430 (94d:62034)]. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **61** 483. MR1680294 https://doi.org/10.1111/1467-9868.00189

GRAHAM, B. S., PINTO, C. C. D. X. and EGEL, D. (2016). Efficient estimation of data combination models by the method of auxiliary-to-study tilting (AST). *J. Bus. Econom. Statist.* **34** 288–301. MR3475879 https://doi.org/10.1080/07350015.2015.1038544

GRAHAM, E., CARONE, M. and ROTNITZKY, A. (2026). Supplement to "Towards a unified theory for semiparametric data fusion with individual-level data." https://doi.org/10.1214/25-AOS2609SUPP

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

HANSEN, B. (2022). *Econometrics*. Princeton Univ. Press, Princeton, NJ.

HU, W., WANG, R., LI, W. and MIAO, W. (2023). Paradoxes and Resolutions for Semiparametric Fusion of Individual and Summary Data. arXiv Preprint. Available at arXiv:2210.00200.

HÜNERMUND, P. and BAREINBOIM, E. (2025). Causal inference and data fusion in econometrics. *Econom. J.* **28** 41–82. MR4876375 https://doi.org/10.1093/ectj/utad008

IMBENS, G. W. and ANGRIST, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica* **62** 467–475.

JIA, J., GENG, Z. and WANG, M. (2006). Identifiability and estimation of probabilities from multiple databases with incomplete data and sampling selection. In *Structural, Syntactic, and Statistical Pattern Recognition* (D.-Y. Yeung, J. T. Kwok, A. Fred, F. Roli and D. de Ridder, eds.) 792–798. Springer, Berlin.

KIVVA, Y., MOKHTARIAN, E., ETESAMI, J. and KIYAVASH, N. (2022). Revisiting the general identifiability problem. In *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence* 1022–1030. PMLR.

KLEVMARKEN, A. (1982). Missing variables and two-stage least-squares estimation from more than one data set. Working Paper No. 62, IUI Working Paper.

LEE, J. J. R., GHASSAMI, A. and SHPITSER, I. (2024). A general identification algorithm for data fusion problems under systematic selection. In *Proceedings of the Fortieth Conference on Uncertainty in Artificial Intelligence* 2188–2204. PMLR.

LEE, S., CORREA, J. D. and BAREINBOIM, E. (2020). General identifiability with arbitrary surrogate experiments. In *Proceedings of the 35th Uncertainty in Artificial Intelligence Conference* 389–398. PMLR.

LI, S. and LUEDTKE, A. (2023). Efficient estimation under data fusion. *Biometrika* **110** 1041–1054. MR4667439 https://doi.org/10.1093/biomet/asad007

LI, W., MIAO, W. and TCHETGEN, E. T. (2022). Nonparametric inference about mean functionals of nonignorable nonresponse data without identifying the joint distribution. Available at https://doi.org/10.48550/arXiv.2110.05776.

LI, W., MIAO, W. and TCHETGEN TCHETGEN, E. (2023). Non-parametric inference about mean functionals of non-ignorable non-response data without identifying the joint distribution. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **85** 913–935. MR4735035 https://doi.org/10.1093/jrsssb/qkad047

LUEDTKE, A. R. and GILBERT, P. B. (2017). Partial Bridging of Vaccine Efficacy to New Populations. arXiv Preprint. Available at arXiv:1701.06739.

MANELA, D. D. V., YANG, L. and EVANS, R. J. (2024). Testing generalizability in causal inference. In *Proceedings of the Forty-first Conference on Uncertainty in Artificial Intelligence* 2906–2927. PMLR.

MIAO, W. and TCHETGEN TCHETGEN, E. J. (2016). On varieties of doubly robust estimators under missingness not at random with a shadow variable. *Biometrika* **103** 475–482. MR3509900 https://doi.org/10.1093/biomet/asw016

PACINI, D. (2019). Two-sample least squares projection. *Econometric Rev.* **38** 95–123. MR3942463 https://doi.org/10.1080/07474938.2016.1222068

PARK, C., RICHARDSON, D. B. and TCHETGEN TCHETGEN, E. J. (2024). Single proxy control. *Biometrics* **80** Paper No. ujae027. MR4754194 https://doi.org/10.1093/biomtc/ujae027

PEARL, J. and BAREINBOIM, E. (2011). Transportability of causal and statistical relations: A formal approach. In 2011 *IEEE* 11*th International Conference on Data Mining Workshops* 540–547.

PIERCE, B. L. and BURGESS, S. (2013). Efficient design for Mendelian randomization studies: Subsample and 2-sample instrumental variable estimators. *Amer. J. Epidemiol.* **178** 1177–1184.

QIU, H., TCHETGEN TCHETGEN, E. and DOBRIBAN, E. (2024). Efficient and multiply robust risk estimation under general forms of dataset shift. *Ann. Statist.* **52** 1796–1824. MR4804828 https://doi.org/10.1214/24-aos2422

RIDDER, G. and MOFFITT, R. (2007). Chapter 75 the econometrics of data combination. In *Handbook of Econometrics* (J. J. Heckman and E. E. Leamer, eds.) **6** 5469–5547. Elsevier.

ROBINS, J. M. and ROTNITZKY, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology*: *Methodological Issues* (N. P. Jewell, K. Dietz and V. T. Farewell, eds.) 297–331. Birkhäuser, Boston, MA.

ROBINS, J. M., ROTNITZKY, A. and ZHAO, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89** 846–866. MR1294730

RUDOLPH, K. E. and VAN DER LAAN, M. J. (2017). Robust estimation of encouragement design intervention effects transported across sites. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1509–1525. MR3731673 https://doi.org/10.1111/rssb.12213

SHI, X., PAN, Z. and MIAO, W. (2023). Data integration in causal inference. *Wiley Interdiscip. Rev.*: *Comput. Stat.* **15** Paper No. e1581. MR4544393 https://doi.org/10.1002/wics.1581

SHU, H. and TAN, Z. (2020). Improved methods for moment restriction models with data combination and an application to two-sample instrumental variable estimation. *Canad. J. Statist.* **48** 259–284. MR4095263 https://doi.org/10.1002/cjs.11530

SUN, B. and MIAO, W. (2021). On semiparametric instrumental variable estimation of average treatment effects through data fusion. *Statist. Sinica* **32** 569–590. MR4393475 https://doi.org/10.1017/s0956792520000406

TSIATIS, A. A. (2006). *Semiparametric Theory and Missing Data. Springer Series in Statistics*. Springer, New York. MR2233926

VAN DER LAAN, M. J. (2008). Estimation based on case-control designs with known prevalence probability. *Int. J. Biostat.* **4** Art. 17. MR2443193 https://doi.org/10.2202/1557-4679.1114

VAN DER LAAN, M. J. and ROBINS, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. *Springer Series in Statistics*. Springer, New York, NY.

VAN DER LAAN, M. J. and RUBIN, D. (2006). Targeted maximum likelihood learning. *Int. J. Biostat.* **2** Art. 11. MR2306500 https://doi.org/10.2202/1557-4679.1043

VAN DER VAART, A. W. (2000). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

WANG, Z., GHASSAMI, A. and SHPITSER, I. (2024). Identification and estimation for nonignorable missing data: A data fusion approach.

ZHAO, Q., WANG, J., SPILLER, W., BOWDEN, J. and SMALL, D. S. (2019). Two-sample instrumental variable analyses using heterogeneous samples. *Statist. Sci.* **34** 317–333. MR3983331 https://doi.org/10.1214/18-STS692

The Annals of Statistics
2026, Vol. 54, No. 3, 1425–1450
https://doi.org/10.1214/25-AOS2610

A TWO-STEP ESTIMATING APPROACH FOR HEAVY-TAILED AR MODELS WITH NONZERO MEDIAN GARCH-TYPE NOISES

BY RUI SHE[1,a], LINLIN DAI[1,b] AND SHIQING LING[2,c]

[1]*Center of Statistical Research, School of Statistics and Data Science, The Southwestern University of Finance and Economics,* [a]*rshe@swufe.edu.cn,* [b]*ldaiab@swufe.edu.cn*

[2]*Department of Mathematics, The Hong Kong University of Science and Technology,* [c]*maling@ust.hk*

This paper develops a novel two-step estimating procedure for heavy-tailed AR models with nonzero median GARCH-type noises, allowing for time-varying volatility. We first establish the self-weighted quantile regression estimator (SQE) across all quantile levels $\tau \in (0, 1)$ for the AR parameters θ_0. We show that the SQE, less a bias, converges weakly to a Gaussian process at a rate of $n^{-1/2}$. The bias is zero if and only if τ equals τ_0, the probability that the noise is less than zero. Based on the SQE, we propose an approach to estimate τ_0 in the second step and feed the estimated τ_0 back into the SQE to estimate θ_0. Both the estimated τ_0 and θ_0 are shown to be consistent and asymptotically normal. A random weighting bootstrap method is developed to approximate the complex distribution. The problem we study is nonstandard because τ_0 may not be identifiable in conventional quantile regression, and the usual methods cannot verify the existence of the SQE bias. Unlike existing procedures for heavy-tailed time series, our method does not require prior information about the symmetry, tail index, or the parametric form of the noise, nor does it require classical identification conditions, such as zero-mean or zero-median.

REFERENCES

ANGRIST, J., CHERNOZHUKOV, V. and FERNÁNDEZ-VAL, I. (2006). Quantile regression under misspecification, with an application to the U.S. wage structure. *Econometrica* **74** 539–563. MR2207400 https://doi.org/10.1111/j.1468-0262.2006.00671.x

BASRAK, B., DAVIS, R. A. and MIKOSCH, T. (2002). Regular variation of GARCH processes. *Stoch. Process. Appl.* **99** 95–115. MR1894253 https://doi.org/10.1016/S0304-4149(01)00156-9

BIERENS, H. J. (1982). Consistent model specification tests. *J. Econometrics* **20** 105–134. MR0685673 https://doi.org/10.1016/0304-4076(82)90105-1

D'INNOCENZO, E., LUCAS, A., SCHWAAB, B. and ZHANG, X. (2024). Modeling extreme events: Time-varying extreme tail shape. *J. Bus. Econom. Statist.* **42** 903–917. MR4757114 https://doi.org/10.1080/07350015.2023.2260439

DALLA, V., GIRAITIS, L. and ROBINSON, P. M. (2020). Asymptotic theory for time series with changing mean and variance. *J. Econometrics* **219** 281–313. MR4175901 https://doi.org/10.1016/j.jeconom.2020.03.005

DE HAAN, L. and ZHOU, C. (2021). Trends in extreme value indices. *J. Amer. Statist. Assoc.* **116** 1265–1279. MR4309271 https://doi.org/10.1080/01621459.2019.1705307

ENGLE, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50** 987–1007. MR0666121 https://doi.org/10.2307/1912773

ENGLE, R. F. (1990). Discussion: Stock market volatility and the crash of 1987. *Rev. Financ. Stud.* **3** 103–106.

ESCANCIANO, J. C. and GOH, S. C. (2014). Specification analysis of linear quantile models. *J. Econometrics* **178** 495–507. MR3132447 https://doi.org/10.1016/j.jeconom.2013.07.006

ESCANCIANO, J. C. and VELASCO, C. (2010). Specification tests of parametric dynamic conditional quantiles. *J. Econometrics* **159** 209–221. MR2720853 https://doi.org/10.1016/j.jeconom.2010.06.003

FAN, J., QI, L. and XIU, D. (2014). Quasi-maximum likelihood estimation of GARCH models with heavy-tailed likelihoods. *J. Bus. Econom. Statist.* **32** 178–191. MR3207827 https://doi.org/10.1080/07350015.2013.840239

FAN, J. and YAO, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods. Springer Series in Statistics*. Springer, New York. MR1964455 https://doi.org/10.1007/b97702

MSC2020 subject classifications. 62F12, 62F35, 62M10.
Key words and phrases. GARCH, heavy-tailed, self-weighted, quantile, nonstationary.

FRANCQ, C. and ZAKOÏAN, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* **10** 605–637. MR2076065 https://doi.org/10.3150/bj/1093265632

FRANCQ, C. and ZAKOÏAN, J.-M. (2010). *GARCH Models*: *Structure*, *Statistical Inference and Financial Applications*. Wiley, Chichester. MR3186556 https://doi.org/10.1002/9780470670057

FRANCQ, C. and ZAKOÏAN, J.-M. (2012). QML estimation of a class of multivariate asymmetric GARCH models. *Econometric Theory* **28** 179–206. MR2899217 https://doi.org/10.1017/S0266466611000156

FRANCQ, C. and ZAKOÏAN, J.-M. (2013). Estimating the marginal law of a time series with applications to heavy-tailed distributions. *J. Bus. Econom. Statist.* **31** 412–425. MR3173691 https://doi.org/10.1080/07350015.2013.801776

GLOSTEN, L. R., JAGANNATHAN, R. and RUNKLE, D. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *J. Finance* **48** 1779–1801.

HAMILTON, J. D. (1994). *Time Series Analysis*. Princeton Univ. Press, Princeton, NJ. MR1278033

HARVEY, C. R. and SIDDIQUE, A. (2000). Conditional skewness in asset pricing tests. *J. Finance* **55** 1263–1295.

JIANG, F., LI, D. and ZHU, K. (2020). Non-standard inference for augmented double autoregressive models with null volatility coefficients. *J. Econometrics* **215** 165–183. MR4066067 https://doi.org/10.1016/j.jeconom.2019.08.009

JIANG, F., LI, D. and ZHU, K. (2021). Adaptive inference for a semiparametric generalized autoregressive conditional heteroskedasticity model. *J. Econometrics* **224** 306–329. MR4303783 https://doi.org/10.1016/j.jeconom.2020.10.007

KATO, K. (2009). Asymptotics for argmin processes: Convexity arguments. *J. Multivariate Anal.* **100** 1816–1829. MR2535389 https://doi.org/10.1016/j.jmva.2009.02.008

KIM, T.-H. and WHITE, H. (2003). Estimation, inference, and specification testing for possibly misspecified quantile regression. In *Maximum Likelihood Estimation of Misspecified Models*: *Twenty Years Later*. *Adv. Econom.* **17** 107–126. Elsevier/JAI, Amsterdam. MR2528712 https://doi.org/10.1016/S0731-9053(03)17005-3

KOEDIJK, K. G. and KOOL, C. J. M. (1992). Tail estimates of East European exchange rates. *J. Bus. Econom. Statist.* **10** 83–96.

LING, S. (2005). Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 381–393. MR2155344 https://doi.org/10.1111/j.1467-9868.2005.00507.x

LING, S. (2007a). Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. *J. Econometrics* **140** 849–873. MR2408929 https://doi.org/10.1016/j.jeconom.2006.07.016

LING, S. (2007b). A double AR(p) model: Structure and estimation. *Statist. Sinica* **17** 161–175. MR2352507

LING, S. and LI, W. K. (1997). On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *J. Amer. Statist. Assoc.* **92** 1184–1194. MR1482150 https://doi.org/10.2307/2965585

MA, Y., ZHOU, M., PENG, L. and ZHANG, R. (2022). Test for zero median of errors in an ARMA-GARCH model. *Econometric Theory* **38** 536–561. MR4438296 https://doi.org/10.1017/S0266466621000244

MAYER, U., ZÄHLE, H. and ZHOU, Z. (2020). Functional weak limit theorem for a local empirical process of non-stationary time series and its application. *Bernoulli* **26** 1891–1911. MR4091095 https://doi.org/10.3150/19-BEJ1174

NEWEY, W. K. (1985). Maximum likelihood specification testing and conditional moment tests. *Econometrica* **53** 1047–1070. MR0809903 https://doi.org/10.2307/1911011

PATTON, A. J. (2004). On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *J. Financ. Econom.* **2** 130–168.

PHANDOIDAEN, N. and RICHTER, S. (2022). Empirical process theory for nonsmooth functions under functional dependence. *Electron. J. Stat.* **16** 3385–3429. MR4423798 https://doi.org/10.1214/22-ejs2023

PHILLIPS, P. C. B. and XU, K.-L. (2006). Inference in autoregression under heteroskedasticity. *J. Time Series Anal.* **27** 289–308. MR2235847 https://doi.org/10.1111/j.1467-9892.2005.00466.x

POLLARD, D. (1990). *Empirical Processes*: *Theory and Applications*. *NSF-CBMS Regional Conference Series in Probability and Statistics* **2**. IMS, Hayward, CA. MR1089429

RHO, Y. and SHAO, X. (2015). Inference for time series regression models with weakly dependent and heteroscedastic errors. *J. Bus. Econom. Statist.* **33** 444–457. MR3372670 https://doi.org/10.1080/07350015.2014.962698

SHE, R. (2023). Tests of unit root hypothesis with heavy-tailed heteroscedastic noises. *Statist. Sinica* **33** 215–236. MR4527769

SHE, R. (2025). Inference in median AR models with nonstationary and heavy-tailed heteroskedastic noises. *Econometric Theory* 1–29.

SHE, R., DAI, L. and LING, S. (2026). Supplement to "A two-step estimating approach for heavy-tailed AR models with nonzero median GARCH-type noises." https://doi.org/10.1214/25-AOS2610SUPP

TAYLOR, S. (1986). *Modelling Financial Time Series*. Wiley, New York.
VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes. Springer Series in Statistics*. Springer, New York. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2
WEISS, A. A. (1986). Asymptotic theory for ARCH models: Estimation and testing. *Econometric Theory* **2** 107–131.
WU, W. and ZHOU, Z. (2017). Nonparametric inference for time-varying coefficient quantile regression. *J. Bus. Econom. Statist.* **35** 98–109. MR3591532 https://doi.org/10.1080/07350015.2015.1060884
WU, W. B. (2005). On the Bahadur representation of sample quantiles for dependent sequences. *Ann. Statist.* **33** 1934–1963. MR2166566 https://doi.org/10.1214/009053605000000291
WU, W. B. (2007). Strong invariance principles for dependent random variables. *Ann. Probab.* **35** 2294–2320. MR2353389 https://doi.org/10.1214/009117907000000060
WU, W. B. (2008). Empirical processes of stationary sequences. *Statist. Sinica* **18** 313–333. MR2384990
XU, K.-L. and PHILLIPS, P. C. B. (2008). Adaptive estimation of autoregressive models with time-varying variances. *J. Econometrics* **142** 265–280. MR2408736 https://doi.org/10.1016/j.jeconom.2007.06.001
YANG, Y. and LING, S. (2017). Self-weighted LAD-based inference for heavy-tailed threshold autoregressive models. *J. Econometrics* **197** 368–381. MR3614206 https://doi.org/10.1016/j.jeconom.2016.11.009
ZHANG, R. and LING, S. (2015). Asymptotic inference for AR models with heavy-tailed G-GARCH noises. *Econometric Theory* **31** 880–890. MR3377272 https://doi.org/10.1017/S0266466614000632
ZHANG, T. and WU, W. B. (2012). Inference of time-varying regression models. *Ann. Statist.* **40** 1376–1402. MR3015029 https://doi.org/10.1214/12-AOS1010
ZHANG, X., ZHANG, R., LI, Y. and LING, S. (2022). LADE-based inferences for autoregressive models with heavy-tailed G-GARCH(1, 1) noise. *J. Econometrics* **227** 228–240. MR4377182 https://doi.org/10.1016/j.jeconom.2020.06.011
ZHOU, Z. and WU, W. B. (2009). Local linear quantile estimation for nonstationary time series. *Ann. Statist.* **37** 2696–2729. MR2541444 https://doi.org/10.1214/08-AOS636
ZHU, K. (2019). Statistical inference for autoregressive models under heteroscedasticity of unknown form. *Ann. Statist.* **47** 3185–3215. MR4025739 https://doi.org/10.1214/18-AOS1775
ZHU, K. and LING, S. (2011). Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA-GARCH/IGARCH models. *Ann. Statist.* **39** 2131–2163. MR2893864 https://doi.org/10.1214/11-AOS895
ZHU, K. and LING, S. (2015). LADE-based inference for ARMA models with unspecified and heavy-tailed heteroscedastic noises. *J. Amer. Statist. Assoc.* **110** 784–794. MR3367264 https://doi.org/10.1080/01621459.2014.977386
ZHU, Q., LI, G. and XIAO, Z. (2021). Quantile estimation of regression models with GARCH-X errors. *Statist. Sinica* **31** 1261–1284. MR4297703 https://doi.org/10.5705/ss.202019.0003
ZHU, Q., ZHENG, Y. and LI, G. (2018). Linear double autoregression. *J. Econometrics* **207** 162–174. MR3856765 https://doi.org/10.1016/j.jeconom.2018.05.006

The Annals of Statistics
2026, Vol. 54, No. 3, 1451–1473
https://doi.org/10.1214/25-AOS2613

TEST OF INDEPENDENCE USING GENERALIZED DISTANCE CORRELATION

BY JIANQING FAN[1,a], ZHIPENG LOU[2,b] AND DANNA ZHANG[2,c]

[1]*Department of Operations Research and Financial Engineering, Princeton University,* [a]*jqfan@princeton.edu*
[2]*Department of Mathematics, University of California, San Diego,* [b]*zlou@ucsd.edu,* [c]*daz076@ucsd.edu*

We study the fundamental statistical inference concerning the testing of independence between two random vectors. Existing asymptotic theories for test statistics based on distance covariance can only apply to either low-dimensional or high-dimensional settings and require stringent distributional assumptions. In this work we develop a new unified distributional theory of the sample generalized distance covariance that works for random vectors of arbitrary dimensions under fairly mild moment conditions. In particular, a Gaussian approximation result is established with a nonasymptotic error bound, and the asymptotic null distribution of the sample generalized distance covariance is shown to be distributed as a linear combination of independently and identically distributed chi-squared random variables. To estimate the asymptotic null distribution practically, we propose a half-permutation procedure and provide the theoretical justification for its validity. The exact asymptotic distribution of the resampling distribution is derived under general marginal moment conditions, and the proposed procedure is shown to be asymptotically equivalent to the oracle procedure with known marginal distributions.

REFERENCES

BERRETT, T. B., KONTOYIANNIS, I. and SAMWORTH, R. J. (2021). Optimal rates for independence testing via U-statistic permutation tests. *Ann. Statist.* **49** 2457–2490. MR4338371 https://doi.org/10.1214/20-aos2041

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

CSÖRGŐ, S. (1985). Testing for independence by the empirical characteristic function. *J. Multivariate Anal.* **16** 290–299. MR0793494 https://doi.org/10.1016/0047-259X(85)90022-3

DAVIS, R. A., MATSUI, M., MIKOSCH, T. and WAN, P. (2018). Applications of distance correlation to time series. *Bernoulli* **24** 3087–3116. MR3779711 https://doi.org/10.3150/17-BEJ955

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

ESSEEN, C.-G. (1945). Fourier analysis of distribution functions. A mathematical study of the Laplace–Gaussian law. *Acta Math.* **77** 1–125. MR0014626 https://doi.org/10.1007/BF02392223

FAN, J., LOU, Z. and ZHANG, D. (2026). Supplement to "Test of independence using generalized distance correlation." https://doi.org/10.1214/25-AOS2613SUPP

FELLER, W. (1991). *An Introduction to Probability Theory and Its Applications*, Vol. **2**. John Wiley & Sons.

FISHER, R. A. (1922). On the interpretation of χ 2 from contingency tables, and the calculation of p. *J. R. Stat. Soc.* **85** 87–94.

FOKIANOS, K. and PITSILLOU, M. (2018). Testing independence for multivariate time series via the auto-distance correlation matrix. *Biometrika* **105** 337–352. MR3804406 https://doi.org/10.1093/biomet/asx082

FUKUMIZU, K., GRETTON, A., SUN, X. and SCHÖLKOPF, B. (2007). Kernel measures of conditional dependence. *Adv. Neural Inf. Process. Syst.* **20**.

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

GRETTON, A. (2015). A simpler condition for consistency of a kernel independence test. arXiv preprint. Available at arXiv:1501.06103.

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

Key words and phrases. Gaussian approximation, generalized distance correlation, half-permutation, test of independence.

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

GRETTON, A., FUKUMIZU, K., TEO, C., SONG, L., SCHÖLKOPF, B. and SMOLA, A. (2007). A kernel statistical test of independence. *Adv. Neural Inf. Process. Syst.* **20**.

HAN, Q. and SHEN, Y. (2024). Generalized kernel distance covariance in high dimensions: Non-null CLTs and power universality. *Inf. Inference* **13** Paper No. iaae017, 43. MR4784486 https://doi.org/10.1093/imaiai/iaae017

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

HUO, X. and SZÉKELY, G. J. (2016). Fast computing for distance covariance. *Technometrics* **58** 435–447. MR3556612 https://doi.org/10.1080/00401706.2015.1054435

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

KIM, I., BALAKRISHNAN, S. and WASSERMAN, L. (2022). Minimax optimality of permutation tests. *Ann. Statist.* **50** 225–251. MR4382015 https://doi.org/10.1214/21-aos2103

KONG, Y., LI, D., FAN, Y. and LV, J. (2017). Interaction pursuit in high-dimensional multi-response regression via distance correlation. *Ann. Statist.* **45** 897–922. MR3650404 https://doi.org/10.1214/16-AOS1474

LEE, C. E. and SHAO, X. (2018). Martingale difference divergence matrix and its application to dimension reduction for stationary multivariate time series. *J. Amer. Statist. Assoc.* **113** 216–229. MR3803459 https://doi.org/10.1080/01621459.2016.1240083

LI, R., ZHONG, W. and ZHU, L. (2012). Feature screening via distance correlation learning. *J. Amer. Statist. Assoc.* **107** 1129–1139. MR3010900 https://doi.org/10.1080/01621459.2012.695654

LINDEBERG, J. W. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Z.* **15** 211–225. MR1544569 https://doi.org/10.1007/BF01494395

LYONS, R. (2013). Distance covariance in metric spaces. *Ann. Probab.* **41** 3284–3305. MR3127883 https://doi.org/10.1214/12-AOP803

MASSART, P. (1990). The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *Ann. Probab.* **18** 1269–1283. MR1062069

MATTESON, D. S. and TSAY, R. S. (2017). Independent component analysis via distance covariance. *J. Amer. Statist. Assoc.* **112** 623–637. MR3671757 https://doi.org/10.1080/01621459.2016.1150851

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

PAN, W., WANG, X., ZHANG, H., ZHU, H. and ZHU, J. (2020). Ball covariance: A generic measure of dependence in Banach space. *J. Amer. Statist. Assoc.* **115** 307–317. MR4078465 https://doi.org/10.1080/01621459.2018.1543600

PEARSON, K. (1895). Vii. note on regression and inheritance in the case of two parents. *Proc. R. Soc. Lond.* **58** 240–242.

PFISTER, N., BÜHLMANN, P., SCHÖLKOPF, B. and PETERS, J. (2018). Kernel-based tests for joint independence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 5–31. MR3744710 https://doi.org/10.1111/rssb.12235

RINDT, D., SEJDINOVIC, D. and STEINSALTZ, D. (2021). Consistency of permutation tests of independence using distance covariance, HSIC and dHSIC. *Stat* **10** Paper No. e364, 10. MR4235612 https://doi.org/10.1002/sta4.364

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

SEJDINOVIC, D., SRIPERUMBUDUR, B., GRETTON, A. and FUKUMIZU, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.* **41** 2263–2291. MR3127866 https://doi.org/10.1214/13-AOS1140

SEN, A. and SEN, B. (2014). Testing independence and goodness-of-fit in linear models. *Biometrika* **101** 927–942. MR3286926 https://doi.org/10.1093/biomet/asu026

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0595165

SHAO, X. and ZHANG, J. (2014). Martingale difference correlation and its use in high-dimensional variable screening. *J. Amer. Statist. Assoc.* **109** 1302–1318. MR3265698 https://doi.org/10.1080/01621459.2014.887012

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. (2009). Brownian distance covariance. *Ann. Appl. Stat.* **3** 1236–1265. MR2752127 https://doi.org/10.1214/09-AOAS312

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. and RIZZO, M. L. (2014). Partial distance correlation with methods for dissimilarities. *Ann. Statist.* **42** 2382–2412. MR3269983 https://doi.org/10.1214/14-AOS1255

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

TASKINEN, S., KANKAINEN, A. and OJA, H. (2003). Sign test of independence between two random vectors. *Statist. Probab. Lett.* **62** 9–21. MR1965367 https://doi.org/10.1016/S0167-7152(02)00399-1

WAN, P. and DAVIS, R. A. (2022). Goodness-of-fit testing for time series models via distance covariance. *J. Econometrics* **227** 4–24. MR4377171 https://doi.org/10.1016/j.jeconom.2020.05.008

WANG, G., ZHU, K. and SHAO, X. (2022). Testing for the martingale difference hypothesis in multivariate time series models. *J. Bus. Econom. Statist.* **40** 980–994. MR4439266 https://doi.org/10.1080/07350015.2021.1889568

WANG, X., PAN, W., HU, W., TIAN, Y. and ZHANG, H. (2015). Conditional distance correlation. *J. Amer. Statist. Assoc.* **110** 1726–1734. MR3449068 https://doi.org/10.1080/01621459.2014.993081

WILKS, S. S. (1935). On the independence of k sets of normally distributed statistical variables. *Econometrica* 309–326.

XU, M., ZHANG, D. and WU, W. B. (2014). l^2 asymptotics for high-dimensional data. arXiv preprint. Available at arXiv:1405.7244.

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

ZHONG, W. and ZHU, L. (2015). An iterative approach to distance correlation-based sure independence screening. *J. Stat. Comput. Simul.* **85** 2331–2345. MR3339301 https://doi.org/10.1080/00949655.2014.928820

ZHOU, Z. (2012). Measuring nonlinear dependence in time-series, a distance correlation approach. *J. Time Series Anal.* **33** 438–457. MR2915095 https://doi.org/10.1111/j.1467-9892.2011.00780.x

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

The Annals of Statistics
2026, Vol. 54, No. 3, 1474–1501
https://doi.org/10.1214/25-AOS2614

ATTAINABILITY OF TWO-POINT TESTING RATES FOR FINITE-SAMPLE LOCATION ESTIMATION

BY SPENCER COMPTON[a] AND GREGORY VALIANT[b]

Department of Computer Science, Stanford University, [a]*comptons@stanford.edu,* [b]*valiant@stanford.edu*

Le Cam's two-point testing method yields perhaps the simplest lower bound for estimating the mean of a distribution: roughly, if it is impossible to well distinguish a distribution centered at μ from the same distribution centered at $\mu + \Delta$, then it is impossible to estimate the mean by better than $\Delta/2$. It is setting-dependent, whether or not a nearly matching upper bound is attainable. We study the conditions under which the two-point testing lower bound can be attained for univariate mean estimation; both in the setting of *location estimation* (where the distribution is known up to translation) and *adaptive location estimation* (unknown distribution). Roughly, we will say an estimate nearly attains the two-point testing lower bound if it incurs error that is at most polylogarithmically larger than the *Hellinger modulus of continuity* for $\tilde{\Omega}(n)$ samples.

Adaptive location estimation is particularly interesting, as some distributions admit much better guarantees than sub-Gaussian rates (e.g., $\mathrm{Unif}(\mu - 1, \mu + 1)$ permit error $\Theta(\frac{1}{n})$, while the sub-Gaussian rate is $\Theta(\frac{1}{\sqrt{n}})$), yet it is not obvious whether these rates may be adaptively attained by one unified approach. Our main result designs an algorithm that nearly attains the two-point testing rate for mixtures of symmetric, log-concave distributions with a common mean. Moreover, this algorithm runs in near-linear time and is parameter-free. In contrast, we show the two-point testing rate is not nearly attainable, even for symmetric, unimodal distributions.

We complement this with results for location estimation, showing the two-point testing rate is nearly attainable for unimodal distributions but unattainable for symmetric distributions.

REFERENCES

[1] BARTL, D. and MENDELSON, S. (2023). On a variance dependent Dvoretzky–Kiefer–Wolfowitz inequality. Available at arXiv:2308.04757.

[2] BERAN, R. (1978). An efficient and robust adaptive estimator of location. *Ann. Statist.* **6** 292–313. MR0518885

[3] BHATT, A., NAZER, B., ORDENTLICH, O. and POLYANSKIY, Y. (2021). Information-distilling quantizers. *IEEE Trans. Inf. Theory* **67** 2472–2487. MR4282367 https://doi.org/10.1109/TIT.2021.3059338

[4] 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

[5] BLANCHARD, M. and VORACEK, V. (2024). Tight bounds for local Glivenko–Cantelli. In *International Conference on Algorithmic Learning Theory* 179–220. PMLR.

[6] BOUSQUET, O., BOUCHERON, S. and LUGOSI, G. (2003). Introduction to statistical learning theory. In *Summer School on Machine Learning* 169–207. Springer, Berlin.

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

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

[9] CHAN, S.-O., DIAKONIKOLAS, I., SERVEDIO, R. A. and SUN, X. (2014). Efficient density estimation via piecewise polynomial approximation. In *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing* 604–613.

MSC2020 subject classifications. 62G99.
Key words and phrases. Location estimation, adaptive estimators, log-concave, shape constraint.

[10] CHIERICHETTI, F., DASGUPTA, A., KUMAR, R. and LATTANZI, S. (2014). Learning entangled single-sample Gaussians. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms* 511–522. ACM, New York. MR3376398 https://doi.org/10.1137/1.9781611973402.38

[11] COMPTON, S. and VALIANT, G. (2024). Near-optimal mean estimation with unknown, heteroskedastic variances. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing* 194–200.

[12] COMPTON, S. and VALIANT, G. (2026). Supplement to "Attainability of two-point testing rates for finite-sample location estimation." https://doi.org/10.1214/25-AOS2614SUPPA, https://doi.org/10.1214/25-AOS2614SUPPB

[13] DALALYAN, A. S., GOLUBEV, G. K. and TSYBAKOV, A. B. (2006). Penalized maximum likelihood and semiparametric second-order efficiency. *Ann. Statist.* **34** 169–201. MR2275239 https://doi.org/10.1214/009053605000000895

[14] DASGUPTA, S., HSU, D. J. and MONTELEONI, C. (2007). A general agnostic active learning algorithm. *Adv. Neural Inf. Process. Syst.* **20**.

[15] DEVROYE, L., LATTANZI, S., LUGOSI, G. and ZHIVOTOVSKIY, N. (2023). On mean estimation for heteroscedastic random variables. *Ann. Inst. Henri Poincaré Probab. Stat.* **59** 1–20. MR4533719 https://doi.org/10.1214/21-aihp1239

[16] DEVROYE, L. and LUGOSI, G. (2001). *Combinatorial Methods in Density Estimation*. *Springer Series in Statistics*. Springer, New York. MR1843146 https://doi.org/10.1007/978-1-4613-0125-7

[17] DIAKONIKOLAS, I., KANE, D. M. and LIU, S. (2024). Testing closeness of multivariate distributions via Ramsey theory. In *STOC'24—Proceedings of the 56th Annual ACM Symposium on Theory of Computing* 340–347. ACM, New York. MR4764818 https://doi.org/10.1145/3618260.3649657

[18] DIAKONIKOLAS, I., KANE, D. M. and NIKISHKIN, V. (2015). Testing identity of structured distributions. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms* 1841–1854. SIAM, Philadelphia, PA. MR3451147 https://doi.org/10.1137/1.9781611973730.123

[19] DIAKONIKOLAS, I., KANE, D. M. and NIKISHKIN, V. (2015). Optimal algorithms and lower bounds for testing closeness of structured distributions. In 2015 *IEEE 56th Annual Symposium on Foundations of Computer Science—FOCS* 2015 1183–1202. IEEE Computer Soc., Los Alamitos, CA. MR3473364 https://doi.org/10.1109/FOCS.2015.76

[20] DIAKONIKOLAS, I., KANE, D. M. and NIKISHKIN, V. (2017). Near-optimal closeness testing of discrete histogram distributions. In *44th International Colloquium on Automata, Languages, and Programming* (*ICALP* 2017). Schloss-Dagstuhl-Leibniz Zentrum Für Informatik.

[21] DIAKONIKOLAS, I., KANE, D. M. and PEEBLES, J. (2019). Testing identity of multidimensional histograms. In *Conference on Learning Theory* 1107–1131. PMLR.

[22] DONOHO, D. L. and LIU, R. C. (1987). Geometrizing rates of convergence. *Ann. Statist.*

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

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

[25] DUCHI, J. C. and RUAN, F. (2024). The right complexity measure in locally private estimation: It is not the Fisher information. *Ann. Statist.* **52** 1–51. MR4718406 https://doi.org/10.1214/22-aos2227

[26] FOSTER, D. J., KAKADE, S. M., QIAN, J. and RAKHLIN, A. (2021). The statistical complexity of interactive decision making. Available at arXiv:2112.13487.

[27] GERBER, P. R., JIANG, T., POLYANSKIY, Y. and SUN, R. (2025). Density estimation using the perceptron. *J. Mach. Learn. Res.* **26** 160. MR4962914

[28] GUPTA, S., HOPKINS, S. and PRICE, E. (2024). Beyond Catoni: Sharper rates for heavy-tailed and robust mean estimation. In *The Thirty Seventh Annual Conference on Learning Theory* 2232–2269. PMLR.

[29] GUPTA, S., LEE, J., PRICE, E. and VALIANT, P. (2022). Finite-sample maximum likelihood estimation of location. *Adv. Neural Inf. Process. Syst.* **35** 30139–30149.

[30] GUPTA, S., LEE, J., PRICE, E. and VALIANT, P. (2024). Minimax-optimal location estimation. *Adv. Neural Inf. Process. Syst.* **36**.

[31] GUPTA, S., LEE, J. C. and PRICE, E. (2023). Finite-sample symmetric mean estimation with Fisher information rate. In *The Thirty Sixth Annual Conference on Learning Theory* 4777–4830. PMLR.

[32] HAN, Q. and WELLNER, J. A. (2016). Approximation and estimation of s-concave densities via Rényi divergences. *Ann. Statist.* **44** 1332–1359. MR3485962 https://doi.org/10.1214/15-AOS1408

[33] JUDITSKY, A. B. and NEMIROVSKI, A. S. (2009). Nonparametric estimation by convex programming. *Ann. Statist.* **37** 2278–2300. MR2543692 https://doi.org/10.1214/08-AOS654

[34] KAO, Y.-C., XU, M. and ZHANG, C.-H. (2024). Choosing the p in L_p loss: Adaptive rates for symmetric mean estimation. In *The Thirty Seventh Annual Conference on Learning Theory* 2795–2839. PMLR.

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

[36] KUR, G., DAGAN, Y. and RAKHLIN, A. (2019). Optimality of maximum likelihood for log-concave density estimation and bounded convex regression. Available at arXiv:1903.05315.

[37] LAHA, N. (2019). Location estimation for symmetric log-concave densities. Available at arXiv:1911.06225.
[38] LE CAM, L. (2012). *Asymptotic Methods in Statistical Decision Theory*. Springer, Berlin.
[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] LECAM, L. (1973). Convergence of estimates under dimensionality restrictions. *Ann. Statist.* **1** 38–53. MR0334381
[41] LEE, J. C. and VALIANT, P. (2022). Optimal sub-Gaussian mean estimation in very high dimensions. In 13*th Innovations in Theoretical Computer Science Conference* (*ITCS* 2022). Schloss-Dagstuhl-Leibniz Zentrum Für Informatik.
[42] LEE, J. C. H. and VALIANT, P. (2022). Optimal sub-Gaussian mean estimation in $\mathbb{R}$. In 2021 *IEEE* 62*nd Annual Symposium on Foundations of Computer Science—FOCS* 2021 672–683. IEEE Computer Soc., Los Alamitos, CA. MR4399724 https://doi.org/10.1109/FOCS52979.2021.00071
[43] LIANG, Y. and YUAN, H. (2020). Learning entangled single-sample Gaussians in the subset-of-signals model. In *Conference on Learning Theory* 2712–2737. PMLR.
[44] PENSIA, A., JOG, V. and LOH, P.-L. (2022). Estimating location parameters in sample-heterogeneous distributions. *Inf. Inference* **11** 959–1036. MR4492472 https://doi.org/10.1093/imaiai/iaab013
[45] PENSIA, A., JOG, V. and LOH, P.-L. (2024). Communication-constrained hypothesis testing: Optimality, robustness, and reverse data processing inequalities. *IEEE Trans. Inf. Theory* **70** 389–414. MR4692681 https://doi.org/10.1109/tit.2023.3334024
[46] POLYANSKIY, Y. and WU, Y. (2019). Dualizing Le Cam's method for functional estimation, with applications to estimating the unseens. Available at arXiv:1902.05616.
[47] POLYANSKIY, Y. and WU, Y. (2025). *Information Theory*: *From Coding to Learning*. Cambridge Univ. Press, Cambridge.
[48] REEVE, H. W. (2024). A short proof of the Dvoretzky–Kiefer–Wolfowitz–Massart inequality. Available at arXiv:2403.16651.
[49] SACKS, J. (1975). An asymptotically efficient sequence of estimators of a location parameter. *Ann. Statist.* **3** 285–298. MR0359174
[50] STEIN, C. et al. (1956). Efficient nonparametric testing and estimation. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* **1** 187–195.
[51] STONE, C. J. (1975). Adaptive maximum likelihood estimators of a location parameter. *Ann. Statist.* **3** 267–284. MR0362669
[52] VAN EEDEN, C. (1970). Efficiency-robust estimation of location. *Ann. Math. Statist.* **41** 172–181. MR0263194 https://doi.org/10.1214/aoms/1177697197
[53] VAN DER VAART, A. (2002). The statistical work of Lucien Le Cam. *Ann. Statist.* **30** 631–682.
[54] VAN DER VAART, A. W. (2000). *Asymptotic Statistics* 3. Cambridge Univ. Press, Cambridge.
[55] VAPNIK, V. N. and CHERVONENKIS, A. Y. (2015). On the uniform convergence of relative frequencies of events to their probabilities. In *Measures of Complexity*: *Festschrift for Alexey Chervonenkis* 11–30. Springer, Cham. MR3408730
[56] YUAN, H. and LIANG, Y. (2020). Learning entangled single-sample distributions via iterative trimming. In *International Conference on Artificial Intelligence and Statistics* 2666–2676. PMLR.

The Annals of Statistics
2026, Vol. 54, No. 3, 1502–1522
https://doi.org/10.1214/25-AOS2615

MINIMAX OPTIMAL SERIATION IN POLYNOMIAL TIME

BY YANN ISSARTEL[1,a], CHRISTOPHE GIRAUD[2,b] AND NICOLAS VERZELEN[3,c]

[1]*LTCI, Télécom Paris, Institut Polytechnique de Paris,* [a]*yann.issartel@telecom-paris.fr*
[2]*Laboratoire de Mathématiques d'Orsay, Université Paris-Saclay,* [b]*christophe.giraud@universite-paris-saclay.fr*
[3]*MISTEA, INRAE, Institut Agro, Université Montpellier,* [c]*Nicolas.Verzelen@inrae.fr*

We consider the seriation problem, whose goal is to recover a hidden ordering from a noisy observation of a permuted Robinson matrix. We establish sharp minimax rates under average-Lipschitz conditions that strictly extend the bi-Lipschitz framework of Giraud, Issartel and Verzelen (*Electron. J. Stat.* (2023) **17** 1587–1662). We further design a polynomial-time algorithm that attains these optimal rates, thereby resolving two open questions raised in Giraud, Issartel and Verzelen. Finally, our analysis extends to a broader class of matrices beyond those generated by exact permutations.

REFERENCES

[1] ATKINS, J. E., BOMAN, E. G. and HENDRICKSON, B. (1998). A spectral algorithm for seriation and the consecutive ones problem. *SIAM J. Comput.* **28** 297–310. MR1630473 https://doi.org/10.1137/S0097539795285771

[2] BARNARD, S. T., POTHEN, A. and SIMON, H. (1995). A spectral algorithm for envelope reduction of sparse matrices. *Numer. Linear Algebra Appl.* **2** 317–334. MR1349276 https://doi.org/10.1002/nla.1680020402

[3] BERENFELD, C., CARPENTIER, A. and VERZELEN, N. (2024). Seriation of Toeplitz and latent position matrices: Optimal rates and computational trade-offs. arXiv preprint. Available at arXiv:2408.10004.

[4] BRAVERMAN, M. and MOSSEL, E. (2009). Sorting from noisy information. arXiv preprint. Available at arXiv:0910.1191.

[5] CAI, T. T. and MA, R. (2024). Matrix reordering for noisy disordered matrices: Optimality and computationally efficient algorithms. *IEEE Trans. Inf. Theory* **70** 509–531. MR4692686 https://doi.org/10.1109/tit.2023.3305538

[6] CARMONA, M., CHEPOI, V., NAVES, G. and PRÉA, P. (2024). Modules in Robinson spaces. *SIAM J. Discrete Math.* **38** 190–224. MR4685945 https://doi.org/10.1137/22M1494348

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

[8] CHEN, P., GAO, C. and ZHANG, A. Y. (2022). Optimal full ranking from pairwise comparisons. *Ann. Statist.* **50** 1775–1805. MR4441140 https://doi.org/10.1214/22-aos2175

[9] CHEN, Y., FAN, J., MA, C. and WANG, K. (2019). Spectral method and regularized MLE are both optimal for top-K ranking. *Ann. Statist.* **47** 2204–2235. MR3953449 https://doi.org/10.1214/18-AOS1745

[10] CHEPOI, V. and SESTON, M. (2009). An approximation algorithm for l_∞-fitting Robinson structures to distances. In *STACS* 2009: 26*th International Symposium on Theoretical Aspects of Computer Science*. *LIPIcs. Leibniz Int. Proc. Inform.* **3** 265–276. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR2870658

[11] CHESHIRE, J. and ISSARTEL, Y. (2025). Active seriation: Efficient ordering recovery with statistical guarantees. In *The Thirty-Ninth Annual Conference on Neural Information Processing Systems*.

[12] COLLIER, O. and DALALYAN, A. S. (2016). Minimax rates in permutation estimation for feature matching. *J. Mach. Learn. Res.* **17** Paper No. 6, 31. MR3482926

[13] DIACONIS, P. and GRAHAM, R. L. (2018). Spearman's footrule as a measure of disarray. *J. Roy. Statist. Soc. Ser. B, Methodol.* **39** 262–268. MR0652736

MSC2020 subject classifications. Primary 62G05; secondary 62C20, 62H20, 15B05, 90C27.

Key words and phrases. Seriation, Robinson matrices, minimax rates, distance-based algorithms, Lipschitz condition, Toeplitz matrices, Kendall tau, Frobenius norm, permutation estimation.

[14] ELSON, J., KARP, R. M., PAPADIMITRIOU, C. H. and SHENKER, S. (2004). Global synchronization in sensornets. In *LATIN* 2004: *Theoretical Informatics* (M. Farach-Colton, ed.). *Lecture Notes in Computer Science* **2976** 609–624. Springer, Berlin. MR2095234 https://doi.org/10.1007/978-3-540-24698-5_63

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

[16] FOGEL, F., D'ASPREMONT, A. and VOJNOVIC, M. (2014). SerialRank: Spectral ranking using seriation. In *Advances in Neural Information Processing Systems* (Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence and K. Q. Weinberger, eds.) 27. Curran Associates, Red Hook.

[17] FOGEL, F., JENATTON, R., BACH, F. and D'ASPREMONT, A. (2015). Convex relaxations for permutation problems. *SIAM J. Matrix Anal. Appl.* **36** 1465–1488. MR3418224 https://doi.org/10.1137/130947362

[18] FULKERSON, D. R. and GROSS, O. A. (1965). Incidence matrices and interval graphs. *Pacific J. Math.* **15** 835–855. MR0186421

[19] GALSTYAN, T., MINASYAN, A. and DALALYAN, A. S. (2022). Optimal detection of the feature matching map in presence of noise and outliers. *Electron. J. Stat.* **16** 5720–5750. MR4500636 https://doi.org/10.1214/22-ejs2076

[20] GARRIGA, G. C., JUNTTILA, E. and MANNILA, H. (2011). Banded structure in binary matrices. *Knowl. Inf. Syst.* **28** 197–226.

[21] GIRAUD, C., ISSARTEL, Y. and VERZELEN, N. (2023). Localization in 1D non-parametric latent space models from pairwise affinities. *Electron. J. Stat.* **17** 1587–1662. MR4612174 https://doi.org/10.1214/23-ejs2134

[22] GIRIDHAR, A. and KUMAR, P. R. (2006). Distributed clock synchronization over wireless networks: Algorithms and analysis. In *Proceedings of the* 45*th IEEE Conference on Decision and Control* 4915–4920.

[23] HÜTTER, J.-C., MAO, C., RIGOLLET, P. and ROBEVA, E. (2020). Estimation of Monge matrices. *Bernoulli* **26** 3051–3080. MR4140537 https://doi.org/10.3150/20-BEJ1215

[24] ISSARTEL, Y. (2021). On the estimation of network complexity: Dimension of graphons. *J. Mach. Learn. Res.* **22** Paper No. 191, 62. MR4329770

[25] ISSARTEL, Y., GIRAUD, C. and VERZELEN, N. (2026). Supplement to "Minimax optimal seriation in polynomial time." https://doi.org/10.1214/25-AOS2615SUPP

[26] JANSSEN, J. and SMITH, A. (2022). Reconstruction of line-embeddings of graphons. *Electron. J. Stat.* **16** 331–407. MR4361745 https://doi.org/10.1214/21-ejs1940

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

[28] MAO, C., WEED, J. and RIGOLLET, P. (2018). Minimax rates and efficient algorithms for noisy sorting. In *Algorithmic Learning Theory* 2018. *Proc. Mach. Learn. Res.* (*PMLR*) **83** 27. PMLR. MR3857331

[29] MASSART, P. (2007). *Concentration Inequalities and Model Selection. Lecture Notes in Math.* **1896**. Springer, Berlin. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003. With a foreword by Jean Picard. MR2319879

[30] NATIK, A. and SMITH, A. (2026). Consistency of spectral seriation. *ESAIM Probab. Stat.* **30** 120–161. MR5036760 https://doi.org/10.1051/ps/2025018

[31] PRÉA, P. and FORTIN, D. (2014). An optimal algorithm to recognize Robinsonian dissimilarities. *J. Classification* **31** 351–385. MR3277710 https://doi.org/10.1007/s00357-014-9150-2

[32] RECANATI, A., BRÜLS, T. and D'ASPREMONT, A. (2017). A spectral algorithm for fast de novo layout of uncorrected long nanopore reads. *Bioinformatics* **33** 3188–3194.

[33] RECANATI, A., KERDREUX, T. and D'ASPREMONT, A. (2018). Reconstructing latent orderings by spectral clustering. arXiv preprint. Available at arXiv:1807.07122.

[34] RIGOLLET, P. and HÜTTER, J.-C. (2023). High-dimensional statistics. arXiv preprint. Available at arXiv:2310.19244.

[35] ROBINSON, W. S. (1951). A method for chronologically ordering archaeological deposits. *Amer. Antiq.* **16** 293–301.

[36] ROCHA, I., JANSSEN, J. and KALYANIWALLA, N. (2018). Recovering the structure of random linear graphs. *Linear Algebra Appl.* **557** 234–264. MR3848269 https://doi.org/10.1016/j.laa.2018.07.029

[37] 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. With a foreword by Sara van de Geer. MR3837109 https://doi.org/10.1017/9781108231596

The Annals of Statistics
2026, Vol. 54, No. 3, 1523–1539
https://doi.org/10.1214/25-AOS2616

DUAL INDUCTION CLT FOR HIGH-DIMENSIONAL *m*-DEPENDENT DATA

BY HEEJONG BONG[1,a], ARUN KUMAR KUCHIBHOTLA[2,b] AND ALESSANDRO RINALDO[3,c]

[1]*Department of Statistics, Purdue University,* [a]*bong0@purdue.edu*
[2]*Department of Statistics and Data Science, Carnegie Mellon University,* [b]*arunku@stat.cmu.edu*
[3]*Department of Statistics and Data Science, University of Texas at Austin,* [c]*alessandro.rinaldo@austin.utexas.edu*

We derive novel and sharp high-dimensional Berry–Esseen bounds for the sum of m-dependent random vectors over the class of hyperrectangles exhibiting only a poly-logarithmic dependence in the dimension. Our results hold under minimal assumptions, such as nondegenerate covariances and finite third moments, and exhibit an optimal sample complexity of order $m^{(q-1)/(q-2)}/\sqrt{n}$. Aside from logarithmic terms, the resulting rates match the optimal rates established in the univariate case. When specialized to the sums of independent nondegenerate random vectors, our results produce sharp and, in some cases, optimal rates under the weakest possible conditions. We develop a novel inductive relationship between anticoncentration inequalities and Berry–Esseen bounds inspired by the classical Lindeberg swapping method and the concentration inequality approach for dependent data that may be of independent interest.

REFERENCES

ANDREWS, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59** 817–858. https://doi.org/10.2307/2938229

BELLONI, A., CHERNOZHUKOV, V., CHETVERIKOV, D., HANSEN, C. B. and KATO, K. (2018). High-dimensional econometrics and regularized GMM cemmap working paper No. CWP35/18, Centre for Microdata Methods and Practice (cemmap), London. https://doi.org/10.1920/wp.cem.2018.3518.

BENTKUS, V. (2003). Dependence of the Berry–Esseen bound on the dimension. *J. Statist. Plann. Inference* **113** 385–402. https://doi.org/10.1016/S0378-3758(02)00094-0

BENTKUS, V. (2005). A Lyapunov-type bound in $\mathbb{R}^d$. *Theory Probab. Appl.* **49** 311–323. https://doi.org/10.1137/S0040585X97981123

BENTKUS, V., GÖTZE, F., PAULAUSKAS, V. and RAČKAUSKAS, A. (2000). The accuracy of Gaussian approximation in Banach spaces. In *Limit Theorems of Probability Theory* (Y. V. Prokhorov and V. Statulevičius, eds.) 25–111. Springer, Berlin. https://doi.org/10.1007/978-3-662-04172-7_2

BENTKUS, V. Y. (1987). Large deviations in Banach spaces. *Theory Probab. Appl.* **31** 627–632. https://doi.org/10.1137/1131085

BERK, K. N. (1973). A central limit theorem for m-dependent random variables with unbounded m. *Ann. Probab.* **1** 352–354. https://doi.org/10.1214/aop/1176996992

BONG, H., KUCHIBHOTLA, A. K. and RINALDO, A. (2026). Supplement to "Dual induction CLT for high-dimensional m-dependent data." https://doi.org/10.1214/25-AOS2616SUPP

CHANG, J., CHEN, X. and WU, M. (2024). Central limit theorems for high dimensional dependent data. *Bernoulli* **30** 712–742. https://doi.org/10.3150/23-BEJ1614

CHEN, L. H. Y. and SHAO, Q.-M. (2004). Normal approximation under local dependence. *Ann. Probab.* **32** 1985–2028. https://doi.org/10.1214/009117904000000450

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. https://doi.org/10.1214/13-AOS1161

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. https://doi.org/10.1214/16-AOP1113

MSC2020 subject classifications. Primary 60B12; secondary 60F05.

Key words and phrases. High-dimensional inference, central limit theorem, Berry–Esseen bound, m-dependence structure.

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2018). Inference on causal and structural parameters using many moment inequalities. *Rev. Econ. Stud.* **86** 1867–1900. https://doi.org/10.1093/restud/rdy065

CHERNOZHUKOV, V., CHETVERIKOV, D., KATO, K. and KOIKE, Y. (2022). Improved central limit theorem and bootstrap approximations in high dimensions. *Ann. Statist.* **50** 2562–2586. https://doi.org/10.1214/22-AOS2193

CHERNOZHUKOV, V., CHETVERIKOV, D., KATO, K. and KOIKE, Y. (2023). High-dimensional data bootstrap. *Annu. Rev. Stat. Appl.* **10** 427–449. https://doi.org/10.1146/annurev-statistics-040120-022239

CHERNOZHUKOV, V., CHETVERIKOV, D. and KOIKE, Y. (2023). Nearly optimal central limit theorem and bootstrap approximations in high dimensions. *Ann. Appl. Probab.* **33** 2374–2425. https://doi.org/10.1214/22-AAP1870

CHIANG, H. D., KATO, K. and SASAKI, Y. (2023). Inference for high-dimensional exchangeable arrays. *J. Amer. Statist. Assoc.* **118** 1595–1605. https://doi.org/10.1080/01621459.2021.2000868

DAS, D. and LAHIRI, S. (2021). Central limit theorem in high dimensions: The optimal bound on dimension growth rate. *Trans. Amer. Math. Soc.* **374** 6991–7009. https://doi.org/10.1090/tran/8459

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

FANG, X. and KOIKE, Y. (2020). Large-dimensional central limit theorem with fourth-moment error bounds on convex sets and balls. arXiv preprint. Available at arXiv:2009.00339.

FANG, X. and KOIKE, Y. (2021). High-dimensional central limit theorems by Stein's method. *Ann. Appl. Probab.* **31** 1660–1686. https://doi.org/10.1214/20-AAP1629

FANG, X., KOIKE, Y., LIU, S.-H. and YI-KUN, Z. (2023). High-dimensional central limit theorems by Stein's method in the degenerate case. arXiv preprint. Available at arXiv:2305.17365.

KOCK, A. B. and PREINERSTORFER, D. (2025). High-dimensional Gaussian approximations for robust means. arXiv preprint. Available at arXiv:2504.08435.

KOIKE, Y. (2021). Notes on the dimension dependence in high-dimensional central limit theorems for hyperrectangles. *Jpn. J. Stat. Data Sci.* **4** 257–297. https://doi.org/10.1007/s42081-020-00096-7

KOIKE, Y. (2023). High-dimensional central limit theorems for homogeneous sums. *J. Theoret. Probab.* **36** 1–45. https://doi.org/10.1007/s10959-022-01156-2

KOJEVNIKOV, D. and SONG, K. (2022). A Berry–Esseen bound for vector-valued martingales. *Statist. Probab. Lett.* **186** 109448. https://doi.org/10.1016/j.spl.2022.109448

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

KUCHIBHOTLA, A. K. and RINALDO, A. (2020). High-dimensional CLT for sums of non-degenerate random Vectors. $n^{-(1/2)}$-rate. arXiv preprint. Available at arXiv:2009.13673.

KURISU, D., KATO, K. and SHAO, X. (2024). Gaussian approximation and spatially dependent wild bootstrap for high-dimensional spatial data. *J. Amer. Statist. Assoc.* **119** 1820–1832. https://doi.org/10.1080/01621459.2023.2218578

LOPES, M. E. (2022). Central limit theorem and bootstrap approximation in high dimensions: Near $1/\sqrt{n}$ rates via implicit smoothing. *Ann. Statist.* **50** 2492–2513. https://doi.org/10.1214/22-AOS2184

LOPES, M. E., LIN, Z. and MÜLLER, H.-G. (2020). Bootstrapping max statistics in high dimensions: Near-parametric rates under weak variance decay and application to functional and multinomial data. *Ann. Statist.* **48** 1214–1229. https://doi.org/10.1214/19-AOS1844

NAZAROV, F. (2003). On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a Gaussian measure. In *Geometric Aspects of Functional Analysis*: *Israel Seminar* 2001–2002 (V. D. Milman and G. Schechtman, eds.) 169–187. Springer, Berlin. https://doi.org/10.1007/978-3-540-36428-3_15

NEWEY, W. K. and WEST, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55** 703–708. https://doi.org/10.2307/1913610

NORVAIŠA, R. and PAULAUSKAS, V. (1991). Rate of convergence in the central limit theorem for empirical processes. *J. Theoret. Probab.* **4** 511–534. MR1115160 https://doi.org/10.1007/BF01210322

PAULAUSKAS, V. and RAČKAUSKAS, A. A. (2012). *Approximation Theory in the Central Limit Theorem*: *Exact Results in Banach Spaces* 32. Springer, Berlin. https://doi.org/10.1007/978-94-011-7798-6

RAIČ, M. (2019). A multivariate Berry–Esseen theorem with explicit constants. *Bernoulli* **25** 2824–2853. MR4003566 https://doi.org/10.3150/18-BEJ1072

RINOTT, Y. and ROTAR, V. (1999). Some bounds on the rate of convergence in the CLT for martingales. I. *Theory Probab. Appl.* **43** 604–619. https://doi.org/10.1137/S0040585X97977148

ROMANO, J. P. and WOLF, M. (2000). A more general central limit theorem for m-dependent random variables with unbounded m. *Statist. Probab. Lett.* **47** 115–124. MR1747098 https://doi.org/10.1016/S0167-7152(99)00146-7

SAZONOV, V. V. (1981). *Normal Approximation—Some Recent Advances*. *Lecture Notes in Math.* **879**. Springer, Berlin. MR0643968

SENATOV, V. V. (1998). *Normal Approximation*: *New Results*, *Methods and Problems*. de Gruyter, Berlin. https://doi.org/10.1515/9783110933666

SHERGIN, V. V. (1980). On the convergence rate in the central limit theorem for m-dependent random variables. *Theory Probab. Appl*. **24** 782–796. https://doi.org/10.1137/1124090

ZHANG, D. and WU, W. B. (2017). Gaussian approximation for high dimensional time series. *Ann*. *Statist*. **45** 1895–1919. https://doi.org/10.1214/16-AOS1512

ZHANG, H. and CHEN, S. X. (2020). Concentration inequalities for statistical inference. arXiv preprint. Available at arXiv:2011.02258.

ZHANG, X. and CHENG, G. (2014). Bootstrapping high dimensional time series. arXiv preprint. Available at arXiv:1406.1037.

ZHANG, X. and CHENG, G. (2018). Gaussian approximation for high dimensional vector under physical dependence. *Bernoulli* **24** 2640–2675. https://doi.org/10.3150/17-BEJ939

ZHILOVA, M. (2020). Nonclassical Berry–Esseen inequalities and accuracy of the bootstrap. *Ann*. *Statist*. **48** 1922–1939. https://doi.org/10.1214/18-AOS1802

The Annals of Statistics
2026, Vol. 54, No. 3, 1540–1565
https://doi.org/10.1214/25-AOS2617

STATISTICAL INFERENCE IN TENSOR COMPLETION: OPTIMAL UNCERTAINTY QUANTIFICATION AND STATISTICAL-TO-COMPUTATIONAL GAPS

BY WANTENG MA[1,a] AND DONG XIA[2,b]

[1]*Department of Statistics and Data Science, University of Pennsylvania,* [a]*wanteng@wharton.upenn.edu*
[2]*Department of Mathematics, Hong Kong University of Science and Technology,* [b]*madxia@ust.hk*

This paper presents a simple yet efficient method for statistical inference of tensor linear forms using incomplete and noisy observations. Under the Tucker low-rank tensor model and the missing-at-random assumption, we utilize an appropriate initial estimate along with a debiasing technique followed by a one-step power iteration to construct an asymptotically normal test statistic. This method is suitable for various statistical inference tasks, including constructing confidence intervals, inference under heteroskedastic and sub-exponential noise, and simultaneous testing. We demonstrate that the estimator achieves the Cramér–Rao lower bound on Riemannian manifolds, indicating its optimality in uncertainty quantification. We comprehensively examine statistical-to-computational gaps and investigate the impact of initialization on the minimal conditions regarding sample sizes and signal-to-noise ratios required for accurate inference. Our findings show that, with independent initialization, statistically optimal sample sizes and signal-to-noise ratios are sufficient for accurate inference. Conversely, if only dependent initialization is available, computationally optimal sample sizes and signal-to-noise ratios still guarantee asymptotic normality without the need for data-splitting. We present the phase transition between computational and statistical limits. Numerical simulation results align with the theoretical findings.

REFERENCES

[1] ABSIL, P.-A., MAHONY, R. and SEPULCHRE, R. (2008). *Optimization Algorithms on Matrix Manifolds*. Princeton Univ. Press, Princeton, NJ. With a foreword by Paul Van Dooren. MR2364186 https://doi.org/10.1515/9781400830244

[2] ABSIL, P.-A. and MALICK, J. (2012). Projection-like retractions on matrix manifolds. *SIAM J. Optim.* **22** 135–158. MR2902688 https://doi.org/10.1137/100802529

[3] AMJAD, M. J. and SHAH, D. (2017). Censored demand estimation in retail. *Proc. ACM Meas. Anal. Comput. Syst.* **1** 1–28.

[4] BADINGER, H. and EGGER, P. (2013). Estimation and testing of higher-order spatial autoregressive panel data error component models. *J. Geogr. Syst.* **15** 453–489.

[5] BAHADORI, M. T., YU, Q. R. and LIU, Y. (2014). Fast multivariate spatio-temporal analysis via low rank tensor learning. *Adv. Neural Inf. Process. Syst.* **27**.

[6] BAI, J. and NG, S. (2021). Matrix completion, counterfactuals, and factor analysis of missing data. *J. Amer. Statist. Assoc.* **116** 1746–1763. MR4353711 https://doi.org/10.1080/01621459.2021.1967163

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

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

[9] BOUMAL, N. (2023). *An Introduction to Optimization on Smooth Manifolds*. Cambridge Univ. Press, Cambridge. MR4533407

[10] CAHAN, E., BAI, J. and NG, S. (2023). Factor-based imputation of missing values and covariances in panel data of large dimensions. *J. Econometrics* **233** 113–131. MR4554729 https://doi.org/10.1016/j.jeconom.2022.01.006

MSC2020 subject classifications. Primary 62E17, 62H10; secondary 62F30.
Key words and phrases. Asymptotic distribution, tensor completion, confidence interval.

[11] CAI, C., LI, G., POOR, H. V. and CHEN, Y. (2019). Nonconvex low-rank tensor completion from noisy data. In *Advances in Neural Information Processing Systems* (H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox and R. Garnett, eds.). **32**. Curran Associates, Red Hook.

[12] CAI, C., POOR, H. V. and CHEN, Y. (2023). Uncertainty quantification for nonconvex tensor completion: Confidence intervals, heteroscedasticity and optimality. *IEEE Trans. Inf. Theory* **69** 407–452. MR4544966 https://doi.org/10.1109/tit.2022.3205781

[13] CAI, J.-F., LI, J. and XIA, D. (2022). Provable tensor-train format tensor completion by Riemannian optimization. *J. Mach. Learn. Res.* **23** Paper No. [123], 77. MR4577075

[14] CAI, J.-F., LI, J. and XIA, D. (2023). Generalized low-rank plus sparse tensor estimation by fast Riemannian optimization. *J. Amer. Statist. Assoc.* **118** 2588–2604. MR4681606 https://doi.org/10.1080/01621459.2022.2063131

[15] CAI, J.-F., LI, J. and XIA, D. (2023). Online Tensor Learning: Computational and Statistical Trade-offs, Adaptivity and Optimal Regret. arXiv preprint. Available at arXiv:2306.03372.

[16] CAI, T. T., LIANG, T. and RAKHLIN, A. (2016). Geometric inference for general high-dimensional linear inverse problems. *Ann. Statist.* **44** 1536–1563. MR3519932 https://doi.org/10.1214/15-AOS1426

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

[18] CARPENTIER, A., KLOPP, O., LÖFFLER, M. and NICKL, R. (2018). Adaptive confidence sets for matrix completion. *Bernoulli* **24** 2429–2460. MR3779691 https://doi.org/10.3150/17-BEJ933

[19] CARROLL, J. D. and CHANG, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart–Young" decomposition. *Psychometrika* **35** 283–319.

[20] CHEN, H., RASKUTTI, G. and YUAN, M. (2019). Non-convex projected gradient descent for generalized low-rank tensor regression. *J. Mach. Learn. Res.* **20** Paper No. 5, 37. MR3911412

[21] 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

[22] CHEN, X., HE, Z. and SUN, L. (2019). A Bayesian tensor decomposition approach for spatiotemporal traffic data imputation. *Transp. Res., Part C, Emerg. Technol.* **98** 73–84.

[23] CHEN, Y., CHI, Y., FAN, J., MA, C. et al. (2021). Spectral methods for data science: A statistical perspective. *Found. Trends Mach. Learn.* **14** 566–806.

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

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

[26] CHERNOZHUKOV, V., HANSEN, C., LIAO, Y. and ZHU, Y. (2023). Inference for low-rank models. *Ann. Statist.* **51** 1309–1330. MR4630950 https://doi.org/10.1214/23-aos2293

[27] CHOI, J. and YUAN, M. (2024). Matrix completion when missing is not at random and its applications in causal panel data models. *J. Amer. Statist. Assoc.* 1–24. https://doi.org/10.1080/01621459.2024.2380105

[28] DING, L. and CHEN, Y. (2020). Leave-one-out approach for matrix completion: Primal and dual analysis. *IEEE Trans. Inf. Theory* **66** 7274–7301. MR4173640 https://doi.org/10.1109/TIT.2020.2992769

[29] FARIAS, V., LI, A. A. and PENG, T. (2022). Uncertainty quantification for low-rank matrix completion with heterogeneous and sub-exponential noise. In *International Conference on Artificial Intelligence and Statistics* 1179–1189. PMLR.

[30] GANDY, S., RECHT, B. and YAMADA, I. (2011). Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Probl.* **27** 025010, 19. MR2765628 https://doi.org/10.1088/0266-5611/27/2/025010

[31] HAN, F., WANG, H., ZHANG, G., HAN, H., SONG, B., LI, L., MOORE, W., LU, H., ZHAO, H. et al. (2015). Texture feature analysis for computer-aided diagnosis on pulmonary nodules. *J. Digit. Imag.* **28** 99–115.

[32] HAN, R., WILLETT, R. and ZHANG, A. R. (2022). An optimal statistical and computational framework for generalized tensor estimation. *Ann. Statist.* **50** 1–29. MR4382094 https://doi.org/10.1214/21-AOS2061

[33] JAGANNATH, A., LOPATTO, P. and MIOLANE, L. (2020). Statistical thresholds for tensor PCA. *Ann. Appl. Probab.* **30** 1910–1933. MR4132641 https://doi.org/10.1214/19-AAP1547

[34] JAIN, P. and OH, S. (2014). Provable tensor factorization with missing data. *Adv. Neural Inf. Process. Syst.* **27**.

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

[36] JAVANMARD, A. and MONTANARI, A. (2018). Debiasing the Lasso: Optimal sample size for Gaussian designs. *Ann. Statist.* **46** 2593–2622. MR3851749 https://doi.org/10.1214/17-AOS1630
[37] JING, B.-Y., LI, T., LYU, Z. and XIA, D. (2021). Community detection on mixture multilayer networks via regularized tensor decomposition. *Ann. Statist.* **49** 3181–3205. MR4352527 https://doi.org/10.1214/21-aos2079
[38] KELLEY, T. L. (1928). *Crossroads in the Mind of Man*: *A Study of Differentiable Mental Abilities*. Stanford Univ. press, Stanford.
[39] KOCH, O. and LUBICH, C. (2010). Dynamical tensor approximation. *SIAM J. Matrix Anal. Appl.* **31** 2360–2375. MR2685162 https://doi.org/10.1137/09076578X
[40] KOLDA, T. G. and BADER, B. W. (2009). Tensor decompositions and applications. *SIAM Rev.* **51** 455–500. MR2535056 https://doi.org/10.1137/07070111X
[41] KRESSNER, D., STEINLECHNER, M. and VANDEREYCKEN, B. (2014). Low-rank tensor completion by Riemannian optimization. *BIT* **54** 447–468. MR3223510 https://doi.org/10.1007/s10543-013-0455-z
[42] LAFOND, J. (2015). Low rank matrix completion with exponential family noise. In *Conference on Learning Theory* 1224–1243. PMLR.
[43] LIU, J., MUSIALSKI, P., WONKA, P. and YE, J. (2012). Tensor completion for estimating missing values in visual data. *IEEE Trans. Pattern Anal. Mach. Intell.* **35** 208–220.
[44] LUO, Y. and ZHANG, A. R. (2024). Tensor-on-tensor regression: Riemannian optimization, over-parameterization, statistical-computational gap and their interplay. *Ann. Statist.* **52** 2583–2612. MR4842819 https://doi.org/10.1214/24-aos2396
[45] LYU, X., SUN, W. W., WANG, Z., LIU, H., YANG, J. and CHENG, G. (2019). Tensor graphical model: Non-convex optimization and statistical inference. *IEEE Trans. Pattern Anal. Mach. Intell.* **42** 2024–2037.
[46] LYU, Z., XIA, D. and ZHANG, Y. (2023). Latent space model for higher-order networks and generalized tensor decomposition. *J. Comput. Graph. Statist.* **32** 1320–1336. MR4669249 https://doi.org/10.1080/10618600.2022.2164289
[47] MA, C., WANG, K., CHI, Y. and CHEN, Y. (2020). Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion, and blind deconvolution. *Found. Comput. Math.* **20** 451–632. MR4099988 https://doi.org/10.1007/s10208-019-09429-9
[48] MA, W., DU, L., XIA, D. and YUAN, M. (2023). Multiple Testing of Linear Forms for Noisy Matrix Completion. arXiv preprint. Available at arXiv:2312.00305.
[49] MA, W. and XIA, D. (2026). Supplement to "Statistical inference in tensor completion: Optimal uncertainty quantification and statistical-to-computational gaps." https://doi.org/10.1214/25-AOS2617SUPP
[50] NANDPURU, H. B., SALANKAR, S. and BORA, V. (2014). MRI brain cancer classification using support vector machine. In 2014 *IEEE Students' Conference on Electrical, Electronics and Computer Science* 1–6. IEEE Press, New York.
[51] OSELEDETS, I. V. (2011). Tensor-train decomposition. *SIAM J. Sci. Comput.* **33** 2295–2317. MR2837533 https://doi.org/10.1137/090752286
[52] POTECHIN, A. and STEURER, D. (2017). Exact tensor completion with sum-of-squares. In *Conference on Learning Theory* 1619–1673. PMLR.
[53] QIU, Y., ZHOU, G., ZHAO, Q. and XIE, S. (2024). Noisy tensor completion via low-rank tensor ring. *IEEE Trans. Neural Netw. Learn. Syst.* **35** 1127–1141. MR4694100 https://doi.org/10.1109/tnnls.2022.3181378
[54] RAO, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0346957
[55] REN, Z. and BARBER, R. F. (2024). Derandomised knockoffs: Leveraging *e*-values for false discovery rate control. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **86** 122–154. MR4754084 https://doi.org/10.1093/jrsssb/qkad085
[56] RICHARD, E. and MONTANARI, A. (2014). A statistical model for tensor PCA. *Adv. Neural Inf. Process. Syst.* **27**.
[57] ROGERS, M., LI, L. and RUSSELL, S. J. (2013). Multilinear dynamical systems for tensor time series. *Adv. Neural Inf. Process. Syst.* **26**.
[58] SEMERCI, O., HAO, N., KILMER, M. E. and MILLER, E. L. (2014). Tensor-based formulation and nuclear norm regularization for multienergy computed tomography. *IEEE Trans. Image Process.* **23** 1678–1693. MR3191324 https://doi.org/10.1109/TIP.2014.2305840
[59] SHI, J., KATEHAKIS, M. N., MELAMED, B. and XIA, Y. (2014). Production-inventory systems with lost sales and compound Poisson demands. *Oper. Res.* **62** 1048–1063. MR3269418 https://doi.org/10.1287/opre.2014.1299
[60] SMITH, S. T. (1994). Optimization techniques on Riemannian manifolds. In *Hamiltonian and Gradient Flows, Algorithms and Control. Fields Inst. Commun.* **3** 113–136. Amer. Math. Soc., Providence, RI. MR1297990 https://doi.org/10.1090/fic/003/09

[61] SMITH, S. T. (2005). Covariance, subspace, and intrinsic Cramér-Rao bounds. *IEEE Trans. Signal Process.* **53** 1610–1630. MR2148599 https://doi.org/10.1109/TSP.2005.845428

[62] STEINLECHNER, M. (2016). Riemannian optimization for high-dimensional tensor completion. *SIAM J. Sci. Comput.* **38** S461–S484. MR3565572 https://doi.org/10.1137/15M1010506

[63] SUN, Y., FLAMMARION, N. and FAZEL, M. (2019). Escaping from saddle points on Riemannian manifolds. *Adv. Neural Inf. Process. Syst.* **32**.

[64] TANDEL, G. S., BALESTRIERI, A., JUJARAY, T., KHANNA, N. N., SABA, L. and SURI, J. S. (2020). Multiclass magnetic resonance imaging brain tumor classification using artificial intelligence paradigm. *Comput. Biol. Med.* **122** 103804.

[65] TESAŘ, L., SHIMIZU, A., SMUTEK, D., KOBATAKE, H. and NAWANO, S. (2008). Medical image analysis of 3D CT images based on extension of Haralick texture features. *Comput. Med. Imaging Graph.* **32** 513–520.

[66] TONG, T., MA, C., PRATER-BENNETTE, A., TRIPP, E. and CHI, Y. (2022). Scaling and scalability: Provable nonconvex low-rank tensor estimation from incomplete measurements. *J. Mach. Learn. Res.* **23** Paper No. [163], 77. MR4577115

[67] TU, L. W. (2017). *Differential Geometry*: *Connections*, *Curvature*, *and Characteristic Classes*. *Graduate Texts in Mathematics* **275**. Springer, Cham. MR3585539 https://doi.org/10.1007/978-3-319-55084-8

[68] TUCKER, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika* **31** 279–311. MR0205395 https://doi.org/10.1007/BF02289464

[69] USCHMAJEW, A. (2012). Local convergence of the alternating least squares algorithm for canonical tensor approximation. *SIAM J. Matrix Anal. Appl.* **33** 639–652. MR2970223 https://doi.org/10.1137/110843587

[70] USCHMAJEW, A. and VANDEREYCKEN, B. (2013). The geometry of algorithms using hierarchical tensors. *Linear Algebra Appl.* **439** 133–166. MR3045227 https://doi.org/10.1016/j.laa.2013.03.016

[71] VANDEREYCKEN, B. (2013). Low-rank matrix completion by Riemannian optimization. *SIAM J. Optim.* **23** 1214–1236. MR3069099 https://doi.org/10.1137/110845768

[72] 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

[73] WANG, H., CHEN, J. and WEI, K. (2021). Implicit Regularization and Entrywise Convergence of Riemannian Optimization for Low Tucker-Rank Tensor Completion. arXiv preprint. Available at arXiv:2108.07899.

[74] WANG, H., CHEN, J. and WEI, K. (2023). Implicit regularization and entrywise convergence of Riemannian optimization for low Tucker-rank tensor completion. *J. Mach. Learn. Res.* **24** Paper No. [347], 84. MR4690296

[75] WANG, M. and LI, L. (2020). Learning from binary multiway data: Probabilistic tensor decomposition and its statistical optimality. *J. Mach. Learn. Res.* **21** Paper No. 154, 38. MR4209440

[76] WEN, X., SUN, W. W. and ZHANG, Y. (2023). Online tensor inference. arXiv preprint. Available at arXiv:2312.17111.

[77] XIA, D. (2019). Confidence region of singular subspaces for low-rank matrix regression. *IEEE Trans. Inf. Theory* **65** 7437–7459. MR4030894 https://doi.org/10.1109/TIT.2019.2924900

[78] XIA, D. (2021). Normal approximation and confidence region of singular subspaces. *Electron. J. Stat.* **15** 3798–3851. MR4298986 https://doi.org/10.1214/21-ejs1876

[79] XIA, D. and YUAN, M. (2021). Statistical inferences of linear forms for noisy matrix completion. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 58–77. MR4220984 https://doi.org/10.1111/rssb.12400

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

[81] XIA, D., ZHANG, A. R. and ZHOU, Y. (2022). Inference for low-rank tensors—no need to debias. *Ann. Statist.* **50** 1220–1245. MR4404934 https://doi.org/10.1214/21-aos2146

[82] YEREDOR, A. and HAARDT, M. (2019). Maximum likelihood estimation of a low-rank probability mass tensor from partial observations. *IEEE Signal Process. Lett.* **26** 1551–1555.

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

[84] 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

[85] ZHANG, A. R., CAI, T. T. and WU, Y. (2022). Heteroskedastic PCA: Algorithm, optimality, and applications. *Ann. Statist.* **50** 53–80. MR4382008 https://doi.org/10.1214/21-aos2074

[86] ZHANG, C.-H. and ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 217–242. MR3153940 https://doi.org/10.1111/rssb.12026

[87] ZHANG, Y., BI, X., TANG, N. and QU, A. (2021). Dynamic tensor recommender systems. *J. Mach. Learn. Res.* **22** Paper No. 65, 35. MR4253758

The Annals of Statistics
2026, Vol. 54, No. 3, 1566–1593
https://doi.org/10.1214/26-AOS2618

A NOVEL STATISTICAL APPROACH TO ANALYZE IMAGE CLASSIFICATION

BY JUNTONG CHEN[1,a], SOPHIE LANGER[2,b] AND JOHANNES SCHMIDT-HIEBER[3,c]

[1]*School of Mathematical Sciences, Xiamen University,* [a]*juntong.chen@xmu.edu.cn*
[2]*Fakultät für Mathematik, Ruhr-Universität Bochum,* [b]*s.langer@rub.de*
[3]*Department of Applied Mathematics, University of Twente,* [c]*a.j.schmidt-hieber@utwente.nl*

The recent statistical theory of neural networks focuses on nonparametric denoising problems that treat randomness as additive noise. Variability in image classification datasets does, however, not originate from additive noise but from variation of the shape and other characteristics of the same object across different images. To address this problem, we introduce a tractable model for supervised image classification. While from the function estimation point of view, every pixel in an image is a variable, and large images lead to high-dimensional function recovery tasks suffering from the curse of dimensionality, increasing the number of pixels in the proposed image deformation model enhances the image resolution and makes the object classification problem easier. We introduce and theoretically analyze three approaches. Two methods combine image alignment with a one-nearest neighbor classifier. Under a separation condition, it is shown that perfect classification is possible. The third method fits a convolutional neural network (CNN) to the data. We derive a rate for the misclassification error that depends on the sample size and the complexity of the deformation class. An empirical study corroborates the theoretical findings.

REFERENCES

[1] ALLEN-ZHU, Z., LI, Y. and LIANG, Y. (2019). Learning and generalization in overparameterized neural networks, going beyond two layers. In *Neural Information Processing Systems*.
[2] ALLEN-ZHU, Z., LI, Y. and ZHAO, S. (2019). A convergence theory for deep learning via over-parameterization. In *International Conference on Machine Learning* 242–252.
[3] ANSELMI, F., ROSASCO, L. and POGGIO, T. (2016). On invariance and selectivity in representation learning. *Inf. Inference* **5** 134–158. MR3516856 https://doi.org/10.1093/imaiai/iaw009
[4] ARLOT, S. and BARTLETT, P. L. (2011). Margin-adaptive model selection in statistical learning. *Bernoulli* **17** 687–713. MR2787611 https://doi.org/10.3150/10-BEJ288
[5] ASHBURNER, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage* **38** 95–113.
[6] BARTLETT, P. L. (1998). The sample complexity of pattern classification with neural networks: The size of the weights is more important than the size of the network. *IEEE Trans. Inf. Theory* **44** 525–536. MR1607706 https://doi.org/10.1109/18.661502
[7] BARTLETT, P. L., FOSTER, D. J. and TELGARSKY, M. J. (2017). Spectrally-normalized margin bounds for neural networks. In *Neural Information Processing Systems*.
[8] BARTLETT, P. L., JORDAN, M. I. and MCAULIFFE, J. D. (2006). Convexity, classification, and risk bounds. *J. Amer. Statist. Assoc.* **101** 138–156. MR2268032 https://doi.org/10.1198/016214505000000907
[9] 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
[10] BOUCHERON, S., BOUSQUET, O. and LUGOSI, G. (2005). Theory of classification: A survey of some recent advances. *ESAIM Probab. Stat.* **9** 323–375. MR2182250 https://doi.org/10.1051/ps:2005018
[11] BRUNA, J. (2013). Scattering representations for recognition. PhD thesis, Ecole Polytechnique.
[12] BRUNA, J. and MALLAT, S. (2011). Classification with scattering operators. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition* 1561–1566.

MSC2020 subject classifications. Primary 62H30; secondary 62G05.
Key words and phrases. Image classification, convolutional neural networks, statistical learning, functional data, misclassification error.

[13] BRUNA, J. and MALLAT, S. (2013). Invariant scattering convolution networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **35** 1872–1886.
[14] CHEN, J., LANGER, S. and SCHMIDT-HIEBER, J. (2026). Supplement to "A novel statistical approach to analyze image classification." https://doi.org/10.1214/26-AOS2618SUPP
[15] CHEN, J., LIU, Y., WEI, S., BIAN, Z., SUBRAMANIAN, S., CARASS, A., PRINCE, J. L. and DU, Y. (2025). A survey on deep learning in medical image registration: New technologies, uncertainty, evaluation metrics, and beyond. *Med. Image Anal.* **100** 103385.
[16] COHEN, T. and WELLING, M. (2016). Group equivariant convolutional networks. In *International Conference on Machine Learning* 2990–2999.
[17] COVER, T. and HART, P. (1967). Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **13** 21–27.
[18] DELAIGLE, A. and HALL, P. (2012). Achieving near perfect classification for functional data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 267–286. MR2899863 https://doi.org/10.1111/j.1467-9868.2011.01003.x
[19] DRYDEN, I. L. and MARDIA, K. V. (2016). *Statistical Shape Analysis with Applications in R*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR3559734 https://doi.org/10.1002/9781119072492
[20] DU, S. S., LEE, J. D., LI, H., WANG, L. and ZHAI, X. (2019). Gradient descent finds global minima of deep neural networks. In *International Conference on Machine Learning* 1675–1685.
[21] DU, S. S., ZHAI, X., PÓCZOS, B. and SINGH, A. (2019). Gradient descent provably optimizes over-parameterized neural networks. In *International Conference on Learning Representations*.
[22] GLUCKMAN, J. (2006). Scale variant image pyramids. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition* 1069–1075.
[23] GONZALEZ, R. C. and WOODS, R. E. (2008). *Digital Image Processing*. Pearson, New York.
[24] GRENANDER, U. (1976–1981). *Lectures in Pattern Theory I, II and III: Pattern Analysis, Pattern Synthesis and Regular Structures*. Springer, New York.
[25] GRENANDER, U. (1993). *General Pattern Theory: A Mathematical Study of Regular Structures. Oxford Mathematical Monographs*. Clarendon, Oxford. MR1270904
[26] HARALICK, R. M. and SHAPIRO, L. G. (1985). Image segmentation techniques. *Comput. Graph. Image Process.* **29** 100–132.
[27] HASHEMI, M. (2019). Enlarging smaller images before inputting into convolutional neural network: Zero-padding vs. interpolation. *J. Big Data* **6** 98.
[28] JACQUES, J. and PREDA, C. (2014). Functional data clustering: A survey. *Adv. Data Anal. Classif.* **8** 231–255. MR3253859 https://doi.org/10.1007/s11634-013-0158-y
[29] KHATRI, N., DASGUPTA, A., SHEN, Y., ZHONG, X. and SHIH, F. Y. (2022). Perspective transformation layer. In *International Conference on Computational Science and Computational Intelligence* 1395–1401.
[30] KIM, Y., OHN, I. and KIM, D. (2021). Fast convergence rates of deep neural networks for classification. *Neural Netw.* **138** 179–197.
[31] KNEIP, A. and RAMSAY, J. O. (2008). Combining registration and fitting for functional models. *J. Amer. Statist. Assoc.* **103** 1155–1165. MR2528838 https://doi.org/10.1198/016214508000000517
[32] KOHLER, M., KRZYŻAK, A. and WALTER, B. (2022). On the rate of convergence of image classifiers based on convolutional neural networks. *Ann. Inst. Statist. Math.* **74** 1085–1108. MR4498394 https://doi.org/10.1007/s10463-022-00828-4
[33] KOHLER, M. and LANGER, S. (2025). Statistical theory for image classification using deep convolutional neural network with cross-entropy loss under the hierarchical max-pooling model. *J. Statist. Plann. Inference* **234** Paper No. 106188, 9. MR4760228 https://doi.org/10.1016/j.jspi.2024.106188
[34] KORMAN, S., REICHMAN, D., TSUR, G. and AVIDAN, S. (2013). FasT-match: Fast affine template matching. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition* 2331–2338.
[35] KRIZHEVSKY, A., SUTSKEVER, I. and HINTON, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Commun. ACM* **60** 84–90.
[36] LECUN, Y., BENGIO, Y. and HINTON, G. E. (2015). Deep learning. *Nature* **521** 436–444.
[37] LI, Y. and LIANG, Y. (2018). Learning overparameterized neural networks via stochastic gradient descent on structured data. In *Neural Information Processing Systems*.
[38] LIU, X. and MÜLLER, H.-G. (2004). Functional convex averaging and synchronization for time-warped random curves. *J. Amer. Statist. Assoc.* **99** 687–699. MR2090903 https://doi.org/10.1198/016214504000000999
[39] LONG, P. M. and SEDGHI, H. (2020). Generalization bounds for deep convolutional neural networks. In *International Conference on Learning Representations*.
[40] LU, Y., HERBEI, R. and KURTEK, S. (2017). Bayesian registration of functions with a Gaussian process prior. *J. Comput. Graph. Statist.* **26** 894–904. MR3765353 https://doi.org/10.1080/10618600.2017.1336444
[41] MALLAT, S. (2012). Group invariant scattering. *Comm. Pure Appl. Math.* **65** 1331–1398. MR2957703 https://doi.org/10.1002/cpa.21413

[42] MARCOS, D., VOLPI, M. and TUIA, D. (2016). Learning rotation invariant convolutional filters for texture classification. In *International Conference on Pattern Recognition* 2012–2017.
[43] MARRON, J. S., RAMSAY, J. O., SANGALLI, L. M. and SRIVASTAVA, A. (2015). Functional data analysis of amplitude and phase variation. *Statist. Sci.* **30** 468–484. MR3432837 https://doi.org/10.1214/15-STS524
[44] MASSART, P. and NÉDÉLEC, É. (2006). Risk bounds for statistical learning. *Ann. Statist.* **34** 2326–2366. MR2291502 https://doi.org/10.1214/009053606000000786
[45] MINAEE, S., BOYKOV, Y. Y., PORIKLI, F., PLAZA, A. J., KEHTARNAVAZ, N. and TERZOPOULOS, D. (2022). Image segmentation using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **44** 3523–3542.
[46] MUMFORD, D. (1996). Pattern theory: A unifying perspective. In *Perception as Bayesian Inference* 25–62. Cambridge Univ. Press, Cambridge.
[47] MUMFORD, D. (2000). The statistical description of visual signals. Unpublished manuscript.
[48] MUMFORD, D. and DESOLNEUX, A. (2010). *Pattern Theory*: *The Stochastic Analysis of Real-World Signals*. *Applying Mathematics*. AK Peters, Natick, MA. MR2723182
[49] NEYSHABUR, B., LI, Z., BHOJANAPALLI, S., LECUN, Y. and SREBRO, N. (2018). Towards understanding the role of over-parametrization in generalization of neural networks. In *International Conference on Learning Representations*.
[50] PANARETOS, V. M. and ZEMEL, Y. (2016). Amplitude and phase variation of point processes. *Ann. Statist.* **44** 771–812. MR3476617 https://doi.org/10.1214/15-AOS1387
[51] PARK, D., RAMANAN, D. and FOWLKES, C. C. (2010). Multiresolution models for object detection. In *European Conference on Computer Vision* 241–254.
[52] RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2168993
[53] RAWAT, W. and WANG, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural Comput.* **29** 2352–2449. MR3866781 https://doi.org/10.1162/neco_a_00990
[54] ROSSI, F. and VILLA, N. (2006). Support vector machine for functional data classification. *Neurocomputing* **69** 730–742.
[55] SCHMIDHUBER, J. (2015). Deep learning in neural networks: An overview. *Neural Netw.* **61** 85–117.
[56] SIMARD, P. Y., STEINKRAUS, D. and PLATT, J. C. (2003). Best practices for convolutional neural networks applied to visual document analysis. In *International Conference on Document Analysis and Recognition* 958–963.
[57] SOATTO, S. (2009). Actionable information in vision. In *IEEE International Conference on Computer Vision* 2138–2145.
[58] SOTIRAS, A., DAVATZIKOS, C. and PARAGIOS, N. (2013). Deformable medical image registration: A survey. *IEEE Trans. Med. Imag.* **32** 1153–1190.
[59] SRIVASTAVA, A., KLASSEN, E., JOSHI, S. H. and JERMYN, I. H. (2011). Shape analysis of elastic curves in Euclidean spaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **33** 1415–1428.
[60] SRIVASTAVA, A., WU, W., KURTEK, S., KLASSEN, E. and MARRON, J. S. (2011). Registration of functional data using Fisher–Rao metric. arXiv preprint. Available at arXiv:1103.3817.
[61] SZELISKI, R. (2010). *Computer Vision*: *Algorithms and Applications*. Springer, Berlin.
[62] TANG, R. and MÜLLER, H.-G. (2008). Pairwise curve synchronization for functional data. *Biometrika* **95** 875–889. MR2461217 https://doi.org/10.1093/biomet/asn047
[63] TARASIUK, P. and PRYCZEK, M. (2016). Geometric transformations embedded into convolutional neural networks. *J. Appl. Comput. Sci.* **24** 33–48.
[64] TUCKER, J. D., WU, W. and SRIVASTAVA, A. (2013). Generative models for functional data using phase and amplitude separation. *Comput. Statist. Data Anal.* **61** 50–66. MR3063000 https://doi.org/10.1016/j.csda.2012.12.001
[65] VAN NOORD, N. and POSTMA, E. (2017). Learning scale-variant and scale-invariant features for deep image classification. *Pattern Recognit.* **61** 583–592.
[66] VAPNIK, V. N. (1998). *Statistical Learning Theory*. *Adaptive and Learning Systems for Signal Processing*, *Communications*, *and Control*. Wiley, New York. A Wiley-Interscience Publication. MR1641250
[67] WANG, J.-L., CHIOU, J.-M. and MÜLLER, H.-G. (2016). Functional data analysis. *Annu. Rev. Stat. Appl.* **3** 257–295.
[68] YAMASHITA, R., NISHIO, M., DO, R. K. G. and TOGASHI, K. (2018). Convolutional neural networks: An overview and application in radiology. *Insights Imaging* **9** 611–629.
[69] YAN, X., YANG, J., YUMER, E., GUO, Y. and LEE, H. (2016). Perspective transformer nets: Learning single-view 3D object reconstruction without 3D supervision. In *Neural Information Processing Systems*.
[70] YANG, X., KWITT, R., STYNER, M. and NIETHAMMER, M. (2017). Quicksilver: Fast predictive image registration—a deep learning approach. *NeuroImage* **158** 378–396.

[71] YOUNES, L., MICHOR, P. W., SHAH, J. and MUMFORD, D. (2008). A metric on shape space with explicit geodesics. *Atti Accad. Naz. Lincei*, *Rend. Lincei*, *Mat. Appl.* **19** 25–57. MR2383560 https://doi.org/10.4171/RLM/506

[72] ZHANG, T. (2004). Statistical behavior and consistency of classification methods based on convex risk minimization. *Ann. Statist.* **32** 56–85. MR2051001 https://doi.org/10.1214/aos/1079120130

[73] ZOU, D., CAO, Y., ZHOU, D. and GU, Q. (2020). Gradient descent optimizes over-parameterized deep ReLU networks. *Mach. Learn.* **109** 467–492. MR4075425 https://doi.org/10.1007/s10994-019-05839-6

The Annals of Statistics
2026, Vol. 54, No. 3, 1594–1621
https://doi.org/10.1214/26-AOS2619

APPROXIMATE INDEPENDENCE OF PERMUTATION MIXTURES

BY YANJUN HAN[a] AND JONATHAN NILES-WEED[b]

Courant Institute of Mathematical Sciences and Center for Data Science, New York University, [a]*yanjunhan@nyu.edu,* [b]*jnw@cims.nyu.edu*

We prove bounds on statistical distances between high-dimensional exchangeable mixture distributions (which we call *permutation mixtures*) and their i.i.d. counterparts. Our results are based on a novel method for controlling χ^2 divergences between exchangeable mixtures, which is tighter than the existing methods of moments or cumulants. At a technical level, a key innovation in our proofs is a new Maclaurin-type inequality for elementary symmetric polynomials of variables that sum to zero and an upper bound on permanents of doubly-stochastic positive semidefinite matrices. We obtain as a corollary a new de Finetti-style theorem (in the language of Diaconis and Freedman, 1987) as well as several new statistical results, including a differential privacy guarantee for the "shuffled privacy model" with Gaussian noise and improved generic consistency guarantees for empirical Bayes procedures in compound decision problems.

REFERENCES

[1] ANARI, N., GURVITS, L., GHARAN, S. O. and SABERI, A. (2017). Simply exponential approximation of the permanent of positive semidefinite matrices. In 58*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 914–925. IEEE Comput. Soc., Los Alamitos, CA. MR3734292

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

[3] BANACH, S. (1938). Über homogene Polynome in (L^2). *Studia Math.* **7** 36–44.

[4] 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

[5] BANKS, J., MOORE, C., VERSHYNIN, R., VERZELEN, N. and XU, J. (2018). Information-theoretic bounds and phase transitions in clustering, sparse PCA, and submatrix localization. *IEEE Trans. Inf. Theory* **64** 4872–4994. MR3819345 https://doi.org/10.1109/tit.2018.2810020

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

[7] BARBER, R. F., CANDÈS, E. J., RAMDAS, A. and TIBSHIRANI, R. J. (2023). Conformal prediction beyond exchangeability. *Ann. Statist.* **51** 816–845. MR4601003 https://doi.org/10.1214/23-aos2276

[8] BARBER, R. F., CANDÈS, E. J., RAMDAS, A. and TIBSHIRANI, R. J. (2024). De Finetti's theorem and related results for infinite weighted exchangeable sequences. *Bernoulli* **30** 3004–3028. MR4779857 https://doi.org/10.3150/23-bej1704

[9] BOBKOV, S. G. (2005). Generalized symmetric polynomials and an approximate de Finetti representation. *J. Theoret. Probab.* **18** 399–412. MR2137450 https://doi.org/10.1007/s10959-005-3509-6

[10] BROWN, L. D. and GREENSHTEIN, E. (2009). Nonparametric empirical Bayes and compound decision approaches to estimation of a high-dimensional vector of normal means. *Ann. Statist.* **37** 1685–1704. MR2533468 https://doi.org/10.1214/08-AOS630

[11] CAI, T. T. and LOW, M. G. (2011). Testing composite hypotheses, Hermite polynomials and optimal estimation of a nonsmooth functional. *Ann. Statist.* **39** 1012–1041. MR2816346 https://doi.org/10.1214/10-AOS849

[12] CARLEN, E., LIEB, E. H. and LOSS, M. (2006). An inequality of Hadamard type for permanents. *Methods Appl. Anal.* **13** 1–18. MR2275869 https://doi.org/10.4310/MAA.2006.v13.n1.a1

MSC2020 subject classifications. Primary 62B10; secondary 62C12.
Key words and phrases. Permutation mixture, de Finetti's theorem, empirical Bayes, method of moments.

[13] COLLIER, O., COMMINGES, L. and TSYBAKOV, A. B. (2017). Minimax estimation of linear and quadratic functionals on sparsity classes. *Ann. Statist.* **45** 923–958. MR3662444 https://doi.org/10.1214/15-AOS1432

[14] COPAS, J. B. (1969). Compound decisions and empirical Bayes. (With discussion). *J. Roy. Statist. Soc. Ser. B, Methodol.* **31** 397–425. MR0269013

[15] DE FINETTI, B. (1929). Funzione caratteristica di un fenomeno aleatorio. In *Atti del Congresso Internazionale dei Matematici*: *Bologna del* 3 *al* 10 *de settembre di* 1928 179–190.

[16] DIACONIS, P. (1977). Finite forms of de Finetti's theorem on exchangeability. *Synthese* **36** 271–281. Foundations of probability and statistics, II. MR0517222 https://doi.org/10.1007/BF00486116

[17] DIACONIS, P. and FREEDMAN, D. (1980). Finite exchangeable sequences. *Ann. Probab.* **8** 745–764. MR0577313

[18] DIACONIS, P. and FREEDMAN, D. (1987). A dozen de Finetti-style results in search of a theory. *Ann. Inst. Henri Poincaré Probab. Stat.* **23** 397–423. MR0898502

[19] DING, Y. (2022). Statistical Inference and Estimation in High Dimensions. PhD thesis, Copyright—Database copyright ProQuest LLC;. ProQuest does not claim copyright in the individual underlying works. Last updated—2023-04-25.

[20] DONOHO, D. L. and JOHNSTONE, I. M. (1994). Minimax risk over l_p-balls for l_q-error. *Probab. Theory Related Fields* **99** 277–303. MR1278886 https://doi.org/10.1007/BF01199026

[21] DORON, D., HATAMI, P. and HOZA, W. M. (2020). Log-seed pseudorandom generators via iterated restrictions. In 35*th Computational Complexity Conference*. *LIPIcs*. *Leibniz Int. Proc. Inform.* **169** Art. No. 6, 36. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR4129268

[22] 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

[23] EFRON, B. (2019). Bayes, oracle Bayes and empirical Bayes. *Statist. Sci.* **34** 177–201. MR3983318 https://doi.org/10.1214/18-STS674

[24] EGORYCHEV, G. P. (1981). The solution of van der Waerden's problem for permanents. *Adv. Math.* **42** 299–305. MR0642395 https://doi.org/10.1016/0001-8708(81)90044-X

[25] 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

[26] FALIKMAN, D. I. (1981). Proof of the van der Waerden conjecture on the permanent of a doubly stochastic matrix. *Mat. Zametki* **29** 931–938, 957. MR0625097

[27] FELDMAN, V., MCMILLAN, A. and TALWAR, K. (2022). Hiding among the clones: A simple and nearly optimal analysis of privacy amplification by shuffling. In 2021 *IEEE* 62*nd Annual Symposium on Foundations of Computer Science—FOCS* 2021 954–964. IEEE Comput. Soc., Los Alamitos, CA. MR4399749 https://doi.org/10.1109/FOCS52979.2021.00096

[28] FLAJOLET, P. and SEDGEWICK, R. (2009). *Analytic Combinatorics*. Cambridge Univ. Press, Cambridge.

[29] GAVALAKIS, L. and KONTOYIANNIS, I. (2021). An information-theoretic proof of a finite de Finetti theorem. *Electron. Commun. Probab.* **26** Paper No. 68, 5. MR4346872 https://doi.org/10.1214/21-ecp428

[30] GIRGIS, A. M., DATA, D., DIGGAVI, S., SURESH, A. T. and KAIROUZ, P. (2021). On the Renyi differential privacy of the shuffle model. In *Proceedings of the* 2021 *ACM SIGSAC Conference on Computer and Communications Security* 2321–2341.

[31] GOPALAN, P. and YEHUDAYOFF, A. (2014). Inequalities and tail bounds for elementary symmetric polynomials. In *Electronic Colloquium on Computational Complexity* (*ECCC*) **21** 1.

[32] GREENSHTEIN, E. and RITOV, Y. (2009). Asymptotic efficiency of simple decisions for the compound decision problem. In *Optimality*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **57** 266–275. IMS, Beachwood, OH. MR2681676 https://doi.org/10.1214/09-LNMS5716

[33] GREENSHTEIN, E. and RITOV, Y. (2019). Comment: Empirical Bayes, compound decisions and exchangeability [MR3983318]. *Statist. Sci.* **34** 224–228. MR3983324 https://doi.org/10.1214/19-STS709

[34] GURVITS, L. (2003). Classical deterministic complexity of Edmond's problem and quantum entanglement. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing* 10–19. ACM, New York. MR2121068 https://doi.org/10.1145/780542.780545

[35] HAN, Y. and NILES-WEED, J. (2026). Supplement to "Approximate independence of permutation mixtures." https://doi.org/10.1214/26-AOS2619SUPP

[36] HAN, Y., NILES-WEED, J., SHEN, Y. and WU, Y. (2025). Besting Good-Turing: Optimality of nonparametric maximum likelihood for distribution estimation. arXiv preprint. Available at arXiv:2509.07355.

[37] HANNAN, J. F. (1953). Asymptotic solutions of compound decision problems. Technical Report, North Carolina State Univ. Dept. Statistics.
[38] HANNAN, J. F. and ROBBINS, H. (1955). Asymptotic solutions of the compound decision problem for two completely specified distributions. *Ann. Math. Statist.* **26** 37–51. MR0067444 https://doi.org/10.1214/aoms/1177728591
[39] HARDT, M. and PRICE, E. (2015). Tight bounds for learning a mixture of two Gaussians [extended abstract]. In *STOC'*15*—Proceedings of the* 2015 *ACM Symposium on Theory of Computing* 753–760. ACM, New York. MR3388255
[40] HEWITT, E. and SAVAGE, L. J. (1955). Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* **80** 470–501. MR0076206 https://doi.org/10.2307/1992999
[41] HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* **19** 293–325. MR0026294 https://doi.org/10.1214/aoms/1177730196
[42] INGSTER, Y. I. and SUSLINA, I. A. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models. Lecture Notes in Statistics* **169**. Springer, New York. MR1991446 https://doi.org/10.1007/978-0-387-21580-8
[43] ISSERLIS, L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika* **12** 134–139.
[44] JAIN, V., RISTESKI, A. and KOEHLER, F. (2019). Mean-field approximation, convex hierarchies, and the optimality of correlation rounding: A unified perspective. In *STOC'*19*—Proceedings of the* 51*st Annual ACM SIGACT Symposium on Theory of Computing* 1226–1236. ACM, New York. MR4003424 https://doi.org/10.1145/3313276.3316299
[45] JIANG, W. and ZHANG, C.-H. (2009). General maximum likelihood empirical Bayes estimation of normal means. *Ann. Statist.* **37** 1647–1684. MR2533467 https://doi.org/10.1214/08-AOS638
[46] JIAO, J., HAN, Y. and WEISSMAN, T. (2018). Minimax estimation of the L_1 distance. *IEEE Trans. Inf. Theory* **64** 6672–6706. MR3860754 https://doi.org/10.1109/TIT.2018.2846245
[47] JOHNSON, O., GAVALAKIS, L. and KONTOYIANNIS, I. (2025). Relative entropy bounds for sampling with and without replacement. *J. Appl. Probab.* **62** 1578–1593. MR4990808 https://doi.org/10.1017/jpr.2025.27
[48] KUNISKY, D. (2025). Low coordinate degree algorithms I: Universality of computational thresholds for hypothesis testing. *Ann. Statist.* **53** 774–801. MR4900166 https://doi.org/10.1214/24-aos2484
[49] 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
[50] LACKER, D., MUKHERJEE, S. and YEUNG, L. C. (2024). Mean field approximations via log-concavity. *Int. Math. Res. Not. IMRN* **7** 6008–6042. MR4728726 https://doi.org/10.1093/imrn/rnad302
[51] LE CAM, L. (1960). *Locally Asymptotically Normal Families of Distributions. Certain Approximations to Families of Distributions and Their Use in the Theory of Estimation and Testing Hypotheses. Univ. California Publ. Statist.* **3** 37–98. MR0126903
[52] LEDOUX, M. (2001). *The Concentration of Measure Phenomenon. Mathematical Surveys and Monographs* **89**. Amer. Math. Soc., Providence, RI. MR1849347 https://doi.org/10.1090/surv/089
[53] LEPSKI, O., NEMIROVSKI, A. and SPOKOINY, V. (1999). On estimation of the L_r norm of a regression function. *Probab. Theory Related Fields* **113** 221–253. MR1670867 https://doi.org/10.1007/s004409970006
[54] MACLAURIN, C. (1730). IV. A second letter from mr. Colin McLaurin, professor of mathematicks in the University of Edinburgh and FRS to Martin Folkes, Esq; concerning the roots of equations, with the demonstration of other rules in algebra; being the continuation of the letter published in the philosophical transactions. *Philos. Trans. R. Soc. Lond.* **36** 59–96.
[55] MAKUR, A. (2020). Coding theorems for noisy permutation channels. *IEEE Trans. Inf. Theory* **66** 6723–6748. MR4173604 https://doi.org/10.1109/TIT.2020.3009468
[56] MALAMUD, S. M. (2005). Inverse spectral problem for normal matrices and the Gauss–Lucas theorem. *Trans. Amer. Math. Soc.* **357** 4043–4064. MR2159699 https://doi.org/10.1090/S0002-9947-04-03649-9
[57] MARCUS, M. and MINC, H. (1965). Generalized matrix functions. *Trans. Amer. Math. Soc.* **116** 316–329. MR0194445 https://doi.org/10.2307/1994120
[58] MEKA, R., REINGOLD, O. and TAL, A. (2019). Pseudorandom generators for width-3 branching programs. In *STOC'*19*—Proceedings of the* 51*st Annual ACM SIGACT Symposium on Theory of Computing* 626–637. ACM, New York. MR4003370 https://doi.org/10.1145/3313276.3316319
[59] MINC, H. (1984). *Permanents* **6**. Cambridge Univ. Press, Cambridge.

[60] PARISI, G. (1988). *Statistical Field Theory*. *Frontiers in Physics* **66**. Benjamin-Cummings, Reading, MA. With a foreword by David Pines. MR1007980

[61] PEREIRA, R. (2003). Differentiators and the geometry of polynomials. *J. Math. Anal. Appl.* **285** 336–348. MR2000158 https://doi.org/10.1016/S0022-247X(03)00465-7

[62] PERRY, A., WEIN, A. S., BANDEIRA, A. S. and MOITRA, A. (2018). Optimality and sub-optimality of PCA I: Spiked random matrix models. *Ann. Statist.* **46** 2416–2451. MR3845022 https://doi.org/10.1214/17-AOS1625

[63] POLYANSKIY, Y. and WU, Y. (2025). *Information Theory*: *From Coding to Learning*. Cambridge Univ. Press, Cambridge.

[64] ROBBINS, H. (1951). Asymptotically subminimax solutions of compound statistical decision problems. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 1950 131–148. Univ. California Press, Berkeley, CA. MR0044803

[65] ROOS, B. (2015). On Bobkov's approximate de Finetti representation via approximation of permanents of complex rectangular matrices. *Proc. Amer. Math. Soc.* **143** 1785–1796. MR3314090 https://doi.org/10.1090/S0002-9939-2014-12429-4

[66] RUSH, C., SKERMAN, F., WEIN, A. S. and YANG, D. (2023). Is it easier to count communities than find them? In 14*th Innovations in Theoretical Computer Science Conference*. *LIPIcs. Leibniz Int. Proc. Inform.* **251** Art. No. 94, 23. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR4587108 https://doi.org/10.4230/lipics.itcs.2023.94

[67] 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

[68] STAM, A. J. (1978). Distance between sampling with and without replacement. *Stat. Neerl.* **32** 81–91. MR0518630 https://doi.org/10.1111/j.1467-9574.1978.tb01387.x

[69] TANG, J. and POLYANSKIY, Y. (2023). Capacity of noisy permutation channels. *IEEE Trans. Inf. Theory* **69** 4145–4162. MR4613538 https://doi.org/10.1109/tit.2023.3247812

[70] TAO, T. (2023). A Maclaurin type inequality. arXiv preprint. Available at arXiv:2310.05328.

[71] TIBSHIRANI, R. J., BARBER, R. F., CANDES, E. and RAMDAS, A. (2019). Conformal prediction under covariate shift. *Adv. Neural Inf. Process. Syst.* **32**.

[72] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. Revised and extended from the 2004 French original. Translated by Vladimir Zaiats. MR2724359 https://doi.org/10.1007/b13794

[73] VAN DER WAERDEN, B. L. (1926). Aufgabe 45. *Jahresber. Dtsch. Math.-Ver.* **35** 23.

[74] WEINSTEIN, A. (2021). On the minimum attainable risk in permutation invariant problems. arXiv preprint. Available at arXiv:2110.06250v2.

[75] WU, Y. and YANG, P. (2016). Minimax rates of entropy estimation on large alphabets via best polynomial approximation. *IEEE Trans. Inf. Theory* **62** 3702–3720. MR3506758 https://doi.org/10.1109/TIT.2016.2548468

[76] WU, Y. and YANG, P. (2020). Optimal estimation of Gaussian mixtures via denoised method of moments. *Ann. Statist.* **48** 1981–2007. MR4134783 https://doi.org/10.1214/19-AOS1873

[77] ZHANG, C.-H. (2003). Compound decision theory and empirical Bayes methods. *Ann. Statist.* 379–390.

[78] ZHANG, C.-H. (2012). Minimax ℓ_q risk in ℓ_p balls. In *Contemporary Developments in Bayesian Analysis and Statistical Decision Theory*: *A Festschrift for William E. Strawderman*. *Inst. Math. Stat.* (*IMS*) *Collect.* **8** 78–89. IMS, Beachwood, OH. MR3202504 https://doi.org/10.1214/11-IMSCOLL806

The Annals of Statistics
2026, Vol. 54, No. 3, 1622–1647
https://doi.org/10.1214/26-AOS2631
This research was funded, in whole or in part, by Austrian Science Fund (FWF), 10.55776 / P34808.

QUASI-MONTE CARLO CONFIDENCE INTERVALS USING QUANTILES OF RANDOMIZED NETS

BY ZEXIN PAN[a]

Institute of Fundamental and Transdisciplinary Research, Zhejiang University, [a]*zep002@zju.edu.cn*

Recent advances in quasi-Monte Carlo integration have shown that for linearly scrambled digital net estimators, the convergence rate can be dramatically improved by taking the median rather than the mean of multiple independent replicates. In this work, we demonstrate that the quantiles of such estimators can be used to construct confidence intervals with asymptotically valid coverage for high-dimensional integrals. By analyzing the error distribution for a class of infinitely differentiable integrands, we prove that as the sample size increases, the integration error decomposes into an asymptotically symmetric component and a vanishing remainder. Consequently, the asymptotic error distribution is symmetric about zero, ensuring that a quantile-based interval constructed from independent replicates captures the true integral with probability converging to a nominal level determined by the binomial distribution.

REFERENCES

[1] BASU, K. and MUKHERJEE, R. (2017). Asymptotic normality of scrambled geometric net quadrature. *Ann. Statist.* **45** 1759–1788. MR3670195 https://doi.org/10.1214/16-AOS1508

[2] CLANCY, N., DING, Y., HAMILTON, C., HICKERNELL, F. J. and ZHANG, Y. (2014). The cost of deterministic, adaptive, automatic algorithms: Cones, not balls. *J. Complexity* **30** 21–45. MR3132868 https://doi.org/10.1016/j.jco.2013.09.002

[3] DICK, J. (2011). Higher order scrambled digital nets achieve the optimal rate of the root mean square error for smooth integrands. *Ann. Statist.* **39** 1372–1398. MR2850206 https://doi.org/10.1214/11-AOS880

[4] DICK, J., KUO, F. Y. and SLOAN, I. H. (2013). High-dimensional integration: The quasi-Monte Carlo way. *Acta Numer.* **22** 133–288. MR3038697 https://doi.org/10.1017/S0962492913000044

[5] DICK, J. and PILLICHSHAMMER, F. (2010). *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*. Cambridge Univ. Press, Cambridge. MR2683394 https://doi.org/10.1017/CBO9780511761188

[6] EFRON, B. (1981). Nonparametric standard errors and confidence intervals. *Canad. J. Statist.* **9** 139–172. With discussion and a reply by the author. MR0640014 https://doi.org/10.2307/3314608

[7] ERDÖS, P. (1945). On a lemma of Littlewood and Offord. *Bull. Amer. Math. Soc.* **51** 898–902. MR0014608 https://doi.org/10.1090/S0002-9904-1945-08454-7

[8] GNEWUCH, M., KRITZER, P., OWEN, A. B. and PAN, Z. (2024). Computable error bounds for quasi-Monte Carlo using points with non-negative local discrepancy. *Inf. Inference* **13** Paper No. iaae021, 23. MR4791441 https://doi.org/10.1093/imaiai/iaae021

[9] GOBET, E., LERASLE, M. and MÉTIVIER, D. (2026). Accelerated convergence of error quantiles using robust randomized quasi Monte Carlo methods. *J. Complexity* **92** Paper No. 101989, 31. MR4953486 https://doi.org/10.1016/j.jco.2025.101989

[10] GRIEWANK, A., KUO, F. Y., LEÖVEY, H. and SLOAN, I. H. (2018). High dimensional integration of kinks and jumps—smoothing by preintegration. *J. Comput. Appl. Math.* **344** 259–274. MR3825515 https://doi.org/10.1016/j.cam.2018.04.009

[11] JAIN, A., HICKERNELL, F. J., OWEN, A. B. and SOROKIN, A. G. (2026). Empirical Bernstein and betting confidence intervals for randomized quasi-Monte Carlo. *Inf. Inference* **15** Paper No. iaag003, 22. MR5040202 https://doi.org/10.1093/imaiai/iaag003

[12] JOE, S. and KUO, F. Y. (2008). Constructing Sobol' sequences with better two-dimensional projections. *SIAM J. Sci. Comput.* **30** 2635–2654. MR2429482 https://doi.org/10.1137/070709359

MSC2020 subject classifications. Primary 65D30, 65C05; secondary 62E20.
Key words and phrases. Numerical integration, randomized quasi-Monte Carlo, uncertainty quantification.

[13] KRANTZ, S. G. and PARKS, H. R. (2002). *A Primer of Real Analytic Functions*, 2nd ed. *Birkhäuser Advanced Texts*: *Basler Lehrbücher*. [*Birkhäuser Advanced Texts*: *Basel Textbooks*]. Birkhäuser, Boston, MA. MR1916029 https://doi.org/10.1007/978-0-8176-8134-0

[14] L'ECUYER, P., NAKAYAMA, M. K., OWEN, A. B. and TUFFIN, B. (2023). Confidence intervals for randomized quasi-Monte Carlo estimators. In *Proceedings of the* 2023 *Winter Simulation Conference* 445–456.

[15] LIU, S. and OWEN, A. B. (2023). Preintegration via active subspace. *SIAM J. Numer. Anal.* **61** 495–514. MR4558754 https://doi.org/10.1137/22M1479129

[16] LOH, W.-L. (2003). On the asymptotic distribution of scrambled net quadrature. *Ann. Statist.* **31** 1282–1324. MR2001651 https://doi.org/10.1214/aos/1059655914

[17] MATOUŠEK, J. (1998). On the L_2-discrepancy for anchored boxes. *J. Complexity* **14** 527–556. MR1659004 https://doi.org/10.1006/jcom.1998.0489

[18] NAKAYAMA, M. K. and TUFFIN, B. (2024). Sufficient conditions for central limit theorems and confidence intervals for randomized quasi–Monte Carlo methods. *ACM Trans. Model. Comput. Simul.* **34** Art. 13, 38. MR4773452 https://doi.org/10.1145/3643847

[19] NIEDERREITER, H. (1988). Low-discrepancy and low-dispersion sequences. *J. Number Theory* **30** 51–70. MR0960233 https://doi.org/10.1016/0022-314X(88)90025-X

[20] OWEN, A. B. (1995). Randomly permuted (t, m, s)-nets and (t, s)-sequences. In *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing* (*Las Vegas*, *NV*, 1994) (H. Niederreiter and P. J. S. Shiue, eds.). *Lect. Notes Stat.* **106** 299–317. Springer, New York. MR1445791 https://doi.org/10.1007/978-1-4612-2552-2_19

[21] OWEN, A. B. (2025). Better bootstrap t confidence intervals for the mean. Preprint.

[22] OWEN, A. B. (2026). Error Estimation for Quasi-Monte Carlo. In *Monte Carlo and Quasi-Monte Carlo* 2024. *Springer Proc. Math. Stat.* **522** 95–125. Springer, Cham. MR5077154 https://doi.org/10.1007/978-3-032-10590-5_4

[23] PAN, Z. (2024). Super-polynomial accuracy of randomized nets using the median-of-means. PhD thesis, Stanford Univ.

[24] PAN, Z. (2026). Supplement to "Quasi-Monte Carlo confidence intervals using quantiles of randomized nets." https://doi.org/10.1214/26-AOS2631SUPP

[25] PAN, Z. (2026). Automatic optimal-rate convergence of randomized nets using median-of-means. *Math. Comp.* **95** 1415–1446. MR5033395 https://doi.org/10.1090/mcom/4093

[26] PAN, Z. and OWEN, A. B. (2023). Super-polynomial accuracy of one dimensional randomized nets using the median of means. *Math. Comp.* **92** 805–837. MR4524109 https://doi.org/10.1090/mcom/3791

[27] PAN, Z. and OWEN, A. B. (2024). Super-polynomial accuracy of multidimensional randomized nets using the median-of-means. *Math. Comp.* **93** 2265–2289. MR4759375 https://doi.org/10.1090/mcom/3880

[28] PAN, Z. and OWEN, A. B. (2025). Skewness of a randomized quasi-Monte Carlo estimate. *J. Complexity* **90** Paper No. 101956, 21. MR4902061 https://doi.org/10.1016/j.jco.2025.101956

[29] SOBOL', I. M. (1967). Distribution of points in a cube and approximate evaluation of integrals. *Zh. Vychisl. Mat. Mat. Fiz.* **7** 784–802. MR0219238

[30] SURJANOVIC, S. and BINGHAM, D. (2013). Virtual library of simulation experiments: Test functions and datasets. Available at https://www.sfu.ca/~ssurjano/.

[31] SUZUKI, K. and YOSHIKI, T. (2016). Formulas for the Walsh coefficients of smooth functions and their application to bounds on the Walsh coefficients. *J. Approx. Theory* **205** 1–24. MR3471153 https://doi.org/10.1016/j.jat.2015.12.002

[32] WIART, J., LEMIEUX, C. and DONG, G. Y. (2021). On the dependence structure and quality of scrambled (t, m, s)-nets. *Monte Carlo Methods Appl.* **27** 1–26. MR4223857 https://doi.org/10.1515/mcma-2020-2079

The Annals of Statistics
2026, Vol. 54, No. 3, 1648
https://doi.org/10.1214/25-AOS2499
Main article: https://doi.org/10.1214/24-AOS2432

ERRATUM: QUANTILE PROCESSES AND THEIR APPLICATIONS IN FINITE POPULATIONS

BY ANURAG DEY[a] AND PROBAL CHAUDHURI[b]

Statistics and Mathematics unit, Indian Statistical Institute, [a]*deyanuragsaltlake64@gmail.com,* [b]*probalchaudhuri@gmail.com*

This erratum contains some corrections to Dey and Chaudhuri (*Ann. Statist.* **52** (2024) 2194–2216).

REFERENCES

[1] DEY, A. and CHAUDHURI, P. (2024). Quantile processes and their applications in finite populations. *Ann. Statist.* **52** 2194–2216. MR4829485 https://doi.org/10.1214/24-aos2432

MSC2020 subject classifications. Primary 62D05, 60F05; secondary 60B05, 60B10.

The Annals of Statistics
2026, Vol. 54, No. 3, 1649–1653
https://doi.org/10.1214/25-AOS2611
Main article: https://doi.org/10.1214/aos/1176347258

ERRATUM: EDGEWORTH EXPANSIONS FOR LINEAR RANK STATISTICS

BY WALTER SCHNELLER[a]

Fakultät Angewandte Natur- und Geisteswissenschaften, Technische Hochschule Würzburg-Schweinfurt,
[a]*walter.schneller@thws.de*

This erratum contains a correction of the incorrect defined distribution on page 1109, last two lines, in Schneller (*Ann. Statist.* **17** (1989) 1103–1123). Additionally, some typographical errors are corrected.

REFERENCES

BARBOUR, A. D. (1986). Asymptotic expansions based on smooth functions in the central limit theorem. *Probab. Theory Related Fields* **72** 289–303. MR0836279 https://doi.org/10.1007/BF00699108

BOLTHAUSEN, E. (1984). An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete* **66** 379–386. MR0751577 https://doi.org/10.1007/BF00533704

SCHNELLER, W. (1987). Edgeworth–Entwicklungen für lineare Rangstatistiken. Ph.D. thesis, Technische Universität Berlin.

SCHNELLER, W. (1989). Edgeworth expansions for linear rank statistics. *Ann. Statist.* **17** 1103–1123. MR1015140 https://doi.org/10.1214/aos/1176347258

SCHNELLER, W. (2025). Edgeworth Expansions for Linear Rank Statistics Using Stein's Method. [English translation of an updated version of the author's Ph.D. thesis from 1987]. Available at arXiv:2511.12187.

SCHNELLER, W. (2026a). Supplement A to "Erratum: Edgeworth expansions for linear rank statistics." https://doi.org/10.1214/25-AOS2611SUPPA

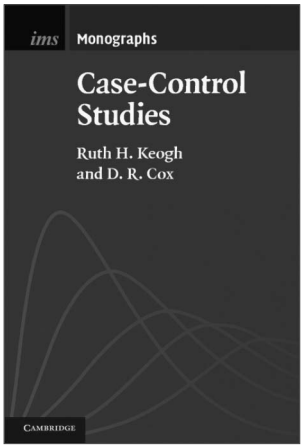
SCHNELLER, W. (2026b). Supplement B to "Erratum: Edgeworth Expansions for Linear Rank Statistics." https://doi.org/10.1214/25-AOS2611SUPPB

MSC2020 subject classifications. 60F05, 62E20.

Key words and phrases. Edgeworth expansion, Stein's method, linear rank statistic, combinatorial method of Bolthausen, conditions of van Zwet.

The Institute of Mathematical Statistics presents

IMS MONOGRAPHS



Case-Control Studies

Ruth H. Keogh
and D. R. Cox

The case-control approach is a powerful method for investigating factors that may explain a particular event. It is extensively used in epidemiology to study disease incidence, one of the best-known examples being Bradford Hill and Doll's investigation of the possible connection between cigarette smoking and lung cancer. More recently, case-control studies have been increasingly used in other fields, including sociology and econometrics.

With a particular focus on statistical analysis, this book is ideal for applied and theoretical statisticians wanting an up-to-date introduction to the field. It covers the fundamentals of case-control study design and analysis as well as more recent developments, including two-stage studies, case-only studies and methods for case-control sampling in time. The latter have important applications in large prospective cohorts which require case-control sampling designs to make efficient use of resources. More theoretical background is provided in an appendix for those new to the field.

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

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

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