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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

Vol. 52, No. 6—December 2024

THE ANNALS OF STATISTICS Vol. 52, No. 6, pp. 2483–2978 December 2024

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

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

IMS OFFICERS

IMS EDITORS

The Annals of Statistics **[ISSN 0090-5364 (print); ISSN 2168-8966 (online)],** Volume 52, Number 6, December 2024. Published bimonthly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, OH 44094, USA. Periodicals postage paid at Cleveland, Ohio, and at additional mailing offices.

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

The Annals of Statistics
2024, Vol. 52, No. 6, 2483–2505
https://doi.org/10.1214/23-AOS2261

TESTING NETWORK CORRELATION EFFICIENTLY VIA COUNTING TREES

BY CHENG MAO[1,a], YIHONG WU[2,b], JIAMING XU[3,c] AND SOPHIE H. YU[4,d]

[1]*The School of Mathematics, Georgia Institute of Technology,* [a]*cheng.mao@math.gatech.edu*
[2]*Department of Statistics and Data Science, Yale University,* [b]*yihong.wu@yale.edu*
[3]*The Fuqua School of Business, Duke University,* [c]*jx77@duke.edu*
[4]*Operations Information and Decisions Department, The Wharton School, University of Pennsylvania,* [d]*hysophie@wharton.upenn.edu*

We propose a new procedure for testing whether two networks are edge-correlated through some latent vertex correspondence. The test statistic is based on counting the cooccurrences of signed trees for a family of nonisomorphic trees. When the two networks are Erdős–Rényi random graphs $\mathcal{G}(n,q)$ that are either independent or correlated with correlation coefficient ρ, our test runs in $n^{2+o(1)}$ time and succeeds with high probability as $n\to\infty$, provided that $n\min\{q,1-q\}\geq n^{-o(1)}$ and $\rho^2>\alpha\approx 0.338$, where α is Otter's constant so that the number of unlabeled trees with K edges grows as $(1/\alpha)^K$. This significantly improves the prior work in terms of statistical accuracy, running time and graph sparsity.

REFERENCES

[1] ALON, N., DAO, P., HAJIRASOULIHA, I., HORMOZDIARI, F. and SAHINALP, S. C. (2008). Biomolecular network motif counting and discovery by color coding. *Bioinformatics* **24** i241–i249. https://doi.org/10.1093/bioinformatics/btn163

[2] ALON, N., YUSTER, R. and ZWICK, U. (1995). Color-coding. *J. Assoc. Comput. Mach.* **42** 844–856. MR1411787 https://doi.org/10.1145/210332.210337

[3] ARVIND, V. and RAMAN, V. (2002). Approximation algorithms for some parameterized counting problems. In *Algorithms and Computation. Lecture Notes in Computer Science* **2518** 453–464. Springer, Berlin. MR2079050 https://doi.org/10.1007/3-540-36136-7_40

[4] BANERJEE, D. (2018). Contiguity and non-reconstruction results for planted partition models: The dense case. *Electron. J. Probab.* **23** 1–28. MR3771755 https://doi.org/10.1214/17-EJP128

[5] BANERJEE, D. and MA, Z. (2017). Optimal hypothesis testing for stochastic block models with growing degrees. arXiv preprint. Available at arXiv:1705.05305.

[6] BARAK, B., CHOU, C.-N., LEI, Z., SCHRAMM, T. and SHENG, Y. (2019). (Nearly) Efficient algorithms for the graph matching problem on correlated random graphs. In *Advances in Neural Information Processing Systems* 9186–9194.

[7] BAYATI, M., GLEICH, D. F., SABERI, A. and WANG, Y. (2013). Message-passing algorithms for sparse network alignment. *ACM Trans. Knowl. Discov. Data* **7** 1–31.

[8] BERG, A. C., BERG, T. L. and MALIK, J. (2005). Shape matching and object recognition using low distortion correspondences. In 2005 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (*CVPR'*05) **1** 26–33. IEEE, New York.

[9] BUBECK, S., DING, J., ELDAN, R. and RÁCZ, M. Z. (2016). Testing for high-dimensional geometry in random graphs. *Random Structures Algorithms* **49** 503–532. MR3545825 https://doi.org/10.1002/rsa.20633

[10] COLBOURN, C. J. and BOOTH, K. S. (1981). Linear time automorphism algorithms for trees, interval graphs, and planar graphs. *SIAM J. Comput.* **10** 203–225. MR0605613 https://doi.org/10.1137/0210015

[11] COUR, T., SRINIVASAN, P. and SHI, J. (2007). Balanced graph matching. In *Advances in Neural Information Processing Systems* 313–320.

MSC2020 subject classifications. Primary 62H15; secondary 05C80, 05C85, 68Q87.
Key words and phrases. Detecting network correlation, correlated Erdős–Rényi graphs, counting signed trees, graph matching, low-degree polynomials.

[12] CULLINA, D. and KIYAVASH, N. (2016). Improved achievability and converse bounds for Erdős–Rényi graph matching. arXiv preprint. Available at arXiv:1602.01042.

[13] CULLINA, D. and KIYAVASH, N. (2017). Exact alignment recovery for correlated Erdős–Rényi graphs. arXiv preprint. Available at arXiv:1711.06783.

[14] DING, J., MA, Z., WU, Y. and XU, J. (2021). Efficient random graph matching via degree profiles. *Probab. Theory Related Fields* **179** 29–115. MR4221654 https://doi.org/10.1007/s00440-020-00997-4

[15] DINNEEN, M. J. (2015). *Constant Time Generation of Free Trees*. Univ. Auckland Lecture, Auckland.

[16] FAN, Z., MAO, C., WU, Y. and XU, J. (2023). Spectral graph matching and regularized quadratic relaxations I: the Gaussian model. arxiv preprint. Available at arXiv:1907.08880.

[17] FAN, Z., MAO, C., WU, Y. and XU, J. (2023). Spectral graph matching and regularized quadratic relaxations II: Erdős–Rényi graphs and universality. *Found. Comput. Math.* **23** 1567–1617. MR4649431 https://doi.org/10.1007/s10208-022-09575-7

[18] FEIZI, S., QUON, G., RECAMONDE-MENDOZA, M., MÉDARD, M., KELLIS, M. and JADBABAIE, A. (2020). Spectral alignment of graphs. *IEEE Trans. Netw. Sci. Eng.* **7** 1182–1197. MR4149994 https://doi.org/10.1109/tnse.2019.2913233

[19] FRIEZE, A. and KAROŃSKI, M. (2016). *Introduction to Random Graphs*. Cambridge Univ. Press, Cambridge. MR3675279 https://doi.org/10.1017/CBO9781316339831

[20] GANASSALI, L. and MASSOULIÉ, L. (2020). From tree matching to sparse graph alignment. arXiv preprint. Available at arXiv:2002.01258.

[21] GANASSALI, L., MASSOULIÉ, L. and LELARGE, M. (2022). Correlation detection in trees for planted graph alignment. In *13th Innovations in Theoretical Computer Science Conference*. *LIPIcs. Leibniz Int. Proc. Inform.* **215** Art. No. 74, 8. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR4459445

[22] GAO, C. and LAFFERTY, J. (2017). Testing network structure using relations between small subgraph probabilities. arXiv preprint. Available at arXiv:1704.06742.

[23] GAO, C. and LAFFERTY, J. (2017). Testing for global network structure using small subgraph statistics. arXiv preprint. Available at arXiv:1710.00862.

[24] HAGHIGHI, A. D., NG, A. Y. and MANNING, C. D. (2005). Robust textual inference via graph matching. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*. 387–394. Association for Computational Linguistics, Cambridge.

[25] HALL, G. and MASSOULIÉ, L. (2023). Partial recovery in the graph alignment problem. *Oper. Res.* **71** 259–272. MR4585566

[26] HARARY, F. and PRINS, G. (1959). The number of homeomorphically irreducible trees, and other species. *Acta Math.* **101** 141–162. MR0101846 https://doi.org/10.1007/BF02559543

[27] HOPKINS, S. (2018). Statistical inference and the sum of squares method. PhD thesis, Cornell University.

[28] HOPKINS, S. B. and STEURER, D. (2017). Efficient Bayesian estimation from few samples: Community detection and related problems. In *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 379–390. IEEE Comp. Soc., Los Alamitos, CA. MR3734245 https://doi.org/10.1109/FOCS.2017.42

[29] JANSON, S., ŁUCZAK, T. and RUCINSKI, A. (2000). *Random Graphs*. *Wiley-Interscience Series in Discrete Mathematics and Optimization*. Wiley-Interscience, New York. MR1782847 https://doi.org/10.1002/9781118032718

[30] JIN, J., KE, Z. T. and LUO, S. (2021). Optimal adaptivity of signed-polygon statistics for network testing. *Ann. Statist.* **49** 3408–3433. MR4352535 https://doi.org/10.1214/21-aos2089

[31] 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

[32] LYZINSKI, V., FISHKIND, D., FIORI, M., VOGELSTEIN, J., PRIEBE, C. and SAPIRO, G. (2016). Graph matching: Relax at your own risk. *IEEE Trans. Pattern Anal. Mach. Intell.* **38** 60–73.

[33] MAO, C., RUDELSON, M. and TIKHOMIROV, K. (2021). Random graph matching with improved noise robustness. In *Proceedings of Thirty Fourth Conference on Learning Theory*. *Proceedings of Machine Learning Research* **134** 3296–3329.

[34] MAO, C., RUDELSON, M. and TIKHOMIROV, K. (2023). Exact matching of random graphs with constant correlation. *Probab. Theory Related Fields* **186** 327–389. MR4586222 https://doi.org/10.1007/s00440-022-01184-3

[35] MAO, C., WU, Y., XU, J. and YU, S. H. (2024). Supplement to "Testing network correlation efficiently via counting trees." https://doi.org/10.1214/23-AOS2261SUPP

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

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

[38] NARAYANAN, A. and SHMATIKOV, V. (2009). De-anonymizing social networks. In 2009 30*th IEEE Symposium on Security and Privacy* 173–187. IEEE, New York.

[39] OEIS Number of trees with n unlabeled nodes, entry A000055 in the on-line encyclopedia of integer sequences. Available at https://oeis.org/A000055.

[40] O'DONNELL, R. (2014). *Analysis of Boolean Functions*. Cambridge Univ. Press, New York. MR3443800 https://doi.org/10.1017/CBO9781139814782

[41] OTTER, R. (1948). The number of trees. *Ann. of Math.* (2) **49** 583–599. MR0025715 https://doi.org/10.2307/1969046

[42] PEDARSANI, P. and GROSSGLAUSER, M. (2011). On the privacy of anonymized networks. In *Proceedings of the* 17*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 1235–1243.

[43] PICCIOLI, G., SEMERJIAN, G., SICURO, G. and ZDEBOROVÁ, L. (2022). Aligning random graphs with a sub-tree similarity message-passing algorithm. *J. Stat. Mech. Theory Exp.* 6. Paper No. 063401, 44. MR4483763 https://doi.org/10.1088/1742-5468/ac70d2

[44] SLOTA, G. M. and MADDURI, K. (2013). Fast approximate subgraph counting and enumeration. In 2013 42*nd International Conference on Parallel Processing* 210–219. IEEE, New York.

[45] STEPHENS, M. A. (1979). Vector correlation. *Biometrika* **66** 41–48. MR0529146 https://doi.org/10.1093/biomet/66.1.41

[46] WRIGHT, R. A., RICHMOND, B., ODLYZKO, A. and MCKAY, B. D. (1986). Constant time generation of free trees. *SIAM J. Comput.* **15** 540–548. MR0837603 https://doi.org/10.1137/0215039

[47] WU, Y., XU, J. and YU, S. H. (2022). Settling the sharp reconstruction thresholds of random graph matching. *IEEE Trans. Inf. Theory* **68** 5391–5417. MR4476403 https://doi.org/10.1109/tit.2022.3169005

[48] WU, Y., XU, J. and YU, S. H. (2023). Testing correlation of unlabeled random graphs. *Ann. Appl. Probab.* **33** 2519–2558. MR4612649 https://doi.org/10.1214/22-aap1786

[49] YU, S. H. (2022). Source code for synthetic experiments in "Testing network correlation efficiently via counting trees". Availabe at https://github.com/SophieYu1014/fasciaGraphSimilarity-1. Accessed: 2022-08-12.

The Annals of Statistics
2024, Vol. 52, No. 6, 2506–2528
https://doi.org/10.1214/24-AOS2373

NON-INDEPENDENT COMPONENT ANALYSIS

BY GEERT MESTERS[1,a] AND PIOTR ZWIERNIK[2,b]

[1]*Department of Economics and Business, Universitat Pompeu Fabra,* [a]*geert.mesters@upf.edu*
[2]*Department of Statistical Sciences, University or Toronto,* [b]*piotr.zwiernik@utoronto.ca*

A seminal result in the ICA literature states that for $AY = \varepsilon$, if the components of ε are independent and at most one is Gaussian, then A is identified up to sign and permutation of its rows (*Signal Process.* **36** (1994)). In this paper we study to which extent the independence assumption can be relaxed by replacing it with restrictions on higher order moment or cumulant tensors of ε. We document new conditions that establish identification for several non-independent component models, for example, common variance models, and propose efficient estimation methods based on the identification results. We show that in situations where independence cannot be assumed the efficiency gains can be significant relative to methods that rely on independence.

REFERENCES

AMARI, S.-I., CICHOCKI, A. and YANG, H. (1996). A new learning algorithm for blind signal separation. *Adv. Neural Inf. Process. Syst.* **8** 757–763.

ASAI, M., MCALEER, M. and YU, J. (2006). Multivariate stochastic volatility: A review. *Econometric Rev.* **25** 145–175. MR2256285 https://doi.org/10.1080/07474930600713564

AZZALINI, A. and GENTON, M. G. (2008). Robust likelihood methods based on the skew-t and related distributions. *Int. Stat. Rev.* **76** 106–129.

BACH, F. R. and JORDAN, M. I. (2003). Kernel independent component analysis. *J. Mach. Learn. Res.* **3** 1–48. MR1966051 https://doi.org/10.1162/153244303768966085

BACH, F. R. and JORDAN, M. I. (2004). Beyond independent components: Trees and clusters. *J. Mach. Learn. Res.* **4** 1205–1233. MR2103627 https://doi.org/10.1162/jmlr.2003.4.7-8.1205

BACK, A. D. and WEIGEND, A. S. (1997). A first application of independent component analysis to extracting structure from stock returns. *Int. J. Neural Syst.* **8** 473–484. https://doi.org/10.1142/s0129065797000458

BEKAERT, G., ENGSTROM, E. and ERMOLOV, A. (2021). Macro risks and the term structure of interest rates. *J. Financ. Econ.* **141** 479–504.

BEKAERT, G., ENGSTROM, E. and ERMOLOV, A. (2022). Identifying aggregate demand and supply shocks using sign restrictions and higher-order moments. working paper.

CARDOSO, J.-F. (1989). Source separation using higher order moments. In *International Conference on Acoustics, Speech, and Signal Processing*, **4** 2109–2112.

CARDOSO, J.-F. and SOULOUMIAC, A. (1993). Blind beamforming for non-Gaussian signals. *IEE Proc., F, Radar Signal Process.* **140**.

CARDOSO, J. F. (1999). High-order contrasts for independent component analysis. *Neural Comput.* **11** 157–192. https://doi.org/10.1162/089976699300016863

CHAMBERLAIN, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *J. Econometrics* **34** 305–334. MR0888070 https://doi.org/10.1016/0304-4076(87)90015-7

CHEN, A. and BICKEL, P. J. (2006). Efficient independent component analysis. *Ann. Statist.* **34** 2825–2855. MR2329469 https://doi.org/10.1214/009053606000000939

COMON, P. (1994). Independent component analysis, a new concept? *Signal Process.* **36**.

COMON, P. and JUTTEN, C. (2010). *Handbook of Blind Source Separation*. Academic Press, Oxford.

DARMOIS, G. (1953). Analyse générale des liaisons stochastiques. Etude particulière de l'analyse factorielle linéaire. *Rev. Inst. Int. Stat.* **21** 2–8. MR0061322

DAVIS, R. and NG, S. (2023). Time series estimation of the dynamic effects of disaster-type shocks. *J. Econometrics* **235** 180–201. MR4580426 https://doi.org/10.1016/j.jeconom.2022.02.009

DRAUTZBURG, T. and WRIGHT, J. H. (2023). Refining set-identification in VARs through independence. *J. Econometrics* **235** 1827–1847. MR4602934 https://doi.org/10.1016/j.jeconom.2023.01.011

MSC2020 subject classifications. 15A69, 62H99.
Key words and phrases. Independent component analysis, identifiability, cumulants, tensors.

ERICKSON, T., JIANG, C. H. and WHITED, T. M. (2014). Minimum distance estimation of the errors-in-variables model using linear cumulant equations. *J. Econometrics* **183** 211–221. MR3276390 https://doi.org/10.1016/j.jeconom.2014.05.011

ERIKSSON, J. and KOIVUNEN, V. (2003). Identifiability and separability of linear ICA models revisited. In *4th International Symposium on Independent Components Analysis and Blind Source Separation* (*ICA*2003) **1** 23–27.

FORBES, F. and WRAITH, D. (2014). A new family of multivariate heavy-tailed distributions with variable marginal amounts of tailweight: Application to robust clustering. *Stat. Comput.* **24** 971–984. MR3253848 https://doi.org/10.1007/s11222-013-9414-4

GEARY, R. C. (1942). Inherent relation between random variables. *Proc. R. Ir. Acad. Sect. A* **47** 63–76. MR0006654

GUAY, A. (2021). Identification of structural vector autoregressions through higher unconditional moments. *J. Econometrics* **225** 27–46. MR4314059 https://doi.org/10.1016/j.jeconom.2020.10.006

HALL, A. R. (2005). *Generalized Method of Moments*. *Advanced Texts in Econometrics*. Oxford Univ. Press, Oxford. MR2135106

HALLIN, M. and MEHTA, C. (2015). *R*-estimation for asymmetric independent component analysis. *J. Amer. Statist. Assoc.* **110** 218–232. MR3338498 https://doi.org/10.1080/01621459.2014.909316

HAN, F. and LIU, H. (2018). ECA: High-dimensional elliptical component analysis in non-Gaussian distributions. *J. Amer. Statist. Assoc.* **113** 252–268. MR3803462 https://doi.org/10.1080/01621459.2016.1246366

HANSEN, B. E. and LEE, S. (2021). Inference for iterated GMM under misspecification. *Econometrica* **89** 1419–1447. MR4325186 https://doi.org/10.3982/ecta16274

HANSEN, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50** 1029–1054. MR0666123 https://doi.org/10.2307/1912775

HANSEN, L. P., HEATON, J. and YARON, A. (1996). Finite-sample properties of some alternative GMM estimators. *J. Bus. Econom. Statist.* **14** 262–280.

HASTIE, T. and TIBSHIRANI, R. (2002). Independent components analysis through product density estimation. In *Proceedings of the 15th International Conference on Neural Information Processing Systems*. *NIPS'02* 665–672. MIT Press, Cambridge, MA, USA.

HAYASHI, F. (2000). *Econometrics*. Princeton Univ. Press, Princeton, NJ. MR1881537

HYVÄRINEN, A. (1999a). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.* **10** 626–634. https://doi.org/10.1109/72.761722

HYVÄRINEN, A. (1999b). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.* **10** 626–634. https://doi.org/10.1109/72.761722

HYVÄRINEN, A. (2013). Independent component analysis: Recent advances. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **371** 20110534. MR3005670 https://doi.org/10.1098/rsta.2011.0534

HYVÄRINEN, A., HOYER, P. O. and INKI, M. (2001). Topographic independent component analysis. *Neural Comput.* **13** 1527–1558. https://doi.org/10.1162/089976601750264992

HYVÄRINEN, A., KARHUNEN, J. and OJA, E. (2001). *Independent Component Analysis*. Wiley, New York.

ILMONEN, P., NORDHAUSEN, K., OJA, H. and OLLILA, E. (2010). A new performance index for ICA: Properties, computation and asymptotic analysis. In *Latent Variable Analysis and Signal Separation* (V. Vigneron, V. Zarzoso, E. Moreau, R. Gribonval and E. Vincent, eds.) 229–236. Springer, Berlin, Heidelberg.

ILMONEN, P. and PAINDAVEINE, D. (2011). Semiparametrically efficient inference based on signed ranks in symmetric independent component models. *Ann. Statist.* **39** 2448–2476. MR2906874 https://doi.org/10.1214/11-AOS906

KELKER, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā Ser. A* **32** 419–438. MR0287628

KILIAN, L. and LÜTKEPOHL, H. (2017). *Structural Vector Autoregressive Analysis*. Cambridge Univ. Press, Cambridge.

LANNE, M. and LUOTO, J. (2021). GMM estimation of non-Gaussian structural vector autoregression. *J. Bus. Econom. Statist.* **39** 69–81. MR4187176 https://doi.org/10.1080/07350015.2019.1629940

LEE, A. and MESTERS, G. (2021). Robust non-Gaussian identification and inference for simultaneous equations. Working Paper.

LEWBEL, A., SCHENNACH, S. M. and ZHANG, L. (2021). Identification of a triangular two equation system without instruments. working paper.

LIM, L.-H. (2021). Tensors in computations. *Acta Numer.* **30** 555–764. MR4298222 https://doi.org/10.1017/S0962492921000076

LUDVIGSON, S. C., MA, S. and NG, S. (2021). Uncertainty and business cycles: Exogenous impulse or endogenous response? *Am. Econ. J. Macroecon.* **13**.

LUKACS, E. (1958). Some extensions of a theorem of Marcinkiewicz. *Pacific J. Math.* **8** 487–501. MR0101543

MARCINKIEWICZ, J. (1939). Sur une propriété de la loi de Gauß. *Math. Z.* **44** 612–618. MR1545791 https://doi.org/10.1007/BF01210677

MARRON, J. S. and WAND, M. P. (1992). Exact mean integrated squared error. *Ann. Statist.* **20** 712–736. MR1165589 https://doi.org/10.1214/aos/1176348653

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

MCCULLAGH, P. (2018). *Tensor Methods in Statistics*: *Monographs on Statistics and Applied Probability*. CRC Press/CRC, Boca Raton.

MESTERS, G. and ZWIERNIK, P. (2024). Supplement to "Non-independent component analysis." https://doi.org/10.1214/24-AOS2373SUPP

MEYER-BASE, A., AUER, D. and WISMUELLER, A. (2003). Topographic independent component analysis for fMRI signal detection. In *Proceedings of the International Joint Conference on Neural Networks*, 2003. **1** 601–605. IEEE, Los Alamitos.

MEYER-BÄSE, A., LANGE, O., WISMÜLLER, A. and RITTER, H. (2004). Model-free functional MRI analysis using topographic independent component analysis. *Int. J. Neural Syst.* **14** 217–228.

MONTIEL OLEA, J. L., PLAGBORG-MØLLER, M. and QIAN, E. (2022). SVAR identification from higher moments: Has the simultaneous causality problem been solved? *AEA Pap. Proc.* **112** 481–85.

PALMER, J. A., KREUTZ-DELGADO, K., RAO, B. D. and MAKEIG, S. (2007). Modeling and estimation of dependent subspaces with non-radially symmetric and skewed densities. In *Independent Component Analysis and Signal Separation*: 7*th International Conference*, *ICA* 2007, *London*, *UK*, *September* 9-12, 2007. *Proceedings* 7 97–104. Springer, Berlin.

ROSSELL, D. and ZWIERNIK, P. (2021). Dependence in elliptical partial correlation graphs. *Electron. J. Stat.* **15** 4236–4263. MR4312203 https://doi.org/10.1214/21-ejs1891

SAMWORTH, R. J. and YUAN, M. (2012). Independent component analysis via nonparametric maximum likelihood estimation. *Ann. Statist.* **40** 2973–3002. MR3097966 https://doi.org/10.1214/12-AOS1060

SCHENNACH, S. M. (2021). Measurement systems. *J. Econ. Lit.*. forthcoming.

SHIMIZU, S., HOYER, P. O., HYVÄRINEN, A. and KERMINEN, A. (2006). A linear non-Gaussian acyclic model for causal discovery. *J. Mach. Learn. Res.* **7** 2003–2030. MR2274431

SKITOVIČ, V. P. (1953). On a property of the normal distribution. *Dokl. Akad. Nauk SSSR* **89** 217–219. MR0055597

SPEED, T. P. (1983). Cumulants and partition lattices. *Aust. J. Stat.* **25** 378–388. MR0725217

VELASCO, C. (2023). Identification and estimation of structural VARMA models using higher order dynamics. *J. Bus. Econom. Statist.* **41** 819–832. MR4600851 https://doi.org/10.1080/07350015.2022.2075000

VOGEL, D. and FRIED, R. (2011). Elliptical graphical modelling. *Biometrika* **98** 935–951. MR2860334 https://doi.org/10.1093/biomet/asr037

ZWIERNIK, P. (2012). L-cumulants, L-cumulant embeddings and algebraic statistics. *J. Algebr. Stat.* **3** 11–43. MR3016419 https://doi.org/10.18409/jas.v3i1.14

ZWIERNIK, P. (2016). *Semialgebraic Statistics and Latent Tree Models*. *Monographs on Statistics and Applied Probability* **146**. CRC Press/CRC, Boca Raton, FL. MR3379921

The Annals of Statistics
2024, Vol. 52, No. 6, 2529–2556
https://doi.org/10.1214/24-AOS2387

ESTIMATION OF THE SPECTRAL MEASURE FROM CONVEX COMBINATIONS OF REGULARLY VARYING RANDOM VECTORS

BY MARCO OESTING[1,a] AND OLIVIER WINTENBERGER[2,b]

[1]*Stuttgart Center for Simulation Science and Institute of Stochastics and Applications, University of Stuttgart,* [a]*marco.oesting@mathematik.uni-stuttgart.de*

[2]*LPSM, Sorbonne Université,* [b]*olivier.wintenberger@sorbonne-universite.fr*

The extremal dependence structure of a regularly varying random vector $\boldsymbol{X}$ is fully described by its limiting spectral measure. In this paper, we investigate how to recover characteristics of the measure, such as extremal coefficients, from the extremal behaviour of convex combinations of components of $\boldsymbol{X}$. Our considerations result in a class of new estimators of moments of the corresponding combinations for the spectral vector. We show asymptotic normality by means of a functional limit theorem and, focusing on the estimation of extremal coefficients, we verify that the minimal asymptotic variance can be achieved by a plug-in estimator using subsampling bootstrap. We illustrate the benefits of our approach on simulated and real data.

REFERENCES

BASRAK, B., DAVIS, R. A. and MIKOSCH, T. (2002). A characterization of multivariate regular variation. *Ann. Appl. Probab.* **12** 908–920. MR1925445 https://doi.org/10.1214/aoap/1031863174

BERNARD, E., NAVEAU, P., VRAC, M. and MESTRE, O. (2013). Clustering of maxima: Spatial dependencies among heavy rainfall in France. *J. Climate* **26** 7929–7937.

BOMAN, J. and LINDSKOG, F. (2009). Support theorems for the Radon transform and Cramér–Wold theorems. *J. Theoret. Probab.* **22** 683–710. MR2530109 https://doi.org/10.1007/s10959-008-0151-0

BURITICÁ, G. and NAVEAU, P. (2023). Stable sums to infer high return levels of multivariate rainfall time series. *Environmetrics* **34** Paper No. e2782, 18. MR4594682

CAPÉRAÀ, P., FOUGÈRES, A.-L. and GENEST, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* **84** 567–577. MR1603985 https://doi.org/10.1093/biomet/84.3.567

COOLEY, D., NAVEAU, P. and PONCET, P. (2006). Variograms for spatial max-stable random fields. In *Dependence in Probability and Statistics*. *Lect. Notes Stat.* **187** 373–390. Springer, New York. MR2283264 https://doi.org/10.1007/0-387-36062-X_17

COOLEY, D. and THIBAUD, E. (2019). Decompositions of dependence for high-dimensional extremes. *Biometrika* **106** 587–604. MR3992391 https://doi.org/10.1093/biomet/asz028

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

DREES, H. and HUANG, X. (1998). Best attainable rates of convergence for estimators of the stable tail dependence function. *J. Multivariate Anal.* **64** 25–46. MR1619974 https://doi.org/10.1006/jmva.1997.1708

DREES, H. and SABOURIN, A. (2021). Principal component analysis for multivariate extremes. *Electron. J. Stat.* **15** 908–943. MR4255291 https://doi.org/10.1214/21-ejs1803

EINMAHL, J. H. J., DE HAAN, L. and HUANG, X. (1993). Estimating a multidimensional extreme-value distribution. *J. Multivariate Anal.* **47** 35–47. MR1239104 https://doi.org/10.1006/jmva.1993.1069

EINMAHL, J. H. J., DE HAAN, L. and PITERBARG, V. I. (2001). Nonparametric estimation of the spectral measure of an extreme value distribution. *Ann. Statist.* **29** 1401–1423. MR1873336 https://doi.org/10.1214/aos/1013203459

EINMAHL, J. H. J., DE HAAN, L. and SINHA, A. K. (1997). Estimating the spectral measure of an extreme value distribution. *Stochastic Process. Appl.* **70** 143–171. MR1475660 https://doi.org/10.1016/S0304-4149(97)00065-3

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

MSC2020 subject classifications. Primary 62G32; secondary 62G30.
Key words and phrases. Asymptotic statistics, extremal coefficient, moment estimator, multivariate extremes.

GENEST, C. and SEGERS, J. (2009). Rank-based inference for bivariate extreme-value copulas. *Ann. Statist.* **37** 2990–3022. MR2541453 https://doi.org/10.1214/08-AOS672

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

KLÜPPELBERG, C. and PERGAMENCHTCHIKOV, S. (2007). Extremal behaviour of models with multivariate random recurrence representation. *Stochastic Process. Appl.* **117** 432–456. MR2305380 https://doi.org/10.1016/j.spa.2006.09.001

MAINIK, G. and RÜSCHENDORF, L. (2010). On optimal portfolio diversification with respect to extreme risks. *Finance Stoch.* **14** 593–623. MR2738025 https://doi.org/10.1007/s00780-010-0122-z

OESTING, M. and WINTENBERGER, O. (2024). Supplement to "Estimation of the Spectral Measure from Convex Combinations of Regularly Varying Random Vectors." https://doi.org/10.1214/24-AOS2387SUPP

PICKANDS, J. III (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3** 119–131. MR0423667

SCHLATHER, M. and TAWN, J. A. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika* **90** 139–156. MR1966556 https://doi.org/10.1093/biomet/90.1.139

SMITH, R. L. (1990). Max–stable processes and spatial extremes. Unpublished manuscript.

STUPFLER, G. (2019). On a relationship between randomly and non-randomly thresholded empirical average excesses for heavy tails. *Extremes* **22** 749–769. MR4031856 https://doi.org/10.1007/s10687-019-00351-5

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

The Annals of Statistics
2024, Vol. 52, No. 6, 2557–2582
https://doi.org/10.1214/24-AOS2395

STATISTICAL COMPLEXITY AND OPTIMAL ALGORITHMS FOR NONLINEAR RIDGE BANDITS

BY NIVED RAJARAMAN[1,a], YANJUN HAN[2,d], JIANTAO JIAO[1,b] AND KANNAN RAMCHANDRAN[1,c]

[1]*Department of Electrical Engineering and Computer Sciences, University of California, Berkely,* [a]*nived.rajaraman@berkeley.edu,* [b]*jiantao@eecs.berkeley.edu,* [c]*kannanr@eecs.berkeley.edu*
[2]*Courant Institute of Mathematical Sciences and Center for Data Science, New York University,* [d]*yanjunhan@nyu.edu*

We consider the sequential decision-making problem where the mean outcome is a nonlinear function of the chosen action. Compared with the linear model, two curious phenomena arise in nonlinear models: first, in addition to the "learning phase" with a standard parametric rate for estimation or regret, there is an "burn-in period" with a fixed cost determined by the nonlinear function; second, achieving the smallest burn-in cost requires new exploration algorithms. For a special family of nonlinear functions named ridge functions in the literature, we derive upper and lower bounds on the optimal burn-in cost, and in addition, on the entire learning trajectory during the burn-in period via differential equations. In particular, a two-stage algorithm that first finds a good initial action and then treats the problem as locally linear is statistically optimal. In contrast, several classical algorithms, such as UCB and algorithms relying on regression oracles, are provably suboptimal.

REFERENCES

[1] ABBASI-YADKORI, Y., PÁL, D. and SZEPESVÁRI, C. (2011). Improved algorithms for linear stochastic bandits. *Adv. Neural Inf. Process. Syst.* **24**.

[2] AGARWAL, A., DUDÍK, M., KALE, S., LANGFORD, J. and SCHAPIRE, R. (2012). Contextual bandit learning with predictable rewards. In *Artificial Intelligence and Statistics* 19–26. PMLR.

[3] AGARWAL, A., FOSTER, D. P., HSU, D. J., KAKADE, S. M. and RAKHLIN, A. (2011). Stochastic convex optimization with bandit feedback. *Adv. Neural Inf. Process. Syst.* **24**.

[4] AGARWAL, A., HSU, D., KALE, S., LANGFORD, J., LI, L. and SCHAPIRE, R. (2014). Taming the monster: A fast and simple algorithm for contextual bandits. In *International Conference on Machine Learning* 1638–1646. PMLR.

[5] AGARWAL, A., WAINWRIGHT, M. J., BARTLETT, P. and RAVIKUMAR, P. (2009). Information-theoretic lower bounds on the oracle complexity of convex optimization. *Adv. Neural Inf. Process. Syst.* **22**.

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

[7] AUER, P., CESA-BIANCHI, N., FREUND, Y. and SCHAPIRE, R. E. (2002/03). The nonstochastic multiarmed bandit problem. *SIAM J. Comput.* **32** 48–77. MR1954855 https://doi.org/10.1137/S0097539701398375

[8] BACHOC, F., CESARI, T. and GERCHINOVITZ, S. (2021). Instance-dependent bounds for zeroth-order Lipschitz optimization with error certificates. *Adv. Neural Inf. Process. Syst.* **34** 24180–24192.

[9] BARTROFF, J., LAI, T. L. and SHIH, M.-C. (2012). *Sequential Experimentation in Clinical Trials: Design and Analysis* **298**. Springer, Berlin.

[10] BERRY, D. A., CHEN, R. W., ZAME, A., HEATH, D. C. and SHEPP, L. A. (1997). Bandit problems with infinitely many arms. *Ann. Statist.* **25** 2103–2116. MR1474085 https://doi.org/10.1214/aos/1069362389

[11] BILLARD, A. and KRAGIC, D. (2019). Trends and challenges in robot manipulation. *Science* **364**. https://doi.org/10.1126/science.aat8414

[12] BLOT, W. J. and MEETER, D. A. (1973). Sequential experimental design procedures. *J. Amer. Statist. Assoc.* **68** 586–593. MR0359209

MSC2020 subject classifications. Primary 62L12; secondary 62C20, 62K05.
Key words and phrases. Bandit problems, regret bounds, adaptive sampling, sequential estimation, ridge functions, minimax rate.

[13] BLUMENTHAL, S. (1976). Sequential estimation of the largest normal mean when the variance is known. *Ann. Statist.* **4** 1077–1087. MR0431479

[14] BONALD, T. and PROUTIERE, A. (2013). Two-target algorithms for infinite-armed bandits with Bernoulli rewards. *Adv. Neural Inf. Process. Syst.* **26**.

[15] BOUTTIER, C., CESARI, T., DUCOFFE, M. and GERCHINOVITZ, S. (2020). Regret analysis of the Piyavskii-Shubert algorithm for global Lipschitz optimization. ArXiv preprint arXiv:2002.02390.

[16] BUBECK, S. and ELDAN, R. (2016). Multi-scale exploration of convex functions and bandit convex optimization. In *Conference on Learning Theory* 583–589. PMLR.

[17] BUBECK, S., LEE, Y. T. and ELDAN, R. (2017). Kernel-based methods for bandit convex optimization. In *STOC'*17*—Proceedings of the* 49*th Annual ACM SIGACT Symposium on Theory of Computing* 72–85. ACM, New York. MR3678172 https://doi.org/10.1145/3055399.3055403

[18] BURNAŠEV, M. (1980). Sequential discrimination of hypotheses with control of observations. *Math. USSR, Izv.* **15** 419.

[19] CHEN, F., MEI, S. and BAI, Y. (2022). Unified algorithms for RL with decision-estimation coefficients: No-regret, PAC, and reward-free learning. ArXiv preprint arXiv:2209.11745.

[20] CHERNOFF, H. (1959). Sequential design of experiments. *Ann. Math. Stat.* **30** 755–770. MR0108874 https://doi.org/10.1214/aoms/1177706205

[21] CHU, W., LI, L., REYZIN, L. and SCHAPIRE, R. (2011). Contextual bandits with linear payoff functions. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics* 208–214. JMLR Workshop and Conference Proceedings.

[22] CSISZÁR, I. (1972). A class of measures of informativity of observation channels. *Period. Math. Hungar.* **2** 191–213. MR0335152 https://doi.org/10.1007/BF02018661

[23] DANI, V., HAYES, T. P. and KAKADE, S. M. (2008). Stochastic linear optimization under bandit feedback. *Conf. Learn. Theory* 355–366.

[24] DU, S., KAKADE, S., LEE, J., LOVETT, S., MAHAJAN, G., SUN, W. and WANG, R. (2021). Bilinear classes: A structural framework for provable generalization in RL. In *International Conference on Machine Learning* 2826–2836. PMLR.

[25] DUDIK, M., HSU, D., KALE, S., KARAMPATZIAKIS, N., LANGFORD, J., REYZIN, L. and ZHANG, T. (2011). Efficient optimal learning for contextual bandits. In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence* 169–178.

[26] FILIPPI, S., CAPPE, O., GARIVIER, A. and SZEPESVÁRI, C. (2010). Parametric bandits: The generalized linear case. *Adv. Neural Inf. Process. Syst.* **23**.

[27] FLAXMAN, A. D., KALAI, A. T. and MCMAHAN, H. B. (2005). Online convex optimization in the bandit setting: Gradient descent without a gradient. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms* 385–394. ACM, New York. MR2298287

[28] FOSTER, D. and RAKHLIN, A. (2020). Beyond UCB: Optimal and efficient contextual bandits with regression oracles. In *International Conference on Machine Learning* 3199–3210. PMLR.

[29] FOSTER, D. J., GENTILE, C., MOHRI, M. and ZIMMERT, J. (2020). Adapting to misspecification in contextual bandits. *Adv. Neural Inf. Process. Syst.* **33** 11478–11489.

[30] FOSTER, D. J., GOLOWICH, N. and HAN, Y. (2023). Tight guarantees for interactive decision making with the decision-estimation coefficient. In *The Thirty Sixth Annual Conference on Learning Theory* 3969–4043. PMLR.

[31] FOSTER, D. J., GOLOWICH, N., QIAN, J., RAKHLIN, A. and SEKHARI, A. (2022). A note on model-free reinforcement learning with the decision-estimation coefficient. ArXiv preprint arXiv:2211.14250.

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

[33] FOSTER, D. J., RAKHLIN, A., SEKHARI, A. and SRIDHARAN, K. (2022). On the complexity of adversarial decision making. *Adv. Neural Inf. Process. Syst.* **35** 35404–35417.

[34] GARIVIER, A., MÉNARD, P. and STOLTZ, G. (2019). Explore first, exploit next: The true shape of regret in bandit problems. *Math. Oper. Res.* **44** 377–399. MR3959077 https://doi.org/10.1287/moor.2017.0928

[35] GITTINS, J. C. (1979). Bandit processes and dynamic allocation indices. *J. Roy. Statist. Soc. Ser. B* **41** 148–164.

[36] HANSEN, P., JAUMARD, B. and LU, S.-H. (1991). On the number of iterations of Piyavskiĭ's global optimization algorithm. *Math. Oper. Res.* **16** 334–350. MR1106805 https://doi.org/10.1287/moor.16.2.334

[37] HÄRDLE, W., MÜLLER, M., SPERLICH, S. and WERWATZ, A. (2004). *Nonparametric and Semiparametric Models. Springer Series in Statistics*. Springer, New York. MR2061786 https://doi.org/10.1007/978-3-642-17146-8

[38] HUANG, B., HUANG, K., KAKADE, S., LEE, J. D., LEI, Q., WANG, R. and YANG, J. (2021). Optimal gradient-based algorithms for non-concave bandit optimization. *Adv. Neural Inf. Process. Syst.* **34** 29101–29115.

[39] JAMIESON, K. G., NOWAK, R. and RECHT, B. (2012). Query complexity of derivative-free optimization. *Adv. Neural Inf. Process. Syst.* **25**.

[40] JIANG, N., KRISHNAMURTHY, A., AGARWAL, A., LANGFORD, J. and SCHAPIRE, R. E. (2017). Contextual decision processes with low Bellman rank are PAC-learnable. In *International Conference on Machine Learning* 1704–1713. PMLR.

[41] JIN, C., LIU, Q. and MIRYOOSEFI, S. (2021). Bellman Eluder dimension: New rich classes of RL problems, and sample-efficient algorithms. *Adv. Neural Inf. Process. Syst.* **34** 13406–13418.

[42] KALASHNIKOV, D., IRPAN, A., PASTOR, P., IBARZ, J., HERZOG, A., JANG, E., QUILLEN, D., HOLLY, E., KALAKRISHNAN, M. et al. (2018). Qt-opt: Scalable deep reinforcement learning for vision-based robotic manipulation. ArXiv preprint arXiv:1806.10293.

[43] KLEINBERG, R. (2004). Nearly tight bounds for the continuum-armed bandit problem. *Adv. Neural Inf. Process. Syst.* **17**.

[44] KOBER, J., BAGNELL, J. A. and PETERS, J. (2013). Reinforcement learning in robotics: A survey. *Int. J. Robot. Res.* **32** 1238–1274.

[45] KOLESNIKOV, A. V. and MILMAN, E. (2016). Riemannian metrics on convex sets with applications to Poincaré and log-Sobolev inequalities. *Calc. Var. Partial Differential Equations* **55** 77. MR3514409 https://doi.org/10.1007/s00526-016-1018-3

[46] KRISHNAMURTHY, S. K., HADAD, V. and ATHEY, S. (2021). Adapting to misspecification in contextual bandits with offline regression oracles. In *International Conference on Machine Learning* 5805–5814. PMLR.

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

[48] LATTIMORE, T. (2019). Improved regret for zeroth-order adversarial bandit convex optimisation. *Math. Stat. Learn.* **2** 311–334. MR4165267 https://doi.org/10.4171/msl/17

[49] LATTIMORE, T. (2021). Minimax regret for bandit convex optimisation of ridge functions. ArXiv preprint arXiv:2106.00444.

[50] LATTIMORE, T. (2022). Minimax regret for partial monitoring: Infinite outcomes and rustichini's regret. In *Conference on Learning Theory* 1547–1575. PMLR.

[51] LATTIMORE, T. and GYORGY, A. (2021). Mirror descent and the information ratio. In *Conference on Learning Theory* 2965–2992. PMLR.

[52] LATTIMORE, T. and HAO, B. (2021). Bandit phase retrieval. *Adv. Neural Inf. Process. Syst.* **34** 18801–18811.

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

[54] LOGAN, B. F. and SHEPP, L. A. (1975). Optimal reconstruction of a function from its projections. *Duke Math. J.* **42** 645–659. MR0397240

[55] MINSKY, M. (1961). Steps toward artificial intelligence. *Proc. IRE* **49** 8–30. MR0134428

[56] OKAMOTO, I., AMARI, S. and TAKEUCHI, K. (1991). Asymptotic theory of sequential estimation: Differential geometrical approach. *Ann. Statist.* **19** 961–981. MR1105855 https://doi.org/10.1214/aos/1176348131

[57] RAGINSKY, M. and RAKHLIN, A. (2011). Information-based complexity, feedback and dynamics in convex programming. *IEEE Trans. Inf. Theory* **57** 7036–7056. MR2882278 https://doi.org/10.1109/TIT.2011.2154375

[58] RAJARAMAN, N., HAN, Y., JIAO, J. and RAMCHANDRAN, K. (2024). Supplement to "Statistical Complexity and Optimal Algorithms for Non-linear Ridge Bandits." https://doi.org/10.1214/24-AOS2395SUPP

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

[60] RUSSO, D. and VAN ROY, B. (2013). Eluder dimension and the sample complexity of optimistic exploration. *Adv. Neural Inf. Process. Syst.* **26**.

[61] RUSSO, D. and VAN ROY, B. (2014). Learning to optimize via information-directed sampling. *Adv. Neural Inf. Process. Syst.* **27**.

[62] SHAMIR, O. (2013). On the complexity of bandit and derivative-free stochastic convex optimization. In *Conference on Learning Theory* 3–24. PMLR.

[63] SIMCHI-LEVI, D. and XU, Y. (2022). Bypassing the monster: A faster and simpler optimal algorithm for contextual bandits under realizability. *Math. Oper. Res.* **47** 1904–1931. MR4506358

[64] SUN, W., JIANG, N., KRISHNAMURTHY, A., AGARWAL, A. and LANGFORD, J. (2019). Model-based RL in contextual decision processes: PAC bounds and exponential improvements over model-free approaches. In *Conference on Learning Theory* 2898–2933. PMLR.

[65] SUTTON, R. S. (1984). *Temporal Credit Assignment in Reinforcement Learning*. Univ. Massachusetts Amherst, Amherst.

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

[67] TATIKONDA, S. and MITTER, S. (2009). The capacity of channels with feedback. *IEEE Trans*. *Inf*. *Theory* **55** 323–349. MR2589700 https://doi.org/10.1109/TIT.2008.2008147

[68] VILLAR, S. S., BOWDEN, J. and WASON, J. (2015). Multi-armed bandit models for the optimal design of clinical trials: Benefits and challenges. *Statist*. *Sci*. **30** 199–215. MR3353103 https://doi.org/10.1214/14-STS504

[69] WAGENMAKER, A. J., CHEN, Y., SIMCHOWITZ, M., DU, S. and JAMIESON, K. (2022). Reward-free RL is no harder than reward-aware RL in linear Markov decision processes. In *International Conference on Machine Learning* 22430–22456. PMLR.

[70] WALD, A. and WOLFOWITZ, J. (1948). Optimum character of the sequential probability ratio test. *Ann*. *Math*. *Stat*. **19** 326–339. MR0026779 https://doi.org/10.1214/aoms/1177730197

[71] WANG, R., SALAKHUTDINOV, R. R. and YANG, L. (2020). Reinforcement learning with general value function approximation: Provably efficient approach via bounded eluder dimension. *Adv*. *Neural Inf*. *Process*. *Syst*. **33** 6123–6135.

[72] WANG, Y., AUDIBERT, J.-Y. and MUNOS, R. (2008). Algorithms for infinitely many-armed bandits. *Adv*. *Neural Inf*. *Process*. *Syst*. **21**.

[73] WANG, Y., BAHARAV, T., HAN, Y., JIAO, J. and TSE, D. (2022). Beyond the best: Estimating distribution functionals in infinite-armed bandits. *Adv*. *Neural Inf*. *Process*. *Syst*. **35**.

[74] WANG, Y., WANG, R. and KAKADE, S. (2021). An exponential lower bound for linearly realizable MDP with constant suboptimality gap. *Adv*. *Neural Inf*. *Process*. *Syst*. **34** 9521–9533.

[75] WEISZ, G., AMORTILA, P. and SZEPESVÁRI, C. (2021). Exponential lower bounds for planning in MDPs with linearly-realizable optimal action-value functions. In *Algorithmic Learning Theory* 1237–1264. PMLR.

[76] ZHU, H., GUPTA, A., RAJESWARAN, A., LEVINE, S. and KUMAR, V. (2019). Dexterous manipulation with deep reinforcement learning: Efficient, general, and low-cost. In 2019 *International Conference on Robotics and Automation* (*ICRA*) 3651–3657. IEEE, Los Alamitos.

The Annals of Statistics
2024, Vol. 52, No. 6, 2583–2612
https://doi.org/10.1214/24-AOS2396

TENSOR-ON-TENSOR REGRESSION: RIEMANNIAN OPTIMIZATION, OVER-PARAMETERIZATION, STATISTICAL-COMPUTATIONAL GAP AND THEIR INTERPLAY

BY YUETIAN LUO[1,a] AND ANRU R. ZHANG[2,b]

[1]*Data Science Institute, University of Chicago,* [a]*yuetian@uchicago.edu*
[2]*Department of Biostatistics & Bioinformatics and Department of Computer Science, Duke University,* [b]*anru.zhang@duke.edu*

We study the tensor-on-tensor regression, where the goal is to connect tensor responses to tensor covariates with a low Tucker rank parameter tensor/matrix without prior knowledge of its intrinsic rank. We propose the Riemannian gradient descent (RGD) and Riemannian Gauss–Newton (RGN) methods and cope with the challenge of unknown rank by studying the effect of rank over-parameterization. We provide the first convergence guarantee for the general tensor-on-tensor regression by showing that RGD and RGN respectively converge linearly and quadratically to a statistically optimal estimate in both rank correctly-parameterized and over-parameterized settings. Our theory reveals an intriguing phenomenon: Riemannian optimization methods naturally adapt to over-parameterization without modifications to their implementation. We also prove the statistical-computational gap in scalar-on-tensor regression by a direct low-degree polynomial argument. Our theory demonstrates a "blessing of statistical-computational gap" phenomenon: in a wide range of scenarios in tensor-on-tensor regression for tensors of order three or higher, the computationally required sample size matches what is needed by moderate rank over-parameterization when considering computationally feasible estimators, while there are no such benefits in the matrix settings. This shows moderate rank over-parameterization is essentially "cost-free" in terms of sample size in tensor-on-tensor regression of order three or higher. Finally, we conduct simulation studies to show the advantages of our proposed methods and to corroborate our theoretical findings.

REFERENCES

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

AHMED, T., RAJA, H. and BAJWA, W. U. (2020). Tensor regression using low-rank and sparse Tucker decompositions. *SIAM J. Math. Data Sci.* **2** 944–966. MR4161310 https://doi.org/10.1137/19M1299335

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

BANDEIRA, A. S., KUNISKY, D. and WEIN, A. S. (2020). Computational hardness of certifying bounds on constrained PCA problems. In 11*th Innovations in Theoretical Computer Science Conference*. *LIPIcs. Leibniz Int. Proc. Inform.* **151** Art. No. 78, 29. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR4048181

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

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

BARBER, R. F. and HA, W. (2018). Gradient descent with non-convex constraints: Local concavity determines convergence. *Inf. Inference* **7** 755–806. MR4023770 https://doi.org/10.1093/imaiai/iay002

MSC2020 subject classifications. Primary 62H15; secondary 62C20.
Key words and phrases. Tensor-on-tensor regression, over-parameterization, Riemannian optimization, statistical-computational gaps, low-degree polynomials.

BARTLETT, P. L., LONG, P. M., LUGOSI, G. and TSIGLER, A. (2020). Benign overfitting in linear regression. *Proc. Natl. Acad. Sci. USA* **117** 30063–30070. MR4263288 https://doi.org/10.1073/pnas.1907378117

BARTLETT, P. L., MONTANARI, A. and RAKHLIN, A. (2021). Deep learning: A statistical viewpoint. *Acta Numer.* **30** 87–201. MR4295218 https://doi.org/10.1017/S0962492921000027

BELKIN, M. (2021). Fit without fear: Remarkable mathematical phenomena of deep learning through the prism of interpolation. *Acta Numer.* **30** 203–248. MR4298218 https://doi.org/10.1017/S0962492921000039

BELKIN, M., HSU, D., MA, S. and MANDAL, S. (2019). Reconciling modern machine-learning practice and the classical bias-variance trade-off. *Proc. Natl. Acad. Sci. USA* **116** 15849–15854. MR3997901 https://doi.org/10.1073/pnas.1903070116

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

BOUMAL, N. (2023). *An Introduction to Optimization on Smooth Manifolds*. Cambridge Univ. Press, Cambridge. MR4533407

BOUMAL, N. and ABSIL, P.-A. (2011). Rtrmc: A Riemannian trust-region method for low-rank matrix completion. In *Advances in Neural Information Processing Systems* 406–414.

BREIDING, P. and VANNIEUWENHOVEN, N. (2018). A Riemannian trust region method for the canonical tensor rank approximation problem. *SIAM J. Optim.* **28** 2435–2465. MR3852721 https://doi.org/10.1137/17M114618X

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

BRESLER, G. and HUANG, B. (2022). The algorithmic phase transition of random k-SAT for low degree polynomials. In 2021 *IEEE* 62*nd Annual Symposium on Foundations of Computer Science—FOCS* 2021 298–309. IEEE Comput. Soc., Los Alamitos, CA. MR4399691

CAI, J.-F., LI, J. and XIA, D. (2022). Provable tensor-train format tensor completion by Riemannian optimization. *J. Mach. Learn. Res.* **23** 5365–5441. MR4577075

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

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

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

CHOO, D. and D'ORSI, T. (2021). The complexity of sparse tensor pca. *Adv. Neural Inf. Process. Syst.* **34**.

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

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

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

DIAKONIKOLAS, I., KANE, D. M., LUO, Y. and ZHANG, A. (2023). Statistical and computational limits for tensor-on-tensor association detection. In *The Thirty Sixth Annual Conference on Learning Theory* 5260–5310. PMLR.

DING, L., JIANG, L., CHEN, Y., QU, Q. and ZHU, Z. (2021a). Rank overspecified robust matrix recovery: Subgradient method and exact recovery. *Adv. Neural Inf. Process. Syst.* **34** 26767–26778.

DING, Y., KUNISKY, D., WEIN, A. S. and BANDEIRA, A. S. (2021b). The average-case time complexity of certifying the restricted isometry property. *IEEE Trans. Inf. Theory* **67** 7355–7361. MR4345126 https://doi.org/10.1109/TIT.2021.3112823

DING, Y., KUNISKY, D., WEIN, A. S. and BANDEIRA, A. S. (2024). Subexponential-time algorithms for sparse PCA. *Found. Comput. Math.* **24** 865–914. MR4760356 https://doi.org/10.1007/s10208-023-09603-0

DONG, S., GAO, B., GUAN, Y. and GLINEUR, F. (2022). New Riemannian preconditioned algorithms for tensor completion via polyadic decomposition. *SIAM J. Matrix Anal. Appl.* **43** 840–866. MR4426891 https://doi.org/10.1137/21M1394734

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

ELDÉN, L. and SAVAS, B. (2009). A Newton–Grassmann method for computing the best multilinear rank-(r_1, r_2, r_3) approximation of a tensor. *SIAM J. Matrix Anal. Appl.* **31** 248–271. MR2496418 https://doi.org/10.1137/070688316

FAN, J., YANG, Z. and YU, M. (2023). Understanding implicit regularization in over-parameterized single index model. *J. Amer. Statist. Assoc.* **118** 2315–2328. MR4681585 https://doi.org/10.1080/01621459.2022.2044824

GAHROOEI, M. R., YAN, H., PAYNABAR, K. and SHI, J. (2021). Multiple tensor-on-tensor regression: An approach for modeling processes with heterogeneous sources of data. *Technometrics* **63** 147–159. MR4251490 https://doi.org/10.1080/00401706.2019.1708463

GE, R., REN, Y., WANG, X. and ZHOU, M. (2021). Understanding deflation process in over-parametrized tensor decomposition. *Adv. Neural Inf. Process. Syst.* **34**.

GUHANIYOGI, R., QAMAR, S. and DUNSON, D. B. (2017). Bayesian tensor regression. *J. Mach. Learn. Res.* **18** Paper No. 79, 31. MR3714242

GUNASEKAR, S., WOODWORTH, B. E., BHOJANAPALLI, S., NEYSHABUR, B. and SREBRO, N. (2017). Implicit regularization in matrix factorization. *Adv. Neural Inf. Process. Syst.* **30**.

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

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

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

HEIDEL, G. and SCHULZ, V. (2018). A Riemannian trust-region method for low-rank tensor completion. *Numer. Linear Algebra Appl.* **25** e2175, 16. MR3890978 https://doi.org/10.1002/nla.2175

HILLAR, C. J. and LIM, L.-H. (2013). Most tensor problems are NP-hard. *J. ACM* **60** Art. 45, 39. MR3144915 https://doi.org/10.1145/2512329

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

HOPKINS, S. (2018). Statistical inference and the sum of squares method. Ph.D. thesis. MR3864930

HOPKINS, S. B., KOTHARI, P. K., POTECHIN, A., RAGHAVENDRA, P., SCHRAMM, T. and STEURER, D. (2017). The power of sum-of-squares for detecting hidden structures. In 58*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 720–731. IEEE Comput. Soc., Los Alamitos, CA. MR3734275 https://doi.org/10.1109/FOCS.2017.72

HOPKINS, S. B. and STEURER, D. (2017). Efficient Bayesian estimation from few samples: Community detection and related problems. In 58*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2017 379–390. IEEE Comput. Soc., Los Alamitos, CA. MR3734245 https://doi.org/10.1109/FOCS.2017.42

HOU, T. Y., LI, Z. and ZHANG, Z. (2020). Fast global convergence for low-rank matrix recovery via Riemannian gradient descent with random initialization. ArXiv preprint. Available at arXiv:2012.15467.

HUANG, W. and HAND, P. (2018). Blind deconvolution by a steepest descent algorithm on a quotient manifold. *SIAM J. Imaging Sci.* **11** 2757–2785. MR3882949 https://doi.org/10.1137/17M1151390

ISHTEVA, M., DE LATHAUWER, L., ABSIL, P.-A. and VAN HUFFEL, S. (2009). Differential-geometric Newton method for the best rank-(R_1, R_2, R_3) approximation of tensors. *Numer. Algorithms* **51** 179–194. MR2505840 https://doi.org/10.1007/s11075-008-9251-2

JIANG, L., CHEN, Y. and DING, L. (2023). Algorithmic regularization in model-free overparametrized asymmetric matrix factorization. *SIAM J. Math. Data Sci.* **5** 723–744. MR4626339 https://doi.org/10.1137/22M1519833

KASAI, H. and MISHRA, B. (2016). Low-rank tensor completion: A Riemannian manifold preconditioning approach. In *International Conference on Machine Learning* 1012–1021. PMLR.

KESHAVAN, R. H., MONTANARI, A. and OH, S. (2010). Matrix completion from a few entries. *IEEE Trans. Inf. Theory* **56** 2980–2998. MR2683452 https://doi.org/10.1109/TIT.2010.2046205

KOCH, O. and LUBICH, C. (2010). Dynamical tensor approximation. *SIAM J. Matrix Anal. Appl.* **31** 2360–2375. MR2685162 https://doi.org/10.1137/09076578X

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

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

KRESSNER, D., STEINLECHNER, M. and VANDEREYCKEN, B. (2016). Preconditioned low-rank Riemannian optimization for linear systems with tensor product structure. *SIAM J. Sci. Comput.* **38** A2018–A2044. MR3519141 https://doi.org/10.1137/15M1032909

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

LEVIN, E., KILEEL, J. and BOUMAL, N. (2023). Finding stationary points on bounded-rank matrices: A geometric hurdle and a smooth remedy. *Math. Program.* **199** 831–864. MR4578384 https://doi.org/10.1007/s10107-022-01851-2

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

LI, Y., MA, T. and ZHANG, H. (2018). Algorithmic regularization in over-parameterized matrix sensing and neural networks with quadratic activations. In *Conference on Learning Theory* 2–47. PMLR.

LI, Z., LUO, Y. and LYU, K. (2020). Towards resolving the implicit bias of gradient descent for matrix factorization: Greedy low-rank learning. In *International Conference on Learning Representations*.

LIU, Y., LIU, J. and ZHU, C. (2020). Low-rank tensor train coefficient array estimation for tensor-on-tensor regression. *IEEE Trans. Neural Netw. Learn. Syst.* **31** 5402–5411. MR4189257 https://doi.org/10.1109/tnnls.2020.2967022

LLOSA, C. and MAITRA, R. (2022). Reduced-rank tensor-on-tensor regression and tensor-variate analysis of variance. *IEEE Trans. Pattern Anal. Mach. Intell.*

LOCK, E. F. (2018). Tensor-on-tensor regression. *J. Comput. Graph. Statist.* **27** 638–647. MR3863764 https://doi.org/10.1080/10618600.2017.1401544

LÖFFLER, M., WEIN, A. S. and BANDEIRA, A. S. (2022). Computationally efficient sparse clustering. *Inf. Inference* **11** 1255–1286. MR4526323 https://doi.org/10.1093/imaiai/iaac019

LUO, Y., HUANG, W., LI, X. and ZHANG, A. (2024). Recursive importance sketching for rank constrained least squares: Algorithms and high-order convergence. *Oper. Res.* **72** 237–256. MR4705836

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

LUO, Y. and ZHANG, A. R. (2023). Low-rank tensor estimation via Riemannian Gauss–Newton: Statistical optimality and second-order convergence. *J. Mach. Learn. Res.* **24** Paper No. 381, 48. MR4720837 https://doi.org/10.59277/pra-ser.a.24.4.09

LUO, Y. and ZHANG, A. R. (2024). Supplement to "Tensor-on-Tensor Regression: Riemannian Optimization, over-parameterization, Statistical-computational Gap, and Their Interplay." https://doi.org/10.1214/24-AOS2396SUPP

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

MA, J. and FATTAHI, S. (2023). Global convergence of sub-gradient method for robust matrix recovery: Small initialization, noisy measurements, and over-parameterization. *J. Mach. Learn. Res.* **24** Paper No. [96], 84. MR4582518

MAO, C. and WEIN, A. S. (2021). Optimal spectral recovery of a planted vector in a subspace. ArXiv preprint. Available at arXiv:2105.15081.

MEYER, G., BONNABEL, S. and SEPULCHRE, R. (2011). Linear regression under fixed-rank constraints: A Riemannian approach. In *Proceedings of the* 28*th International Conference on Machine Learning*.

MISHRA, B., MEYER, G., BONNABEL, S. and SEPULCHRE, R. (2014). Fixed-rank matrix factorizations and Riemannian low-rank optimization. *Comput. Statist.* **29** 591–621. MR3261830 https://doi.org/10.1007/s00180-013-0464-z

MU, C., HUANG, B., WRIGHT, J. and GOLDFARB, D. (2014). Square deal: Lower bounds and improved relaxations for tensor recovery. In *ICML* 73–81.

OLIKIER, G. and ABSIL, P.-A. (2023). An apocalypse-free first-order low-rank optimization algorithm with at most one rank reduction attempt per iteration. *SIAM J. Matrix Anal. Appl.* **44** 1421–1435. MR4644394 https://doi.org/10.1137/22M1518256

RABUSSEAU, G. and KADRI, H. (2016). Low-rank regression with tensor responses. *Adv. Neural Inf. Process. Syst.* **29**.

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

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

RAZIN, N., MAMAN, A. and COHEN, N. (2021). Implicit regularization in tensor factorization. In *International Conference on Machine Learning* 8913–8924. PMLR.

RECHT, B., FAZEL, M. and PARRILO, P. A. (2010). Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.* **52** 471–501. MR2680543 https://doi.org/10.1137/070697835

RICHARD, E. and MONTANARI, A. (2014). A statistical model for tensor pca. *Adv. Neural Inf. Process. Syst.* 2897–2905.

SAVAS, B. and LIM, L.-H. (2010). Quasi-Newton methods on Grassmannians and multilinear approximations of tensors. *SIAM J. Sci. Comput.* **32** 3352–3393. MR2746624 https://doi.org/10.1137/090763172

SCHNEIDER, R. and USCHMAJEW, A. (2015). Convergence results for projected line-search methods on varieties of low-rank matrices via Łojasiewicz inequality. *SIAM J. Optim.* **25** 622–646. MR3323551 https://doi.org/10.1137/140957822

SOLTANOLKOTABI, M., JAVANMARD, A. and LEE, J. D. (2019). Theoretical insights into the optimization landscape of over-parameterized shallow neural networks. *IEEE Trans. Inf. Theory* **65** 742–769. MR3904911 https://doi.org/10.1109/TIT.2018.2854560

SORBER, L., VAN BAREL, M. and DE LATHAUWER, L. (2013). Optimization-based algorithms for tensor decompositions: Canonical polyadic decomposition, decomposition in rank-$(L_r, L_r, 1)$ terms, and a new generalization. *SIAM J. Optim.* **23** 695–720. MR3044107 https://doi.org/10.1137/120868323

STEINLECHNER, M. (2016). Riemannian optimization for high-dimensional tensor completion. *SIAM J. Sci. Comput.* **38** S461–S484. MR3565572 https://doi.org/10.1137/15M1010506

STÖGER, D. and SOLTANOLKOTABI, M. (2021). Small random initialization is akin to spectral learning: Optimization and generalization guarantees for overparameterized low-rank matrix reconstruction. *Adv. Neural Inf. Process. Syst.* **34**.

SUN, W. W. and LI, L. (2017). STORE: Sparse tensor response regression and neuroimaging analysis. *J. Mach. Learn. Res.* **18** Paper No. 135, 37. MR3763769

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

TUCKER, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika* **31** 279–311. MR0205395 https://doi.org/10.1007/BF02289464

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

VANDEREYCKEN, B. (2013). Low-rank matrix completion by Riemannian optimization. *SIAM J. Optim.* **23** 1214–1236. MR3069099 https://doi.org/10.1137/110845768

VANNIEUWENHOVEN, N., VANDEBRIL, R. and MEERBERGEN, K. (2012). A new truncation strategy for the higher-order singular value decomposition. *SIAM J. Sci. Comput.* **34** A1027–A1052. MR2914314 https://doi.org/10.1137/110836067

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

YU, R. and LIU, Y. (2016). Learning from multiway data: Simple and efficient tensor regression. In *International Conference on Machine Learning* 373–381. PMLR.

WEI, K., CAI, J.-F., CHAN, T. F. and LEUNG, S. (2016). Guarantees of Riemannian optimization for low rank matrix recovery. *SIAM J. Matrix Anal. Appl.* **37** 1198–1222. MR3543156 https://doi.org/10.1137/15M1050525

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

XIA, D., 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

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

ZHANG, J., FATTAHI, S. and ZHANG, R. (2021). Preconditioned gradient descent for over-parameterized nonconvex matrix factorization. *Adv. Neural Inf. Process. Syst.* **34**.

ZHENG, Q. and LAFFERTY, J. (2015). A convergent gradient descent algorithm for rank minimization and semidefinite programming from random linear measurements. In *Advances in Neural Information Processing Systems* 109–117.

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

ZHUO, J., KWON, J., HO, N. and CARAMANIS, C. (2024). On the computational and statistical complexity of over-parameterized matrix sensing. *J. Mach. Learn. Res.* **25** Paper No. [169], 47. MR4777411

The Annals of Statistics
2024, Vol. 52, No. 6, 2613–2640
https://doi.org/10.1214/24-AOS2408

TIME-UNIFORM CENTRAL LIMIT THEORY AND ASYMPTOTIC CONFIDENCE SEQUENCES

BY IAN WAUDBY-SMITH[1,a], DAVID ARBOUR[2,d], RITWIK SINHA[2,e], EDWARD H. KENNEDY[1,b] AND AADITYA RAMDAS[1,c]

[1]*Department of Statistics & Data Science, Carnegie Mellon University,* [a]*ian@ianws.com,* [b]*edward@stat.cmu.edu,* [c]*aramdas@cmu.edu*

[2]*Adobe Research, Adobe Inc.,* [d]*darbour26@gmail.com,* [e]*ritwik.sinha@gmail.com*

Confidence intervals based on the central limit theorem (CLT) are a cornerstone of classical statistics. Despite being only asymptotically valid, they are ubiquitous because they permit statistical inference under weak assumptions and can often be applied to problems even when nonasymptotic inference is impossible. This paper introduces time-uniform analogues of such asymptotic confidence intervals, adding to the literature on confidence sequences (CS)—sequences of confidence intervals that are uniformly valid over time—which provide valid inference at arbitrary stopping times and incur no penalties for "peeking" at the data, unlike classical confidence intervals which require the sample size to be fixed in advance. Existing CSs in the literature are nonasymptotic, enjoying finite-sample guarantees but not the aforementioned broad applicability of asymptotic confidence intervals. This work provides a definition for "asymptotic CSs" and a general recipe for deriving them. Asymptotic CSs forgo nonasymptotic validity for CLT-like versatility and (asymptotic) time-uniform guarantees. While the CLT approximates the distribution of a sample average by that of a Gaussian for a fixed sample size, we use strong invariance principles (stemming from the seminal 1960s work of Strassen) to uniformly approximate the entire sample average process by an implicit Gaussian process. As an illustration, we derive asymptotic CSs for the average treatment effect in observational studies (for which nonasymptotic bounds are essentially impossible to derive even in the fixed-time regime) as well as randomized experiments, enabling causal inference in sequential environments.

REFERENCES

[1] BAHADUR, R. R. and SAVAGE, L. J. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *Ann. Math. Stat.* **27** 1115–1122. MR0084241 https://doi.org/10.1214/aoms/1177728077

[2] BIBAUT, A., KALLUS, N. and LINDON, M. (2022). Near-optimal non-parametric sequential tests and confidence sequences with possibly dependent observations. arXiv preprint. Available at arXiv:2212.14411.

[3] BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd ed. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. A Wiley-Interscience Publication. MR1324786

[4] BREIMAN, L. (1996). Stacked regressions. *Mach. Learn.* **24** 49–64.

[5] CHATTERJEE, S. (2012). A new approach to strong embeddings. *Probab. Theory Related Fields* **152** 231–264. MR2875758 https://doi.org/10.1007/s00440-010-0321-8

[6] 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

[7] DARLING, D. A. and ROBBINS, H. (1967). Confidence sequences for mean, variance, and median. *Proc. Natl. Acad. Sci. USA* **58** 66–68. MR0215406 https://doi.org/10.1073/pnas.58.1.66

[8] DÍAZ MUÑOZ, I. and VAN DER LAAN, M. (2012). Population intervention causal effects based on stochastic interventions. *Biometrics* **68** 541–549. MR2959621 https://doi.org/10.1111/j.1541-0420.2011.01685.x

MSC2020 subject classifications. Primary 62L10, 62G20; secondary 60F17, 62D20.
Key words and phrases. Sequential, nonparametric inference, strong approximation, observational study.

[9] EFRON, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* **58** 403–417. MR0312660 https://doi.org/10.1093/biomet/58.3.403

[10] EINMAHL, U. (2009). A new strong invariance principle for sums of independent random vectors. *J. Math. Sci.* **163** 311–327. MR2749123 https://doi.org/10.1007/s10958-009-9676-8

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

[12] HOWARD, S. R. and RAMDAS, A. (2022). Sequential estimation of quantiles with applications to A/B testing and best-arm identification. *Bernoulli* **28** 1704–1728. MR4411508 https://doi.org/10.3150/21-bej1388

[13] HOWARD, S. R., RAMDAS, A., MCAULIFFE, J. and SEKHON, J. (2020). Time-uniform Chernoff bounds via nonnegative supermartingales. *Probab. Surv.* **17** 257–317. MR4100718 https://doi.org/10.1214/18-PS321

[14] HOWARD, S. R., RAMDAS, A., MCAULIFFE, J. and SEKHON, J. (2021). Time-uniform, nonparametric, nonasymptotic confidence sequences. *Ann. Statist.* **49** 1055–1080. MR4255119 https://doi.org/10.1214/20-aos1991

[15] JOHARI, R., KOOMEN, P., PEKELIS, L. and WALSH, D. (2017). Peeking at A/B tests: Why it matters, and what to do about it. In *Proceedings of the* 23*rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 1517–1525.

[16] JOHNSON, A. and POLLARD, T. (2018). sepsis3–mimic. https://doi.org/10.5281/zenodo.1256723

[17] JOHNSON, A. E., POLLARD, T. J., SHEN, L., LEHMAN, L. H., FENG, M., GHASSEMI, M., MOODY, B., SZOLOVITS, P., CELI, L. A. et al. (2016). MIMIC-III, a freely accessible critical care database. *Sci. Data* **3** 160035.

[18] KENNEDY, E. H. (2016). Semiparametric theory and empirical processes in causal inference. In *Statistical Causal Inferences and Their Applications in Public Health Research*. *ICSA Book Ser. Stat.* 141–167. Springer, Berlin. MR3617956

[19] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131. MR0375412 https://doi.org/10.1007/BF00533093

[20] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrsch. Verw. Gebiete* **34** 33–58. MR0402883 https://doi.org/10.1007/BF00532688

[21] LAI, T. L. (1976). Boundary crossing probabilities for sample sums and confidence sequences. *Ann. Probab.* **4** 299–312. MR0405578 https://doi.org/10.1214/aop/1176996135

[22] LAI, T. L. (1976). On confidence sequences. *Ann. Statist.* **4** 265–280. MR0395103

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

[24] MAJOR, P. (1976). The approximation of partial sums of independent RV's. *Z. Wahrsch. Verw. Gebiete* **35** 213–220. MR0415743 https://doi.org/10.1007/BF00532673

[25] MARCINKIEWICZ, J. and ZYGMUND, A. (1937). Sur les fonctions indépendantes. *Fund. Math.* **29** 60–90.

[26] MORROW, G. and PHILIPP, W. (1982). An almost sure invariance principle for Hilbert space valued martingales. *Trans. Amer. Math. Soc.* **273** 231–251. MR0664040 https://doi.org/10.2307/1999203

[27] PACE, L. and SALVAN, A. (2020). Likelihood, replicability and Robbins' confidence sequences. *Int. Stat. Rev.* **88** 599–615. MR4180669 https://doi.org/10.1111/insr.12355

[28] POLLARD, T. J. and JOHNSON, A. E. (2016). The MIMIC-III Clinical Database. https://doi.org/10.13026/C2XW26

[29] POLLEY, E. C. and VAN DER LAAN, M. J. (2010). *Super Learner in Prediction*. *U.C. Berkeley Division of Biostatistics Working Paper Series* **222**.

[30] ROBBINS, H. (1970). Statistical methods related to the law of the iterated logarithm. *Ann. Math. Stat.* **41** 1397–1409. MR0277063 https://doi.org/10.1214/aoms/1177696786

[31] ROBBINS, H. and SIEGMUND, D. (1970). Boundary crossing probabilities for the Wiener process and sample sums. *Ann. Math. Stat.* **41** 1410–1429. MR0277059 https://doi.org/10.1214/aoms/1177696787

[32] ROBINS, J., LI, L., TCHETGEN, E. and VAN DER VAART, A. (2008). Higher order influence functions and minimax estimation of nonlinear functionals. In *Probability and Statistics*: *Essays in Honor of David A. Freedman*. *Inst. Math. Stat.* (*IMS*) *Collect.* **2** 335–421. IMS, Beachwood, OH. MR2459958 https://doi.org/10.1214/193940307000000527

[33] 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

[34] ROTNITZKY, A., ROBINS, J. M. and SCHARFSTEIN, D. O. (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *J. Amer. Statist. Assoc.* **93** 1321–1339. MR1666631 https://doi.org/10.2307/2670049

[35] SCHARFSTEIN, D. O., ROTNITZKY, A. and ROBINS, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J. Amer. Statist. Assoc.* **94** 1096–1120. With comments and a rejoinder by the authors. MR1731478 https://doi.org/10.2307/2669923

[36] SHAHN, Z., SHAPIRO, N. I., TYLER, P. D., TALMOR, D. and LI-WEI, H. L. (2020). Fluid-limiting treatment strategies among sepsis patients in the ICU: A retrospective causal analysis. *J. Crit. Care* **24** 1–9.

[37] SINGER, M., DEUTSCHMAN, C. S., SEYMOUR, C. W., SHANKAR-HARI, M., ANNANE, D., BAUER, M., BELLOMO, R., BERNARD, G. R., CHICHE, J.-D. et al. (2016). The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* **315** 801–810.

[38] SKOROKHOD, A. (1961). *Research on the Theory of Random Processes*. Kiev Univ., Kiev.

[39] STRASSEN, V. (1964). An invariance principle for the law of the iterated logarithm. *Z. Wahrsch. Verw. Gebiete* **3** 211–226. MR0175194 https://doi.org/10.1007/BF00534910

[40] STRASSEN, V. (1967). Almost sure behavior of sums of independent random variables and martingales. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability* (*Berkeley, Calif.*, 1965/66), *Vol. II*: *Contributions to Probability Theory, Part* 1 315–343. Univ. California Press, Berkeley, CA. MR0214118

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

[42] TSYBAKOV, A. B. (2003). Optimal rates of aggregation. In *Learning Theory and Kernel Machines* 303–313. Springer, Berlin.

[43] VAN WALRAVEN, C., AUSTIN, P. C., JENNINGS, A., QUAN, H. and FORSTER, A. J. (2009). A modification of the elixhauser comorbidity measures into a point system for hospital death using administrative data. *Med. Care* **47** 626–633. https://doi.org/10.1097/MLR.0b013e31819432e5

[44] VAN DER LAAN, M. J., POLLEY, E. C. and HUBBARD, A. E. (2007). Super learner. *Stat. Appl. Genet. Mol. Biol.* **6** Art. 25, 23. MR2349918 https://doi.org/10.2202/1544-6115.1309

[45] VAN DER LAAN, M. J. and ROSE, S. (2011). *Targeted Learning*: *Causal Inference for Observational and Experimental Data*. *Springer Series in Statistics*. Springer, New York. MR2867111 https://doi.org/10.1007/978-1-4419-9782-1

[46] VAN DER VAART, A. W., DUDOIT, S. and VAN DER LAAN, M. J. (2006). Oracle inequalities for multi-fold cross validation. *Statist. Decisions* **24** 351–371. MR2305112 https://doi.org/10.1524/stnd.2006.24.3.351

[47] WANG, H. and RAMDAS, A. (2023). Catoni-style confidence sequences for heavy-tailed mean estimation. *Stochastic Process. Appl.* **163** 168–202. MR4610125 https://doi.org/10.1016/j.spa.2023.05.007

[48] 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

[49] WAUDBY-SMITH, I., ARBOUR, D., SINHA, R., KENNEDY, E. H. and RAMDAS, A. (2024). Supplement to "Time-uniform central limit theory and asymptotic confidence sequences." https://doi.org/10.1214/24-AOS2408SUPPA, https://doi.org/10.1214/24-AOS2408SUPPB

[50] WAUDBY-SMITH, I. and RAMDAS, A. (2024). Estimating means of bounded random variables by betting. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **86** 1–27. MR4716192 https://doi.org/10.1093/jrsssb/qkad009

[51] YU, M., LU, W. and SONG, R. (2020). A new framework for online testing of heterogeneous treatment effect. In *Proceedings of the AAAI Conference on Artificial Intelligence*, *Vol.* 34 10310–10317.

[52] ZHENG, W. and VAN DER LAAN, M. J. (2010). *Asymptotic Theory for Cross-Validated Targeted Maximum Likelihood Estimation*. *U.C. Berkeley Division of Biostatistics Working Paper Series* **273**.

The Annals of Statistics
2024, Vol. 52, No. 6, 2641–2667
https://doi.org/10.1214/24-AOS2412

TENSOR FACTOR MODEL ESTIMATION BY ITERATIVE PROJECTION

BY YUEFENG HAN[1,a], RONG CHEN[2,b], DAN YANG[3,d] AND CUN-HUI ZHANG[2,c]

[1]*Department of Applied and Computational Mathematics and Statistics, University of Notre Dame,* [a]*yuefeng.han@nd.edu*
[2]*Department of Statistics, Rutgers University,* [b]*rongchen@stat.rutgers.edu,* [c]*czhang@stat.rutgers.edu*
[3]*Faculty of Business and Economics, The University of Hong Kong,* [d]*dyanghku@hku.hk*

Tensor time series, which is a time series consisting of tensorial observations, has become ubiquitous. It typically exhibits high dimensionality. One approach for dimension reduction is to use a factor model structure, in a form similar to Tucker tensor decomposition, except that the time dimension is treated as a dynamic process with a time dependent structure. In this paper, we introduce two approaches to estimate such a tensor factor model by using iterative orthogonal projections of the original tensor time series. These approaches extend the existing estimation procedures and improve the estimation accuracy and convergence rate significantly as proven in our theoretical investigation. Our algorithms are similar to the higher-order orthogonal projection method for tensor decomposition, but with significant differences due to the need to unfold tensors in the iterations and the use of autocorrelation. Consequently, our analysis is significantly different from the existing ones. Computational and statistical lower bounds are derived to prove the optimality of the sample size requirement and convergence rate for the proposed methods. Simulation study is conducted to further illustrate the statistical properties of these estimators.

REFERENCES

AHN, S. C. and HORENSTEIN, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica* **81** 1203–1227. MR3064065 https://doi.org/10.3982/ECTA8968

ALON, N., KRIVELEVICH, M. and SUDAKOV, B. (1998). Finding a large hidden clique in a random graph. *Random Structures Algorithms* **13** 457–466.

ALTER, O. and GOLUB, G. H. (2005). Reconstructing the pathways of a cellular system from genome-scale signals by using matrix and tensor computations. *Proc. Natl. Acad. Sci. USA* **102** 17559–17564.

AMES, B. P. W. and VAVASIS, S. A. (2011). Nuclear norm minimization for the planted clique and biclique problems. *Math. Program.* **129** 69–89. MR2831403 https://doi.org/10.1007/s10107-011-0459-x

ANANDKUMAR, A., GE, R., HSU, D. and KAKADE, S. M. (2014). A tensor approach to learning mixed membership community models. *J. Mach. Learn. Res.* **15** 2239–2312. MR3231594

BAI, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* **71** 135–171. MR1956857 https://doi.org/10.1111/1468-0262.00392

BAI, J. and NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70** 191–221. MR1926259 https://doi.org/10.1111/1468-0262.00273

BAI, J. and NG, S. (2007). Determining the number of primitive shocks in factor models. *J. Bus. Econom. Statist.* **25** 52–60. MR2338870 https://doi.org/10.1198/073500106000000413

BERTHET, Q. and RIGOLLET, P. (2013a). Complexity theoretic lower bounds for sparse principal component detection. In *Conference on Learning Theory* 1046–1066. PMLR.

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

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

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

MSC2020 subject classifications. Primary 62H25, 62H12; secondary 62R07.
Key words and phrases. High-dimensional tensor data, factor model, orthogonal projection, time series, Tucker decomposition.

Cai, T. T., Liang, T. and Rakhlin, A. (2017). Computational and statistical boundaries for submatrix localization in a large noisy matrix. *Ann. Statist.* **45** 1403–1430. MR3670183 https://doi.org/10.1214/16-AOS1488

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

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica* **51** 1281–1304. MR0736050 https://doi.org/10.2307/1912275

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. and Fan, J. (2023). Statistical inference for high-dimensional matrix-variate factor models. *J. Amer. Statist. Assoc.* **118** 1038–1055. MR4595475 https://doi.org/10.1080/01621459.2021.1970569

Chen, 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, E. Y., Xia, D., Cai, C. and Fan, J. (2024). Semi-parametric tensor factor analysis by iteratively projected singular value decomposition. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **86** 793–823. MR4792086 https://doi.org/10.1093/jrsssb/qkae001

Chen, R., Yang, D. and Zhang, C.-H. (2022a). Factor models for high-dimensional tensor time series. *J. Amer. Statist. Assoc.* **117** 94–116. MR4399070 https://doi.org/10.1080/01621459.2021.1912757

Chen, R., Yang, D. and Zhang, C.-H. (2022b). Rejoinder. *J. Amer. Statist. Assoc.* **117** 128–132.

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

Dekel, Y., Gurel-Gurevich, O. and Peres, Y. (2014). Finding hidden cliques in linear time with high probability. *Combin. Probab. Comput.* **23** 29–49. MR3197965 https://doi.org/10.1017/S096354831300045X

Deshpande, Y. and Montanari, A. (2015). Finding hidden cliques of size $\sqrt{N/e}$ in nearly linear time. *Found. Comput. Math.* **15** 1069–1128. MR3371378 https://doi.org/10.1007/s10208-014-9215-y

Fan, J., Liao, Y. and Mincheva, M. (2011). High-dimensional covariance matrix estimation in approximate factor models. *Ann. Statist.* **39** 3320–3356. MR3012410 https://doi.org/10.1214/11-AOS944

Fan, J., Liao, Y. and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 603–680. MR3091653 https://doi.org/10.1111/rssb.12016

Fan, J., Liao, Y. and Wang, W. (2016). Projected principal component analysis in factor models. *Ann. Statist.* **44** 219–254. MR3449767 https://doi.org/10.1214/15-AOS1364

Fan, J., Liu, H. and Wang, W. (2018). Large covariance estimation through elliptical factor models. *Ann. Statist.* **46** 1383–1414. MR3819104 https://doi.org/10.1214/17-AOS1588

Feige, U. and Krauthgamer, R. (2000). Finding and certifying a large hidden clique in a semirandom graph. *Random Structures Algorithms* **16** 195–208. MR1742351 https://doi.org/10.1002/(SICI)1098-2418(200003)16:2<195::AID-RSA5>3.3.CO;2-1

Feige, U. and Ron, D. (2010). Finding hidden cliques in linear time. In 21*st International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithms* (*AofA*'10). *Discrete Math. Theor. Comput. Sci. Proc., AM* 189–203. Assoc. Discrete Math. Theor. Comput. Sci., Nancy. MR2735341

Feldman, V., Grigorescu, E., Reyzin, L., Vempala, S. S. and Xiao, Y. (2017). Statistical algorithms and a lower bound for detecting planted cliques. *J. ACM* **64** 8. MR3664576 https://doi.org/10.1145/3046674

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Rev. Econ. Stat.* **82** 540–554.

Foster, G. (1996). Time series analysis by projection. II. Tensor methods for time series analysis. *Astron. J.* **111** 555.

Gao, C., Ma, Z. and Zhou, H. H. (2017). Sparse CCA: Adaptive estimation and computational barriers. *Ann. Statist.* **45** 2074–2101. MR3718162 https://doi.org/10.1214/16-AOS1519

Hajek, B., Wu, Y. and Xu, J. (2015). Computational lower bounds for community detection on random graphs. In *Conference on Learning Theory* 899–928. PMLR.

Hallin, M. and Liška, R. (2007). Determining the number of factors in the general dynamic factor model. *J. Amer. Statist. Assoc.* **102** 603–617. MR2325115 https://doi.org/10.1198/016214506000001275

Han, Y., Chen, R., Yang, D. and Zhang, C.-H (2024). Supplement to "Tensor Factor Model Estimation by Iterative Projection." https://doi.org/10.1214/24-AOS2412SUPP

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

HAN, Y., ZHANG, C.-H. and CHEN, R. (2021). CP factor model for dynamic tensors. ArXiv Preprint. Available at arXiv:2110.15517.

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

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.

LIU, Y., SHANG, F., FAN, W., CHENG, J. and CHENG, H. (2014). Generalized higher-order orthogonal iteration for tensor decomposition and completion. In *Advances in Neural Information Processing Systems* 1763–1771.

LUO, Y. and ZHANG, A. R. (2020). Open problem: Average-case hardness of hypergraphic planted clique detection. In *Conference on Learning Theory* 3852–3856. PMLR.

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

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

OMBERG, L., GOLUB, G. H. and ALTER, O. (2007). A tensor higher-order singular value decomposition for integrative analysis of DNA microarray data from different studies. *Proc. Natl. Acad. Sci. USA* **104** 18371–18376.

OUYANG, J. and YUAN, M. (2022). Comments on “Factor models for high-dimensional tensor time series” [4399070]. *J. Amer. Statist. Assoc.* **117** 124–127. MR4399073 https://doi.org/10.1080/01621459.2022.2028630

PAN, J. and YAO, Q. (2008). Modelling multiple time series via common factors. *Biometrika* **95** 365–379. MR2521589 https://doi.org/10.1093/biomet/asn009

PANANJADY, A. and SAMWORTH, R. J. (2022). Isotonic regression with unknown permutations: Statistics, computation and adaptation. *Ann. Statist.* **50** 324–350. MR4382019 https://doi.org/10.1214/21-aos2107

PEÑA, D. and BOX, G. E. P. (1987). Identifying a simplifying structure in time series. *J. Amer. Statist. Assoc.* **82** 836–843. MR0909990

ROGERS, M., LI, L. and RUSSELL, S. J. (2013). Multilinear dynamical systems for tensor time series. *Adv. Neural Inf. Process. Syst.* **26** 2634–2642.

SHEEHAN, B. N. and SAAD, Y. (2007). Higher order orthogonal iteration of tensors (HOOI) and its relation to PCA and GLRAM. In *Proceedings of the* 2007 *SIAM International Conference on Data Mining* 355–365. SIAM, Philadelphia.

STOCK, J. H. and WATSON, M. W. (2002). Forecasting using principal components from a large number of predictors. *J. Amer. Statist. Assoc.* **97** 1167–1179. MR1951271 https://doi.org/10.1198/016214502388618960

SUN, W. W. and LI, L. (2017). STORE: Sparse tensor response regression and neuroimaging analysis. *J. Mach. Learn. Res.* **18** 135. MR3763769

TUCKER, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika* **31** 279–311. MR0205395 https://doi.org/10.1007/BF02289464

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

WANG, D., ZHENG, Y. and LI, G. (2024). High-dimensional low-rank tensor autoregressive time series modeling. *J. Econometrics* **238** 105544. MR4663431 https://doi.org/10.1016/j.jeconom.2023.105544

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

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

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

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

The Annals of Statistics
2024, Vol. 52, No. 6, 2668–2691
https://doi.org/10.1214/24-AOS2417

STATISTICAL INFERENCE FOR FOUR-REGIME SEGMENTED REGRESSION MODELS

BY HAN YAN[1,a] AND SONG XI CHEN[2,b]

[1]*Guanghua School of Management, Peking University,* [a]*hanyan@stu.pku.edu.cn*
[2]*Department of Statistics and Data Science, Tsinghua University,* [b]*sxchen@tsinghua.edu.cn*

Segmented regression models offer model flexibility and interpretability as compared to the global parametric and the nonparametric models, and yet are challenging in both estimation and inference. We consider a four-regime segmented model for temporally dependent data with segmenting boundaries depending on multivariate covariates with nondiminishing boundary effects. A mixed integer quadratic programming algorithm is formulated to facilitate the least square estimation of the regression and the boundary parameters. The rates of convergence and the asymptotic distributions of the least square estimators are obtained for the regression and the boundary coefficients, respectively. We propose a smoothed regression bootstrap to facilitate inference on the parameters and a model selection procedure to select the most suitable model within the model class with at most four segments. Numerical simulations and a case study on air pollution in Beijing are conducted to demonstrate the proposed approach, which shows that the segmented models with three or four regimes are suitable for the modeling of the meteorological effects on the $PM_{2.5}$ concentration.

REFERENCES

[1] AUERBACH, A. J. and GORODNICHENKO, Y. (2012). Measuring the output responses to fiscal policy. *Amer. Econ. J.: Econ. Policy* **4** 1–27.

[2] BAI, J. and PERRON, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66** 47–78. MR1616121 https://doi.org/10.2307/2998540

[3] 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

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

[5] CARD, D., MAS, A. and ROTHSTEIN, J. (2008). Tipping and the dynamics of segregation. *Q. J. Econ.* **123** 177–218.

[6] CHAN, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Ann. Statist.* **21** 520–533. MR1212191 https://doi.org/10.1214/aos/1176349040

[7] CHERNOZHUKOV, V. and FERNÁNDEZ-VAL, I. (2011). Inference for extremal conditional quantile models, with an application to market and birthweight risks. *Rev. Econ. Stud.* **78** 559–589. MR2808129 https://doi.org/10.1093/restud/rdq020

[8] CHERNOZHUKOV, V. and HONG, H. (2004). Likelihood estimation and inference in a class of nonregular econometric models. *Econometrica* **72** 1445–1480. MR2077489 https://doi.org/10.1111/j.1468-0262.2004.00540.x

[9] DAVIES, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74** 33–43. MR0885917 https://doi.org/10.1093/biomet/74.1.33

[10] FAN, J. and YAO, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* **85** 645–660. MR1665822 https://doi.org/10.1093/biomet/85.3.645

[11] FRIEDMAN, J. H. (1991). Multivariate adaptive regression splines. *Ann. Statist.* **19** 1–141. MR1091842 https://doi.org/10.1214/aos/1176347963

[12] GONZALO, J. and PITARAKIS, J.-Y. (2002). Estimation and model selection based inference in single and multiple threshold models. *J. Econometrics* **110** 319–352.

MSC2020 subject classifications. Primary 62J99, 62H12; secondary 62F12.
Key words and phrases. Mixed integer programming, segmented model, smoothed regression bootstrap, temporal dependence, threshold regression.

[13] GONZALO, J. and WOLF, M. (2005). Subsampling inference in threshold autoregressive models. *J. Econometrics* **127** 201–224. MR2156333 https://doi.org/10.1016/j.jeconom.2004.08.004
[14] GYÖRFI, L., HÄRDLE, W., SARDA, P. and VIEU, P. (1989). *Nonparametric Curve Estimation from Time Series*. *Lecture Notes in Statistics* **60**. Springer, Berlin. MR1027837 https://doi.org/10.1007/978-1-4612-3686-3
[15] HANSEN, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* **64** 413–430. MR1375740 https://doi.org/10.2307/2171789
[16] HANSEN, B. E. (2000). Sample splitting and threshold estimation. *Econometrica* **68** 575–603. MR1769379 https://doi.org/10.1111/1468-0262.00124
[17] HÄRDLE, W., HOROWITZ, J. and KREISS, J.-P. (2003). Bootstrap methods for time series. *Int. Stat. Rev.* **71** 435–459.
[18] HSING, T. (1995). On the asymptotic independence of the sum and rare values of weakly dependent stationary random variables. *Stochastic Process. Appl.* **60** 49–63. MR1362318 https://doi.org/10.1016/0304-4149(95)00054-2
[19] JIANG, Z., DU, C., JABLENSKY, A., LIANG, H., LU, Z., MA, Y. and TEO, K. L. (2014). Analysis of schizophrenia data using a nonlinear threshold index logistic model. *PLoS ONE* **9** e109454.
[20] KHALILI, A. and CHEN, J. (2007). Variable selection in finite mixture of regression models. *J. Amer. Statist. Assoc.* **102** 1025–1038. MR2411662 https://doi.org/10.1198/016214507000000590
[21] KNIGHT, K. (1999). Epi-convergence and stochastic equisemicontinuity. Preprint.
[22] LEE, S., LIAO, Y., SEO, M. H. and SHIN, Y. (2021). Factor-driven two-regime regression. *Ann. Statist.* **49** 1656–1678. MR4298876 https://doi.org/10.1214/20-aos2017
[23] LI, D. and LING, S. (2012). On the least squares estimation of multiple-regime threshold autoregressive models. *J. Econometrics* **167** 240–253. MR2885449 https://doi.org/10.1016/j.jeconom.2011.11.006
[24] LIU, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models. *Ann. Statist.* **16** 1696–1708. MR0964947 https://doi.org/10.1214/aos/1176351062
[25] MEYER, R. M. (1973). A Poisson-type limit theorem for mixing sequences of dependent "rare" events. *Ann. Probab.* **1** 480–483. MR0350816 https://doi.org/10.1214/aop/1176996941
[26] 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
[27] POTTER, S. M. (1995). A nonlinear approach to US GNP. *J. Appl. Econometrics* **10** 109–125.
[28] RESNICK, S. I. (2008). *Extreme Values, Regular Variation and Point Processes*. *Springer Series in Operations Research and Financial Engineering*. Springer, New York. MR2364939
[29] SCHWARTZ, P. F., GENNINGS, C. and CHINCHILLI, V. M. (1995). Threshold models for combination data from reproductive and developmental experiments. *J. Amer. Statist. Assoc.* **90** 862–870.
[30] SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464. MR0468014
[31] SEIJO, E. and SEN, B. (2011). Change-point in stochastic design regression and the bootstrap. *Ann. Statist.* **39** 1580–1607. MR2850213 https://doi.org/10.1214/11-AOS874
[32] TONG, H. (1983). *Threshold Models in Nonlinear Time Series Analysis*. *Lecture Notes in Statistics* **21**. Springer, New York. MR0717388 https://doi.org/10.1007/978-1-4684-7888-4
[33] YAN, H. and CHEN, S. X. (2024). Supplement to "Statistical inference for four-regime segmented regression models." https://doi.org/10.1214/24-AOS2417SUPP
[34] YU, P. (2014). The bootstrap in threshold regression. *Econometric Theory* **30** 676–714. MR3205610 https://doi.org/10.1017/S0266466614000012
[35] YU, P. and FAN, X. (2021). Threshold regression with a threshold boundary. *J. Bus. Econom. Statist.* **39** 953–971. MR4319684 https://doi.org/10.1080/07350015.2020.1740712
[36] YU, P. and PHILLIPS, P. C. B. (2018). Threshold regression with endogeneity. *J. Econometrics* **203** 50–68. MR3758327 https://doi.org/10.1016/j.jeconom.2017.09.007
[37] ZEILEIS, A., HOTHORN, T. and HORNIK, K. (2008). Model-based recursive partitioning. *J. Comput. Graph. Statist.* **17** 492–514. MR2439970 https://doi.org/10.1198/106186008X319331

The Annals of Statistics
2024, Vol. 52, No. 6, 2692–2713
https://doi.org/10.1214/24-AOS2426

STEREOGRAPHIC MARKOV CHAIN MONTE CARLO

BY JUN YANG[1,a], KRZYSZTOF ŁATUSZYŃSKI[2,b] AND GARETH O. ROBERTS[2,c]

[1]*Department of Mathematical Sciences, University of Copenhagen,* [a]*jy@math.ku.dk*
[2]*Department of Statistics, University of Warwick,* [b]*k.g.latuszynski@warwick.ac.uk,* [c]*gareth.o.roberts@warwick.ac.uk*

High-dimensional distributions, especially those with heavy tails, are notoriously difficult for off-the-shelf MCMC samplers: the combination of unbounded state spaces, diminishing gradient information, and local moves results in empirically observed "stickiness" and poor theoretical mixing properties—lack of geometric ergodicity. In this paper, we introduce a new class of MCMC samplers that map the original high-dimensional problem in Euclidean space onto a sphere and remedy these notorious mixing problems. In particular, we develop random-walk Metropolis type algorithms as well as versions of the Bouncy Particle Sampler that are uniformly ergodic for a large class of light and heavy-tailed distributions and also empirically exhibit rapid convergence in high dimensions. In the best scenario, the proposed samplers can enjoy the "blessings of dimensionality" that the convergence is faster in higher dimensions.

REFERENCES

ANDRIEU, C., DOBSON, P. and WANG, A. Q. (2021). Subgeometric hypocoercivity for piecewise-deterministic Markov process Monte Carlo methods. *Electron. J. Probab.* **26** 1–26. MR4269208 https://doi.org/10.1214/21-EJP643

ANDRIEU, C., DURMUS, A., NÜSKEN, N. and ROUSSEL, J. (2021). Hypocoercivity of piecewise deterministic Markov process-Monte Carlo. *Ann. Appl. Probab.* **31** 2478–2517. MR4332703 https://doi.org/10.1214/20-aap1653

BERNARD, E. P., KRAUTH, W. and WILSON, D. B. (2009). Event-chain Monte Carlo algorithms for hard-sphere systems. *Phys. Rev. E* **80** 056704.

BIERKENS, J., FEARNHEAD, P. and ROBERTS, G. (2019). The zig-zag process and super-efficient sampling for Bayesian analysis of big data. *Ann. Statist.* **47** 1288–1320. MR3911113 https://doi.org/10.1214/18-AOS1715

BIERKENS, J., KAMATANI, K. and ROBERTS, G. O. (2022). High-dimensional scaling limits of piecewise deterministic sampling algorithms. *Ann. Appl. Probab.* **32** 3361–3407. MR4497848 https://doi.org/10.1214/21-aap1762

BOUCHARD-CÔTÉ, A., VOLLMER, S. J. and DOUCET, A. (2018). The bouncy particle sampler: A nonreversible rejection-free Markov chain Monte Carlo method. *J. Amer. Statist. Assoc.* **113** 855–867. MR3832232 https://doi.org/10.1080/01621459.2017.1294075

CHRISTENSEN, O. F., ROBERTS, G. O. and ROSENTHAL, J. S. (2005). Scaling limits for the transient phase of local Metropolis–Hastings algorithms. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 253–268. MR2137324 https://doi.org/10.1111/j.1467-9868.2005.00500.x

COTTER, S. L., ROBERTS, G. O., STUART, A. M. and WHITE, D. (2013). MCMC methods for functions: Modifying old algorithms to make them faster. *Statist. Sci.* **28** 424–446. MR3135540 https://doi.org/10.1214/13-STS421

COXETER, H. S. M. (1961). *Introduction to Geometry*. Wiley, New York. MR0123930

DAVIS, M. H. A. (1984). Piecewise-deterministic Markov processes: A general class of nondiffusion stochastic models. *J. Roy. Statist. Soc. Ser. B* **46** 353–388. MR0790622

DELIGIANNIDIS, G., BOUCHARD-CÔTÉ, A. and DOUCET, A. (2019). Exponential ergodicity of the bouncy particle sampler. *Ann. Statist.* **47** 1268–1287. MR3911112 https://doi.org/10.1214/18-AOS1714

DELIGIANNIDIS, G., PAULIN, D., BOUCHARD-CÔTÉ, A. and DOUCET, A. (2021). Randomized Hamiltonian Monte Carlo as scaling limit of the bouncy particle sampler and dimension-free convergence rates. *Ann. Appl. Probab.* **31** 2612–2662. MR4350970 https://doi.org/10.1214/20-aap1659

MSC2020 subject classifications. 60J20, 60J25, 65C05.
Key words and phrases. Random walk Metropolis, piecewise deterministic Markov processes, stereographic projection, uniform ergodicity, heavy tailed distributions, blessings of dimensionality.

DURMUS, A., GUILLIN, A. and MONMARCHÉ, P. (2020). Geometric ergodicity of the bouncy particle sampler. *Ann. Appl. Probab.* **30** 2069–2098. MR4149523 https://doi.org/10.1214/19-AAP1552

GIROLAMI, M. and CALDERHEAD, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 123–214. MR2814492 https://doi.org/10.1111/j.1467-9868.2010.00765.x

JARNER, S. F. and HANSEN, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.* **85** 341–361. MR1731030 https://doi.org/10.1016/S0304-4149(99)00082-4

JARNER, S. F. and ROBERTS, G. O. (2007). Convergence of heavy-tailed Monte Carlo Markov chain algorithms. *Scand. J. Stat.* **34** 781–815. MR2396939 https://doi.org/10.1111/j.1467-9469.2007.00557.x

JOHNSON, L. T. and GEYER, C. J. (2012). Variable transformation to obtain geometric ergodicity in the random-walk Metropolis algorithm. *Ann. Statist.* **40** 3050–3076. MR3097969 https://doi.org/10.1214/12-AOS1048

KAMATANI, K. (2018). Efficient strategy for the Markov chain Monte Carlo in high-dimension with heavy-tailed target probability distribution. *Bernoulli* **24** 3711–3750. MR3788187 https://doi.org/10.3150/17-BEJ976

LIE, H. C., RUDOLF, D., SPRUNGK, B. and SULLIVAN, T. J. (2023). Dimension-independent Markov chain Monte Carlo on the sphere. *Scand. J. Stat.* **50** 1818–1858. MR4677377 https://doi.org/10.1111/sjos.12653

MANGOUBI, O. and SMITH, A. (2018). Rapid mixing of geodesic walks on manifolds with positive curvature. *Ann. Appl. Probab.* **28** 2501–2543. MR3843835 https://doi.org/10.1214/17-AAP1365

MENGERSEN, K. L. and TWEEDIE, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24** 101–121. MR1389882 https://doi.org/10.1214/aos/1033066201

MIJATOVIĆ, A., MRAMOR, V. and URIBE BRAVO, G. (2018). Projections of spherical Brownian motion. *Electron. Commun. Probab.* **23** 1–12. MR3852266 https://doi.org/10.1214/18-ECP162

OLLIVIER, Y. (2009). Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.* **256** 810–864. MR2484937 https://doi.org/10.1016/j.jfa.2008.11.001

POMPE, E., HOLMES, C. and ŁATUSZYŃSKI, K. (2020). A framework for adaptive MCMC targeting multimodal distributions. *Ann. Statist.* **48** 2930–2952. MR4152629 https://doi.org/10.1214/19-AOS1916

ROBERTS, G. O., GELMAN, A. and GILKS, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* **7** 110–120. MR1428751 https://doi.org/10.1214/aoap/1034625254

ROBERTS, G. O. and ROSENTHAL, J. S. (2009). Examples of adaptive MCMC. *J. Comput. Graph. Statist.* **18** 349–367. MR2749836 https://doi.org/10.1198/jcgs.2009.06134

ROBERTS, G. O. and TWEEDIE, R. L. (1996a). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83** 95–110. MR1399158 https://doi.org/10.1093/biomet/83.1.95

ROBERTS, G. O. and TWEEDIE, R. L. (1996b). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** 341–363. MR1440273 https://doi.org/10.2307/3318418

SHERLOCK, C., FEARNHEAD, P. and ROBERTS, G. O. (2010). The random walk Metropolis: Linking theory and practice through a case study. *Statist. Sci.* **25** 172–190. MR2789988 https://doi.org/10.1214/10-STS327

VASDEKIS, G. and ROBERTS, G. O. (2022). A note on the polynomial ergodicity of the one-dimensional zig-zag process. *J. Appl. Probab.* **59** 895–903. MR4480086 https://doi.org/10.1017/jpr.2021.97

VASDEKIS, G. and ROBERTS, G. O. (2023). Speed up zig-zag. *Ann. Appl. Probab.* **33** 4693–4746. MR4674062 https://doi.org/10.1214/23-aap1930

YANG, J., ŁATUSZYŃSKI, K. and ROBERTS, G. O. (2024). Supplement to "Stereographic Markov chain Monte Carlo." https://doi.org/10.1214/24-AOS2426SUPP

YANG, J., ROBERTS, G. O. and ROSENTHAL, J. S. (2020). Optimal scaling of random-walk Metropolis algorithms on general target distributions. *Stochastic Process. Appl.* **130** 6094–6132. MR4140028 https://doi.org/10.1016/j.spa.2020.05.004

YANG, J. and ROSENTHAL, J. S. (2023). Complexity results for MCMC derived from quantitative bounds. *Ann. Appl. Probab.* **33** 1459–1500. MR4564431 https://doi.org/10.1214/22-aap1846

ZAPPA, E., HOLMES-CERFON, M. and GOODMAN, J. (2018). Monte Carlo on manifolds: Sampling densities and integrating functions. *Comm. Pure Appl. Math.* **71** 2609–2647. MR3869037 https://doi.org/10.1002/cpa.21783

The Annals of Statistics
2024, Vol. 52, No. 6, 2714–2737
https://doi.org/10.1214/24-AOS2429

SKEWED BERNSTEIN–VON MISES THEOREM AND SKEW-MODAL APPROXIMATIONS

BY DANIELE DURANTE[a], FRANCESCO POZZA[b] AND BOTOND SZABO[c]

Department of Decision Sciences and Institute for Data Science and Analytics, Bocconi University,
[a]*daniele.durante@unibocconi.it,* [b]*francesco.pozza2@unibocconi.it,* [c]*botond.szabo@unibocconi.it*

Gaussian deterministic approximations are routinely employed in Bayesian statistics to ease inference when the target posterior is intractable. While these approximations are justified, in asymptotic regimes, by Bernstein–von Mises type results, in practice the expected Gaussian behavior might poorly represent the actual shape of the target posterior, thus affecting approximation accuracy. Motivated by these considerations, we derive an improved class of closed-form and valid deterministic approximations of posterior distributions that arise from a novel treatment of a third-order version of the Laplace method yielding approximations within a tractable family of skew-symmetric distributions. Under general assumptions accounting for misspecified models and non-i.i.d. settings, such a family of approximations is shown to have a total variation distance from the target posterior whose convergence rate improves by at least one order of magnitude the one achieved by the Gaussian from the classical Bernstein–von Mises theorem. Specializing this result to the case of regular parametric models shows that the same accuracy improvement can be also established for the posterior expectation of polynomially bounded functions. Unlike available higher-order approximations based on, for example, Edgeworth expansions, our results prove that it is possible to derive closed-form and valid densities which provide a more accurate, yet similarly tractable, alternative to Gaussian approximations of the target posterior, while inheriting its limiting frequentist properties. We strengthen these arguments by developing a practical skew-modal approximation for both joint and marginal posteriors which preserves the guarantees of its theoretical counterpart by replacing the unknown model parameters with the corresponding maximum a posteriori estimate. Simulation studies and real-data applications confirm that our theoretical results closely match the empirical gains observed in practice.

REFERENCES

ANCESCHI, N., FASANO, A., DURANTE, D. and ZANELLA, G. (2023). Bayesian conjugacy in probit, tobit, multinomial probit and extensions: A review and new results. *J. Amer. Statist. Assoc.* **118** 1451–1469. MR4595508 https://doi.org/10.1080/01621459.2023.2169150

ARELLANO-VALLE, R. B. and AZZALINI, A. (2006). On the unification of families of skew-normal distributions. *Scand. J. Stat.* **33** 561–574. MR2298065 https://doi.org/10.1111/j.1467-9469.2006.00503.x

AZZALINI, A. and CAPITANIO, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 367–389. MR1983753 https://doi.org/10.1111/1467-9868.00391

BERNSTEIN, S. (1917). Theory of Probability, Moskow.

BICKEL, P. J. and KLEIJN, B. J. K. (2012). The semiparametric Bernstein–von Mises theorem. *Ann. Statist.* **40** 206–237. MR3013185 https://doi.org/10.1214/11-AOS921

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

BLEI, D. M., KUCUKELBIR, A. and MCAULIFFE, J. D. (2017). Variational inference: A review for statisticians. *J. Amer. Statist. Assoc.* **112** 859–877. MR3671776 https://doi.org/10.1080/01621459.2017.1285773

MSC2020 subject classifications. Primary 62F15; secondary 62E20, 62E17.
Key words and phrases. Bernstein–von Mises theorem, deterministic approximation, skew-symmetric distribution.

BOCHKINA, N. A. and GREEN, P. J. (2014). The Bernstein–von Mises theorem and nonregular models. *Ann. Statist.* **42** 1850–1878. MR3262470 https://doi.org/10.1214/14-AOS1239

BOUCHERON, S. and GASSIAT, E. (2009). A Bernstein–von Mises theorem for discrete probability distributions. *Electron. J. Stat.* **3** 114–148. MR2471588 https://doi.org/10.1214/08-EJS262

CASTILLO, I. and NICKL, R. (2014). On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. *Ann. Statist.* **42** 1941–1969. MR3262473 https://doi.org/10.1214/14-AOS1246

CASTILLO, I. and ROUSSEAU, J. (2015). A Bernstein–von Mises theorem for smooth functionals in semiparametric models. *Ann. Statist.* **43** 2353–2383. MR3405597 https://doi.org/10.1214/15-AOS1336

CHALLIS, E. and BARBER, D. (2012). Affine independent variational inference. *Adv. Neural Inf. Process. Syst.* **25** 1–9.

CHOPIN, N. and RIDGWAY, J. (2017). Leave Pima Indians alone: Binary regression as a benchmark for Bayesian computation. *Statist. Sci.* **32** 64–87. MR3634307 https://doi.org/10.1214/16-STS581

CONSONNI, G. and MARIN, J.-M. (2008). Mean-field variational approximate Bayesian inference for latent variable models. *Comput. Statist. Data Anal.* **52** 790–798. MR2418528 https://doi.org/10.1016/j.csda.2006.10.028

DEHAENE, G. and BARTHELMÉ, S. (2018). Expectation propagation in the large data limit. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 199–217. MR3744718 https://doi.org/10.1111/rssb.12241

DURANTE, D. (2019). Conjugate Bayes for probit regression via unified skew-normal distributions. *Biometrika* **106** 765–779. MR4031198 https://doi.org/10.1093/biomet/asz034

DURANTE, D., POZZA, F. and SZABO, B. (2024). Supplement to "Skewed Bernstein–von Mises theorem and skew-modal approximations." https://doi.org/10.1214/24-AOS2429SUPPA, https://doi.org/10.1214/24-AOS2429SUPPB

DURANTE, D. and RIGON, T. (2019). Conditionally conjugate mean-field variational Bayes for logistic models. *Statist. Sci.* **34** 472–485. MR4017524 https://doi.org/10.1214/19-STS712

FASANO, A. and DURANTE, D. (2022). A class of conjugate priors for multinomial probit models which includes the multivariate normal one. *J. Mach. Learn. Res.* **23** Paper No. [30], 26. MR4420755

FASANO, A., DURANTE, D. and ZANELLA, G. (2022). Scalable and accurate variational Bayes for high-dimensional binary regression models. *Biometrika* **109** 901–919. MR4519107 https://doi.org/10.1093/biomet/asac026

GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. and RUBIN, D. B. (2014). *Bayesian Data Analysis*, 3rd ed. *Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. MR3235677

JOHNSON, R. A. (1970). Asymptotic expansions associated with posterior distributions. *Ann. Math. Stat.* **41** 851–864. MR0263198 https://doi.org/10.1214/aoms/1177696963

KASPRZAK, M. J., GIORDANO, R. and BRODERICK, T. (2022). How good is your Gaussian approximation of the posterior? Finite-sample computable error bounds for a variety of useful divergences. Available at arXiv:2209.14992.

KASS, R. E., TIERNEY, L. and KADANE, J. B. (1990). The validity of posterior expansions based on Laplace's method. In *Bayesian and Likelihood Methods in Statistics and Econometrics: Essays in Honor of George A. Barnard* 473–487.

KATSEVICH, A. (2024). The Laplace approximation accuracy in high dimensions: a refined analysis and new skew adjustment. Available at arXiv:2306.07262.

KATSEVICH, A. and RIGOLLET, P. (2024). On the approximation accuracy of Gaussian variational inference. *Ann. Statist.* **52** 1384–1409. MR4804813 https://doi.org/10.1214/24-aos2393

KLEIJN, B. J. K. and VAN DER VAART, A. W. (2012). The Bernstein–Von-Mises theorem under misspecification. *Electron. J. Stat.* **6** 354–381. MR2988412 https://doi.org/10.1214/12-EJS675

KOERS, G., SZABO, B. and VAN DER VAART, A. (2023). Misspecified Bernstein–Von Mises theorem for hierarchical models. Available at arXiv:2308.07803.

KOLASSA, J. E. and KUFFNER, T. A. (2020). On the validity of the formal Edgeworth expansion for posterior densities. *Ann. Statist.* **48** 1940–1958. MR4134781 https://doi.org/10.1214/19-AOS1871

LAPLACE, P. S. (1810). *Théorie Analytique des Probabilités*, 3rd ed. Courcier, Paris.

LE CAM, L. and YANG, G. L. (1990). *Asymptotics in Statistics: Some Basic Concepts*. *Springer Series in Statistics*. Springer, New York. MR1066869 https://doi.org/10.1007/978-1-4684-0377-0

LECAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. *Univ. Calif. Publ. Stat.* **1** 277–329. MR0054913

MA, Y. and GENTON, M. G. (2004). Flexible class of skew-symmetric distributions. *Scand. J. Stat.* **31** 459–468. MR2087837 https://doi.org/10.1111/j.1467-9469.2004.03_007.x

MCCULLAGH, P. (2018). *Tensor Methods in Statistics*, 2nd ed. Dover Publications, New York.

MINKA, T. P. (2001). Expectation propagation for approximate Bayesian inference. *Proc. Uncertainty Artif. Intell.* **17** 362–369.

OPPER, M. and ARCHAMBEAU, C. (2009). The variational Gaussian approximation revisited. *Neural Comput.* **21** 786–792. MR2478318 https://doi.org/10.1162/neco.2008.08-07-592

PACE, L. and SALVAN, A. (1997). *Principles of Statistical Inference*: *From a Neo-Fisherian Perspective*. *Advanced Series on Statistical Science & Applied Probability* **4**. World Scientific, River Edge, NJ. MR1476674

PANOV, M. and SPOKOINY, V. (2015). Finite sample Bernstein–von Mises theorem for semiparametric problems. *Bayesian Anal*. **10** 665–710. MR3420819 https://doi.org/10.1214/14-BA926

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

RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **71** 319–392. MR2649602 https://doi.org/10.1111/j.1467-9868.2008.00700.x

SPOKOINY, V. (2025). Inexact Laplace approximation and the use of posterior mean in Bayesian inference. *Bayesian Anal*. **20** 1303–1330. MR4832249 https://doi.org/10.1214/23-BA1391

SPOKOINY, V. and PANOV, M. (2021). Accuracy of Gaussian approximation for high-dimensional posterior distribution. *Bernoulli* (in print).

TAN, L. S. (2024). Variational inference based on a subclass of closed skew normals. *J. Comput. Graph. Statist.* (in print).

TIERNEY, L. and KADANE, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc*. **81** 82–86. MR0830567

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

VEHTARI, A., GELMAN, A., SIVULA, T., JYLÄNKI, P., TRAN, D., SAHAI, S., BLOMSTEDT, P., CUNNINGHAM, J. P., SCHIMINOVICH, D. et al. (2020). Expectation propagation as a way of life: A framework for Bayesian inference on partitioned data. *J. Mach. Learn. Res*. **21** Paper No. 17, 53. MR4071200

VENABLES, W. N. and RIPLEY, B. D. (2002). *Modern Applied Statistics with S*, 4th ed. Springer, New York.

VON MISES, R. (1931). *Wahrscheinlichkeitsrechnung*. Springer, Berlin.

WANG, C. and BLEI, D. M. (2013). Variational inference in nonconjugate models. *J. Mach. Learn. Res*. **14** 1005–1031. MR3063617

WANG, J., BOYER, J. and GENTON, M. G. (2004). A skew-symmetric representation of multivariate distributions. *Statist. Sinica* **14** 1259–1270. MR2126352

WANG, Y. and BLEI, D. M. (2019). Frequentist consistency of variational Bayes. *J. Amer. Statist. Assoc*. **114** 1147–1161. MR4011769 https://doi.org/10.1080/01621459.2018.1473776

WENG, R. C. (2010). A Bayesian Edgeworth expansion by Stein's identity. *Bayesian Anal*. **5** 741–763. MR2740155 https://doi.org/10.1214/10-BA526

The Annals of Statistics
2024, Vol. 52, No. 6, 2738–2766
https://doi.org/10.1214/24-AOS2442

DEEP NEURAL NETWORKS FOR NONPARAMETRIC INTERACTION MODELS WITH DIVERGING DIMENSION

BY SOHOM BHATTACHARYA[1,a], JIANQING FAN[2,b] AND DEBARGHYA MUKHERJEE[3,c]

[1]*Department of Statistics, University of Florida,* [a]*bhattacharya.s@ufl.edu*
[2]*Department of Operations Research and Financial Engineering, Princeton University,* [b]*jqfan@princeton.edu*
[3]*Department of Mathematics and Statistics, Boston University,* [c]*mdeb@bu.edu*

Deep neural networks have achieved tremendous success due to their representation power and adaptation to low-dimensional structures. Their potential for estimating structured regression functions has been recently established in the literature. However, most of the studies require the input dimension to be fixed, and consequently, they ignore the effect of dimension on the rate of convergence and hamper their applications to modern big data with high dimensionality. In this paper, we bridge this gap by analyzing a k-way nonparametric interaction model in both growing dimension scenarios (d grows with n but at a slower rate) and in high dimension ($d \gtrsim n$). In the latter case, sparsity assumptions and associated regularization are required to obtain optimal convergence rates. A new challenge in diverging dimension setting is in calculation mean-square error; the covariance terms among estimated additive components are an order of magnitude larger than those of the variances and can deteriorate statistical properties without proper care. We introduce a critical debiasing technique to amend the problem. We show that under certain standard assumptions, debiased deep neural networks achieve a minimax optimal rate both in terms of (n, d). Our proof techniques rely crucially on a novel debiasing technique that makes the covariances of additive components negligible in the mean-square error calculation. In addition, we establish the matching lower bounds.

REFERENCES

[1] ANTHONY, M. and BARTLETT, P. L. (1999). *Neural Network Learning*: *Theoretical Foundations*. Cambridge Univ. Press, Cambridge. MR1741038 https://doi.org/10.1017/CBO9780511624216
[2] ANTONIADIS, A. and FAN, J. (2001). Regularization of wavelet approximations. *J. Amer. Statist. Assoc.* **96** 939–967. MR1946364 https://doi.org/10.1198/016214501753208942
[3] BARRON, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans. Inf. Theory* **39** 930–945. MR1237720 https://doi.org/10.1109/18.256500
[4] BARTLETT, P. L., HARVEY, N., LIAW, C. and MEHRABIAN, A. (2019). Nearly-tight VC-dimension and pseudodimension bounds for piecewise linear neural networks. *J. Mach. Learn. Res.* **20** Paper No. 63. MR3960917
[5] BAUER, B. and KOHLER, M. (2019). On deep learning as a remedy for the curse of dimensionality in nonparametric regression. *Ann. Statist.* **47** 2261–2285. MR3953451 https://doi.org/10.1214/18-AOS1747
[6] BELLONI, A. and CHERNOZHUKOV, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli* **19** 521–547. MR3037163 https://doi.org/10.3150/11-BEJ410
[7] BHATTACHARYA, S., FAN, J. and MUKHERJEE, D. (2024). Supplement to “Deep neural networks for nonparametric interaction models with diverging dimension.” https://doi.org/10.1214/24-AOS2442SUPP
[8] 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
[9] BICKEL, P. J., RITOV, Y., TSYBAKOV, A. B. et al. (2010). Hierarchical selection of variables in sparse high-dimensional regression. *IMS Collect.* **6** 28.

MSC2020 subject classifications. 62G08, 62M45.
Key words and phrases. Deep neural networks, high-dimensional statistics, nonparametric interaction model, minimax rate, sparse nonparametric components.

[10] BIEN, J., TAYLOR, J. and TIBSHIRANI, R. (2013). A LASSO for hierarchical interactions. *Ann. Statist.* **41** 1111–1141. MR3113805 https://doi.org/10.1214/13-AOS1096
[11] BREIMAN, L. (2001). Random forests. *Mach. Learn.* **45** 5–32.
[12] BUJA, A., HASTIE, T. and TIBSHIRANI, R. (1989). Linear smoothers and additive models. *Ann. Statist.* **17** 453–555. MR0994249 https://doi.org/10.1214/aos/1176347115
[13] CANDES, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *Ann. Statist.* **35** 2313–2351. MR2382644 https://doi.org/10.1214/009053606000001523
[14] FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications. Monographs on Statistics and Applied Probability* **66**. CRC Press, London. MR1383587
[15] FAN, J. and GU, Y. (2022). Factor Augmented Sparse Throughput Deep ReLU Neural Networks for High Dimensional Regression. ArXiv preprint. Available at arXiv:2210.02002.
[16] FAN, J., GU, Y. and ZHOU, W.-X. (2022). How do noise tails impact on deep ReLU networks? ArXiv preprint. Available at 2203.10418.
[17] FAN, J., HÄRDLE, W. and MAMMEN, E. (1998). Direct estimation of low-dimensional components in additive models. *Ann. Statist.* **26** 943–971. MR1635422 https://doi.org/10.1214/aos/1024691083
[18] FAN, J., LI, R., ZHANG, C.-H. and ZOU, H. (2020). *Statistical Foundations of Data Science*. CRC Press/CRC, Boca Raton, FL.
[19] FAN, J. and LV, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 849–911. MR2530322 https://doi.org/10.1111/j.1467-9868.2008.00674.x
[20] FAN, J. and SONG, R. (2010). Sure independence screening in generalized linear models with NP-dimensionality. *Ann. Statist.* **38** 3567–3604. MR2766861 https://doi.org/10.1214/10-AOS798
[21] FREUND, Y., SCHAPIRE, R. E. et al. (1996). Experiments with a new boosting algorithm. In *icml* **96** 148–156. Citeseer.
[22] FRIEDMAN, J. H. and STUETZLE, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76** 817–823. MR0650892
[23] GOODFELLOW, I., BENGIO, Y. and COURVILLE, A. (2016). *Deep Learning. Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR3617773
[24] HAN, Q. and WELLNER, J. A. (2019). Convergence rates of least squares regression estimators with heavy-tailed errors. *Ann. Statist.* **47** 2286–2319. MR3953452 https://doi.org/10.1214/18-AOS1748
[25] HAO, N. and ZHANG, H. H. (2014). Interaction screening for ultrahigh-dimensional data. *J. Amer. Statist. Assoc.* **109** 1285–1301. MR3265697 https://doi.org/10.1080/01621459.2014.881741
[26] HÄRDLE, W., HALL, P. and ICHIMURA, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21** 157–178. MR1212171 https://doi.org/10.1214/aos/1176349020
[27] HORNIK, K., STINCHCOMBE, M. and WHITE, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Netw.* **2** 359–366.
[28] HOROWITZ, J. L. and MAMMEN, E. (2007). Rate-optimal estimation for a general class of nonparametric regression models with unknown link functions. *Ann. Statist.* **35** 2589–2619. MR2382659 https://doi.org/10.1214/009053607000000415
[29] KOHLER, M. and KRZYŻAK, A. (2005). Adaptive regression estimation with multilayer feedforward neural networks. *J. Nonparametr. Stat.* **17** 891–913. MR2192165 https://doi.org/10.1080/10485250500309608
[30] KOHLER, M. and KRZYŻAK, A. (2017). Nonparametric regression based on hierarchical interaction models. *IEEE Trans. Inf. Theory* **63** 1620–1630. MR3625984 https://doi.org/10.1109/TIT.2016.2634401
[31] KOHLER, M. and LANGER, S. (2021). On the rate of convergence of fully connected deep neural network regression estimates. *Ann. Statist.* **49** 2231–2249. MR4319248 https://doi.org/10.1214/20-aos2034
[32] KOLTCHINSKII, V. and YUAN, M. (2010). Sparsity in multiple kernel learning. *Ann. Statist.* **38** 3660–3695. MR2766864 https://doi.org/10.1214/10-AOS825
[33] KPOTUFE, S. and DASGUPTA, S. (2012). A tree-based regressor that adapts to intrinsic dimension. *J. Comput. System Sci.* **78** 1496–1515. MR2926146 https://doi.org/10.1016/j.jcss.2012.01.002
[34] KRIZHEVSKY, A., HINTON, G. et al. (2009). Learning multiple layers of features from tiny images.
[35] LECUN, Y. (1998). The MNIST database of handwritten digits. Available at http://yann.lecun.com/exdb/mnist/.
[36] LIN, Y. and ZHANG, H. H. (2006). Component selection and smoothing in multivariate nonparametric regression. *Ann. Statist.* **34** 2272–2297. MR2291500 https://doi.org/10.1214/009053606000000722
[37] LU, J., SHEN, Z., YANG, H. and ZHANG, S. (2021). Deep network approximation for smooth functions. *SIAM J. Math. Anal.* **53** 5465–5506. MR4319100 https://doi.org/10.1137/20M134695X
[38] MHASKAR, H. N. (1996). Neural networks for optimal approximation of smooth and analytic functions. *Neural Comput.* **8** 164–177.

[39] NOVEMBRE, J., JOHNSON, T., BRYC, K., KUTALIK, Z., BOYKO, A. R., AUTON, A., INDAP, A., KING, K. S., BERGMANN, S. et al. (2008). Genes mirror geography within Europe. *Nature* **456** 98–101.

[40] RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2012). Minimax-optimal rates for sparse additive models over kernel classes via convex programming. *J. Mach. Learn. Res.* **13** 389–427. MR2913704

[41] RAVIKUMAR, P., LAFFERTY, J., LIU, H. and WASSERMAN, L. (2009). Sparse additive models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 1009–1030. MR2750255 https://doi.org/10.1111/j.1467-9868.2009.00718.x

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

[43] SHEN, Z., YANG, H. and ZHANG, S. (2022). Deep network approximation: Achieving arbitrary accuracy with fixed number of neurons. *J. Mach. Learn. Res.* **23** Paper No. [276]. MR4577715

[44] SHEN, Z., YANG, H. and ZHANG, S. (2022). Optimal approximation rate of ReLU networks in terms of width and depth. *J. Math. Pures Appl.* (9) **157** 101–135. MR4351074 https://doi.org/10.1016/j.matpur.2021.07.009

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

[46] STONE, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13** 689–705. MR0790566 https://doi.org/10.1214/aos/1176349548

[47] STONE, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation. *Ann. Statist.* **22** 118–184. MR1272079 https://doi.org/10.1214/aos/1176325361

[48] TAN, Z. and ZHANG, C.-H. (2019). Doubly penalized estimation in additive regression with high-dimensional data. *Ann. Statist.* **47** 2567–2600. MR3988766 https://doi.org/10.1214/18-AOS1757

[49] TSYBAKOV, A. B. (2004). *Introduction à L'estimation Non-paramétrique*. *Mathématiques & Applications* (*Berlin*) [*Mathematics & Applications*] **41**. Springer, Berlin. MR2013911

[50] YANG, Y. and DUNSON, D. B. (2016). Bayesian manifold regression. *Ann. Statist.* **44** 876–905. MR3476620 https://doi.org/10.1214/15-AOS1390

[51] YANG, Y. and TOKDAR, S. T. (2015). Minimax-optimal nonparametric regression in high dimensions. *Ann. Statist.* **43** 652–674. MR3319139 https://doi.org/10.1214/14-AOS1289

[52] YAROTSKY, D. (2017). Error bounds for approximations with deep ReLU networks. *Neural Netw.* **94** 103–114. https://doi.org/10.1016/j.neunet.2017.07.002

[53] 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

[54] YUAN, M. and ZHOU, D.-X. (2016). Minimax optimal rates of estimation in high dimensional additive models. *Ann. Statist.* **44** 2564–2593. MR3576554 https://doi.org/10.1214/15-AOS1422

[55] ZHANG, S., SHEN, Z. and YANG, H. (2022). Neural network architecture beyond width and depth. *Adv. Neural Inf. Process. Syst.* **35** 5669–5681.

[56] ZHANG, Y., WAINWRIGHT, M. J. and JORDAN, M. I. (2014). Lower bounds on the performance of polynomial-time algorithms for sparse linear regression. In *Conference on Learning Theory* 921–948. PMLR.

[57] ZHAO, P., ROCHA, G. and YU, B. (2009). The composite absolute penalties family for grouped and hierarchical variable selection. *Ann. Statist.* **37** 3468–3497. MR2549566 https://doi.org/10.1214/07-AOS584

The Annals of Statistics
2024, Vol. 52, No. 6, 2767–2790
https://doi.org/10.1214/24-AOS2444

ON THE STATISTICAL COMPLEXITY OF SAMPLE AMPLIFICATION

BY BRIAN AXELROD[1,a], SHIVAM GARG[2,c], YANJUN HAN[3,d], VATSAL SHARAN[4,e] AND GREGORY VALIANT[1,b]

[1]*Department of Computer Science, Stanford University,* [a]*baxelrod@cs.stanford.edu,* [b]*valiant@cs.stanford.edu*
[2]*Microsoft Research,* [c]*shigarg@microsoft.com*
[3]*Courant Institute of Mathematical Sciences and Center for Data Science, New York University,* [d]*yanjunhan@nyu.edu*
[4]*Department of Computer Science, University of Southern California,* [e]*vsharan@usc.edu*

The "sample amplification" problem formalizes the following question: Given n i.i.d. samples drawn from an unknown distribution P, when is it possible to produce a larger set of $n+m$ samples which cannot be distinguished from $n+m$ i.i.d. samples drawn from P? In this work, we provide a firm statistical foundation for this problem by deriving generally applicable amplification procedures, lower bound techniques and connections to existing statistical notions. Our techniques apply to a large class of distributions including the exponential family, and establish a rigorous connection between sample amplification and distribution learning.

REFERENCES

[1] ANTONIOU, A., STORKEY, A. and EDWARDS, H. (2017). Data augmentation generative adversarial networks. Preprint. Available at arXiv:1711.04340.

[2] AXELROD, B.,GARG, S.,HAN, Y.,SHARAN, V.,VALIANT, G. (2024). Supplement to "On the statistical complexity of sample amplification." https://doi.org/10.1214/24-AOS2444SUPP

[3] AXELROD, B., GARG, S., SHARAN, V. and VALIANT, G. (2020). Sample amplification: Increasing dataset size even when learning is impossible. In *International Conference on Machine Learning* 442–451. PMLR.

[4] BAI, Y., KADAVATH, S., KUNDU, S., ASKELL, A., KERNION, J., JONES, A., CHEN, A., GOLDIE, A., MIRHOSEINI, A. et al. (2022). Constitutional AI: Harmlessness from AI feedback. Preprint. Available at arXiv:2212.08073.

[5] BALLY, V. and CARAMELLINO, L. (2014). On the distances between probability density functions. *Electron. J. Probab.* **19** no. 110, 33 pp. MR3296526 https://doi.org/10.1214/EJP.v19-3175

[6] BALLY, V. and CARAMELLINO, L. (2016). Asymptotic development for the CLT in total variation distance. *Bernoulli* **22** 2442–2485. MR3498034 https://doi.org/10.3150/15-BEJ734

[7] BARRON, A. R. (1986). Entropy and the central limit theorem. *Ann. Probab.* **14** 336–342. MR0815975

[8] BASU, D. (1955). On statistics independent of a complete sufficient statistic. *Sankhyā* **15** 377–380. MR0074745 https://doi.org/10.1007/978-1-4419-5825-9_14

[9] BASU, D. (1958). On statistics independent of sufficient statistics. *Sankhyā* **20** 223–226. MR0105758 https://doi.org/10.1007/978-1-4419-5825-9_16

[10] BASU, D. (1959). The family of ancillary statistics. *Sankhyā* **21** 247–256. MR0110115 https://doi.org/10.1007/978-1-4419-5825-9_18

[11] 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

[12] BHATTACHARYA, R. N. and RAO, R. R. (2010). *Normal Approximation and Asymptotic Expansions. Classics in Applied Mathematics* **64**. SIAM, Philadelphia, PA. MR3396213 https://doi.org/10.1137/1.9780898719895.ch1

[13] BOBKOV, S. G., CHISTYAKOV, G. P. and GÖTZE, F. (2014). Berry–Esseen bounds in the entropic central limit theorem. *Probab. Theory Related Fields* **159** 435–478. MR3230000 https://doi.org/10.1007/s00440-013-0510-3

[14] BRENNAN, M. and BRESLER, G. (2019). Optimal average-case reductions to sparse pca: From weak assumptions to strong hardness. In *Conference on Learning Theory* 469–470. PMLR.

MSC2020 subject classifications. Primary 62B15; secondary 62C20.
Key words and phrases. Sample amplification, Le Cam's distance, minimax rate.

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

[16] BROWN, L. D., CAI, T. T., LOW, M. G. and ZHANG, C.-H. (2002). Asymptotic equivalence theory for nonparametric regression with random design. *Ann. Statist.* **30** 688–707. MR1922538 https://doi.org/10.1214/aos/1028674838

[17] BROWN, L. D., CARTER, A. V., LOW, M. G. and ZHANG, C.-H. (2004). Equivalence theory for density estimation, Poisson processes and Gaussian white noise with drift. *Ann. Statist.* **32** 2074–2097. MR2102503 https://doi.org/10.1214/009053604000000012

[18] BROWN, L. D. and LOW, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24** 2384–2398. MR1425958 https://doi.org/10.1214/aos/1032181159

[19] CALIMERI, F., MARZULLO, A., STAMILE, C. and TERRACINA, G. (2017). Biomedical data augmentation using generative adversarial neural networks. In *International Conference on Artificial Neural Networks* 626–634. Springer, Berlin.

[20] CHATZIAGAPI, A., PARASKEVOPOULOS, G., SGOUROPOULOS, D., PANTAZOPOULOS, G., NIKANDROU, M., GIANNAKOPOULOS, T., KATSAMANIS, A., POTAMIANOS, A. and NARAYANAN, S. (2019). Data augmentation using GANs for speech emotion recognition. In *Interspeech* 171–175.

[21] CHEN, D., QI, X., ZHENG, Y., LU, Y. and LI, Z. (2022). Deep data augmentation for weed recognition enhancement: A diffusion probabilistic model and transfer learning based approach. Preprint. Available at arXiv:2210.09509.

[22] CHLAP, P., MIN, H., VANDENBERG, N., DOWLING, J., HOLLOWAY, L. and HAWORTH, A. (2021). A review of medical image data augmentation techniques for deep learning applications. *J. Med. Imag. Radiat. Oncol.* **65** 545–563. https://doi.org/10.1111/1754-9485.13261

[23] CUBUK, E. D., ZOPH, B., MANE, D., VASUDEVAN, V. and LE, Q. V. (2019). Autoaugment: Learning augmentation strategies from data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 113–123.

[24] CUBUK, E. D., ZOPH, B., SHLENS, J. and LE, Q. V. (2020). Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* 702–703.

[25] DIACONIS, P. and YLVISAKER, D. (1979). Conjugate priors for exponential families. *Ann. Statist.* **7** 269–281. MR0520238

[26] FRID-ADAR, M., DIAMANT, I., KLANG, E., AMITAI, M., GOLDBERGER, J. and GREENSPAN, H. (2018). GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* **321** 321–331.

[27] HAN, C., RUNDO, L., ARAKI, R., NAGANO, Y., FURUKAWA, Y., MAURI, G., NAKAYAMA, H. and HAYASHI, H. (2019). Combining noise-to-image and image-to-image GANs: Brain MR image augmentation for tumor detection. *IEEE Access* **7** 156966–156977.

[28] HAN, Y., JIAO, J. and WEISSMAN, T. (2015). Minimax estimation of discrete distributions under ℓ_1 loss. *IEEE Trans. Inf. Theory* **61** 6343–6354. MR3418968 https://doi.org/10.1109/TIT.2015.2478816

[29] HENDRYCKS, D., MU, N., CUBUK, E. D., ZOPH, B., GILMER, J. and LAKSHMINARAYANAN, B. (2020). AugMix: A simple data processing method to improve robustness and uncertainty. In *International Conference on Learning Representations*.

[30] KRIZHEVSKY, A., SUTSKEVER, I. and HINTON, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25**.

[31] LE CAM, L. (1964). Sufficiency and approximate sufficiency. *Ann. Math. Stat.* **35** 1419–1455. MR0207093 https://doi.org/10.1214/aoms/1177700372

[32] LE CAM, L. (1972). Limits of experiments. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume* 1: *Theory of Statistics* The Regents of the University of California.

[33] LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. *Springer Series in Statistics*. Springer, New York. MR0856411 https://doi.org/10.1007/978-1-4612-4946-7

[34] LE CAM, L. and YANG, G. L. (1990). *Asymptotics in Statistics*: *Some Basic Concepts*. *Springer Series in Statistics*. Springer, New York. MR1066869 https://doi.org/10.1007/978-1-4684-0377-0

[35] LEHMANN, E. L. and SCHOLZ, F.-W. (1992). Ancillarity. In *Current Issues in Statistical Inference*: *Essays in Honor of D. Basu*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **17** 32–51. IMS, Hayward, CA. MR1194408 https://doi.org/10.1214/lnms/1215458837

[36] LIESE, F. and MIESCKE, K.-J. (2008). *Statistical Decision Theory*: *Estimation*, *Testing*, *and Selection*. *Springer Series in Statistics*. Springer, New York. MR2421720

[37] LU, Y., SHEN, M., WANG, H., WANG, X., VAN RECHEM, C. and WEI, W. (2023). Machine learning for synthetic data generation: A review. Preprint. Available at arXiv:2302.04062.

[38] LUZI, L., SIAHKOOHI, A., MAYER, P. M., CASCO-RODRIGUEZ, J. and BARANIUK, R. (2022). Boomerang: Local sampling on image manifolds using diffusion models. Preprint. Available at arXiv:2210.12100.
[39] MA, Z. and WU, Y. (2015). Computational barriers in minimax submatrix detection. *Ann. Statist.* **43** 1089–1116. MR3346698 https://doi.org/10.1214/14-AOS1300
[40] MADANI, A., MORADI, M., KARARGYRIS, A. and SYEDA-MAHMOOD, T. (2018). Chest X-ray generation and data augmentation for cardiovascular abnormality classification. In *Medical Imaging* 2018: *Image Processing* **10574** 415–420. SPIE.
[41] 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
[42] PROKHOROV, Y. V. (1952). A local theorem for densities. *Dokl. Akad. Nauk SSSR* **83** 797–800. MR0049501
[43] RAY, K. and SCHMIDT-HIEBER, J. (2019). Asymptotic nonequivalence of density estimation and Gaussian white noise for small densities. *Ann. Inst. Henri Poincaré Probab. Stat.* **55** 2195–2208. MR4029152 https://doi.org/10.1214/18-AIHP946
[44] SANDFORT, V., YAN, K., PICKHARDT, P. J. and SUMMERS, R. M. (2019). Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Nat. Sci. Rep.* **9** 1–9.
[45] SHORTEN, C. and KHOSHGOFTAAR, T. M. (2019). A survey on image data augmentation for deep learning. *J. Big Data* **6** 1–48.
[46] SIMARD, P. Y., STEINKRAUS, D., PLATT, J. C. et al. (2003). Best practices for convolutional neural networks applied to visual document analysis. In *Proceedings of International Conference on Document Analysis and Recognition* **3**. Edinburgh.
[47] SIRAZHDINOV, S. K. and MAMATOV, M. (1962). On convergence in the mean for densities. *Theory Probab. Appl.* **7** 424–428.
[48] TOKOZUME, Y., USHIKU, Y. and HARADA, T. (2018). Between-class learning for image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 5486–5494.
[49] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. MR2724359 https://doi.org/10.1007/b13794
[50] 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
[51] VERMA, V., LAMB, A., BECKHAM, C., NAJAFI, A., MITLIAGKAS, I., LOPEZ-PAZ, D. and BENGIO, Y. (2019). Manifold mixup: Better representations by interpolating hidden states. In *International Conference on Machine Learning* 6438–6447. PMLR.
[52] WALD, A. (1950). *Statistical Decision Functions*. Wiley, New York. MR0036976
[53] WILLIAMS, E. J. (1982). Some classes of conditional inference procedures. *J. Appl. Probab.* **19** 293–303. MR0633198
[54] YI, W., SUN, Y. and HE, S. (2018). Data augmentation using conditional GANs for facial emotion recognition. In 2018 *Progress in Electromagnetics Research Symposium* (*PIERS-Toyama*) 710–714. IEEE.
[55] YI, X., WALIA, E. and BABYN, P. (2019). Generative adversarial network in medical imaging: A review. *Med. Image Anal.* **58** 101552. https://doi.org/10.1016/j.media.2019.101552
[56] YUN, S., HAN, D., OH, S. J., CHUN, S., CHOE, J. and YOO, Y. (2019). Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* 6023–6032.
[57] ZHANG, H., CISSE, M., DAUPHIN, Y. N. and LOPEZ-PAZ, D. (2018). Mixup: Beyond empirical risk minimization. In *International Conference on Learning Representations*.

The Annals of Statistics
2024, Vol. 52, No. 6, 2791–2815
https://doi.org/10.1214/24-AOS2445

CONVEX REGRESSION IN MULTIDIMENSIONS: SUBOPTIMALITY OF LEAST SQUARES ESTIMATORS

BY GIL KUR[1,a], FUCHANG GAO[2,b], ADITYANAND GUNTUBOYINA[3,c] AND BODHISATTVA SEN[4,d]

[1]*Institute for Machine Learning, ETH Zürich,* [a]*gil.kur@inf.ethz.ch*
[2]*Department of Mathematics, University of Idaho,* [b]*fuchang@uidaho.edu*
[3]*Department of Statistics, University of California, Berkeley,* [c]*aditya@stat.berkeley.edu*
[4]*Department of Statistics, Columbia University,* [d]*bodhi@stat.columbia.edu*

Under the usual nonparametric regression model with Gaussian errors, Least Squares Estimators (LSEs) over natural subclasses of convex functions are shown to be suboptimal for estimating a d-dimensional convex function in squared error loss when the dimension d is 5 or larger. The specific function classes considered include: (i) bounded convex functions supported on a polytope (in random design), (ii) Lipschitz convex functions supported on any convex domain (in random design) and (iii) convex functions supported on a polytope (in fixed design). For each of these classes, the risk of the LSE is proved to be of the order $n^{-2/d}$ (up to logarithmic factors) while the minimax risk is $n^{-4/(d+4)}$, when $d \geq 5$. In addition, the first rate of convergence results (worst case and adaptive) for the unrestricted convex LSE are established in fixed design for polytopal domains for all $d \geq 1$. Some new metric entropy results for convex functions are also proved, which are of independent interest.

REFERENCES

[1] AÏT-SAHALIA, Y. and DUARTE, J. (2003). Nonparametric option pricing under shape restrictions. *J. Econometrics* **116** 9–47. MR2002521 https://doi.org/10.1016/S0304-4076(03)00102-7

[2] ALLON, G., BEENSTOCK, M., HACKMAN, S., PASSY, U. and SHAPIRO, A. (2007). Nonparametric estimation of concave production technologies by entropic methods. *J. Appl. Econometrics* **22** 795–816. MR2370975 https://doi.org/10.1002/jae.918

[3] BALÁZS, G. (2016). Convex regression: Theory, practice, and applications Ph.D. thesis, Univ. Alberta.

[4] BALÁZS, G., GYÖRGY, A. and SZEPESVÁRI, C. (2015). Near-optimal max-affine estimators for convex regression. In *AISTATS*.

[5] BALL, K. (1997). An elementary introduction to modern convex geometry. In *Flavors of Geometry*. *Math. Sci. Res. Inst. Publ.* **31** 1–58. Cambridge Univ. Press, Cambridge. MR1491097 https://doi.org/10.2977/prims/1195164788

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

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

[8] BRONŠTEĬN, E. M. (1976). ε-entropy of convex sets and functions. *Sibirsk. Mat. Zh.* **17** 508–514, 715. MR0415155

[9] BRONŠTEĬN, E. M. and IVANOV, L. D. (1975). The approximation of convex sets by polyhedra. *Sibirsk. Mat. Zh.* **16** 1110–1112, 1132. MR0400054

[10] CHATTERJEE, S. (2014). A new perspective on least squares under convex constraint. *Ann. Statist.* **42** 2340–2381. MR3269982 https://doi.org/10.1214/14-AOS1254

[11] CHATTERJEE, S. (2016). An improved global risk bound in concave regression. *Electron. J. Stat.* **10** 1608–1629. MR3522655 https://doi.org/10.1214/16-EJS1151

MSC2020 subject classifications. 62G08.
Key words and phrases. Adaptive risk bounds, bounded convex regression, Dudley's entropy bound, Lipschitz convex regression, lower bounds on the risk of least squares estimators, metric entropy, nonparametric maximum likelihood estimation, Sudakov minoration.

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

[13] CHEN, W. and MAZUMDER, R. (2020). Subgradient Regularized Multivariate Convex Regression at Scale. Preprint. Available at arXiv:2005.11588.

[14] CHEN, Y. and WELLNER, J. A. (2016). On convex least squares estimation when the truth is linear. *Electron. J. Stat.* **10** 171–209. MR3466180 https://doi.org/10.1214/15-EJS1098

[15] DOSS, C. R. (2020). Bracketing numbers of convex and m-monotone functions on polytopes. *J. Approx. Theory* **256** 105425, 33. MR4093045 https://doi.org/10.1016/j.jat.2020.105425

[16] DUDLEY, R. M. (1967). The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Funct. Anal.* **1** 290–330. MR0220340 https://doi.org/10.1016/0022-1236(67)90017-1

[17] DÜMBGEN, L., FREITAG, S. and JONGBLOED, G. (2004). Consistency of concave regression with an application to current-status data. *Math. Methods Statist.* **13** 69–81. MR2078313

[18] GAO, F. and WELLNER, J. A. (2007). Entropy estimate for high-dimensional monotonic functions. *J. Multivariate Anal.* **98** 1751–1764. MR2392431 https://doi.org/10.1016/j.jmva.2006.09.003

[19] GAO, F. and WELLNER, J. A. (2017). Entropy of convex functions on $\mathbb{R}^d$. *Constr. Approx.* **46** 565–592. MR3735701 https://doi.org/10.1007/s00365-017-9387-1

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

[21] GROENEBOOM, P. and JONGBLOED, G. (2014). *Nonparametric Estimation Under Shape Constraints: Estimators, Algorithms and Asymptotics. Cambridge Series in Statistical and Probabilistic Mathematics* **38**. Cambridge Univ. Press, New York. MR3445293 https://doi.org/10.1017/CBO9781139020893

[22] GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2001). Estimation of a convex function: Characterizations and asymptotic theory. *Ann. Statist.* **29** 1653–1698. MR1891742 https://doi.org/10.1214/aos/1015345958

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

[24] HAN, Q. (2021). Set structured global empirical risk minimizers are rate optimal in general dimensions. *Ann. Statist.* **49** 2642–2671. MR4338378 https://doi.org/10.1214/21-aos2049

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

[26] HAN, Q. and WELLNER, J. A. (2016). Multivariate convex regression: Global risk bounds and adaptation. Preprint. Available at arXiv:1601.06844.

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

[28] HANSON, D. L. and PLEDGER, G. (1976). Consistency in concave regression. *Ann. Statist.* **4** 1038–1050. MR0426273

[29] HILDRETH, C. (1954). Point estimates of ordinates of concave functions. *J. Amer. Statist. Assoc.* **49** 598–619. MR0065093

[30] KESHAVARZ, A., WANG, Y. and BOYD, S. (2011). Imputing a convex objective function. In *Intelligent Control* (*ISIC*), 2011 *IEEE International Symposium on* 613–619. IEEE.

[31] KUOSMANEN, T. (2008). Representation theorem for convex nonparametric least squares. *Econom. J.* **11** 308–325.

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

[33] KUR, G., GAO, F., GUNTUBOYINA, A. and SEN, B. (2024). Supplement to "Convex regression in multidimensions: Suboptimality of least squares estimators." https://doi.org/10.1214/24-AOS2445SUPP

[34] LIM, E. (2014). On convergence rates of convex regression in multiple dimensions. *INFORMS J. Comput.* **26** 616–628. MR3246615 https://doi.org/10.1287/ijoc.2013.0587

[35] LIM, E. and GLYNN, P. W. (2012). Consistency of multidimensional convex regression. *Oper. Res.* **60** 196–208. MR2911667 https://doi.org/10.1287/opre.1110.1007

[36] MAMMEN, E. (1991). Nonparametric regression under qualitative smoothness assumptions. *Ann. Statist.* **19** 741–759. MR1105842 https://doi.org/10.1214/aos/1176348118

[37] MATZKIN, R. L. (1991). Semiparametric estimation of monotone and concave utility functions for polychotomous choice models. *Econometrica* **59** 1315–1327. MR1133036 https://doi.org/10.2307/2938369

[38] MAZUMDER, R., CHOUDHURY, A., IYENGAR, G. and SEN, B. (2019). A computational framework for multivariate convex regression and its variants. *J. Amer. Statist. Assoc.* **114** 318–331. MR3941257 https://doi.org/10.1080/01621459.2017.1407771

[39] NEMIROVSKI, A. S. (2000). Topics in nonparametric statistics. In *Lecture on Probability Theory and Statistics*, *École D'Été de Probabilitiés de Saint-Flour XXVIII*-1998. *Lecture Notes in Math*. 1738. Springer, Berlin, Germany.

[40] NEMIROVSKIĬ, A. S., POLYAK, B. T. and TSYBAKOV, A. B. (1984). Signal processing by the nonparametric maximum likelihood method. *Problemy Peredachi Informatsii* **20** 29–46. MR0791733

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

[42] SEIJO, E. and SEN, B. (2011). Nonparametric least squares estimation of a multivariate convex regression function. *Ann. Statist.* **39** 1633–1657. MR2850215 https://doi.org/10.1214/10-AOS852

[43] TALAGRAND, M. (1996). A new look at independence. *Ann. Probab.* **24** 1–34. MR1387624 https://doi.org/10.1214/aop/1042644705

[44] TORIELLO, A., NEMHAUSER, G. and SAVELSBERGH, M. (2010). Decomposing inventory routing problems with approximate value functions. *Naval Res. Logist.* **57** 718–727. MR2762313 https://doi.org/10.1002/nav.20433

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

[46] 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

[47] VARIAN, H. R. (1982). The nonparametric approach to demand analysis. *Econometrica* **50** 945–973. MR0666119 https://doi.org/10.2307/1912771

[48] VARIAN, H. R. (1984). The nonparametric approach to production analysis. *Econometrica* **52** 579–597. MR0740302 https://doi.org/10.2307/1913466

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

The Annals of Statistics
2024, Vol. 52, No. 6, 2816–2850
https://doi.org/10.1214/24-AOS2446

NOISY RECOVERY FROM RANDOM LINEAR OBSERVATIONS: SHARP MINIMAX RATES UNDER ELLIPTICAL CONSTRAINTS

BY REESE PATHAK[1,a], MARTIN J. WAINWRIGHT[2,b] AND LIN XIAO[3,c]

[1]*EECS, University of California, Berkeley,* [a]*pathakr@berkeley.edu*
[2]*EECS, Mathematics, Statistics and Data Science Center, Massachusetts Institute of Technology,* [b]*wainwrigwork@gmail.com*
[3]*Meta FAIR,* [c]*linx@meta.com*

Estimation problems with constrained parameter spaces arise in various settings. In many of these problems, the observations available to the statistician can be modelled as arising from the noisy realization of the image of a random linear operator; an important special case is random design regression. We derive sharp rates of estimation for arbitrary compact elliptical parameter sets and demonstrate how they depend on the distribution of the random linear operator. Our main result is a functional that characterizes the minimax rate of estimation in terms of the noise level, the law of the random operator, and elliptical norms that define the error metric and the parameter space. This nonasymptotic result is sharp up to an explicit universal constant, and it becomes asymptotically exact as the radius of the parameter space is allowed to grow. We demonstrate the generality of the result by applying it to both parametric and nonparametric regression problems.

REFERENCES

[1] ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR1990662

[2] ANTONIADIS, A., PENSKY, M. and SAPATINAS, T. (2014). Nonparametric regression estimation based on spatially inhomogeneous data: Minimax global convergence rates and adaptivity. *ESAIM Probab. Stat.* **18** 1–41. MR3143732 https://doi.org/10.1051/ps/2012024

[3] ARONSZAJN, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68** 337–404. MR0051437 https://doi.org/10.2307/1990404

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

[5] BELITSER, E. N. and LEVIT, B. Y. (1995). On minimax filtering over ellipsoids. *Math. Methods Statist.* **4** 259–273. MR1355248

[6] BERKSON, J. (1950). Are there two regressions? *J. Amer. Statist. Assoc.* **45** 164–180.

[7] BERRY, J. C. (1990). Minimax estimation of a bounded normal mean vector. *J. Multivariate Anal.* **35** 130–139. MR1084946 https://doi.org/10.1016/0047-259X(90)90020-I

[8] BHATIA, R. (2007). *Positive Definite Matrices*. *Princeton Series in Applied Mathematics*. Princeton Univ. Press, Princeton, NJ. MR2284176

[9] BICKEL, P. J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Ann. Statist.* **9** 1301–1309. MR0630112

[10] BORWEIN, J. M. and ZHUANG, D. (1986). On Fan's minimax theorem. *Math. Program.* **34** 232–234. MR0838482 https://doi.org/10.1007/BF01580587

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

[12] BREIMAN, L. and FREEDMAN, D. (1983). How many variables should be entered in a regression equation? *J. Amer. Statist. Assoc.* **78** 131–136. MR0696857

[13] BROWN, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Stat.* **42** 855–903. MR0286209 https://doi.org/10.1214/aoms/1177693318

[14] CARROLL, R. J., RUPPERT, D. and STEFANSKI, L. A. (1995). *Measurement Error in Nonlinear Models*. *Monographs on Statistics and Applied Probability* **63**. CRC Press, London. MR1630517 https://doi.org/10.1007/978-1-4899-4477-1

MSC2020 subject classifications. Primary 62C20; secondary 62G08.
Key words and phrases. Random design, nonparametric regression, covariate shift, minimax rates.

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

[16] CUCKER, F. and SMALE, S. (2002). On the mathematical foundations of learning. *Bull. Amer. Math. Soc.* (*N.S.*) **39** 1–49. MR1864085 https://doi.org/10.1090/S0273-0979-01-00923-5

[17] DICKER, L. H. (2016). Ridge regression and asymptotic minimax estimation over spheres of growing dimension. *Bernoulli* **22** 1–37. MR3449775 https://doi.org/10.3150/14-BEJ609

[18] DONOHO, D. L. (1994). Statistical estimation and optimal recovery. *Ann. Statist.* **22** 238–270. MR1272082 https://doi.org/10.1214/aos/1176325367

[19] 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

[20] DONOHO, D. L., LIU, R. C. and MACGIBBON, B. (1990). Minimax risk over hyperrectangles, and implications. *Ann. Statist.* **18** 1416–1437. MR1062717 https://doi.org/10.1214/aos/1176347758

[21] FOURDRINIER, D., STRAWDERMAN, W. E. and WELLS, M. T. (2018). *Shrinkage Estimation*. *Springer Series in Statistics*. Springer, Cham. MR3887633 https://doi.org/10.1007/978-3-030-02185-6

[22] GAÏFFAS, S. (2005). Convergence rates for pointwise curve estimation with a degenerate design. *Math. Methods Statist.* **14** 1–27. MR2158069

[23] GAÏFFAS, S. (2007). Sharp estimation in sup norm with random design. *Statist. Probab. Lett.* **77** 782–794. MR2369683 https://doi.org/10.1016/j.spl.2006.11.017

[24] GAÏFFAS, S. (2007). On pointwise adaptive curve estimation based on inhomogeneous data. *ESAIM Probab. Stat.* **11** 344–364. MR2339297 https://doi.org/10.1051/ps:2007023

[25] GAÏFFAS, S. (2009). Uniform estimation of a signal based on inhomogeneous data. *Statist. Sinica* **19** 427–447. MR2514170

[26] GOGOLASHVILI, D. (2022). Importance weighting correction of regularized least-squares for covariate and target shifts. https://doi.org/10.48550/ARXIV.2210.09709

[27] GOGOLASHVILI, D., ZECCHIN, M., KANAGAWA, M., KOUNTOURIS, M. and FILIPPONE, M. (2023). When is importance weighting correction needed for covariate shift adaptation? https://doi.org/10.48550/ARXIV.2303.04020

[28] GOLDENSHLUGER, A. and TSYBAKOV, A. (2001). Adaptive prediction and estimation in linear regression with infinitely many parameters. *Ann. Statist.* **29** 1601–1619. MR1891740 https://doi.org/10.1214/aos/1015345956

[29] GOLDENSHLUGER, A. and TSYBAKOV, A. (2003). Optimal prediction for linear regression with infinitely many parameters. *J. Multivariate Anal.* **84** 40–60. MR1965822 https://doi.org/10.1016/S0047-259X(02)00006-4

[30] GOLUBEV, G. K. (1990). Quasilinear estimates for signals in L_2. *Problemy Peredachi Informatsii* **26** 19–24. MR1051584

[31] GUILLOU, A. and KLUTCHNIKOFF, N. (2011). Minimax pointwise estimation of an anisotropic regression function with unknown density of the design. *Math. Methods Statist.* **20** 30–57. MR2811030 https://doi.org/10.3103/S1066530711010030

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

[33] HSU, D., KAKADE, S. M. and ZHANG, T. (2014). Random design analysis of ridge regression. *Found. Comput. Math.* **14** 569–600. MR3201956 https://doi.org/10.1007/s10208-014-9192-1

[34] HSU, D. and SABATO, S. (2016). Loss minimization and parameter estimation with heavy tails. *J. Mach. Learn. Res.* **17** Paper No. 18, 40 pp. MR3491112

[35] IBRAGIMOV, I. A. and KHAS'MINSKIĬ, R. Z. (1980). Nonparametric regression estimation. *Dokl. Akad. Nauk SSSR* **252** 780–784. MR0580829

[36] JOHNSTONE, I. M. (2019). Gaussian estimation: Sequence and wavelet models. Book manuscript.

[37] JUDITSKY, A. and NEMIROVSKI, A. (2018). Near-optimality of linear recovery in Gaussian observation scheme under $\|\cdot\|_2^2$-loss. *Ann. Statist.* **46** 1603–1629. MR3819111 https://doi.org/10.1214/17-AOS1596

[38] KAC, M., MURDOCK, W. L. and SZEGÖ, G. (1953). On the eigenvalues of certain Hermitian forms. *J. Ration. Mech. Anal.* **2** 767–800. MR0059482 https://doi.org/10.1512/iumj.1953.2.52034

[39] KOH, P. W., SAGAWA, S., MARKLUND, H., XIE, S. M., ZHANG, M., BALSUBRAMANI, A., HU, W., YASUNAGA, M., PHILLIPS, R. L. et al. (2021). Wilds: A benchmark of in-the-wild distribution shifts. In *International Conference on Machine Learning* 5637–5664. PMLR.

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

[41] LECUÉ, G. and MENDELSON, S. (2016). Performance of empirical risk minimization in linear aggregation. *Bernoulli* **22** 1520–1534. MR3474824 https://doi.org/10.3150/15-BEJ701

[42] LEHMANN, E. L. and CASELLA, G. (1998). *Theory of Point Estimation*, 2nd ed. *Springer Texts in Statistics*. Springer, New York. MR1639875

[43] LIU, M., ZHANG, Y., LIAO, K. P. and CAI, T. (2020). Augmented transfer regression learning with semi-non-parametric nuisance models.

[44] LUGOSI, G. and MENDELSON, S. (2019). Mean estimation and regression under heavy-tailed distributions: A survey. *Found. Comput. Math.* **19** 1145–1190. MR4017683 https://doi.org/10.1007/s10208-019-09427-x

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

[46] MARCHAND, E. (1993). Estimation of a multivariate mean with constraints on the norm. *Canad. J. Statist.* **21** 359–366. MR1254283 https://doi.org/10.2307/3315700

[47] MARCHAND, E. and STRAWDERMAN, W. E. (2004). Estimation in restricted parameter spaces: A review. In *A Festschrift for Herman Rubin*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **45** 21–44. IMS, Beachwood, OH. MR2126884 https://doi.org/10.1214/lnms/1196285377

[48] MELKMAN, A. A. and RITOV, Y. (1987). Minimax estimation of the mean of a general distribution when the parameter space is restricted. *Ann. Statist.* **15** 432–442. MR0885749 https://doi.org/10.1214/aos/1176350278

[49] MENDELSON, S. (2015). Learning without concentration. *J. ACM* **62** Art. 21, 25 pp. MR3367000 https://doi.org/10.1145/2699439

[50] MOURTADA, J. (2020). Contributions à l'apprentissage statistique: Estimation de densité, agrégation d'experts et forêts aléatoires. Theses, Institut Polytechnique de Paris.

[51] MOURTADA, J. (2022). Exact minimax risk for linear least squares, and the lower tail of sample covariance matrices. *Ann. Statist.* **50** 2157–2178. MR4474486 https://doi.org/10.1214/22-aos2181

[52] OLIVEIRA, R. I. (2016). The lower tail of random quadratic forms with applications to ordinary least squares. *Probab. Theory Related Fields* **166** 1175–1194. MR3568047 https://doi.org/10.1007/s00440-016-0738-9

[53] PATHAK, R., MA, C. and WAINWRIGHT, M. (2022). A new similarity measure for covariate shift with applications to nonparametric regression. In *Proceedings of the* 39*th International Conference on Machine Learning* (K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu and S. Sabato, eds.). *Proceedings of Machine Learning Research* **162** 17517–17530. PMLR.

[54] PATHAK, R., WAINWRIGHT, M. J. and XIAO, L. (2024). Supplement to "Noisy recovery from random linear observations: Sharp minimax rates under elliptical constraints." https://doi.org/10.1214/24-AOS2446SUPP

[55] PINSKER, M. S. (1980). Optimal filtration of square-integrable signals in Gaussian noise. *Problemy Peredachi Informatsii* **16** 52–68. MR0624591

[56] ROBBINS, H. (1956). An empirical Bayes approach to statistics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, *Vol. I* 157–163. Univ. California Press, Berkeley-Los Angeles, CA. MR0084919

[57] SCHMIDT-HIEBER, J. and ZAMOLODTCHIKOV, P. (2024). Local convergence rates of the nonparametric least squares estimator with applications to transfer learning. *Bernoulli* **30** 1845–1877. MR4746591 https://doi.org/10.3150/23-bej1655

[58] SIMCHOWITZ, M., AJAY, A., AGRAWAL, P. and KRISHNAMURTHY, A. (2023). Statistical learning under heterogenous distribution shift. https://doi.org/10.48550/ARXIV.2302.13934

[59] STEIN, C. (1960). Multiple regression. In *Contributions to Probability and Statistics*. *Stanford Studies in Mathematics and Statistics* **2** 424–443. Stanford Univ. Press, Stanford, CA. MR0120718

[60] STEINWART, I. and SCOVEL, C. (2012). Mercer's theorem on general domains: On the interaction between measures, kernels, and RKHSs. *Constr. Approx.* **35** 363–417. MR2914365 https://doi.org/10.1007/s00365-012-9153-3

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

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

[63] TWEEDIE, M. C. K. (1947). Functions of a statistical variate with given means, with special reference to Laplacian distributions. *Proc. Camb. Philos. Soc.* **43** 41–49. MR0018381

[64] WANG, K. (2023). Pseudo-labeling for kernel ridge regression under covariate shift. https://doi.org/10.48550/ARXIV.2302.10160

[65] YANG, Y., PILANCI, M. and WAINWRIGHT, M. J. (2017). Randomized sketches for kernels: Fast and optimal nonparametric regression. *Ann. Statist.* **45** 991–1023. MR3662446 https://doi.org/10.1214/16-AOS1472

[66] ZAMIR, R. (1998). A proof of the Fisher information inequality via a data processing argument. *IEEE Trans. Inf. Theory* **44** 1246–1250. MR1616672 https://doi.org/10.1109/18.669301

The Annals of Statistics
2024, Vol. 52, No. 6, 2851–2878
https://doi.org/10.1214/24-AOS2447

THE PROJECTED COVARIANCE MEASURE FOR ASSUMPTION-LEAN VARIABLE SIGNIFICANCE TESTING

BY ANTON RASK LUNDBORG[1,3,a], ILMUN KIM[2,b], RAJEN D. SHAH[3,c] AND RICHARD J. SAMWORTH[3,d]

[1]*Department of Mathematical Sciences, University of Copenhagen,* [a]*arl@math.ku.dk*
[2]*Department of Statistics and Data Science, Department of Applied Statistics, Yonsei University,* [b]*ilmun@yonsei.ac.kr*
[3]*Statistical Laboratory, University of Cambridge,* [c]*r.shah@statslab.cam.ac.uk,* [d]*r.samworth@statslab.cam.ac.uk*

Testing the significance of a variable or group of variables X for predicting a response Y, given additional covariates Z, is a ubiquitous task in statistics. A simple but common approach is to specify a linear model, and then test whether the regression coefficient for X is nonzero. However, when the model is misspecified, the test may have poor power, for example, when X is involved in complex interactions, or lead to many false rejections. In this work, we study the problem of testing the model-free null of conditional mean independence, that is, that the conditional mean of Y given X and Z does not depend on X. We propose a simple and general framework that can leverage flexible nonparametric or machine learning methods, such as additive models or random forests, to yield both robust error control and high power. The procedure involves using these methods to perform regressions, first to estimate a form of projection of Y on X and Z using one-half of the data, and then to estimate the expected conditional covariance between this projection and Y on the remaining half of the data. While the approach is general, we show that a version of our procedure using spline regression achieves what we show is the minimax optimal rate in this nonparametric testing problem. Numerical experiments demonstrate the effectiveness of our approach both in terms of maintaining Type I error control, and power, compared to several existing approaches.

REFERENCES

AÏT-SAHALIA, Y., BICKEL, P. J. and STOKER, T. M. (2001). Goodness-of-fit tests for kernel regression with an application to option implied volatilities. *J. Econometrics* **105** 363–412. MR1873358 https://doi.org/10.1016/S0304-4076(01)00091-4

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

BELLONI, A., CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Some new asymptotic theory for least squares series: Pointwise and uniform results. *J. Econometrics* **186** 345–366. MR3343791 https://doi.org/10.1016/j.jeconom.2015.02.014

BELLONI, A., CHERNOZHUKOV, V. and WANG, L. (2011). Square-root lasso: Pivotal recovery of sparse signals via conic programming. *Biometrika* **98** 791–806. MR2860324 https://doi.org/10.1093/biomet/asr043

BERRETT, T. B., WANG, Y., BARBER, R. F. and SAMWORTH, R. J. (2020). The conditional permutation test for independence while controlling for confounders. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 175–197. MR4060981

BOJER, C. S. and MELDGAARD, J. P. (2021). Kaggle forecasting competitions: An overlooked learning opportunity. *Int. J. Forecast.* **37** 587–603.

BOUSQUET, O. and ELISSEEFF, A. (2002). Stability and generalization. *J. Mach. Learn. Res.* **2** 499–526. MR1929416 https://doi.org/10.1162/153244302760200704

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

MSC2020 subject classifications. 62G10, 62G20, 62H20.
Key words and phrases. Conditional mean independence, hypothesis testing, spline regression, sample splitting, minimax power.

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

CAI, Z., LEI, J. and ROEDER, K. (2022). Model-free prediction test with application to genomics data. *Proc. Natl. Acad. Sci. USA* **119**.

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

CHEN, Q., SYRGKANIS, V. and AUSTERN, M. (2022). Debiased machine learning without sample-splitting for stable estimators. In *Advances in Neural Information Processing Systems* (S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho and A. Oh, eds.) **35** 3096–3109. Curran Associates.

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

COX, D. R. (1975). A note on data-splitting for the evaluation of significance levels. *Biometrika* **62** 441–444. MR0378189 https://doi.org/10.1093/biomet/62.2.441

DAI, B., SHEN, X. and PAN, W. (2024). Significance tests of feature relevance for a black-box learner. *IEEE Trans. Neural Netw. Learn. Syst.* **35** 1898–1911. MR4710268 https://doi.org/10.1109/tnnls.2022.3185742

FAN, Y. and LI, Q. (1996). Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica* **64** 865–890. MR1399221 https://doi.org/10.2307/2171848

FERNÁNDEZ, T. and RIVERA, N. (2024). A general framework for the analysis of kernel-based tests. *J. Mach. Learn. Res.* **25** Paper No. 95. MR4749131

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

GUO, F. R. and SHAH, R. D. (2023). Rank-transformed subsampling: Inference for multiple data splitting and exchangeable p-values. ArXiv preprint. Available at arXiv:2301.02739.

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

HARDT, M., RECHT, B. and SINGER, Y. (2016). Train faster, generalize better: Stability of stochastic gradient descent. In *Proceedings of the* 33*rd International Conference on International Conference on Machine Learning—Volume* 48, *ICML'*16 1225–1234. JMLR.org.

HEINZE-DEML, C., PETERS, J. and MEINSHAUSEN, N. (2018). Invariant causal prediction for nonlinear models. *J. Causal Inference* **6** Art. No. 20170016. MR4335430 https://doi.org/10.1515/jci-2017-0016

ICHIMURA, H. and NEWEY, W. K. (2022). The influence function of semiparametric estimators. *Quant. Econ.* **13** 29–61. MR4399602 https://doi.org/10.3982/qe826

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

INGSTER, Y. I. (1997). Adaptive chi-square tests. *J. Math. Sci.* **99** 1110–1119.

JANKOVÁ, J., SHAH, R. D., BÜHLMANN, P. and SAMWORTH, R. J. (2020). Goodness-of-fit testing in high dimensional generalized linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 773–795. MR4112784 https://doi.org/10.1111/rssb.12371

JIN, Z., YAN, X. and MATTESON, D. S. (2018). Testing for conditional mean independence with covariates through martingale difference divergence. In *Proceedings of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence* 1–12. AUAI Press.

KENNEDY, E. H. (2023). Towards optimal doubly robust estimation of heterogeneous causal effects. *Electron. J. Stat.* **17** 3008–3049. MR4667730 https://doi.org/10.1214/23-ejs2157

KIM, I. and RAMDAS, A. (2024). Dimension-agnostic inference using cross U-statistics. *Bernoulli* **30** 683–711. MR4665594 https://doi.org/10.3150/23-bej1613

LAVERGNE, P. and VUONG, Q. (2000). Nonparametric significance testing. *Econometric Theory* **16** 576–601. MR1790292 https://doi.org/10.1017/S0266466600164059

LEPSKIĬ, O. V. (1991). Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.* **36** 682–697.

LOVÁSZ, L. and VEMPALA, S. (2007). The geometry of logconcave functions and sampling algorithms. *Random Structures Algorithms* **30** 307–358. MR2309621 https://doi.org/10.1002/rsa.20135

LUNDBORG, A. R., KIM, I., SHAH, R. D. and SAMWORTH, R. J. (2024). Supplement to "The Projected covariance measure for assumption-lean variable significance testing." https://doi.org/10.1214/24-AOS2447SUPPA, https://doi.org/10.1214/24-AOS2447SUPPB

LUNDBORG, A. R., SHAH, R. D. and PETERS, J. (2022). Conditional independence testing in Hilbert spaces with applications to functional data analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 1821–1850. MR4515559 https://doi.org/10.1111/rssb.12544

MEINSHAUSEN, N. and BÜHLMANN, P. (2010). Stability selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 417–473. MR2758523 https://doi.org/10.1111/j.1467-9868.2010.00740.x

MEINSHAUSEN, N., MEIER, L. and BÜHLMANN, P. (2009). p-values for high-dimensional regression. *J. Amer. Statist. Assoc.* **104** 1671–1681. MR2750584 https://doi.org/10.1198/jasa.2009.tm08647

MENDELSON, S. and ZHIVOTOVSKIY, N. (2020). Robust covariance estimation under L_4-L_2 norm equivalence. *Ann. Statist.* **48** 1648–1664. MR4124338 https://doi.org/10.1214/19-AOS1862

NEMIROVSKI, A. (2000). Topics in non-parametric statistics. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1998). *Lecture Notes in Math.* **1738** 85–277. Springer, Berlin. MR1775640

NEWEY, W. K. and ROBINS, J. R. (2018). Cross-fitting and fast remainder rates for semiparametric estimation. ArXiv preprint. Available at arXiv:1801.09138.

NEYKOV, M., BALAKRISHNAN, S. and WASSERMAN, L. (2021). Minimax optimal conditional independence testing. *Ann. Statist.* **49** 2151–2177. MR4319245 https://doi.org/10.1214/20-aos2030

NEYMAN, J. (1923). Sur les applications de la théorie des probabilités aux experiences agricoles: Essai des principes. *Rocz. Nauk Rol.* **10** 1–51.

PETERSEN, L. and HANSEN, N. R. (2021). Testing conditional independence via quantile regression based partial copulas. *J. Mach. Learn. Res.* **22** Paper No. 70. MR4253763

RINALDO, A., WASSERMAN, L. and G'SELL, M. (2019). Bootstrapping and sample splitting for high-dimensional, assumption-lean inference. *Ann. Statist.* **47** 3438–3469. MR4025748 https://doi.org/10.1214/18-AOS1784

ROMANO, J. P. (2004). On non-parametric testing, the uniform behaviour of the t-test, and related problems. *Scand. J. Stat.* **31** 567–584. MR2101540 https://doi.org/10.1111/j.1467-9469.2004.00407.x

RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66** 688.

SCHEIDEGGER, C., HÖRRMANN, J. and BÜHLMANN, P. (2022). The weighted generalised covariance measure. *J. Mach. Learn. Res.* **23** Paper No. [273]. MR4577712

SHAH, R. D. and BÜHLMANN, P. (2018). Goodness-of-fit tests for high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 113–135. MR3744714 https://doi.org/10.1111/rssb.12234

SHAH, R. D. and PETERS, J. (2020). The hardness of conditional independence testing and the generalised covariance measure. *Ann. Statist.* **48** 1514–1538. MR4124333 https://doi.org/10.1214/19-AOS1857

SHAH, R. D. and SAMWORTH, R. J. (2013). Variable selection with error control: Another look at stability selection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 55–80. MR3008271 https://doi.org/10.1111/j.1467-9868.2011.01034.x

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

SOLOFF, J. A., BARBER, R. F. and WILLETT, R. (2024). Bagging provides assumption-free stability. *J. Mach. Learn. Res.* **25** Paper No. [131]. MR4749783

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

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

VERDINELLI, I. and WASSERMAN, L. (2024). Decorrelated variable importance. *J. Mach. Learn. Res.* **25** Paper No. [7]. MR4723857

WANG, T. and SAMWORTH, R. J. (2018). High dimensional change point estimation via sparse projection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 57–83. MR3744712 https://doi.org/10.1111/rssb.12243

WANG, Y. and SHAH, R. D. (2020). Debiased Inverse Propensity Score Weighting for Estimation of Average Treatment Effects with High-Dimensional Confounders. ArXiv preprint. Available at arXiv:2011.08661.

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

WASSERMAN, L. and ROEDER, K. (2009). High-dimensional variable selection. *Ann. Statist.* **37** 2178–2201. MR2543689 https://doi.org/10.1214/08-AOS646

WILLIAMSON, B. D., GILBERT, P. B., CARONE, M. and SIMON, N. (2021). Nonparametric variable importance assessment using machine learning techniques. *Biometrics* **77** 9–22. MR4229718 https://doi.org/10.1111/biom.13392

WILLIAMSON, B. D., GILBERT, P. B., SIMON, N. R. and CARONE, M. (2023). A general framework for inference on algorithm-agnostic variable importance. *J. Amer. Statist. Assoc.* **118** 1645–1658. MR4646595 https://doi.org/10.1080/01621459.2021.2003200

WOOD, S. N. (2013). On p-values for smooth components of an extended generalized additive model. *Biometrika* **100** 221–228. MR3034335 https://doi.org/10.1093/biomet/ass048

WOOD, S. N. (2017). *Generalized Additive Models. Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. MR3726911

WRIGHT, M. N. and ZIEGLER, A. (2017). `ranger`: A fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Softw.* **77** 1–17.

ZHANG, K., PETERS, J., JANZING, D. and SCHÖLKOPF, B. (2011). Kernel-based conditional independence test and application in causal discovery. In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*, *UAI'11* 804–813. AUAI Press, Arlington, VA, USA.

ZHANG, L. and JANSON, L. (2020). Floodgate: Inference for model-free variable importance. ArXiv preprint. Available at arXiv:2007.01283.

The Annals of Statistics
2024, Vol. 52, No. 6, 2879–2912
https://doi.org/10.1214/24-AOS2449

DIMENSION FREE RIDGE REGRESSION

BY CHEN CHENG[1,a] AND ANDREA MONTANARI[1,2,b]

[1]*Department of Statistics, Stanford University,* [a]*chencheng@stanford.edu*
[2]*Department of Mathematics, Stanford University,* [b]*montanari@stanford.edu*

Random matrix theory has become a widely useful tool in high-dimensional statistics and theoretical machine learning. However, random matrix theory is largely focused on the proportional asymptotics in which the number of columns grows proportionally to the number of rows of the data matrix. This is not always the most natural setting in statistics where columns correspond to covariates and rows to samples.

With the objective to move beyond the proportional asymptotics, we revisit ridge regression (ℓ_2-penalized least squares) on i.i.d. data $(\boldsymbol{x}_i, y_i)$, $i \le n$, where $\boldsymbol{x}_i$ is a feature vector and $y_i = \langle \boldsymbol{\beta}, \boldsymbol{x}_i \rangle + \varepsilon_i \in \mathbb{R}$ is a response. We allow the feature vector to be high-dimensional, or even infinite-dimensional, in which case it belongs to a separable Hilbert space, and assume either $z_i := \boldsymbol{\Sigma}^{-1/2}\boldsymbol{x}_i$ to have i.i.d. entries, or to satisfy a certain convex concentration property.

Within this setting, we establish nonasymptotic bounds that approximate the bias and variance of ridge regression in terms of the bias and variance of an "equivalent" sequence model (a regression model with diagonal design matrix). The approximation is up to multiplicative factors bounded by $(1 \pm \Delta)$ for some explicitly small Δ.

Previously, such an approximation result was known only in the proportional regime and only up to additive errors: in particular, it did not allow to characterize the behavior of the excess risk when this converges to 0. Our general theory recovers earlier results in the proportional regime (with better error rates). As a new application, we obtain a completely explicit and sharp characterization of ridge regression for Hilbert covariates with regularly varying spectrum. Finally, we analyze the overparametrized near-interpolation setting and obtain sharp "benign overfitting" guarantees.

REFERENCES

ADAMCZAK, R. (2015). A note on the Hanson-Wright inequality for random vectors with dependencies. *Electron. Commun. Probab.* **20** no. 72, 13. MR3407216 https://doi.org/10.1214/ECP.v20-3829

ADVANI, M. S., SAXE, A. M. and SOMPOLINSKY, H. (2020). High-dimensional dynamics of generalization error in neural networks. *Neural Netw.* **132** 428–446. https://doi.org/10.1016/j.neunet.2020.08.022

BAKRY, D., GENTIL, I. and LEDOUX, M. (2014). *Analysis and Geometry of Markov Diffusion Operators. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **348**. Springer, Cham. MR3155209 https://doi.org/10.1007/978-3-319-00227-9

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

BARTLETT, P. L., LONG, P. M., LUGOSI, G. and TSIGLER, A. (2020). Benign overfitting in linear regression. *Proc. Natl. Acad. Sci. USA* **117** 30063–30070. MR4263288 https://doi.org/10.1073/pnas.1907378117

BARTLETT, P. L., MONTANARI, A. and RAKHLIN, A. (2021). Deep learning: A statistical viewpoint. *Acta Numer.* **30** 87–201. MR4295218 https://doi.org/10.1017/S0962492921000027

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

BEAN, D., BICKEL, P. J., EL KAROUI, N. and YU, B. (2013). Optimal M-estimation in high-dimensional regression. *Proc. Natl. Acad. Sci. USA* **110** 14563–14568.

MSC2020 subject classifications. Primary 62J07; secondary 60B20.
Key words and phrases. Ridge regression, Hilbert space, random matrix.

BELKIN, M. (2021). Fit without fear: Remarkable mathematical phenomena of deep learning through the prism of interpolation. *Acta Numer.* **30** 203–248. MR4298218 https://doi.org/10.1017/S0962492921000039

BLOEMENDAL, A., ERDŐS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2014). Isotropic local laws for sample covariance and generalized Wigner matrices. *Electron. J. Probab.* **19** no. 33, 53. MR3183577 https://doi.org/10.1214/ejp.v19-3054

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

BREZIS, H. (2011). *Functional Analysis, Sobolev Spaces and Partial Differential Equations. Universitext*. Springer, New York. MR2759829

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

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

CHENG, C. and MONTANARI, A. (2024). Supplement to "Dimension free ridge regression." https://doi.org/10.1214/24-AOS2449SUPP

DICKER, L. H. (2016). Ridge regression and asymptotic minimax estimation over spheres of growing dimension. *Bernoulli* **22** 1–37. MR3449775 https://doi.org/10.3150/14-BEJ609

DOBRIBAN, E. and WAGER, S. (2018). High-dimensional asymptotics of prediction: Ridge regression and classification. *Ann. Statist.* **46** 247–279. MR3766952 https://doi.org/10.1214/17-AOS1549

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., ELAD, M. and TEMLYAKOV, V. N. (2006). Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inf. Theory* **52** 6–18. MR2237332 https://doi.org/10.1109/TIT.2005.860430

DONOHO, D. L., JOHNSTONE, I. and MONTANARI, A. (2013). Accurate prediction of phase transitions in compressed sensing via a connection to minimax denoising. *IEEE Trans. Inf. Theory* **59** 3396–3433. MR3061255 https://doi.org/10.1109/TIT.2013.2239356

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.

GALAMBOS, J. and SENETA, E. (1973). Regularly varying sequences. *Proc. Amer. Math. Soc.* **41** 110–116. MR0323963 https://doi.org/10.2307/2038824

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

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

KARAMATA, J. (1933). Sur un mode de croissance régulière. Théorèmes fondamentaux. *Bull. Soc. Math. France* **61** 55–62. MR1504998

KNOWLES, A. and YIN, J. (2017). Anisotropic local laws for random matrices. *Probab. Theory Related Fields* **169** 257–352. MR3704770 https://doi.org/10.1007/s00440-016-0730-4

KOEHLER, F., ZHOU, L., SUTHERLAND, D. J. and SREBRO, N. (2021). Uniform convergence of interpolators: Gaussian width, norm bounds and benign overfitting. *Adv. Neural Inf. Process. Syst.* **34** 20657–20668.

LYTOVA, A. and PASTUR, L. (2009). Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Ann. Probab.* **37** 1778–1840. MR2561434 https://doi.org/10.1214/09-AOP452

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

RICHARDS, D., MOURTADA, J. and ROSASCO, L. (2021). Asymptotics of ridge (less) regression under general source condition. In *International Conference on Artificial Intelligence and Statistics* 3889–3897. PMLR.

RUDELSON, M. and VERSHYNIN, R. (2013). Hanson-Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.* **18** no. 82, 9. MR3125258 https://doi.org/10.1214/ECP.v18-2865

TAHERI, H., PEDARSANI, R. and THRAMPOULIDIS, C. (2021). Fundamental limits of ridge-regularized empirical risk minimization in high dimensions. In *International Conference on Artificial Intelligence and Statistics* 2773–2781. PMLR.

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* **58** 267–288. MR1379242

TSIGLER, A. and BARTLETT, P. L. (2023). Benign overfitting in ridge regression. *J. Mach. Learn. Res.* **24** Paper No. [123], 76. MR4583284

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

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

WU, D. and XU, J. (2020). On the optimal weighted *ell*_2 regularization in overparameterized linear regression. *Adv. Neural Inf. Process. Syst.* **33** 10112–10123.

The Annals of Statistics
2024, Vol. 52, No. 6, 2913–2930
https://doi.org/10.1214/24-AOS2451

CHANGE-POINT ANALYSIS WITH IRREGULAR SIGNALS

BY TOBIAS KLEY[1,a], YUHAN PHILIP LIU[2,b], HONGYUAN CAO[3,d] AND WEI BIAO WU[2,c]

[1]*Institute for Mathematical Stochastics, Georg-August-Universität Göttingen,* [a]*tobias.kley@uni-goettingen.de*
[2]*Department of Statistics, University of Chicago,* [b]*yuhanphilipliu@uchicago.edu,* [c]*wbwu@uchicago.edu*
[3]*Department of Statistics, Florida State University,* [d]*hcao@fsu.edu*

This paper considers the problem of testing and estimation of change point where signals after the change point can be highly irregular, which departs from the existing literature that assumes signals after the change point to be piecewise constant or vary smoothly. A two-step approach is proposed to effectively estimate the location of the change point. The first step consists of a preliminary estimation of the change point that allows us to obtain unknown parameters for the second step. In the second step, we use a new procedure to determine the position of the change point. We show that, under suitable conditions, the desirable $\mathcal{O}_{\mathbb{P}}(1)$ rate of convergence of the estimated change point can be obtained. We apply our method to analyze the Baidu search index of COVID-19 related symptoms and find December 8, 2019, to be the starting date of the COVID-19 pandemic.

REFERENCES

AUE, A. and HORVÁTH, L. (2013). Structural breaks in time series. *J. Time Series Anal.* **34** 1–16. MR3008012 https://doi.org/10.1111/j.1467-9892.2012.00819.x

BARANOWSKI, R., CHEN, Y. and FRYZLEWICZ, P. (2019). Narrowest-over-threshold detection of multiple change points and change-point-like features. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 649–672. MR3961502 https://doi.org/10.1111/rssb.12322

BARANOWSKI, R. and FRYZLEWICZ, P. (2019). wbs: Wild Binary Segmentation for Multiple Change-Point Detection. R package version 1.4.

BERKES, I., LIU, W. and WU, W. B. (2014). Komlós-Major-Tusnády approximation under dependence. *Ann. Probab.* **42** 794–817. MR3178474 https://doi.org/10.1214/13-AOP850

BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. *Wiley Series in Probability and Statistics*: *Probability and Statistics*. Wiley, New York. MR1700749 https://doi.org/10.1002/9780470316962

BÜCHER, A., DETTE, H. and HEINRICHS, F. (2021). Are deviations in a gradually varying mean relevant? A testing approach based on sup-norm estimators. *Ann. Statist.* **49** 3583–3617. MR4352542 https://doi.org/10.1214/21-aos2098

BÜHLMANN, P. and KÜNSCH, H. R. (1999). Block length selection in the bootstrap for time series. *Comput. Statist. Data Anal.* **31** 295–310.

CAO, H. and WU, W. B. (2015). Changepoint estimation: Another look at multiple testing problems. *Biometrika* **102** 974–980. MR3431567 https://doi.org/10.1093/biomet/asv031

CAO, H. and WU, W. B. (2022). Testing and estimation for clustered signals. *Bernoulli* **28** 525–547. MR4337715 https://doi.org/10.3150/21-bej1355

CHEN, Y., WANG, T. and SAMWORTH, R. J. (2022). High-dimensional, multiscale online changepoint detection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 234–266. MR4400396 https://doi.org/10.1111/rssb.12447

CSÖRGŐ, M. and HORVÁTH, L. (1997). *Limit Theorems in Change-Point Analysis*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR2743035

DETTE, H., ECKLE, T. and VETTER, M. (2020). Multiscale change point detection for dependent data. *Scand. J. Stat.* **47** 1243–1274. MR4178193 https://doi.org/10.1111/sjos.12465

DETTE, H. and WU, W. (2019). Detecting relevant changes in the mean of nonstationary processes—a mass excess approach. *Ann. Statist.* **47** 3578–3608. MR4025752 https://doi.org/10.1214/19-AOS1811

CENTRE FOR DISEASE CONTROL AND PREVENTION (2022). CDC museum COVID-19 timeline. *website*. Accessed: 2022-05-18.

MSC2020 subject classifications. 62M10, 62G20.
Key words and phrases. Change-point analysis, COVID-19, invariance principle, irregular signals, long-run variance estimate, smoothing, weak convergence.

FRICK, K., MUNK, A. and SIELING, H. (2014). Multiscale change point inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 495–580. MR3210728 https://doi.org/10.1111/rssb.12047

FRYZLEWICZ, P. (2014). Wild binary segmentation for multiple change-point detection. *Ann. Statist.* **42** 2243–2281. MR3269979 https://doi.org/10.1214/14-AOS1245

FRYZLEWICZ, P. (2018). Tail-greedy bottom-up data decompositions and fast multiple change-point detection. *Ann. Statist.* **46** 3390–3421. MR3852656 https://doi.org/10.1214/17-AOS1662

HAWKINS, D. M. (1977). Testing a sequence of observations for a shift in location. *J. Amer. Statist. Assoc.* **72** 180–186. MR0451496

HEINRICHS, F. and DETTE, H. (2021). A distribution free test for changes in the trend function of locally stationary processes. *Electron. J. Stat.* **15** 3762–3797. MR4298981 https://doi.org/10.1214/21-ejs1871

HINKLEY, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika* **57** 1–17. MR0273727 https://doi.org/10.1093/biomet/57.1.1

HORVÁTH, L. and KOKOSZKA, P. (2002). Change-point detection with non-parametric regression. *Statistics* **36** 9–31. MR1906372 https://doi.org/10.1080/02331880210930

HUANG, C., WANG, Y., LI, X., REN, L., ZHAO, J., HU, Y., ZHANG, L., FAN, G., XU, J. et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395** 497–506.

JANDHYALA, V., FOTOPOULOS, S., MACNEILL, I. and LIU, P. (2013). Inference for single and multiple change-points in time series. *J. Time Series Anal.* **34** 423–446. MR3070866 https://doi.org/10.1111/jtsa.12035

KILLICK, R. and ECKLEY, I. A. (2014). Changepoint: An R package for changepoint analysis. *J. Stat. Softw.* **58** 1–19.

KILLICK, R., FEARNHEAD, P. and ECKLEY, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *J. Amer. Statist. Assoc.* **107** 1590–1598. MR3036418 https://doi.org/10.1080/01621459.2012.737745

KILLICK, R., HAYNES, K. and ECKLEY, I. A. (2022). changepoint: an R package for changepoint analysis. R package version 2.2.4.

KLEY, T., LIU, Y. P, CAO, H. and WU, W. B (2024). Supplement to "Change-point analysis with irregular signals." https://doi.org/10.1214/24-AOS2451SUPPA, https://doi.org/10.1214/24-AOS2451SUPPB

LAHIRI, S. N. (1999). Theoretical comparisons of block bootstrap methods. *Ann. Statist.* **27** 386–404. MR1701117 https://doi.org/10.1214/aos/1018031117

MALLIK, A., BANERJEE, M. and SEN, B. (2013). Asymptotics for p-value based threshold estimation in regression settings. *Electron. J. Stat.* **7** 2477–2515. MR3117104 https://doi.org/10.1214/13-EJS845

MALLIK, A., SEN, B., BANERJEE, M. and MICHAILIDIS, G. (2011). Threshold estimation based on a p-value framework in dose-response and regression settings. *Biometrika* **98** 887–900. MR2860331 https://doi.org/10.1093/biomet/asr051

MIES, F. and STELAND, A. (2023). Sequential Gaussian approximation for nonstationary time series in high dimensions. *Bernoulli* **29** 3114–3140. MR4632133 https://doi.org/10.3150/22-bej1577

MÜLLER, H.-G. (1992). Change-points in nonparametric regression analysis. *Ann. Statist.* **20** 737–761. MR1165590 https://doi.org/10.1214/aos/1176348654

NIU, Y. S., HAO, N. and ZHANG, H. (2016). Multiple change-point detection: A selective overview. *Statist. Sci.* **31** 611–623. MR3598742 https://doi.org/10.1214/16-STS587

PAGE, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika* **42** 523–527. MR0072412 https://doi.org/10.1093/biomet/42.3-4.523

PAGE, E. S. (1957). On problems in which a change in a parameter occurs at an unknown point. *Biometrika* **44** 248–252.

PELIGRAD, M. and SHAO, Q. M. (1995). Estimation of the variance of partial sums for ρ-mixing random variables. *J. Multivariate Anal.* **52** 140–157. MR1325375 https://doi.org/10.1006/jmva.1995.1008

PRIESTLEY, M. B. (1988). *Nonlinear and Nonstationary Time Series Analysis*. Academic Press, London. MR0991969

SEN, A. and SRIVASTAVA, M. S. (1975). On tests for detecting change in mean. *Ann. Statist.* **3** 98–108. MR0362649

SHAO, X. and WU, W. B. (2007). Asymptotic spectral theory for nonlinear time series. *Ann. Statist.* **35** 1773–1801. MR2351105 https://doi.org/10.1214/009053606000001479

SIEGMUND, D. (1988). Confidence sets in change-point problems. *Int. Stat. Rev.* **56** 31–48. MR0963139 https://doi.org/10.2307/1403360

TONG, H. (1990). *Nonlinear Time Series*. *Oxford Statistical Science Series* **6**. Clarendon Press, Oxford University Press, New York. MR1079320

VOGT, M. and DETTE, H. (2015). Detecting gradual changes in locally stationary processes. *Ann. Statist.* **43** 713–740. MR3319141 https://doi.org/10.1214/14-AOS1297

WOROBEY, M. (2021). Dissecting the early COVID-19 cases in Wuhan. *Science* **374** 1202–1204.

WORSLEY, K. J. (1986). Confidence regions and test for a change-point in a sequence of exponential family random variables. *Biometrika* **73** 91–104. MR0836437 https://doi.org/10.1093/biomet/73.1.91

WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **102** 14150–14154. MR2172215 https://doi.org/10.1073/pnas.0506715102

WU, W. B. (2011). Asymptotic theory for stationary processes. *Stat*. *Interface* **4** 207–226. MR2812816 https://doi.org/10.4310/SII.2011.v4.n2.a15

ZHOU, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *J*. *Amer*. *Statist*. *Assoc*. **108** 726–740. MR3174655 https://doi.org/10.1080/01621459.2013.787184

The Annals of Statistics
2024, Vol. 52, No. 6, 2931–2955
https://doi.org/10.1214/24-AOS2452

STATISTICAL INFERENCE FOR DECENTRALIZED FEDERATED LEARNING

BY JIA GU[1,a] AND SONG XI CHEN[2,b]

[1]*Center for Data Science, Zhejiang University,* [a]*gujia@zju.edu.cn*
[2]*Department of Statistics and Data Science, Tsinghua University,* [b]*sxchen@tsinghua.edu.cn*

This paper considers decentralized Federated Learning (FL) under heterogeneous distributions among distributed clients or data blocks for the M-estimation. The mean squared error and consensus error across the estimators from different clients via the decentralized stochastic gradient descent algorithm are derived. The asymptotic normality of the Polyak–Ruppert (PR) averaged estimator in the decentralized distributed setting is attained, which shows that its statistical efficiency comes at a cost as it is more restrictive on the number of clients than that in the distributed M-estimation. To overcome the restriction, a one-step estimator is proposed which permits a much larger number of clients while still achieving the same efficiency as the original PR-averaged estimator in the nondistributed setting. The confidence regions based on both the PR-averaged estimator and the proposed one-step estimator are constructed to facilitate statistical inference for decentralized FL.

REFERENCES

ALGHUNAIM, S. A. and YUAN, K. (2022). A unified and refined convergence analysis for non-convex decentralized learning. *IEEE Trans. Signal Process.* **70** 3264–3279. MR4449095 https://doi.org/10.1109/tsp.2022.3184770

APPLE (2019). Private federated learning. NeurIPS 2019 Expo Talk.

BACH, F. (2010). Self-concordant analysis for logistic regression. *Electron. J. Stat.* **4** 384–414. MR2645490 https://doi.org/10.1214/09-EJS521

BACH, F. (2014). Adaptivity of averaged stochastic gradient descent to local strong convexity for logistic regression. *J. Mach. Learn. Res.* **15** 595–627. MR3190851

BICKEL, P. J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70** 428–434. MR0386168

BOTTOU, L., CURTIS, F. E. and NOCEDAL, J. (2018). Optimization methods for large-scale machine learning. *SIAM Rev.* **60** 223–311. MR3797719 https://doi.org/10.1137/16M1080173

BOYD, S., GHOSH, A., PRABHAKAR, B. and SHAH, D. (2006). Randomized gossip algorithms. *IEEE Trans. Inf. Theory* **52** 2508–2530. MR2238556 https://doi.org/10.1109/TIT.2006.874516

CHEN, M. and CUI, S. (2024). Federated learning for autonomous vehicles control. In *Communication Efficient Federated Learning for Wireless Networks* 129–150. Springer, Switzerland. https://doi.org/10.1007/978-3-031-51266-7_6

CHEN, X., LAI, Z., LI, H. et al. (2024). Online statistical inference for stochastic optimization via Kiefer–Wolfowitz methods. *J. Amer. Statist. Assoc.* 1–24. https://doi.org/10.1080/01621459.2023.2296703

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

CHEN, Y., QIN, X., WANG, J. et al. (2020b). FedHealth: A federated transfer learning framework for wearable healthcare. *IEEE Intell. Syst.* **35** 83–93. https://doi.org/10.1109/MIS.2020.2988604

CHOI, W. and KIM, J. (2022). On the convergence of decentralized gradient descent with diminishing stepsize. Revisited.

CHUNG, K. L. (1954). On a stochastic approximation method. *Ann. Math. Stat.* **25** 463–483. MR0064365 https://doi.org/10.1214/aoms/1177728716

FANG, Y., XU, J. and YANG, L. (2018). Online bootstrap confidence intervals for the stochastic gradient descent estimator. *J. Mach. Learn. Res.* **19** 1–21. MR3899780

MSC2020 subject classifications. Primary 62-08; secondary 62L12.
Key words and phrases. Decentralized estimation, decentralized stochastic gradient descent, federated learning, heterogeneity, one-step estimation.

GU, J. and CHEN, S. X. (2023). Distributed statistical inference under heterogeneity. *J. Mach. Learn. Res.* **24** 1–56. MR4720843

GU, J. and CHEN, S. X. (2024). Supplement to "Statistical Inference for Decentralized Federated Learning." https://doi.org/10.1214/24-AOS2452SUPP

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

HARD, A., RAO, K., MATHEWS, R. et al. (2018). Federated learning for mobile keyboard prediction.

JOHNSON, R. and ZHANG, T. (2013). Accelerating stochastic gradient descent using predictive variance reduction. *Adv. Neural Inf. Process. Syst.* **1** 315–323.

KAIROUZ, P., MCMAHAN, H. B., AVENT, B. et al. (2021). Advances and open problems in federated learning. *Found. Trends Mach. Learn.* **14** 1–210. https://doi.org/10.1561/2200000083

KOLOSKOVA, A., STICH, S. and JAGGI, M. (2019). Decentralized stochastic optimization and gossip algorithms with compressed communication. In *Proceedings of the* 36*th International Conference on Machine Learning* **97** 3478–3487.

LAI, T. L. (2003). Stochastic approximation. *Ann. Statist.* **31** 391–406. MR1983535 https://doi.org/10.1214/aos/1051027873

LEE, S., LIAO, Y., SEO, M. H. et al. (2022). Fast and robust online inference with stochastic gradient descent via random scaling. In *The AAAI Conference on Artificial Intelligence*.

LI, T., SAHU, A., TALWALKAR, A. et al. (2020). Federated learning: Challenges, methods, and future directions. *IEEE Signal Process. Mag.* **37** 50–60. https://doi.org/10.1109/MSP.2020.2975749

LI, X., LIANG, J., CHANG, X. et al. (2022). Statistical estimation and online inference via local SGD. In *Conference on Learning Theory* (*COLT*) **178** 1613–1661.

LIAN, X., ZHANG, C., ZHANG, H. et al. (2017). Can decentralized algorithms outperform centralized algorithms? A case study for decentralized parallel stochastic gradient descent. In *Advances in Neural Information Processing Systems* **30**.

MCMAHAN, B., MOORE, E., RAMAGE, D. et al. (2017). Communication-efficient learning of deep networks from decentralized data. *Int. Conf. Artif. Intell. Stat.* **54** 1273–1282.

NEDIĆ, A., OLSHEVSKY, A. and RABBAT, M. G. (2018). Network topology and communication-computation tradeoffs in decentralized optimization. *Proc. IEEE* **106** 953–976. https://doi.org/10.1109/JPROC.2018.2817461

NEDIĆ, A., OLSHEVSKY, A. and SHI, W. (2017). Achieving geometric convergence for distributed optimization over time-varying graphs. *SIAM J. Optim.* **27** 2597–2633. MR3738851 https://doi.org/10.1137/16M1084316

NGUYEN, L. M., NGUYEN, P. H., RICHTÁRIK, P., SCHEINBERG, K., TAKÁČ, M. and VAN DIJK, M. (2019). New convergence aspects of stochastic gradient algorithms. *J. Mach. Learn. Res.* **20** Paper No. 176, 49. MR4048987

PANTELOPOULOS, A. and BOURBAKIS, N. G. (2010). A survey on wearable sensor-based systems for health monitoring and prognosis. *IEEE Trans. Syst. Man Cybern.*, *Part C Appl. Rev.* **40** 1–12. https://doi.org/10.1109/TSMCC.2009.2032660

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

QIAN, N. (1999). On the momentum term in gradient descent learning algorithms. *Neural Netw.* **12** 145–151. https://doi.org/10.1016/S0893-6080(98)00116-6

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

ROBBINS, H. and SIEGMUND, D. (1971). A convergence theorem for non negative almost supermartingales and some applications. In *Optimizing Methods in Statistics* (*Proc. Sympos.*, *Ohio State Univ.*, *Columbus*, *Ohio*, 1971) 233–257. Academic Press, New York. MR0343355

RUPPERT, D. (1988). Efficient estimations from a slowly convergent Robbins–Monro process Technical report, Cornell Univ. Operations Research and Industrial Engineering.

SACKS, J. (1958). Asymptotic distribution of stochastic approximation procedures. *Ann. Math. Stat.* **29** 373–405. MR0098427 https://doi.org/10.1214/aoms/1177706619

STICH, S. U. (2019). Local SGD converges fast and communicates little. In *International Conference on Learning Representations*.

SU, W. J. and ZHU, Y. (2018). Uncertainty Quantification for Online Learning and Stochastic Approximation via Hierarchical Incremental Gradient Descent. arXiv. Available at arXiv:10.48550/ARXIV.1802.04876.

WANG, J., CHARLES, Z., XU, Z. et al. (2021). A field guide to federated optimization.

WANG, J. and JOSHI, G. (2021). Cooperative SGD: A unified framework for the design and analysis of local-update SGD algorithms. *J. Mach. Learn. Res.* **22** 1–50. MR4329792

YUAN, K., ALGHUNAIM, S. A., YING, B. and SAYED, A. H. (2020). On the influence of bias-correction on distributed stochastic optimization. *IEEE Trans. Signal Process.* **68** 4352–4367. MR4144907 https://doi.org/10.1109/TSP.2020.3008605

YUAN, K., LING, Q. and YIN, W. (2016). On the convergence of decentralized gradient descent. *SIAM J. Optim.* **26** 1835–1854. MR3544854 https://doi.org/10.1137/130943170

ZHANG, Y., DUCHI, J. C. and WAINWRIGHT, M. J. (2013). Communication-efficient algorithms for statistical optimization. *J. Mach. Learn. Res.* **14** 3321–3363. MR3144464

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
2024, Vol. 52, No. 6, 2956–2978
https://doi.org/10.1214/24-AOS2469

INCREASING DIMENSION ASYMPTOTICS FOR TWO-WAY CROSSED MIXED EFFECT MODELS

BY ZIYANG LYU[1,a], S.A. SISSON[1,b] AND A.H. WELSH[2,c]

[1]*UNSW Data Science Hub, and School of Mathematics and Statistics, University of New South Wales,* [a]*Ziyang.Lyu@unsw.edu.au,* [b]*Scott.Sisson@unsw.edu.au*

[2]*Research School of Finance, Actuarial Studies and Statistics, Australian National University,* [c]*Alan.Welsh@anu.edu.au*

This paper presents asymptotic results for the maximum likelihood and restricted maximum likelihood (REML) estimators within a two-way crossed mixed effect model, when the number of rows, columns, and the number of observations per cell tend to infinity. The relative growth rate for the number of rows, columns, and cells is unrestricted, whether considered pairwise or collectively. Under very mild conditions (which include moment conditions instead of requiring normality for either the random effects or errors), the estimators are proven to be asymptotically normal, with a structured covariance matrix. We also discuss the case where the number of observations per cell is fixed at 1.

REFERENCES

[1] ANDERSON, T. (1969). *Statistical Inference for Covariance Matrices with Linear Structure*. Academic Press, New York.

[2] BAEY, C., COURNÈDE, P.-H. and KUHN, E. (2019). Asymptotic distribution of likelihood ratio test statistics for variance components in nonlinear mixed effects models. *Comput. Statist. Data Anal.* **135** 107–122. MR3915362 https://doi.org/10.1016/j.csda.2019.01.014

[3] BATTESE, G. E., HARTER, R. M. and FULLER, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.* **83** 28–36.

[4] BATTEY, H. S. and MCCULLAGH, P. (2024). An anomaly arising in the analysis of processes with more than one source of variability. *Biometrika* **111** 677–689. MR4745588 https://doi.org/10.1093/biomet/asad044

[5] BELLIO, R., GHOSH, S., OWEN, A. B. and VARIN, C. (2023). Scalable estimation of probit models with crossed random effects.

[6] BICKEL, P. J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70** 428–434. MR0386168

[7] CRESSIE, N. and LAHIRI, S. N. (1993). The asymptotic distribution of REML estimators. *J. Multivariate Anal.* **45** 217–233. MR1221918 https://doi.org/10.1006/jmva.1993.1034

[8] DAS, K. (1979). Asymptotic optimality of restricted maximum likelihood estimates for the mixed model. *Calcutta Statist. Assoc. Bull.* **28** 125–142. MR0586086 https://doi.org/10.1177/0008068319790108

[9] EKVALL, K. O. and BOTTAI, M. (2022). Confidence regions near singular information and boundary points with applications to mixed models. *Ann. Statist.* **50** 1806–1832. MR4441141 https://doi.org/10.1214/22-aos2177

[10] EKVALL, K. O. and JONES, G. L. (2020). Consistent maximum likelihood estimation using subsets with applications to multivariate mixed models. *Ann. Statist.* **48** 932–952. MR4102682 https://doi.org/10.1214/19-AOS1830

[11] GHOSH, S., HASTIE, T. and OWEN, A. B. (2022). Scalable logistic regression with crossed random effects. *Electron. J. Stat.* **16** 4604–4635. MR4489236 https://doi.org/10.1214/22-ejs2047

[12] GHOSH, S., HASTIE, T. and OWEN, A. B. (2022). Backfitting for large scale crossed random effects regressions. *Ann. Statist.* **50** 560–583. MR4382028 https://doi.org/10.1214/21-aos2121

[13] HARTLEY, H. O. and RAO, J. N. K. (1967). Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika* **54** 93–108. MR0216684 https://doi.org/10.1093/biomet/54.1-2.93

MSC2020 subject classifications. 62E20, 62J05, 62F12.

Key words and phrases. Crossed random effect, maximum likelihood estimator, Kronecker product, variance components, asymptotic independence.

[14] HARVILLE, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *J. Amer. Statist. Assoc.* **72** 320–340. MR0451550

[15] JIANG, J. (1998). Asymptotic properties of the empirical BLUP and BLUE in mixed linear models. *Statist. Sinica* **8** 861–885. MR1651513

[16] JIANG, J. (2013). The subset argument and consistency of MLE in GLMM: Answer to an open problem and beyond. *Ann. Statist.* **41** 177–195. MR3059414 https://doi.org/10.1214/13-AOS1084

[17] JIANG, J. (2017). *Asymptotic Analysis of Mixed Effects Models: Theory, Applications, and Open Problems. Monographs on Statistics and Applied Probability* **155**. CRC Press, Boca Raton, FL. MR3676375

[18] JIANG, J., WAND, M. P. and BHASKARAN, A. (2022). Usable and precise asymptotics for generalized linear mixed model analysis and design. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 55–82. MR4400390

[19] LAIRD, N. M. and WARE, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38** 963–974.

[20] LYU, Z., SISSON, S. and WELSH, A. (2024). Supplement to "Increasing dimension asymptotics for two-way crossed mixed effect models." https://doi.org/10.1214/24-AOS2469SUPPA, https://doi.org/10.1214/24-AOS2469SUPPB

[21] LYU, Z. and WELSH, A. H. (2022). Increasing cluster size asymptotics for nested error regression models. *J. Statist. Plann. Inference* **217** 52–68. MR4292374 https://doi.org/10.1016/j.jspi.2021.07.009

[22] LYU, Z. and WELSH, A. H. (2022). Asymptotics for EBLUPs: Nested error regression models. *J. Amer. Statist. Assoc.* **117** 2028–2042. MR4528487 https://doi.org/10.1080/01621459.2021.1895178

[23] LYU, Z. and WELSH, A. H. (2023). Small area estimation using EBLUPs under the nested error regression model. *Statist. Sinica*.

[24] MENICTAS, M., CREDICO, G. D. and WAND, M. P. (2023). Streamlined variational inference for linear mixed models with crossed random effects. *J. Comput. Graph. Statist.* **32** 99–115. MR4552940 https://doi.org/10.1080/10618600.2022.2096622

[25] MILLER, J. J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *Ann. Statist.* **5** 746–762. MR0448661

[26] ORTEGA, J. M. and RHEINBOLDT, W. C. (1973). *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York.

[27] PATEFIELD, W. M. (1977). On the maximized likelihood function. *Sankhyā, Ser. B* **39** 92–96. MR0652317

[28] RICHARDSON, A. M. and WELSH, A. H. (1994). Asymptotic properties of restricted maximum likelihood (REML) estimates for hierarchical mixed linear models. *Aust. J. Stat.* **36** 31–43. MR1309503 https://doi.org/10.1111/j.1467-842x.1994.tb00636.x

[29] SCOTT, A. J. and HOLT, D. (1982). The effect of two-stage sampling on ordinary least squares methods. *J. Amer. Statist. Assoc.* **77** 848–854.

[30] SEARLE, S. R. and HENDERSON, H. V. (1979). Dispersion matrices for variance components models. *J. Amer. Statist. Assoc.* **74** 465–470. MR0548045

[31] STERN, S. E. and WELSH, A. H. (2000). Likelihood inference for small variance components. *Canad. J. Statist.* **28** 517–532. MR1793108 https://doi.org/10.2307/3315962

[32] YOON, H.-J. and WELSH, A. H. (2020). On the effect of ignoring correlation in the covariates when fitting linear mixed models. *J. Statist. Plann. Inference* **204** 18–34. MR3961926 https://doi.org/10.1016/j.jspi.2019.04.001

VOLUME 52

2024

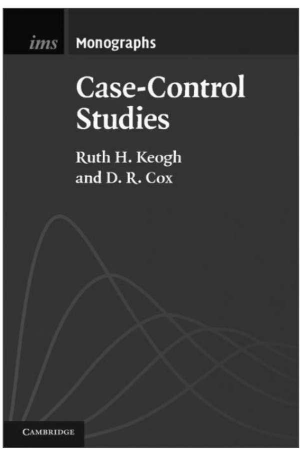
Articles

Erratum

Correction

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.