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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

Vol. 53, No. 1—February 2025

THE ANNALS OF STATISTICS Vol. 53, No. 1, pp. 1–456 February 2025

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

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

The Annals of Statistics **[ISSN 0090-5364 (print); ISSN 2168-8966 (online)],** Volume 53, Number 1, February 2025. 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
2025, Vol. 53, No. 1, 1–36
https://doi.org/10.1214/24-AOS2431

EMPIRICAL PARTIALLY BAYES MULTIPLE TESTING AND COMPOUND χ^2 DECISIONS

BY NIKOLAOS IGNATIADIS[1,a] AND BODHISATTVA SEN[2,b]

[1]*Department of Statistics and Data Science Institute, University of Chicago,* [a]*ignat@uchicago.edu*
[2]*Department of Statistics, Columbia University,* [b]*bodhi@stat.columbia.edu*

A common task in high-throughput biology is to screen for associations across thousands of units of interest, for example, genes or proteins. Often, the data for each unit are modeled as Gaussian measurements with unknown mean and variance and are summarized as per-unit sample averages and sample variances. The downstream goal is multiple testing for the means. In this domain, it is routine to "moderate" (i.e., to shrink) the sample variances through parametric empirical Bayes methods before computing p-values for the means. Such an approach is asymmetric in that a prior is posited and estimated for the nuisance parameters (variances) but not the primary parameters (means). Our work initiates the formal study of this paradigm, which we term "empirical partially Bayes multiple testing." In this framework, if the prior for the variances were known, one could proceed by computing p-values conditional on the sample variances—a strategy called partially Bayes inference by Sir David Cox. We show that these conditional p-values satisfy an Eddington/Tweedie-type formula and are approximated at nearly-parametric rates when the prior is estimated by nonparametric maximum likelihood. The estimated p-values can be used with the Benjamini–Hochberg procedure to guarantee asymptotic control of the false discovery rate. Even in the compound setting, wherein the variances are fixed, the approach retains asymptotic type-I error guarantees.

REFERENCES

ARMSTRONG, T. B. (2022). False discovery rate adjustments for average significance level controlling tests. Available at arXiv:2209.13686.

ARMSTRONG, T. B., KOLESÁR, M. and PLAGBORG-MØLLER, M. (2022). Robust empirical Bayes confidence intervals. *Econometrica* **90** 2567–2602. MR4524894 https://doi.org/10.3982/ecta18597

BALDI, P. and LONG, A. D. (2001). A Bayesian framework for the analysis of microarray expression data: Regularized t -test and statistical inferences of gene changes. *Bioinformatics* **17** 509–519. https://doi.org/10.1093/bioinformatics/17.6.509

BANERJEE, T., FU, L. J., JAMES, G. M. and SUN, W. (2021). Nonparametric empirical Bayes estimation on heterogeneous data. Available at arXiv:2002.12586.

BATTEY, H. S., COX, D. R. and LEE, S. H. (2024). On partial likelihood and the construction of factorisable transformations. *Inf. Geom.* **7** S9–S28. MR4682861 https://doi.org/10.1007/s41884-022-00068-8

BAYARRI, M. J. and BERGER, J. O. (2000). p values for composite null models. *J. Amer. Statist. Assoc.* **95** 1127–1142, 1157–1170. MR1804239 https://doi.org/10.2307/2669749

BELOMESTNY, D. and GOLDENSHLUGER, A. (2020). Nonparametric density estimation from observations with multiplicative measurement errors. *Ann. Inst. Henri Poincaré Probab. Stat.* **56** 36–67. MR4058981 https://doi.org/10.1214/18-AIHP954

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

BRENNER MIGUEL, S., COMTE, F. and JOHANNES, J. (2021). Spectral cut-off regularisation for density estimation under multiplicative measurement errors. *Electron. J. Stat.* **15** 3551–3573. MR4298979 https://doi.org/10.1214/21-ejs1870

MSC2020 subject classifications. Primary 62C12, 62C25; secondary 62P10.
Key words and phrases. Partially Bayes inference, conditionality principle, nonparametric maximum likelihood, Benjamini-Hochberg, average significance control.

BUTZER, P. L. and JANSCHE, S. (1999). A self-contained approach to Mellin transform analysis for square integrable functions; applications. *Integral Transforms Spec. Funct.* **8** 175–198. MR1771422 https://doi.org/10.1080/10652469908819226

CASTILLO, I. and ROQUAIN, É. (2020). On spike and slab empirical Bayes multiple testing. *Ann. Statist.* **48** 2548–2574. MR4152112 https://doi.org/10.1214/19-AOS1897

CHEN, J. (2024). Empirical Bayes when estimation precision predicts parameters. Available at arXiv:2212.14444.

COPAS, J. B. (1969). Compound decisions and empirical Bayes. (With discussion.). *J. Roy. Statist. Soc. Ser. B* **31** 397–425. MR0269013

COX, D. R. (2005). *Selected Statistical Papers of Sir David Cox. Vol. II: Foundations of Statistical Inference, Theoretical Statistics, Time Series and Stochastic Processes*. Cambridge Univ. Press, Cambridge. MR2238854

COX, D. R. (1975). A note on partially Bayes inference and the linear model. *Biometrika* **62** 651–654. MR0408131 https://doi.org/10.1093/biomet/62.3.651

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics*. CRC Press, London. MR0370837

CRESSIE, N. (1982). A useful empirical Bayes identity. *Ann. Statist.* **10** 625–629. MR0653538

CUI, X. and CHURCHILL, G. A. (2003). Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol.* **4** 210. https://doi.org/10.1186/gb-2003-4-4-210

CUI, X., HWANG, J. T. G., QIU, J., BLADES, N. J. and CHURCHILL, G. A. (2005). Improved statistical tests for differential gene expression by shrinking variance components estimates. *Biostatistics* **6** 59–75. https://doi.org/10.1093/biostatistics/kxh018

DE MENEZES, R. X., BOER, J. M. and VAN HOUWELINGEN, H. C. (2004). Microarray data analysis: A hierarchical T-test to handle heteroscedasticity. *Appl. Bioinform.* **3** 229–235.

DELMAR, P., ROBIN, S., TRONIK-LE ROUX, D. and DAUDIN, J. J. (2005). Mixture model on the variance for the differential analysis of gene expression data. *J. Roy. Statist. Soc. Ser. C* **54** 31–50. MR2134596 https://doi.org/10.1111/j.1467-9876.2005.00468.x

DICKER, L. H. and ZHAO, S. D. (2016). High-dimensional classification via nonparametric empirical Bayes and maximum likelihood inference. *Biometrika* **103** 21–34. MR3465819 https://doi.org/10.1093/biomet/asv067

DIETRICH, S., OLEŚ, M., LU, J., SELLNER, L., ANDERS, S., VELTEN, B., WU, B., HÜLLEIN, J., DA SILVA LIBERIO, M. et al. (2017). Drug-perturbation-based stratification of blood cancer. *J. Clin. Invest.* **128** 427–445.

DU, L. and HU, I. (2022). An empirical Bayes method for chi-squared data. *J. Amer. Statist. Assoc.* **117** 334–347. MR4399089 https://doi.org/10.1080/01621459.2020.1777137

DU, L. and ZHANG, C. (2014). Single-index modulated multiple testing. *Ann. Statist.* **42** 30–79. MR3226157 https://doi.org/10.1214/14-AOS1222

DYSON, F. (1926). A method for correcting series of parallax observations. *Mon. Not. R. Astron. Soc.* **86** 686.

EBRAHIMPOOR, M. and GOEMAN, J. J. (2021). Inflated false discovery rate due to volcano plots: Problem and solutions. *Brief. Bioinform.* **22**. https://doi.org/10.1093/bib/bbab053

EFRON, B. (2007). Size, power and false discovery rates. *Ann. Statist.* **35** 1351–1377. MR2351089 https://doi.org/10.1214/009053606000001460

EFRON, B. (2010). *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction. Institute of Mathematical Statistics (IMS) Monographs* **1**. Cambridge Univ. Press, Cambridge. MR2724758 https://doi.org/10.1017/CBO9780511761362

EFRON, B. (2011). Tweedie's formula and selection bias. *J. Amer. Statist. Assoc.* **106** 1602–1614. MR2896860 https://doi.org/10.1198/jasa.2011.tm11181

EFRON, B. (2014). Two modeling strategies for empirical Bayes estimation. *Statist. Sci.* **29** 285–301. MR3264543 https://doi.org/10.1214/13-STS455

EFRON, B. and MORRIS, C. (1973). Stein's estimation rule and its competitors—an empirical Bayes approach. *J. Amer. Statist. Assoc.* **68** 117–130. MR0388597

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

EPSTEIN, B. (1948). Some applications of the Mellin transform in statistics. *Ann. Math. Stat.* **19** 370–379. MR0029128 https://doi.org/10.1214/aoms/1177730201

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

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

FU, L., GANG, B., JAMES, G. M. and SUN, W. (2022). Heteroscedasticity-adjusted ranking and thresholding for large-scale multiple testing. *J. Amer. Statist. Assoc.* **117** 1028–1040. MR4436330 https://doi.org/10.1080/01621459.2020.1840992

GENOVESE, C. and WASSERMAN, L. (2002). Operating characteristics and extensions of the false discovery rate procedure. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 499–517. MR1924303 https://doi.org/10.1111/1467-9868.00347

GHOSAL, S. and VAN DER VAART, A. W. (2001). Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.* **29** 1233–1263. MR1873329 https://doi.org/10.1214/aos/1013203453

GREENSHTEIN, E. and RITOV, Y. (2022). Generalized maximum likelihood estimation of the mean of parameters of mixtures. With applications to sampling and to observational studies. *Electron. J. Stat.* **16** 5934–5954. MR4515714 https://doi.org/10.1214/22-ejs2082

GU, J. and KOENKER, R. (2017a). Unobserved heterogeneity in income dynamics: An empirical Bayes perspective. *J. Bus. Econom. Statist.* **35** 1–16. MR3591533 https://doi.org/10.1080/07350015.2015.1052457

GU, J. and KOENKER, R. (2017b). Empirical Bayesball remixed: Empirical Bayes methods for longitudinal data. *J. Appl. Econometrics* **32** 575–599. MR3631958 https://doi.org/10.1002/jae.2530

GU, J. and KOENKER, R. (2023). Invidious comparisons: Ranking and selection as compound decisions. *Econometrica* **91** 1–41. MR4562942

GU, J. and SHEN, S. (2018). Oracle and adaptive false discovery rate controlling methods for one-sided testing: Theory and application in treatment effect evaluation. *Econom. J.* **21** 11–35. MR3769546 https://doi.org/10.1111/ectj.12092

HOUDE, D., BERKOWITZ, S. A. and ENGEN, J. R. (2011). The utility of hydrogen/deuterium exchange mass spectrometry in biopharmaceutical comparability studies. *J. Pharm. Sci.* **100** 2071–2086. https://doi.org/10.1002/jps.22432

HWANG, J. T. G. and LIU, P. (2010). Optimal tests shrinking both means and variances applicable to microarray data analysis. *Stat. Appl. Genet. Mol. Biol.* **9** Art. 36, 35. MR2746019 https://doi.org/10.2202/1544-6115.1587

IGNATIADIS, N. and HUBER, W. (2021). Covariate powered cross-weighted multiple testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 720–751. MR4319999 https://doi.org/10.1111/rssb.12411

IGNATIADIS, N., SAHA, S., SUN, D. L. and MURALIDHARAN, O. (2023). Empirical Bayes mean estimation with nonparametric errors via order statistic regression on replicated data. *J. Amer. Statist. Assoc.* **118** 987–999. MR4595471 https://doi.org/10.1080/01621459.2021.1967164

IGNATIADIS, N. and SEN, B. (2025). Supplement to "Empirical partially Bayes multiple testing and compound χ^2 decisions." https://doi.org/10.1214/24-AOS2431SUPPA, https://doi.org/10.1214/24-AOS2431SUPPB

IGNATIADIS, N. and WAGER, S. (2022). Confidence intervals for nonparametric empirical Bayes analysis. *J. Amer. Statist. Assoc.* **117** 1149–1166. MR4480697 https://doi.org/10.1080/01621459.2021.2008403

IGNATIADIS, N., WANG, R. and RAMDAS, A. (2024). E-values as unnormalized weights in multiple testing. *Biometrika* **111** 417–439. MR4745574 https://doi.org/10.1093/biomet/asad057

JANA, S., POLYANSKIY, Y. and WU, Y. (2022). Optimal empirical Bayes estimation for the Poisson model via minimum-distance methods. Available at arXiv:2209.01328.

JEFFERY, I. B., HIGGINS, D. G. and CULHANE, A. C. (2006). Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. *BMC Bioinform.* **7** 359. https://doi.org/10.1186/1471-2105-7-359

JEWELL, N. P. (1982). Mixtures of exponential distributions. *Ann. Statist.* **10** 479–484. MR0653523

JIANG, W. (2020). On general maximum likelihood empirical Bayes estimation of heteroscedastic IID normal means. *Electron. J. Stat.* **14** 2272–2297. MR4109006 https://doi.org/10.1214/20-EJS1717

JIANG, W. and ZHANG, C.-H. (2009). General maximum likelihood empirical Bayes estimation of normal means. *Ann. Statist.* **37** 1647–1684. MR2533467 https://doi.org/10.1214/08-AOS638

KAMMERS, K., COLE, R. N., TIENGWE, C. and RUCZINSKI, I. (2015). Detecting significant changes in protein abundance. *EuPA Open Proteomics* **7** 11–19. https://doi.org/10.1016/j.euprot.2015.02.002

KERR, M. K., MARTIN, M. and CHURCHILL, G. A. (2000). Analysis of variance for gene expression microarray data. *J. Comput. Biol.* **7** 819–837. https://doi.org/10.1089/10665270050514954

KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Stat.* **27** 887–906. MR0086464 https://doi.org/10.1214/aoms/1177728066

KIM, Y., CARBONETTO, P., STEPHENS, M. and ANITESCU, M. (2020). A fast algorithm for maximum likelihood estimation of mixture proportions using sequential quadratic programming. *J. Comput. Graph. Statist.* **29** 261–273. MR4116040 https://doi.org/10.1080/10618600.2019.1689985

KLAUS, B. and REISENAUER, S. (2018). An end to end workflow for differential gene expression using affymetrix microarrays. *F1000Res.* **5** 1384.

KOENKER, R. and GU, J. (2017). REBayes: An R package for empirical Bayes mixture methods. *J. Stat. Softw.* **82**.

KOENKER, R. and MIZERA, I. (2014). Convex optimization, shape constraints, compound decisions, and empirical Bayes rules. *J. Amer. Statist. Assoc.* **109** 674–685. MR3223742 https://doi.org/10.1080/01621459.2013.869224

KOOPERBERG, C., ARAGAKI, A., STRAND, A. D. and OLSON, J. M. (2005). Significance testing for small microarray experiments. *Stat. Med.* **24** 2281–2298. MR2151706 https://doi.org/10.1002/sim.2109

KWON, Y. and ZHAO, Z. (2023). On F-modelling-based empirical Bayes estimation of variances. *Biometrika* **110** 69–81. MR4565444 https://doi.org/10.1093/biomet/asac019

LAUFER, C., FISCHER, B., BILLMANN, M., HUBER, W. and BOUTROS, M. (2013). Mapping genetic interactions in human cancer cells with RNAi and multiparametric phenotyping. *Nat. Methods* **10** 427–431. https://doi.org/10.1038/nmeth.2436

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

LI, A. and BARBER, R. F. (2019). Multiple testing with the structure-adaptive Benjamini-Hochberg algorithm. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 45–74. MR3904779

LINDSAY, B. G. (1983). The geometry of mixture likelihoods. II. The exponential family. *Ann. Statist.* **11** 783–792. MR0707929 https://doi.org/10.1214/aos/1176346245

LINDSAY, B. G. (1985). Using empirical partially Bayes inference for increased efficiency. *Ann. Statist.* **13** 914–931. MR0803748 https://doi.org/10.1214/aos/1176349646

LINDSAY, B. G. and ROEDER, K. (1993). Uniqueness of estimation and identifiability in mixture models. *Canad. J. Statist.* **21** 139–147. MR1234757 https://doi.org/10.2307/3315807

LÖNNSTEDT, I. (2005). Empirical Bayes methods for DNA microarray data Ph.D. thesis, Matematiska institutionen / Uppsala Univ., Department of Mathematics.

LÖNNSTEDT, I. and SPEED, T. (2002). Replicated microarray data *Statist. Sinica* **12** 31–46. MR1894187

LOVE, M. I., HUBER, W. and ANDERS, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15** 550. https://doi.org/10.1186/s13059-014-0550-8

LU, M. and STEPHENS, M. (2016). Variance adaptive shrinkage (vash): Flexible empirical Bayes estimation of variances. *Bioinformatics* **32** 3428–3434. https://doi.org/10.1093/bioinformatics/btw483

LU, M. and STEPHENS, M. (2019). Empirical Bayes estimation of normal means, accounting for uncertainty in estimated standard errors. Available at arXiv:1901.10679.

MAKSIMOVIC, J., PHIPSON, B. and OSHLACK, A. (2017). A cross-package bioconductor workflow for analysing methylation array data. F1000Res. 5.

MCCARTHY, D. J. and SMYTH, G. K. (2009). Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics* **25** 765–771. https://doi.org/10.1093/bioinformatics/btp053

MCCULLAGH, P. (1990). A note on partially Bayes inference for generalized linear models Technical Report No. 284, Department of Statistics, Univ. Chicago.

MENG, X.-L. (1994). Posterior predictive p-values. *Ann. Statist.* **22** 1142–1160. MR1311969 https://doi.org/10.1214/aos/1176325622

MOSEK APS (2020). The MOSEK optimization suite manual, Version 9.2.

MURALIDHARAN, O. (2012). High dimensional exponential family estimation via empirical Bayes. *Statist. Sinica* **22** 1217–1232. MR2987489

NEWTON, M. A., KENDZIORSKI, C. M., RICHMOND, C. S., BLATTNER, F. R. and TSUI, K. W. (2001). On differential variability of expression ratios: Improving statistical inference about gene expression changes from microarray data. *J. Comput. Biol.* **8** 37–52. https://doi.org/10.1089/106652701300099074

NEWTON, M. A., NOUEIRY, A., SARKAR, D. and AHLQUIST, P. (2004). Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics* **5** 155–176. https://doi.org/10.1093/biostatistics/5.2.155

PHIPSON, B., LEE, S., MAJEWSKI, I. J., ALEXANDER, W. S. and SMYTH, G. K. (2016). Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *Ann. Appl. Stat.* **10** 946–963. MR3528367 https://doi.org/10.1214/16-AOAS920

PLONER, A., CALZA, S., GUSNANTO, A. and PAWITAN, Y. (2006). Multidimensional local false discovery rate for microarray studies. *Bioinformatics* **22** 556–565. https://doi.org/10.1093/bioinformatics/btk013

POLYANSKIY, Y. and WU, Y. (2020). Self-regularizing property of nonparametric maximum likelihood estimator in mixture models. Available at arXiv:2008.08244.

RITCHIE, M. E., PHIPSON, B., WU, D., HU, Y., LAW, C. W., SHI, W. and SMYTH, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43** e47–e47.

ROBBINS, H. (1950). A generalization of the method of maximum likelihood: Estimating a mixing distribution (abstract). *Ann. Math. Stat.* **21** 314–315.

ROBBINS, H. (1951). Asymptotically subminimax solutions of compound statistical decision problems. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 1950 131–148. Univ. California Press, Berkeley–Los Angeles, CA. MR0044803

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

ROBBINS, H. (1982). Estimating many variances. In *Statistical Decision Theory and Related Topics*, *III*, *Vol.* 2 (*West Lafayette*, *Ind.*, 1981) 251–261. Academic Press, New York. MR0705317

ROBINS, J. M., VAN DER VAART, A. and VENTURA, V. (2000). Asymptotic distribution of p values in composite null models. *J. Amer. Statist. Assoc.* **95** 1143–1167, 1171–1172. MR1804240 https://doi.org/10.2307/2669750

ROBINSON, M. D., MCCARTHY, D. J. and SMYTH, G. K. (2010). edgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26** 139–140. https://doi.org/10.1093/bioinformatics/btp616

ROQUAIN, E. and VERZELEN, N. (2022). False discovery rate control with unknown null distribution: Is it possible to mimic the oracle? *Ann. Statist.* **50** 1095–1123. MR4404913 https://doi.org/10.1214/21-aos2141

ROSS, N. (2011). Fundamentals of Stein's method. *Probab. Surv.* **8** 210–293. MR2861132 https://doi.org/10.1214/11-PS182

SMYTH, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3** Art. 3, 29. MR2101454 https://doi.org/10.2202/1544-6115.1027

SMYTH, G. K. (2022). On linear modeling assumptions for genomic technologies. Reply in Bioconductor Support forum. Available at https://web.archive.org/web/20231115170405/https://support.bioconductor.org/p/9142432/.

SOLOFF, J. A. GUNTUBOYINA, A. SEN, B. (2024). Multivariate, heteroscedastic empirical Bayes via nonparametric maximum likelihood. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* qkae040.

STARK, R., GRZELAK, M. and HADFIELD, J. (2019). RNA sequencing: The teenage years. *Nat. Rev. Genet.* **20** 631–656. https://doi.org/10.1038/s41576-019-0150-2

STEPHENS, M. (2017). False discovery rates: A new deal. *Biostatistics* **18** 275–294.

STOREY, J. D., TAYLOR, J. E. and SIEGMUND, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: A unified approach. *J. Roy. Statist. Soc. Ser. B* **66** 187–205.

SUN, W. and CAI, T. T. (2007). Oracle and adaptive compound decision rules for false discovery rate control. *J. Amer. Statist. Assoc.* **102** 901–912.

SUN, W. and MCLAIN, A. C. (2012). Multiple testing of composite null hypotheses in heteroscedastic models. *J. Amer. Statist. Assoc.* **107** 673–687.

TEICHER, H. (1961). Identifiability of mixtures. *Ann. Math. Stat.* **32** 244–248.

TERKELSEN, T., PERNEMALM, M., GROMOV, P., BØRRESEN-DALE, A.-L., KROGH, A., HAAKENSEN, V. D., LETHIÖ, J., PAPALEO, E. and GROMOVA, I. (2021). High-throughput proteomics of breast cancer interstitial fluid: Identification of tumor subtype-specific serologically relevant biomarkers. *Mol. Oncol.* **15** 429–461.

TUSHER, V. G., TIBSHIRANI, R. and CHU, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA* **98** 5116–5121.

VAN HOUWELINGEN, H. C. (2014). The role of empirical Bayes methodology as a leading principle in modern medical statistics. *Biom. J.* **56** 919–932. https://doi.org/10.1002/bimj.201400073

VAN HOUWELINGEN, J. C. (1976). Monotone empirical Bayes tests for the continuous one-parameter exponential family. *Ann. Statist.* **4** 981–989.

WALSH, J. E. (1949). On the "information" lost by using a T-test when the population variance is known. *J. Amer. Statist. Assoc.* **44** 122–125.

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

WEINSTEIN, A., MA, Z., BROWN, L. D. and ZHANG, C.-H. (2018). Group-linear empirical Bayes estimates for a heteroscedastic normal mean. *J. Amer. Statist. Assoc.* **113** 698–710.

WEIS, D. D. (2019). Comment on Houde, D.; Berkowitz, S. A.; Engen, J. R., The utility of hydrogen/deuterium exchange mass spectrometry in biopharmaceutical comparability studies. J. Pharm. Sci. 2011, 100, 2071–2086. *J. Pharm. Sci.* **108** 807–810.

WITTEN, D. M. and TIBSHIRANI, R. (2007). A comparison of fold-change and the t-statistic for microarray data analysis Technical Report, Stanford Univ.

WRIGHT, G. W. and SIMON, R. M. (2003). A random variance model for detection of differential gene expression in small microarray experiments. *Bioinformatics* **19** 2448–2455.

XIE, D. and STEPHENS, M. (2022). Discussion of "Confidence intervals for nonparametric empirical Bayes analysis". *J. Amer. Statist. Assoc.* **117** 1186–1191. https://doi.org/10.1080/01621459.2022.2093727

YU, C. and HOFF, P. D. (2018). Adaptive multigroup confidence intervals with constant coverage. *Biometrika* **105** 319–335.

YU, D., HUBER, W. and VITEK, O. (2013). Shrinkage estimation of dispersion in negative binomial models for RNA-seq experiments with small sample size. *Bioinformatics* **29** 1275–1282. https://doi.org/10.1093/bioinformatics/btt143

ZHANG, C.-H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18** 806–831.

ZHANG, C.-H. (2003). Compound decision theory and empirical Bayes methods: Invited paper. *Ann. Statist.* **31** 379–390.

ZHANG, C.-H. (2009). Generalized maximum likelihood estimation of normal mixture densities. *Statist. Sinica* 1297–1318.

ZHANG, Y., MAKSIMOVIC, J., NASELLI, G., QIAN, J., CHOPIN, M., BLEWITT, M. E., OSHLACK, A. and HARRISON, L. C. (2013). Genome-wide DNA methylation analysis identifies hypomethylated genes regulated by FOXP3 in human regulatory T cells. *Blood* **122** 2823–2836. https://doi.org/10.1182/blood-2013-02-481788

ZHAO, Q., CHEN, Y., WANG, J. and SMALL, D. S. (2019). Powerful three-sample genome-wide design and robust statistical inference in summary-data Mendelian randomization. *Int. J. Epidemiol.* **48** 1478–1492.

ZHENG, Z., MERGAERT, A. M., ONG, I. M., SHELEF, M. A. and NEWTON, M. A. (2021). MixTwice: Large-scale hypothesis testing for peptide arrays by variance mixing. *Bioinformatics* **37** 2637–2643. https://doi.org/10.1093/bioinformatics/btab162

The Annals of Statistics
2025, Vol. 53, No. 1, 37–60
https://doi.org/10.1214/24-AOS2453

CONCENTRATION OF DISCREPANCY-BASED APPROXIMATE BAYESIAN COMPUTATION VIA RADEMACHER COMPLEXITY

BY SIRIO LEGRAMANTI[1,a], DANIELE DURANTE[2,b] AND PIERRE ALQUIER[3,c]

[1]*Department of Economics, University of Bergamo,* [a]*sirio.legramanti@unibg.it*

[2]*Department of Decision Sciences and Institute for Data Science and Analytics, Bocconi University,* [b]*daniele.durante@unibocconi.it*

[3]*Department of Information Systems, Decision Sciences and Statistics, ESSEC Business School,* [c]*alquier@essec.edu*

There has been an increasing interest on summary-free solutions for approximate Bayesian computation (ABC) that replace distances among summaries with discrepancies between the empirical distributions of the observed data and the synthetic samples generated under the proposed parameter values. The success of these strategies has motivated theoretical studies on the limiting properties of the induced posteriors. However, there is still the lack of a theoretical framework for summary-free ABC that *(i)* is unified, instead of discrepancy-specific, *(ii)* does not necessarily require to constrain the analysis to data generating processes and statistical models meeting specific regularity conditions, but rather facilitates the derivation of limiting properties that hold uniformly, and *(iii)* relies on verifiable assumptions that provide more explicit concentration bounds clarifying which factors govern the limiting behavior of the ABC posterior. We address this gap via a novel theoretical framework that introduces the concept of Rademacher complexity in the analysis of the limiting properties for discrepancy-based ABC posteriors, including in non-i.i.d. and misspecified settings. This yields a unified theory that relies on constructive arguments and provides more informative asymptotic results and uniform concentration bounds, even in those settings not covered by current studies. These key advancements are obtained by relating the asymptotic properties of summary-free ABC posteriors to the behavior of the Rademacher complexity associated with the chosen discrepancy within the family of integral probability semimetrics (IPS). The IPS class extends summary-based distances, and also includes the widely implemented Wasserstein distance and maximum mean discrepancy (MMD), among others. As clarified in specialized theoretical analyses of popular IPS discrepancies and via illustrative simulations, this new perspective improves the understanding of summary-free ABC.

REFERENCES

AGRAWAL, R. and HOREL, T. (2021). Optimal bounds between f-divergences and integral probability metrics. *J. Mach. Learn. Res.* **22** Paper No. 128, 59 pp. MR4279779

BARTLETT, P. L., BOUSQUET, O. and MENDELSON, S. (2005). Local Rademacher complexities. *Ann. Statist.* **33** 1497–1537. MR2166554 https://doi.org/10.1214/009053605000000282

BARTLETT, P. L. and MENDELSON, S. (2002). Rademacher and Gaussian complexities: Risk bounds and structural results. *J. Mach. Learn. Res.* **3** 463–482. MR1984026 https://doi.org/10.1162/153244303321897690

BERNTON, E., JACOB, P. E., GERBER, M. and ROBERT, C. P. (2019). Approximate Bayesian computation with the Wasserstein distance. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 235–269. MR3928142 https://doi.org/10.1111/rssb.12312

BIRRELL, J., DUPUIS, P., KATSOULAKIS, M. A., PANTAZIS, Y. and REY-BELLET, L. (2022). (f, Γ)-divergences: Interpolating between f-divergences and integral probability metrics. *J. Mach. Learn. Res.* **23** Paper No. 39, 70 pp. MR4420764

BISSIRI, P. G., HOLMES, C. C. and WALKER, S. G. (2016). A general framework for updating belief distributions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 1103–1130. MR3557191 https://doi.org/10.1111/rssb.12158

MSC2020 subject classifications. Primary 62F15; secondary 62C10, 62E20.

Key words and phrases. ABC, integral probability semimetrics, MMD, Rademacher complexity, Wasserstein distance.

CHÉRIEF-ABDELLATIF, B.-E. and ALQUIER, P. (2020). MMD-Bayes: Robust Bayesian estimation via maximum mean discrepancy. In *Symposium on Advances in Approximate Bayesian Inference*. *Proc. Mach. Learn. Res.* (*PMLR*) **118** 21. PMLR. MR4147569

CHÉRIEF-ABDELLATIF, B.-E. and ALQUIER, P. (2022). Finite sample properties of parametric MMD estimation: Robustness to misspecification and dependence. *Bernoulli* **28** 181–213. MR4337702 https://doi.org/10.3150/21-bej1338

DOUKHAN, P. (1994). *Mixing*: *Properties and Examples*. *Lecture Notes in Statistics* **85**. Springer, New York. MR1312160 https://doi.org/10.1007/978-1-4612-2642-0

DROVANDI, C. and FRAZIER, D. T. (2022). A comparison of likelihood-free methods with and without summary statistics. *Stat*. *Comput*. **32** Paper No. 42, 23 pp. MR4426803 https://doi.org/10.1007/s11222-022-10092-4

DUDLEY, R. M. (2018). *Real Analysis and Probability*. *Cambridge Studies in Advanced Mathematics* **74**. Cambridge Univ. Press, Cambridge. MR1932358 https://doi.org/10.1017/CBO9780511755347

DUDLEY, R. M., GINÉ, E. and ZINN, J. (1991). Uniform and universal Glivenko–Cantelli classes. *J. Theoret. Probab.* **4** 485–510. MR1115159 https://doi.org/10.1007/BF01210321

DUTTA, R., SCHOENGENS, M., PACCHIARDI, L., UMMADISINGU, A., WIDMER, N., KÜNZLI, P., ONNELA, J.-P. and MIRA, A. (2021). ABCpy: A high-performance computing perspective to approximate Bayesian computation. *J. Stat*. *Softw.* **100** 1–38.

FEARNHEAD, P. and PRANGLE, D. (2012). Constructing summary statistics for approximate Bayesian computation: Semi-automatic approximate Bayesian computation. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **74** 419–474. MR2925370 https://doi.org/10.1111/j.1467-9868.2011.01010.x

FORBES, F., NGUYEN, H. D., NGUYEN, T. T. and ARBEL, J. (2021). Approximate Bayesian computation with surrogate posteriors. Hal-03139256v4.

FOURNIER, N. and GUILLIN, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* **162** 707–738. MR3383341 https://doi.org/10.1007/s00440-014-0583-7

FRAZIER, D. T. (2020). Robust and efficient approximate Bayesian computation: A minimum distance approach. Preprint. Available at arXiv:2006.14126.

FRAZIER, D. T., KNOBLAUCH, J. and DROVANDI, C. (2024). The impact of loss estimation on Gibbs measures. Preprint. Available at arXiv:2404.15649.

FRAZIER, D. T., MARTIN, G. M., ROBERT, C. P. and ROUSSEAU, J. (2018). Asymptotic properties of approximate Bayesian computation. *Biometrika* **105** 593–607. MR3842887 https://doi.org/10.1093/biomet/asy027

FRAZIER, D. T., ROBERT, C. P. and ROUSSEAU, J. (2020). Model misspecification in approximate Bayesian computation: Consequences and diagnostics. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **82** 421–444. MR4084170 https://doi.org/10.1111/rssb.12356

FUJISAWA, M., TESHIMA, T., SATO, I. and SUGIYAMA, M. (2021). γ-ABC: Outlier-robust approximate Bayesian computation based on a robust divergence estimator. In *International Conference on Artificial Intelligence and Statistics* 1783–1791. PMLR.

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

GUTMANN, M. U., DUTTA, R., KASKI, S. and CORANDER, J. (2018). Likelihood-free inference via classification. *Stat*. *Comput*. **28** 411–425. MR3747571 https://doi.org/10.1007/s11222-017-9738-6

HOFMANN, T., SCHÖLKOPF, B. and SMOLA, A. J. (2008). Kernel methods in machine learning. *Ann. Statist*. **36** 1171–1220. MR2418654 https://doi.org/10.1214/009053607000000677

JEWSON, J., SMITH, J. Q. and HOLMES, C. (2018). Principles of Bayesian inference using general divergence criteria. *Entropy* **20** Paper No. 442, 25 pp. MR3879894 https://doi.org/10.3390/e20060442

JIANG, B., WU, T.-Y. and WONG, W. H. (2018). Approximate Bayesian computation with Kullback-Leibler divergence as data discrepancy. In *International Conference on Artificial Intelligence and Statistics* 1711–1721. PMLR.

LEGRAMANTI, S., DURANTE, D. and ALQUIER, P. (2025). Supplement to "Concentration of discrepancy-based approximate Bayesian computation via Rademacher complexity." https://doi.org/10.1214/24-AOS2453SUPP

LEI, J. (2020). Convergence and concentration of empirical measures under Wasserstein distance in unbounded functional spaces. *Bernoulli* **26** 767–798. MR4036051 https://doi.org/10.3150/19-BEJ1151

LI, W. and FEARNHEAD, P. (2018). On the asymptotic efficiency of approximate Bayesian computation estimators. *Biometrika* **105** 285–299. MR3804403 https://doi.org/10.1093/biomet/asx078

MARIN, J.-M., PILLAI, N. S., ROBERT, C. P. and ROUSSEAU, J. (2014). Relevant statistics for Bayesian model choice. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **76** 833–859. MR3271169 https://doi.org/10.1111/rssb.12056

MARIN, J.-M., PUDLO, P., ROBERT, C. P. and RYDER, R. J. (2012). Approximate Bayesian computational methods. *Stat*. *Comput*. **22** 1167–1180. MR2992292 https://doi.org/10.1007/s11222-011-9288-2

MATSUBARA, T., KNOBLAUCH, J., BRIOL, F.-X. and OATES, C. J. (2022). Robust generalised Bayesian inference for intractable likelihoods. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **84** 997–1022. MR4460583 https://doi.org/10.1111/rssb.12500

MILLER, J. W. and DUNSON, D. B. (2019). Robust Bayesian inference via coarsening. *J. Amer. Statist. Assoc.* **114** 1113–1125. MR4011766 https://doi.org/10.1080/01621459.2018.1469995

MOHRI, M. and ROSTAMIZADEH, A. (2008). Rademacher complexity bounds for non-i.i.d. processes. In *Advances in Neural Information Processing Systems* **21** 1097–1104.

MUANDET, K., FUKUMIZU, K., SRIPERUMBUDUR, B. and SCHÖLKOPF, B. (2017). Kernel mean embedding of distributions: A review and beyond. *Found. Trends Mach. Learn.* **10** 1–141.

MÜLLER, A. (1997). Integral probability metrics and their generating classes of functions. *Adv. in Appl. Probab.* **29** 429–443. MR1450938 https://doi.org/10.2307/1428011

NGUYEN, H. D., ARBEL, J., LÜ, H. and FORBES, F. (2020). Approximate Bayesian computation via the energy statistic. *IEEE Access* **8** 131683–131698.

PARK, M., JITKRITTUM, W. and SEJDINOVIC, D. (2016). K2-ABC: Approximate Bayesian computation with kernel embeddings. In *International Conference on Artificial Intelligence and Statistics* 398–407. PMLR.

RAMDAS, A., GARCÍA TRILLOS, N. and CUTURI, M. (2017). On Wasserstein two-sample testing and related families of nonparametric tests. *Entropy* **19** Paper No. 47, 15 pp. MR3608466 https://doi.org/10.3390/e19020047

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

SRIPERUMBUDUR, B. K., FUKUMIZU, K., GRETTON, A., SCHÖLKOPF, B. and LANCKRIET, G. R. G. (2012). On the empirical estimation of integral probability metrics. *Electron. J. Stat.* **6** 1550–1599. MR2988458 https://doi.org/10.1214/12-EJS722

TALAGRAND, M. (1994). The transportation cost from the uniform measure to the empirical measure in dimension ≥ 3. *Ann. Probab.* **22** 919–959. MR1288137

VILLANI, C. (2021). *Topics in Optimal Transportation*, 2nd ed. *Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI. MR1964483 https://doi.org/10.1090/gsm/058

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

WANG, Y., KAJI, T. and ROCKOVA, V. (2022). Approximate Bayesian computation via classification. *J. Mach. Learn. Res.* **23** Paper No. 350, 49 pp. MR4577789

WEED, J. and BACH, F. (2019). Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli* **25** 2620–2648. MR4003560 https://doi.org/10.3150/18-BEJ1065

The Annals of Statistics
2025, Vol. 53, No. 1, 61–90
https://doi.org/10.1214/24-AOS2455

ON THE SAMPLE COMPLEXITY OF ENTROPIC OPTIMAL TRANSPORT

BY PHILIPPE RIGOLLET[1,a] AND AUSTIN J. STROMME[2,b]

[1]*Department of Mathematics, Massachusetts Institute of Technology,* [a]*rigollet@math.mit.edu*
[2]*Department of Statistics, ENSAE-CREST,* [b]*austin.stromme@ensae.fr*

We study the sample complexity of entropic optimal transport in high dimensions using computationally efficient plug-in estimators. We significantly advance the state of the art by establishing dimension-free, parametric rates for estimating various quantities of interest, including the entropic regression function, which is a natural analog to the optimal transport map. As an application, we propose a practical model for transfer learning based on entropic optimal transport and establish parametric rates of convergence for nonparametric regression and classification.

REFERENCES

[1] ADAMS, S., DIRR, N., PELETIER, M. A. and ZIMMER, J. (2011). From a large-deviations principle to the Wasserstein gradient flow: A new micro-macro passage. *Comm. Math. Phys.* **307** 791–815. MR2842966 https://doi.org/10.1007/s00220-011-1328-4

[2] ALTSCHULER, J., NILES-WEED, J. and RIGOLLET, P. (2017). Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration. *Adv. Neural Inf. Process. Syst.*

[3] ALTSCHULER, J. M., NILES-WEED, J. and STROMME, A. J. (2022). Asymptotics for semidiscrete entropic optimal transport. *SIAM J. Math. Anal.* **54** 1718–1741. MR4393198 https://doi.org/10.1137/21M1440165

[4] ÁLVAREZ-ESTEBAN, P. C., DEL BARRIO, E., CUESTA-ALBERTOS, J. A. and MATRÁN, C. (2016). A fixed-point approach to barycenters in Wasserstein space. *J. Math. Anal. Appl.* **441** 744–762. MR3491556 https://doi.org/10.1016/j.jmaa.2016.04.045

[5] ASADULAEV, A., KOROTIN, A., EGIAZARIAN, V., MOKROV, P. and BURNAEV, E. (2024). Neural optimal transport with general cost functionals. In *The Twelfth International Conference on Learning Representations*.

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

[7] BAYRAKTAR, E., ECKSTEIN, S. and ZHANG, X. (2025). Stability and sample complexity of divergence regularized optimal transport. *Bernoulli* **31** 213–239. MR4815982 https://doi.org/10.3150/24-bej1725

[8] BERCU, B. and BIGOT, J. (2021). Asymptotic distribution and convergence rates of stochastic algorithms for entropic optimal transportation between probability measures. *Ann. Statist.* **49** 968–987. MR4255115 https://doi.org/10.1214/20-aos1987

[9] BERMAN, R. J. (2020). The Sinkhorn algorithm, parabolic optimal transport and geometric Monge-Ampère equations. *Numer. Math.* **145** 771–836. MR4125978 https://doi.org/10.1007/s00211-020-01127-x

[10] BERNTON, E., GHOSAL, P. and NUTZ, M. (2022). Entropic optimal transport: Geometry and large deviations. *Duke Math. J.* **171** 3363–3400. MR4505361 https://doi.org/10.1215/00127094-2022-0035

[11] BIGOT, J., CAZELLES, E. and PAPADAKIS, N. (2019). Central limit theorems for entropy-regularized optimal transport on finite spaces and statistical applications. *Electron. J. Stat.* **13** 5120–5150. MR4041704 https://doi.org/10.1214/19-EJS1637

[12] CARLIER, G., CHIZAT, L. and LABORDE, M. (2024). Displacement smoothness of entropic optimal transport. *ESAIM Control Optim. Calc. Var.* **30** Paper No. 25, 24 pp. MR4729266 https://doi.org/10.1051/cocv/2024013

[13] CARLIER, G., DUVAL, V., PEYRÉ, G. and SCHMITZER, B. (2017). Convergence of entropic schemes for optimal transport and gradient flows. *SIAM J. Math. Anal.* **49** 1385–1418. MR3635459 https://doi.org/10.1137/15M1050264

MSC2020 subject classifications. 62G05.
Key words and phrases. Optimal transport, entropic regularization, Schrödinger bridge, statistical rates.

[14] CARLIER, G., PEGON, P. and TAMANINI, L. (2023). Convergence rate of general entropic optimal transport costs. *Calc. Var. Partial Differential Equations* **62** Paper No. 116, 28 pp. MR4565039 https://doi.org/10.1007/s00526-023-02455-0

[15] CHIZAT, L., ROUSSILLON, P., LÉGER, F., VIALARD, F.-X. and PEYRÉ, G. (2020). Faster Wasserstein distance estimation with the Sinkhorn divergence. *Adv. Neural Inf. Process. Syst.*

[16] COMINETTI, R. and SAN MARTÍN, J. (1994). Asymptotic analysis of the exponential penalty trajectory in linear programming. *Math. Program.* **67** 169–187. MR1305565 https://doi.org/10.1007/BF01582220

[17] CONFORTI, G. and TAMANINI, L. (2021). A formula for the time derivative of the entropic cost and applications. *J. Funct. Anal.* **280** Paper No. 108964, 48 pp. MR4232667 https://doi.org/10.1016/j.jfa.2021.108964

[18] COURTY, N., FLAMARY, R. and TUIA, D. (2014). Domain adaptation with regularized optimal transport. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD* 2014, *Nancy, France, September* 15–19, 2014. *Proceedings, Part I* 14 274–289. Springer, Berlin.

[19] COURTY, N., FLAMARY, R., TUIA, D. and RAKOTOMAMONJY, A. (2017). Optimal transport for domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39** 1853–1865.

[20] CUTURI, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. *Adv. Neural Inf. Process. Syst.*

[21] CUTURI, M., KLEIN, M. and ABLIN, P. (2023). Monge, Bregman and Occam: Interpretable optimal transport in high-dimensions with feature-sparse maps. In *International Conference on Machine Learning*.

[22] DEB, N., GHOSAL, P. and SEN, B. (2021). Rates of estimation of optimal transport maps using plug-in estimators via barycentric projections. *Adv. Neural Inf. Process. Syst.*

[23] DEL BARRIO, E., GONZÁLEZ SANZ, A., LOUBES, J.-M. and NILES-WEED, J. (2023). An improved central limit theorem and fast convergence rates for entropic transportation costs. *SIAM J. Math. Data Sci.* **5** 639–669. MR4616887 https://doi.org/10.1137/22M149260X

[24] DELALANDE, A. (2022). Nearly tight convergence bounds for semi-discrete entropic optimal transport. In *International Conference on Artificial Intelligence and Statistics*.

[25] DELIGIANNIDIS, G., DE BORTOLI, V. and DOUCET, A. (2024). Quantitative uniform stability of the iterative proportional fitting procedure. *Ann. Appl. Probab.* **34** 501–516. MR4696283 https://doi.org/10.1214/23-aap1970

[26] DEVROYE, L., GYÖRFI, L. and LUGOSI, G. (1996). *A Probabilistic Theory of Pattern Recognition. Applications of Mathematics (New York)* **31**. Springer, New York. MR1383093 https://doi.org/10.1007/978-1-4612-0711-5

[27] DI MARINO, S. and GEROLIN, A. (2020). An optimal transport approach for the Schrödinger bridge problem and convergence of Sinkhorn algorithm. *J. Sci. Comput.* **85** Paper No. 27, 28 pp. MR4164428 https://doi.org/10.1007/s10915-020-01325-7

[28] DVURECHENSKY, P., GASNIKOV, A. and KROSHNIN, A. (2018). Computational optimal transport: Complexity by accelerated gradient descent is better than by Sinkhorn's algorithm. In *International Conference on Machine Learning*.

[29] ECKSTEIN, S. and NUTZ, M. (2022). Quantitative stability of regularized optimal transport and convergence of Sinkhorn's algorithm. *SIAM J. Math. Anal.* **54** 5922–5948. MR4506579 https://doi.org/10.1137/21M145505X

[30] ECKSTEIN, S. and NUTZ, M. (2024). Convergence rates for regularized optimal transport via quantization. *Math. Oper. Res.* **49** 1223–1240. MR4755769

[31] ERBAR, M., MAAS, J. and RENGER, D. R. M. (2015). From large deviations to Wasserstein gradient flows in multiple dimensions. *Electron. Commun. Probab.* **20** Paper No. 89, 12 pp. MR3434206 https://doi.org/10.1214/ECP.v20-4315

[32] FAN, J., LIU, S., MA, S., ZHOU, H.-M. and CHEN, Y. (2023). Neural Monge map estimation and its applications. *Trans. Mach. Learn. Res.*

[33] FLAMARY, R., COURTY, N., TUIA, D. and RAKOTOMAMONJY, A. (2016). Optimal transport for domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39** 1853–1865.

[34] FORROW, A., HÜTTER, J.-C., NITZAN, M., RIGOLLET, P., SCHIEBINGER, G. and WEED, J. (2019). Statistical optimal transport via factored couplings. In *International Conference on Artificial Intelligence and Statistics*.

[35] GENEVAY, A., CHIZAT, L., BACH, F., CUTURI, M. and PEYRÉ, G. (2019). Sample complexity of Sinkhorn divergences. In *International Conference on Artificial Intelligence and Statistics*.

[36] GENEVAY, A., PEYRÉ, G. and CUTURI, M. (2018). Learning generative models with Sinkhorn divergences. In *International Conference on Artificial Intelligence and Statistics*.

[37] GHOSAL, P., NUTZ, M. and BERNTON, E. (2022). Stability of entropic optimal transport and Schrödinger bridges. *J. Funct. Anal.* **283** Paper No. 109622, 22 pp. MR4460341 https://doi.org/10.1016/j.jfa.2022.109622

[38] GOLDFELD, Z., KATO, K., RIOUX, G. and SADHU, R. (2024). Limit theorems for entropic optimal transport maps and Sinkhorn divergence. *Electron. J. Stat.* **18** 980–1041. MR4718466 https://doi.org/10.1214/24-ejs2217

[39] GOLDFELD, Z., KATO, K., RIOUX, G. and SADHU, R. (2024). Statistical inference with regularized optimal transport. *Inf. Inference* **13** Paper No. 13, 68 pp. MR4701828 https://doi.org/10.1093/imaiai/iaad056

[40] GONZÁLEZ-SANZ, A. and HUNDRIESER, S. (2023). Weak limits for empirical entropic optimal transport: Beyond smooth costs. arXiv preprint.

[41] GONZALEZ-SANZ, A., LOUBES, J.-M. and NILES-WEED, J. (2022). Weak limits of entropy regularized optimal transport; potentials, plans and divergences. arXiv preprint.

[42] GROPPE, M. and HUNDRIESER, S. (2024). Lower complexity adaptation for empirical entropic optimal transport. *J. Mach. Learn. Res.* **25** Paper No. 344, 55 pp. MR4829188

[43] GUNSILIUS, F. and XU, Y. (2021). Matching for causal effects via multimarginal unbalanced optimal transport. arXiv preprint.

[44] HARCHAOUI, Z., LIU, L. and PAL, S. (2024). Asymptotics of discrete Schrödinger bridges via chaos decomposition. *Bernoulli* **30** 1945–1970. MR4746595 https://doi.org/10.3150/23-bej1659

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

[46] HOWARD, S., DELIGIANNIDIS, G., REBESCHINI, P. and THORNTON, J. (2024). Differentiable cost-parameterized Monge map estimators. arXiv preprint.

[47] HUNDRIESER, S., KLATT, M. and MUNK, A. (2024). Limit distributions and sensitivity analysis for empirical entropic optimal transport on countable spaces. *Ann. Appl. Probab.* **34** 1403–1468. MR4700262 https://doi.org/10.1214/23-aap1995

[48] HÜTTER, J.-C. and RIGOLLET, P. (2021). Minimax estimation of smooth optimal transport maps. *Ann. Statist.* **49** 1166–1194. MR4255123 https://doi.org/10.1214/20-aos1997

[49] KARIMI, H., NUTINI, J. and SCHMIDT, M. (2016). Linear convergence of gradient and proximal-gradient methods under the Polyak–Łojasiewicz condition. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*.

[50] KLEIN, M., POOLADIAN, A.-A., ABLIN, P., NDIAYE, E., NILES-WEED, J. and CUTURI, M. (2023). Learning costs for structured Monge displacements. arXiv preprint.

[51] KOLTCHINSKII, V. and BEZNOSOVA, O. (2005). Exponential convergence rates in classification. In *Learning Theory. Lecture Notes in Computer Science* **3559** 295–307. Springer, Berlin. MR2203269 https://doi.org/10.1007/11503415_20

[52] LAVENANT, H., ZHANG, S., KIM, Y.-H. and SCHIEBINGER, G. (2024). Toward a mathematical theory of trajectory inference. *Ann. Appl. Probab.* **34** 428–500. MR4696282 https://doi.org/10.1214/23-aap1969

[53] LÉONARD, C. (2012). From the Schrödinger problem to the Monge–Kantorovich problem. *J. Funct. Anal.* **262** 1879–1920. MR2873864 https://doi.org/10.1016/j.jfa.2011.11.026

[54] LÉONARD, C. (2014). A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.* **34** 1533–1574. MR3121631 https://doi.org/10.3934/dcds.2014.34.1533

[55] LUISE, G., SALZO, S., PONTIL, M. and CILIBERTO, C. (2019). Sinkhorn barycenters with free support via Frank–Wolfe algorithm. *Adv. Neural Inf. Process. Syst.*

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

[57] MANOLE, T., BALAKRISHNAN, S., NILES-WEED, J. and WASSERMAN, L. (2024). Plugin estimation of smooth optimal transport maps. *Ann. Statist.* **52** 966–998. MR4784066 https://doi.org/10.1214/24-aos2379

[58] MANOLE, T. and NILES-WEED, J. (2024). Sharp convergence rates for empirical optimal transport with smooth costs. *Ann. Appl. Probab.* **34** 1108–1135. MR4700254 https://doi.org/10.1214/23-aap1986

[59] MASUD, S. B., WERENSKI, M., MURPHY, J. M. and AERON, S. (2021). Multivariate rank via entropic optimal transport: Sample efficiency and generative modeling. arXiv preprint.

[60] MENA, G. and NILES-WEED, J. (2019). Statistical bounds for entropic optimal transport: Sample complexity and the central limit theorem. *Adv. Neural Inf. Process. Syst.*

[61] MIKAMI, T. (2004). Monge's problem with a quadratic cost by the zero-noise limit of h-path processes. *Probab. Theory Related Fields* **129** 245–260. MR2063377 https://doi.org/10.1007/s00440-004-0340-4

[62] NILES-WEED, J. and RIGOLLET, P. (2022). Estimation of Wasserstein distances in the spiked transport model. *Bernoulli* **28** 2663–2688. MR4474558 https://doi.org/10.3150/21-bej1433

[63] NUTZ, M. and WIESEL, J. (2022). Entropic optimal transport: Convergence of potentials. *Probab. Theory Related Fields* **184** 401–424. MR4498514 https://doi.org/10.1007/s00440-021-01096-8

[64] PAL, S. (2024). On the difference between entropic cost and the optimal transport cost. *Ann. Appl. Probab.* **34** 1003–1028. MR4700251 https://doi.org/10.1214/23-aap1983

[65] PEYRÉ, G. and CUTURI, M. (2019). Computational optimal transport. *Found. Trends Mach. Learn.* **11** 355–607.

[66] POOLADIAN, A.-A., DIVOL, V. and NILES-WEED, J. (2023). Minimax estimation of discontinuous optimal transport maps: The semi-discrete case. In *International Conference on Machine Learning*.

[67] POOLADIAN, A.-A., DOMINGO-ENRICH, C., CHEN, R. T. and AMOS, B. (2024). Neural optimal transport with Lagrangian costs. arXiv preprint.

[68] POOLADIAN, A.-A. and NILES-WEED, J. (2021). Entropic estimation of optimal transport maps. arXiv preprint.

[69] RAMDAS, A., GARCÍA TRILLOS, N. and CUTURI, M. (2017). On Wasserstein two-sample testing and related families of nonparametric tests. *Entropy* **19** Paper No. 47, 15 pp. MR3608466 https://doi.org/10.3390/e19020047

[70] RIGOLLET, P. and WEED, J. (2018). Entropic optimal transport is maximum-likelihood deconvolution. *C. R. Math. Acad. Sci. Paris* **356** 1228–1235. MR3907589 https://doi.org/10.1016/j.crma.2018.10.010

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

[72] RIOUX, G., GOLDFELD, Z. and KATO, K. (2023). Entropic Gromov–Wasserstein distances: Stability, algorithms, and distributional limits. arXiv preprint.

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

[74] SCHIEBINGER, G., SHU, J., TABAKA, M., CLEARY, B., SUBRAMANIAN, V., SOLOMON, A., GOULD, J., LIU, S., LIN, S. et al. (2019). Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell* **176** 928–943.

[75] SCHRÖDINGER, E. (1931). Über die Umkehrung der Naturgesetze. *Angew. Chem.* **44** 636–636.

[76] SCHRÖDINGER, E. (1932). Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *Ann. Inst. Henri Poincaré* **2** 269–310. MR1508000

[77] SEGUY, V., DAMODARAN, B. B., FLAMARY, R., COURTY, N., ROLET, A. and BLONDEL, M. (2018). Large-scale optimal transport and mapping estimation. In *International Conference on Learning Representations*.

[78] SINKHORN, R. (1964). A relationship between arbitrary positive matrices and doubly stochastic matrices. *Ann. Math. Stat.* **35** 876–879. MR0161868 https://doi.org/10.1214/aoms/1177703591

[79] SOLOMON, J., DE GOES, F., PEYRÉ, G., CUTURI, M., BUTSCHER, A., NGUYEN, A., DU, T. and GUIBAS, L. (2015). Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Trans. Graph.* **34** 66.

[80] STROMME, A. J. (2023). Minimum intrinsic dimension scaling for entropic optimal transport. arXiv preprint.

[81] STROMME, A. J. (2023). Sampling from a Schrödinger bridge. In *International Conference on Artificial Intelligence and Statistics*.

[82] TOROUS, W., GUNSILIUS, F. and RIGOLLET, P. (2024). An optimal transport approach to estimating causal effects via nonlinear difference-in-differences. *J. Causal Inference* **12** Paper No. 20230004, 26 pp. MR4782049 https://doi.org/10.1515/jci-2023-0004

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

[84] VACHER, A., MUZELLEC, B., BACH, F., VIALARD, F.-X. and RUDI, A. (2024). Optimal estimation of smooth transport maps with kernel SoS. *SIAM J. Math. Data Sci.* **6** 311–342. MR4732265 https://doi.org/10.1137/22M1528847

[85] VACHER, A., MUZELLEC, B., RUDI, A., BACH, F. and VIALARD, F.-X. (2021). A dimension-free computational upper-bound for smooth optimal transport estimation. In *Conference on Learning Theory*.

[86] VILLANI, C. (2003). *Topics in Optimal Transportation. Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI. MR1964483 https://doi.org/10.1090/gsm/058

[87] VILLANI, C. (2009). *Optimal Transport: Old and New*. Springer, Berlin.

[88] WERENSKI, M., MASUD, S. B., MURPHY, J. M. and AERON, S. (2024). On rank energy statistics via optimal transport: Continuity, convergence, and change point detection. *IEEE Trans. Inf. Theory* **70** 5168–5190. MR4762888 https://doi.org/10.1109/tit.2024.3367182

[89] WERENSKI, M., MURPHY, J. M. and AERON, S. (2023). Estimation of entropy-regularized optimal transport maps between non-compactly supported measures. arXiv preprint.

The Annals of Statistics
2025, Vol. 53, No. 1, 91–116
https://doi.org/10.1214/24-AOS2456

DEFLATED HETEROPCA: OVERCOMING THE CURSE OF ILL-CONDITIONING IN HETEROSKEDASTIC PCA

BY YUCHEN ZHOU[1,a] AND YUXIN CHEN[2,b]

[1]*Department of Statistics, University of Illinois Urbana-Champaign,* [a]*yuchenz@illinois.edu*
[2]*Department of Statistics and Data Science, The Wharton School, University of Pennsylvania,* [b]*yuxinc@wharton.upenn.edu*

This paper is concerned with estimating the column subspace of a low-rank matrix $\boldsymbol{X}^\star \in \mathbb{R}^{n_1 \times n_2}$ from contaminated data. How to obtain optimal statistical accuracy while accommodating the widest range of signal-to-noise ratios (SNRs) becomes particularly challenging in the presence of heteroskedastic noise and unbalanced dimensionality (i.e., $n_2 \gg n_1$). While the state-of-the-art algorithm HeteroPCA emerges as a powerful solution for solving this problem, it suffers from "the curse of ill-conditioning," namely, its performance degrades as the condition number of $\boldsymbol{X}^\star$ grows. In order to overcome this critical issue without compromising the range of allowable SNRs, we propose a novel algorithm, called Deflated − HeteroPCA, that achieves near-optimal and condition-number-free theoretical guarantees in terms of both ℓ_2 and $\ell_{2,\infty}$ statistical accuracy. The proposed algorithm divides the spectrum of $\boldsymbol{X}^\star$ into well-conditioned and mutually well-separated subblocks, and applies HeteroPCA to conquer each subblock successively. Further, an application of our algorithm and theory to two canonical examples—the factor model and tensor PCA—leads to remarkable improvement for each application.

REFERENCES

ABBE, E., FAN, J. and WANG, K. (2022). An ℓ_p theory of PCA and spectral clustering. *Ann. Statist.* **50** 2359–2385. MR4474494 https://doi.org/10.1214/22-aos2196

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

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

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

BAI, J. and WANG, P. (2016). Econometric analysis of large factor models. *Ann. Rev. Econ.* **8** 53–80.

BAI, Z. and DING, X. (2012). Estimation of spiked eigenvalues in spiked models. *Random Matrices Theory Appl.* **1** 1150011, 21. MR2934717 https://doi.org/10.1142/S2010326311500110

BALZANO, L., CHI, Y. and LU, Y. M. (2018). Streaming PCA and subspace tracking: The missing data case. *Proc. IEEE Inst. Electr. Electron. Eng.* **106** 1293–1310. https://doi.org/10.1109/JPROC.2018.2847041

BAO, Z., DING, X., WANG, J. and WANG, K. (2022). Statistical inference for principal components of spiked covariance matrices. *Ann. Statist.* **50** 1144–1169. MR4404931 https://doi.org/10.1214/21-aos2143

BAO, Z., DING, X. and WANG, K. (2021). Singular vector and singular subspace distribution for the matrix denoising model. *Ann. Statist.* **49** 370–392. MR4206682 https://doi.org/10.1214/20-AOS1960

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

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

CAI, J.-F., CANDÈS, E. J. and SHEN, Z. (2010). A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* **20** 1956–1982. MR2600248 https://doi.org/10.1137/080738970

MSC2020 subject classifications. Primary 62F10; secondary 62H25.
Key words and phrases. Principal component analysis (PCA), heteroskedastic noise, the curse of ill-conditioning, factor models, tensor PCA.

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

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

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

CHEN, P. H., CHEN, J., YESHURUN, Y., HASSON, U., HAXBY, J. V. and RAMADGE, P. J. (2015). A reduced-dimension fMRI shared response model. *Adv. Neural Inf. Process. Syst.* **2015** 460–468.

CHEN, S., LIU, S. and MA, Z. (2022). Global and individualized community detection in inhomogeneous multilayer networks. *Ann. Statist.* **50** 2664–2693. MR4500621 https://doi.org/10.1214/22-aos2202

CHEN, Y., CHI, Y., FAN, J. and MA, C. (2021). *Spectral Methods for Data Science: A Statistical Perspective. Foundations and Trends® in Machine Learning* **14** 566–806.

CHEN, Y., CHI, Y., FAN, J., MA, C. and YAN, Y. (2020). Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. *SIAM J. Optim.* **30** 3098–3121. MR4167625 https://doi.org/10.1137/19M1290000

CHEN, Y., FAN, J., MA, C. and YAN, Y. (2019). Inference and uncertainty quantification for noisy matrix completion. *Proc. Natl. Acad. Sci. USA* **116** 22931–22937. MR4036123 https://doi.org/10.1073/pnas.1910053116

CHEN, Y., FAN, J., MA, C. and YAN, Y. (2021). Bridging convex and nonconvex optimization in robust PCA: Noise, outliers and missing data. *Ann. Statist.* **49** 2948–2971. MR4338899 https://doi.org/10.1214/21-aos2066

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

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

DOBRIBAN, E. and OWEN, A. B. (2019). Deterministic parallel analysis: An improved method for selecting factors and principal components. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 163–183. MR3904784

DONOHO, D. and GAVISH, M. (2014). Minimax risk of matrix denoising by singular value thresholding. *Ann. Statist.* **42** 2413–2440. MR3269984 https://doi.org/10.1214/14-AOS1257

DONOHO, D., GAVISH, M. and JOHNSTONE, I. (2018). Optimal shrinkage of eigenvalues in the spiked covariance model. *Ann. Statist.* **46** 1742–1778. MR3819116 https://doi.org/10.1214/17-AOS1601

ELSENER, A. and VAN DE GEER, S. (2019). Sparse spectral estimation with missing and corrupted measurements. *Stat* **8** e229, 11. MR3978409 https://doi.org/10.1002/sta4.229

FAN, J., LI, R., ZHANG, C.-H. and ZOU, H. (2020). *Statistical Foundations of Data Science*. CRC Press, Boca Raton.

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., WANG, K., ZHONG, Y. and ZHU, Z. (2021). Robust high-dimensional factor models with applications to statistical machine learning. *Statist. Sci.* **36** 303–327. MR4255196 https://doi.org/10.1214/20-sts785

FENG, O. Y., VENKATARAMANAN, R., RUSH, C., SAMWORTH, R. J. et al. (2022). *A Unifying Tutorial on Approximate Message Passing. Foundations and Trends® in Machine Learning* **15** 335–536.

FLORESCU, L. and PERKINS, W. (2016). Spectral thresholds in the bipartite stochastic block model. In *Conference on Learning Theory* 943–959. PMLR.

GAVISH, M. and DONOHO, D. L. (2014). The optimal hard threshold for singular values is $4/\sqrt{3}$. *IEEE Trans. Inf. Theory* **60** 5040–5053. MR3245370 https://doi.org/10.1109/TIT.2014.2323359

GAVISH, M. and DONOHO, D. L. (2017). Optimal shrinkage of singular values. *IEEE Trans. Inf. Theory* **63** 2137–2152. MR3626861 https://doi.org/10.1109/TIT.2017.2653801

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

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

HONG, D., BALZANO, L. and FESSLER, J. A. (2016). Towards a theoretical analysis of PCA for heteroscedastic data. In 2016 54*th Annual Allerton Conference on Communication, Control, and Computing* (*Allerton*) 496–503. IEEE.

HONG, D., BALZANO, L. and FESSLER, J. A. (2018). Asymptotic performance of PCA for high-dimensional heteroscedastic data. *J. Multivariate Anal.* **167** 435–452. MR3830656 https://doi.org/10.1016/j.jmva.2018.06.002

HONG, D., YANG, F., FESSLER, J. A. and BALZANO, L. (2023). Optimally weighted PCA for high-dimensional heteroscedastic data. *SIAM J. Math. Data Sci.* **5** 222–250. MR4567414 https://doi.org/10.1137/22M1470244

JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961 https://doi.org/10.1214/aos/1009210544

JOHNSTONE, I. M. and PAUL, D. (2018). PCA in high dimensions: An orientation. *Proc. IEEE Inst. Electr. Electron. Eng.* **106** 1277–1292. https://doi.org/10.1109/JPROC.2018.2846730

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

KOLTCHINSKII, V. and GINÉ, E. (2000). Random matrix approximation of spectra of integral operators. *Bernoulli* **6** 113–167. MR1781185 https://doi.org/10.2307/3318636

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

KOLTCHINSKII, V. and XIA, D. (2016). Perturbation of linear forms of singular vectors under Gaussian noise. In *High Dimensional Probability VII. Progress in Probability* **71** 397–423. Springer, Berlin. MR3565274 https://doi.org/10.1007/978-3-319-40519-3_18

KRITCHMAN, S. and NADLER, B. (2008). Determining the number of components in a factor model from limited noisy data. *Chemom. Intell. Lab. Syst.* **94** 19–32.

KRITCHMAN, S. and NADLER, B. (2009). Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory. *IEEE Trans. Signal Process.* **57** 3930–3941. MR2683143 https://doi.org/10.1109/TSP.2009.2022897

LAWLEY, D. N. and MAXWELL, A. E. (1962). Factor analysis as a statistical method. *J. R. Stat. Soc., Ser. D, Stat.* **12** 209–229.

LI, G., FAN, W. and WEI, Y. (2023). Approximate message passing from random initialization with applications to $\mathbb{Z}_2$ synchronization. *Proc. Natl. Acad. Sci. USA* **120** Paper No. e2302930120, 7. MR4637851

LI, G. and WEI, Y. (2022). A non-asymptotic framework for approximate message passing in spiked models. arXiv preprint. Available at arXiv:2208.03313.

LIU, L. T., DOBRIBAN, E. and SINGER, A. (2018). *e*PCA: High dimensional exponential family PCA. *Ann. Appl. Stat.* **12** 2121–2150. MR3875695 https://doi.org/10.1214/18-AOAS1146

LÖFFLER, M., ZHANG, A. Y. and ZHOU, H. H. (2021). Optimality of spectral clustering in the Gaussian mixture model. *Ann. Statist.* **49** 2506–2530. MR4338373 https://doi.org/10.1214/20-aos2044

LOH, P.-L. and WAINWRIGHT, M. J. (2012). High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *Ann. Statist.* **40** 1637–1664. MR3015038 https://doi.org/10.1214/12-AOS1018

LOUNICI, K. (2014). High-dimensional covariance matrix estimation with missing observations. *Bernoulli* **20** 1029–1058. MR3217437 https://doi.org/10.3150/12-BEJ487

MA, C., WANG, K., CHI, Y. and CHEN, Y. (2020). Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion, and blind deconvolution. *Found. Comput. Math.* **20** 451–632. MR4099988 https://doi.org/10.1007/s10208-019-09429-9

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

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

MONTANARI, A. and WU, Y. (2024). Fundamental limits of low-rank matrix estimation with diverging aspect ratios. *Ann. Statist.* **52** 1460–1484. MR4804816 https://doi.org/10.1214/24-aos2400

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

NDAOUD, M. (2022). Sharp optimal recovery in the two component Gaussian mixture model. *Ann. Statist.* **50** 2096–2126. MR4474484 https://doi.org/10.1214/22-aos2178

NDAOUD, M., SIGALLA, S. and TSYBAKOV, A. B. (2022). Improved clustering algorithms for the bipartite stochastic block model. *IEEE Trans. Inf. Theory* **68** 1960–1975. MR4395508 https://doi.org/10.1109/tit.2021.3130683

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

PAVEZ, E. and ORTEGA, A. (2021). Covariance matrix estimation with non uniform and data dependent missing observations. *IEEE Trans. Inf. Theory* **67** 1201–1215. MR4232009 https://doi.org/10.1109/tit.2020.3039118

PERRY, A., WEIN, A. S., BANDEIRA, A. S. and MOITRA, A. (2018). Optimality and sub-optimality of PCA I: Spiked random matrix models. *Ann. Statist.* **46** 2416–2451. MR3845022 https://doi.org/10.1214/17-AOS1625

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

ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. MR2893856 https://doi.org/10.1214/11-AOS887

SRIVASTAVA, P. R., SARKAR, P. and HANASUSANTO, G. A. (2023). A robust spectral clustering algorithm for sub-Gaussian mixture models with outliers. *Oper. Res.* **71** 224–244. MR4560196 https://doi.org/10.1287/opre.2022.2317

TONG, T., MA, C. and CHI, Y. (2021). Accelerating ill-conditioned low-rank matrix estimation via scaled gradient descent. *J. Mach. Learn. Res.* **22** Paper No. 150, 63. MR4318506

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

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

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

XU, X., SHEN, Y., CHI, Y. and MA, C. (2023). The power of preconditioning in overparameterized low-rank matrix sensing. arXiv preprint. Available at arXiv:2302.01186.

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

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

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

ZHANG, A. Y. and ZHOU, H. Y. (2024). Leave-one-out singular subspace perturbation analysis for spectral clustering. *Ann. Statist.* **52** 2004–2033. MR4829478 https://doi.org/10.1214/24-aos2418

ZHAO, L. C., KRISHNAIAH, P. R. and BAI, Z. D. (1986). On detection of the number of signals in presence of white noise. *J. Multivariate Anal.* **20** 1–25. MR0862239 https://doi.org/10.1016/0047-259X(86)90017-5

ZHOU, Y. and CHEN, Y. (2023). Heteroskedastic tensor clustering. arXiv preprint. Available at arXiv:2311.02306.

ZHOU, Y. and CHEN, Y. (2025). Supplement to “Deflated HeteroPCA: Overcoming the curse of ill-conditioning in heteroskedastic PCA.” https://doi.org/10.1214/24-AOS2456SUPPA, https://doi.org/10.1214/24-AOS2456SUPPB

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

ZHU, Z., WANG, T. and SAMWORTH, R. J. (2022). High-dimensional principal component analysis with heterogeneous missingness. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 2000–2031. MR4515564

The Annals of Statistics
2025, Vol. 53, No. 1, 117–143
https://doi.org/10.1214/24-AOS2457

NONLINEAR GLOBAL FRÉCHET REGRESSION FOR RANDOM OBJECTS VIA WEAK CONDITIONAL EXPECTATION

BY SATARUPA BHATTACHARJEE[1,a], BING LI[2,b] AND LINGZHOU XUE[2,c]

[1]*Department of Statistics, University of Florida,* [a]*bhattacharjee.sa@ufl.edu*
[2]*Department of Statistics, Pennsylvania State University,* [b]*bxl9@psu.edu,* [c]*lzxue@psu.edu*

Random objects are complex non-Euclidean data taking values in general metric spaces, possibly devoid of any underlying vector space structure. Such data are becoming increasingly abundant with the rapid advancement in technology. Examples include probability distributions, positive semidefinite matrices and data on Riemannian manifolds. However, except for regression for object-valued response with Euclidean predictors and distribution-on-distribution regression, there has been limited development of a general framework for object-valued response with object-valued predictors in the literature. To fill this gap, we introduce the notion of a weak conditional Fréchet mean based on Carleman operators and then propose a global nonlinear Fréchet regression model through the reproducing kernel Hilbert space (RKHS) embedding. Furthermore, we establish the relationships between the conditional Fréchet mean and the weak conditional Fréchet mean for both Euclidean and object-valued data. We also show that the state-of-the-art global Fréchet regression developed by Petersen and Müller (*Ann. Statist.* **47** (2019) 691–719) emerges as a special case of our method by choosing a linear kernel. We require that the metric space for the predictor admits a reproducing kernel, while the intrinsic geometry of the metric space for the response is utilized to study the asymptotic properties of the proposed estimates. Numerical studies, including extensive simulations and a real application, are conducted to investigate the finite-sample performance.

REFERENCES

[1] AFSARI, B. (2011). Riemannian L^p center of mass: Existence, uniqueness, and convexity. *Proc. Amer. Math. Soc.* **139** 655–673. MR2736346 https://doi.org/10.1090/S0002-9939-2010-10541-5
[2] AMBROSIO, L. and GIGLI, N. (2008). Construction of the parallel transport in the Wasserstein space. *Methods Appl. Anal.* **15** 1–29. MR2482206 https://doi.org/10.4310/MAA.2008.v15.n1.a3
[3] BHATTACHARJEE, S., LI, B. and XUE, L. (2025). Supplement to "Nonlinear global Fréchet regression for random objects via weak conditional expectation." https://doi.org/10.1214/24-AOS2457SUPP
[4] BHATTACHARJEE, S. and MÜLLER, H.-G. (2023). Single index Fréchet regression. *Ann. Statist.* **51** 1770–1798. MR4658576 https://doi.org/10.1214/23-aos2307
[5] BHATTACHARYA, R. and PATRANGENARU, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann. Statist.* **31** 1–29. MR1962498 https://doi.org/10.1214/aos/1046294456
[6] BHATTACHARYA, R. and PATRANGENARU, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *Ann. Statist.* **33** 1225–1259. MR2195634 https://doi.org/10.1214/009053605000000093
[7] BIGOT, J., GOUET, R., KLEIN, T. and LÓPEZ, A. (2017). Geodesic PCA in the Wasserstein space by convex PCA. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 1–26. MR3606732 https://doi.org/10.1214/15-AIHP706
[8] BILLERA, L. J., HOLMES, S. P. and VOGTMANN, K. (2001). Geometry of the space of phylogenetic trees. *Adv. in Appl. Math.* **27** 733–767. MR1867931 https://doi.org/10.1006/aama.2001.0759
[9] CHARLIER, B. (2013). Necessary and sufficient condition for the existence of a Fréchet mean on the circle. *ESAIM Probab. Stat.* **17** 635–649. MR3126155 https://doi.org/10.1051/ps/2012015

MSC2020 subject classifications. Primary 62G05, 62J02; secondary 62G08, 62J99.
Key words and phrases. Random object, object-on-object regression, weak conditional expectation, metric spaces, reproducing kernel Hilbert spaces.

[10] CHEN, Y., GAJARDO, A., FAN, J., ZHONG, Q., DUBEY, P., BHATTACHARJEE, S., HAN, K. and MÜLLER, H. (2020). frechet: Statistical analysis for random objects and non-Euclidean data. R Package Version 0.2. 0.

[11] CHEN, Y., LIN, Z. and MÜLLER, H.-G. (2023). Wasserstein regression. *J. Amer. Statist. Assoc.* **118** 869–882. MR4595462 https://doi.org/10.1080/01621459.2021.1956937

[12] CHEN, Z., BAO, Y., LI, H. and SPENCER JR., B. F. (2019). LQD-RKHS-based distribution-to-distribution regression methodology for restoring the probability distributions of missing SHM data. *Mech. Syst. Signal Process.* **121** 655–674.

[13] DELICADO, P. and VIEU, P. (2017). Choosing the most relevant level sets for depicting a sample of densities. *Comput. Statist.* **32** 1083–1113. MR3703579 https://doi.org/10.1007/s00180-017-0746-y

[14] DI MARZIO, M., PANZERA, A. and TAYLOR, C. C. (2014). Nonparametric regression for spherical data. *J. Amer. Statist. Assoc.* **109** 748–763. MR3223747 https://doi.org/10.1080/01621459.2013.866567

[15] DRYDEN, I. L., KOLOYDENKO, A. and ZHOU, D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Ann. Appl. Stat.* **3** 1102–1123. MR2750388 https://doi.org/10.1214/09-AOAS249

[16] DUBEY, P. and MÜLLER, H.-G. (2019). Fréchet analysis of variance for random objects. *Biometrika* **106** 803–821. MR4031200 https://doi.org/10.1093/biomet/asz052

[17] FRÉCHET, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. Henri Poincaré* **10** 215–310. MR0027464

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

[19] GHODRATI, L. and PANARETOS, V. M. (2022). Distribution-on-distribution regression via optimal transport maps. *Biometrika* **109** 957–974. MR4519110 https://doi.org/10.1093/biomet/asac005

[20] GOLUB, G. H., HEATH, M. and WAHBA, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* **21** 215–223. MR0533250 https://doi.org/10.2307/1268518

[21] HEIN, M. (2009). Robust nonparametric regression with metric-space valued output. *Adv. Neural Inf. Process. Syst.* **22**.

[22] HORN, R. A. and JOHNSON, C. R. (2013). *Matrix Analysis*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2978290

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

[24] HUCKEMANN, S. F. (2015). (Semi-) intrinsic statistical analysis on non-Euclidean spaces. In *Advances in Complex Data Modeling and Computational Methods in Statistics* 103–118. Springer, Berlin.

[25] KELLEY, J. L. (1955). *General Topology*. D. Van Nostrand Co., Inc., Toronto–New York. MR0070144

[26] LE GOUIC, T. and LOUBES, J.-M. (2017). Existence and consistency of Wasserstein barycenters. *Probab. Theory Related Fields* **168** 901–917. MR3663634 https://doi.org/10.1007/s00440-016-0727-z

[27] LE, H. (2001). Locating Fréchet means with application to shape spaces. *Adv. in Appl. Probab.* **33** 324–338. MR1842295 https://doi.org/10.1239/aap/999188316

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

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

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

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

[32] LI, B. and SONG, J. (2022). Dimension reduction for functional data based on weak conditional moments. *Ann. Statist.* **50** 107–128. MR4382010 https://doi.org/10.1214/21-aos2091

[33] MARRON, J. S. and ALONSO, A. M. (2014). Overview of object oriented data analysis. *Biom. J.* **56** 732–753. MR3258083 https://doi.org/10.1002/bimj.201300072

[34] MICCHELLI, C. A., XU, Y. and ZHANG, H. (2006). Universal kernels. *J. Mach. Learn. Res.* **7** 2651–2667. MR2274454

[35] PANARETOS, V. M. and ZEMEL, Y. (2019). Statistical aspects of Wasserstein distances. *Annu. Rev. Stat. Appl.* **6** 405–431. MR3939527 https://doi.org/10.1146/annurev-statistics-030718-104938

[36] PENNEC, X. (2018). Barycentric subspace analysis on manifolds. *Ann. Statist.* **46** 2711–2746. MR3851753 https://doi.org/10.1214/17-AOS1636

[37] PETERSEN, A. and MÜLLER, H.-G. (2016). Functional data analysis for density functions by transformation to a Hilbert space. *Ann. Statist.* **44** 183–218. MR3449766 https://doi.org/10.1214/15-AOS1363

[38] PETERSEN, A. and MÜLLER, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *Ann. Statist.* **47** 691–719. MR3909947 https://doi.org/10.1214/17-AOS1624

[39] QIU, R., YU, Z. and ZHU, R. (2024). Random forest weighted local Fréchet regression with random objects. *J. Mach. Learn. Res.* **25** Paper No. 107, 69 pp. MR4749759

[40] SANG, P. and LI, B. (2022). Nonlinear function-on-function regression by RKHS. Preprint. Available at arXiv:2207.08211.

[41] SEJDINOVIC, D., GRETTON, A., SRIPERUMBUDUR, B. and FUKUMIZU, K. (2012). Hypothesis testing using pairwise distances and associated kernels (with appendix). Preprint. Available at arXiv:1205.0411.

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

[43] SRIPERUMBUDUR, B. K., FUKUMIZU, K. and LANCKRIET, G. R. G. (2011). Universality, characteristic kernels and RKHS embedding of measures. *J. Mach. Learn. Res.* **12** 2389–2410. MR2825431

[44] STURM, K.-T. (2003). Probability measures on metric spaces of nonpositive curvature. In *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces* (*Paris*, 2002). *Contemp. Math.* **338** 357–390. Amer. Math. Soc., Providence, RI. MR2039961 https://doi.org/10.1090/conm/338/06080

[45] TAO, J., LI, B. and XUE, L. (2024). An additive graphical model for discrete data. *J. Amer. Statist. Assoc.* **119** 368–381. MR4713899 https://doi.org/10.1080/01621459.2022.2119983

[46] VAN DER VAART, A. W. and WELLNER, J. A. (2023). *Weak Convergence and Empirical Processes—with Applications to Statistics. Springer Series in Statistics*. Springer, Cham. MR4628026 https://doi.org/10.1007/978-3-031-29040-4

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

[48] WEIDMANN, J. (2012). *Linear Operators in Hilbert Spaces. Graduate Texts in Mathematics* **68**. Springer, New York–Berlin. MR0566954

[49] ZEMEL, Y. and PANARETOS, V. M. (2019). Fréchet means and Procrustes analysis in Wasserstein space. *Bernoulli* **25** 932–976. MR3920362 https://doi.org/10.3150/17-bej1009

[50] ZHANG, Q., LI, B. and XUE, L. (2024). A copula graphical model for multi-attribute data using optimal transport. Preprint. Available at arXiv:2404.06735.

[51] ZHANG, Q., LI, B. and XUE, L. (2024). Nonlinear sufficient dimension reduction for distribution-on-distribution regression. *J. Multivariate Anal.* **202** Paper No. 105302, 17 pp. MR4711112 https://doi.org/10.1016/j.jmva.2024.105302

[52] ZHANG, Q., XUE, L. and LI, B. (2024). Dimension reduction for Fréchet regression. *J. Amer. Statist. Assoc.* **119** 2733–2747. MR4833911 https://doi.org/10.1080/01621459.2023.2277406

[53] ZHOU, Y. and MÜLLER, H.-G. (2022). Network regression with graph Laplacians. *J. Mach. Learn. Res.* **23** Paper No. 320, 41 pp. MR4577759

The Annals of Statistics
2025, Vol. 53, No. 1, 144–169
https://doi.org/10.1214/24-AOS2458

SIMPLEX QUANTILE REGRESSION WITHOUT CROSSING

BY TOMOHIRO ANDO[1,a] AND KER-CHAU LI[2,b]

[1]*Melbourne Business School, University of Melbourne,* [a]*T.Ando@mbs.edu*
[2]*Department of Statistics, University of California, Los Angeles,* [b]*kcli@stat.ucla.edu*

Noncrossing quantile regression with an emphasis on model misspecification is investigated. While many sophisticated methods for noncrossing quantile have been developed under the assumption of correct model specification, model misspecification and extrapolation are two issues rarely considered in the literature. In this paper, a monotonicity representation for quantile regression models is obtained under simplex embedding, which leads to the simplex quantile regression (SQR) method. SQR model advocates the use of the Barycentric coordinate system and is immune to quantile crossing. An innovative concept, the maximum effective region, which defines the allowable region for extrapolation without quantile crossing is introduced. Under model misspecification, the existence of pseudo true quantile regression parameters, the asymptotic property of SQR estimator and asymptotic optimality of cross-validation for model selection and model averaging are established. The advantages of our new approach are also illustrated by numerical results. The proposed method is applied to U.S. financial market data. Our SQR-based strategy of asset pricing analysis allows for head-to-head transparent comparisons of five periods between October 1987 (Black Monday crash period) and February 2020 (COVID-19 pandemic).

REFERENCES

[1] ANDO, T. and BAI, J. (2020). Quantile co-movement in financial markets: A panel quantile model with unobserved heterogeneity. *J. Amer. Statist. Assoc.* **115** 266–279. MR4078462 https://doi.org/10.1080/01621459.2018.1543598

[2] ANDO, T. and LI, K. (2017). A weight-relaxed model averaging approach for high-dimensional generalized linear models. *Ann. Statist.* **45** 2654–2679. MR3737905 https://doi.org/10.1214/17-AOS1538

[3] ANDO, T. and LI, K.-C. (2014). A model-averaging approach for high-dimensional regression. *J. Amer. Statist. Assoc.* **109** 254–265. MR3180561 https://doi.org/10.1080/01621459.2013.838168

[4] ANDO, T. and LI, K.-C. (2025). Supplement to "Simplex quantile regression without crossing." https://doi.org/10.1214/24-AOS2458SUPP

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

[6] BONDELL, H. D., REICH, B. J. and WANG, H. (2010). Noncrossing quantile regression curve estimation. *Biometrika* **97** 825–838. MR2746154 https://doi.org/10.1093/biomet/asq048

[7] CARHART, M. M. (1997). On persistence in mutual fund performance. *J. Finance* **52** 57–82.

[8] CHEN, C. W. S., GERLACH, R. and WEI, D. C. M. (2009). Bayesian causal effects in quantiles: Accounting for heteroscedasticity. *Comput. Statist. Data Anal.* **53** 1993–2007. MR2665090 https://doi.org/10.1016/j.csda.2008.12.014

[9] DE BOOR, C. (1978). *A Practical Guide to Splines*. *Applied Mathematical Sciences* **27**. Springer, New York. MR0507062

[10] FAMA, E. F. and FRENCH, K. R. (1993). Common factors in the returns on stocks and bonds. *J. Financ. Econ.* **33** 3–56.

[11] FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581 https://doi.org/10.1198/016214501753382273

MSC2020 subject classifications. Primary 62J99; secondary 62G20.
Key words and phrases. Asymptotic optimality, cross-validation, extrapolation, model misspecification, quantile noncrossing.

[12] FLYNN, C. J., HURVICH, C. M. and SIMONOFF, J. S. (2013). Efficiency for regularization parameter selection in penalized likelihood estimation of misspecified models. *J. Amer. Statist. Assoc.* **108** 1031–1043. MR3174682 https://doi.org/10.1080/01621459.2013.801775

[13] FRUMENTO, P. and BOTTAI, M. (2017). Parametric modeling of quantile regression coefficient functions with censored and truncated data. *Biometrics* **73** 1179–1188. MR3744532 https://doi.org/10.1111/biom.12675

[14] GERACI, M. and BOTTAI, M. (2007). Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics* **8** 140–154. https://doi.org/10.1093/biostatistics/kxj039

[15] GERLACH, R., CHEN, C. W. S. and LIN, E. M. H. (2016). Bayesian assessment of dynamic quantile forecasts. *J. Forecast.* **35** 751–764. MR3573060 https://doi.org/10.1002/for.2408

[16] GERLACH, R. H., CHEN, C. W. S. and CHAN, N. Y. C. (2011). Bayesian time-varying quantile forecasting for value-at-risk in financial markets. *J. Bus. Econom. Statist.* **29** 481–492. MR2879236 https://doi.org/10.1198/jbes.2010.08203

[17] GEYER, C. J. (1994). On the asymptotics of constrained M-estimation. *Ann. Statist.* **22** 1993–2010. MR1329179 https://doi.org/10.1214/aos/1176325768

[18] GIESSING, A. and HE, X. (2019). On the predictive risk in misspecified quantile regression. *J. Econometrics* **213** 235–260. MR4013223 https://doi.org/10.1016/j.jeconom.2019.04.013

[19] HANSEN, B. E. (2007). Least squares model averaging. *Econometrica* **75** 1175–1189. MR2333497 https://doi.org/10.1111/j.1468-0262.2007.00785.x

[20] HANSEN, B. E. and RACINE, J. S. (2012). Jackknife model averaging. *J. Econometrics* **167** 38–46. MR2885437 https://doi.org/10.1016/j.jeconom.2011.06.019

[21] HE, X. (1997). Quantile curves without crossing. *Amer. Statist.* **51** 186–191.

[22] HE, X. and SHI, P. D. (1994). Convergence rate of B-spline estimators of nonparametric conditional quantile functions. *J. Nonparametr. Stat.* **3** 299–308. MR1291551 https://doi.org/10.1080/10485259408832589

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

[24] KOENKER, R. and BASSETT, G. JR. (1978). Regression quantiles. *Econometrica* **46** 33–50. MR0474644 https://doi.org/10.2307/1913643

[25] KOENKER, R. and NG, P. (2005). Inequality constrained quantile regression. *Sankhyā* **67** 418–440. MR2208897

[26] KOZUMI, H. and KOBAYASHI, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *J. Stat. Comput. Simul.* **81** 1565–1578. MR2851270 https://doi.org/10.1080/00949655.2010.496117

[27] LI, K.-C. (1987). Asymptotic optimality for C_p, C_L, cross-validation and generalized cross-validation: Discrete index set. *Ann. Statist.* **15** 958–975. MR0902239 https://doi.org/10.1214/aos/1176350486

[28] PARKER, T. (2019). Asymptotic inference for the constrained quantile regression process. *J. Econometrics* **213** 174–189. MR4013220 https://doi.org/10.1016/j.jeconom.2019.04.010

[29] PENG, B. and WANG, L. (2015). An iterative coordinate descent algorithm for high-dimensional nonconvex penalized quantile regression. *J. Comput. Graph. Statist.* **24** 676–694. MR3397228 https://doi.org/10.1080/10618600.2014.913516

[30] POLSON, N. G. and SCOTT, J. G. (2013). Data augmentation for non-Gaussian regression models using variance-mean mixtures. *Biometrika* **100** 459–471. MR3068446 https://doi.org/10.1093/biomet/ass081

[31] RAMSAY, J. O. (1988). Monotone regression splines in action. *Statist. Sci.* **3** 425–441.

[32] REICH, B. J., FUENTES, M. and DUNSON, D. B. (2011). Bayesian spatial quantile regression. *J. Amer. Statist. Assoc.* **106** 6–20. MR2816698 https://doi.org/10.1198/jasa.2010.ap09237

[33] RODRIGUES, T., DORTET-BERNADET, J.-L. and FAN, Y. (2019). Pyramid quantile regression. *J. Comput. Graph. Statist.* **28** 732–746. MR4007754 https://doi.org/10.1080/10618600.2019.1575225

[34] RODRIGUES, T., DORTET-BERNADET, J.-L. and FAN, Y. (2019). Simultaneous fitting of Bayesian penalised quantile splines. *Comput. Statist. Data Anal.* **134** 93–109. MR3926488 https://doi.org/10.1016/j.csda.2018.12.009

[35] SHAO, J. (1993). Linear model selection by cross-validation. *J. Amer. Statist. Assoc.* **88** 486–494. MR1224373

[36] SHAO, J. (1997). An asymptotic theory for linear model selection. *Statist. Sinica* **7** 221–264. With comments and a rejoinder by the author. MR1466682

[37] SHARPE, W. F. (1964). Capital asset prices—A theory of market equilibrium under conditions of risk. *J. Finance* **19** 425–442.

[38] SHERWOOD, B. and WANG, L. (2016). Partially linear additive quantile regression in ultra-high dimension. *Ann. Statist.* **44** 288–317. MR3449769 https://doi.org/10.1214/15-AOS1367

[39] SOTTILE, G. and FRUMENTO, P. (2023). Parametric estimation of non-crossing quantile functions. *Stat. Model.* **23** 173–195. MR4582635 https://doi.org/10.1177/1471082X211036517

[40] SRIRAM, K., RAMAMOORTHI, R. V. and GHOSH, P. (2013). Posterior consistency of Bayesian quantile regression based on the misspecified asymmetric Laplace density. *Bayesian Anal.* **8** 479–504. MR3066950 https://doi.org/10.1214/13-BA817

[41] TAN, K. M., WANG, L. and ZHOU, W.-X. (2022). High-dimensional quantile regression: Convolution smoothing and concave regularization. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 205–233. MR4400395 https://doi.org/10.1111/rssb.12485

[42] TANG, Y., SONG, X., WANG, H. J. and ZHU, Z. (2013). Variable selection in high-dimensional quantile varying coefficient models. *J. Multivariate Anal.* **122** 115–132. MR3189311 https://doi.org/10.1016/j.jmva.2013.07.015

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

[44] TOKDAR, S. T. and KADANE, J. B. (2012). Simultaneous linear quantile regression: A semiparametric Bayesian approach. *Bayesian Anal.* **7** 51–72. MR2896712 https://doi.org/10.1214/12-BA702

[45] WANG, L., WU, Y. and LI, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *J. Amer. Statist. Assoc.* **107** 214–222. MR2949353 https://doi.org/10.1080/01621459.2012.656014

[46] WU, Y. and LIU, Y. (2009). Stepwise multiple quantile regression estimation using non-crossing constraints. *Stat. Interface* **2** 299–310. MR2540088 https://doi.org/10.4310/SII.2009.v2.n3.a4

[47] XU, X. and GHOSH, M. (2015). Bayesian variable selection and estimation for group lasso. *Bayesian Anal.* **10** 909–936. MR3432244 https://doi.org/10.1214/14-BA929

[48] YAN, Y. and KOTTAS, A. (2017). A new family of error distributions for Bayesian quantile regression. Working Paper.

[49] YANG, Y. and TOKDAR, S. T. (2017). Joint estimation of quantile planes over arbitrary predictor spaces. *J. Amer. Statist. Assoc.* **112** 1107–1120. MR3735363 https://doi.org/10.1080/01621459.2016.1192545

[50] YU, K. and MOYEED, R. A. (2001). Bayesian quantile regression. *Statist. Probab. Lett.* **54** 437–447. MR1861390 https://doi.org/10.1016/S0167-7152(01)00124-9

[51] YUAN, Y., CHEN, N. and ZHOU, S. (2017). Modeling regression quantile process using monotone B-splines. *Technometrics* **59** 338–350. MR3677965 https://doi.org/10.1080/00401706.2016.1211553

[52] YUE, Y. R. and RUE, H. (2011). Bayesian inference for additive mixed quantile regression models. *Comput. Statist. Data Anal.* **55** 84–96. MR2736538 https://doi.org/10.1016/j.csda.2010.05.006

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

The Annals of Statistics
2025, Vol. 53, No. 1, 170–193
https://doi.org/10.1214/24-AOS2459
This research was funded, in whole or in part, by UKRI, G116786 and EP/X038645.

BAYESIAN NONPARAMETRIC INFERENCE IN MCKEAN–VLASOV MODELS

BY RICHARD NICKL[1,a], GRIGORIOS A. PAVLIOTIS[2,b] AND KOLYAN RAY[2,c]

[1]*Department of Pure Mathematics and Mathematical Statistics, University of Cambridge,* [a]*nickl@maths.cam.ac.uk*
[2]*Department of Mathematics, Imperial College London,* [b]*g.pavliotis@imperial.ac.uk,* [c]*kolyan.ray@imperial.ac.uk*

We consider nonparametric statistical inference on a periodic interaction potential W from noisy discrete space-time measurements of solutions $\rho = \rho_W$ of the nonlinear McKean–Vlasov equation, describing the probability density of the mean field limit of an interacting particle system. We show how Gaussian process priors assigned to W give rise to posterior mean estimators that exhibit fast convergence rates for the implied estimated densities $\bar{\rho}$ towards ρ_W. We further show that if the initial condition ϕ is not too smooth and satisfies a standard deconvolvability condition, then one can consistently infer Sobolev-regular potentials W at convergence rates $N^{-\theta}$ for appropriate $\theta > 0$, where N is the number of measurements. The exponent θ can be taken to approach $1/2$ as the regularity of W increases corresponding to 'near-parametric' models.

REFERENCES

[1] AGAPIOU, S., DASHTI, M. and HELIN, T. (2021). Rates of contraction of posterior distributions based on p-exponential priors. *Bernoulli* **27** 1616–1642. MR4278794 https://doi.org/10.3150/20-bej1285

[2] AGAPIOU, S. and WANG, S. (2024). Laplace priors and spatial inhomogeneity in Bayesian inverse problems. *Bernoulli* **30** 878–910. MR4699538 https://doi.org/10.3150/22-bej1563

[3] AMORINO, C., HEIDARI, A., PILIPAUSKAITĖ, V. and PODOLSKIJ, M. (2023). Parameter estimation of discretely observed interacting particle systems. *Stochastic Process*. *Appl*. **163** 350–386. MR4612302 https://doi.org/10.1016/j.spa.2023.06.011

[4] AMORINO, C., BELOMESTNY, D., PILIPAUSKAITĖ, V., PODOLSKI, M. and ZHOU, S.-Y. (2024). Polynomial rates via deconvolution for nonparametric estimation in McKean-Vlasov SDEs. Preprint. Available at arXiv:2401.04667.

[5] BESKOS, A., GIROLAMI, M., LAN, S., FARRELL, P. E. and STUART, A. M. (2017). Geometric MCMC for infinite-dimensional inverse problems. *J*. *Comput*. *Phys*. **335** 327–351. MR3612501 https://doi.org/10.1016/j.jcp.2016.12.041

[6] BISHWAL, J. P. N. (2011). Estimation in interacting diffusions: Continuous and discrete sampling. *Appl*. *Math*. (*Irvine*) **2** 1154–1158. MR2924983 https://doi.org/10.4236/am.2011.29160

[7] BOGACHEV, V. I., RÖCKNER, M. and SHAPOSHNIKOV, S. V. (2016). Distances between transition probabilities of diffusions and applications to nonlinear Fokker–Planck–Kolmogorov equations. *J*. *Funct*. *Anal*. **271** 1262–1300. MR3522009 https://doi.org/10.1016/j.jfa.2016.05.016

[8] CARRILLO, J. A., CHERTOCK, A. and HUANG, Y. (2015). A finite-volume method for nonlinear nonlocal equations with a gradient flow structure. *Commun*. *Comput*. *Phys*. **17** 233–258. MR3372289 https://doi.org/10.4208/cicp.160214.010814a

[9] CARRILLO, J. A., GVALANI, R. S., PAVLIOTIS, G. A. and SCHLICHTING, A. (2020). Long-time behaviour and phase transitions for the McKean–Vlasov equation on the torus. *Arch*. *Ration*. *Mech*. *Anal*. **235** 635–690. MR4062483 https://doi.org/10.1007/s00205-019-01430-4

[10] CHAZELLE, B., JIU, Q., LI, Q. and WANG, C. (2017). Well-posedness of the limiting equation of a noisy consensus model in opinion dynamics. *J*. *Differ*. *Equ*. **263** 365–397. MR3631310 https://doi.org/10.1016/j.jde.2017.02.036

[11] COMTE, F. and GENON-CATALOT, V. (2023). Nonparametric adaptive estimation for interacting particle systems. *Scand*. *J*. *Stat*. **50** 1716–1755. MR4677374 https://doi.org/10.1111/sjos.12661

MSC2020 subject classifications. Primary 62G20, 62F15; secondary 35Q70, 35Q84.
Key words and phrases. McKean–Vlasov PDE, Bayesian inverse problems, Gaussian processes, interacting particle systems.

[12] 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

[13] DAWSON, D. A. (1983). Critical dynamics and fluctuations for a mean-field model of cooperative behavior. *J. Stat. Phys.* **31** 29–85. MR0711469 https://doi.org/10.1007/BF01010922

[14] DELGADINO, M. G., GVALANI, R. S. and PAVLIOTIS, G. A. (2021). On the diffusive-mean field limit for weakly interacting diffusions exhibiting phase transitions. *Arch. Ration. Mech. Anal.* **241** 91–148. MR4271956 https://doi.org/10.1007/s00205-021-01648-1

[15] DELGADINO, M. G., GVALANI, R. S., PAVLIOTIS, G. A. and SMITH, S. A. (2023). Phase transitions, logarithmic Sobolev inequalities, and uniform-in-time propagation of chaos for weakly interacting diffusions. *Comm. Math. Phys.* **401** 275–323. MR4604897 https://doi.org/10.1007/s00220-023-04659-z

[16] DELLA MAESTRA, L. and HOFFMANN, M. (2022). Nonparametric estimation for interacting particle systems: McKean–Vlasov models. *Probab. Theory Related Fields* **182** 551–613. MR4367954 https://doi.org/10.1007/s00440-021-01044-6

[17] DELLA MAESTRA, L. and HOFFMANN, M. (2023). The LAN property for McKean–Vlasov models in a mean-field regime. *Stochastic Process. Appl.* **155** 109–146. MR4503434 https://doi.org/10.1016/j.spa.2022.10.002

[18] EVANS, L. C. (1998). *Partial Differential Equations*. Amer. Math. Soc., Providence, RI.

[19] FOLLAND, G. B. (1999). *Real Analysis: Modern Techniques and Their Applications*, 2nd ed. *Pure and Applied Mathematics (New York)*. Wiley, New York. MR1681462

[20] FOURNIER, N. and JOURDAIN, B. (2017). Stochastic particle approximation of the Keller–Segel equation and two-dimensional generalization of Bessel processes. *Ann. Appl. Probab.* **27** 2807–2861. MR3719947 https://doi.org/10.1214/16-AAP1267

[21] FRANK, T. D. (2005). *Nonlinear Fokker–Planck Equations: Fundamentals and Applications*. *Springer Series in Synergetics*. Springer, Berlin. MR2118870

[22] GENON-CATALOT, V. and LARÉDO, C. (2021). Parametric inference for small variance and long time horizon McKean–Vlasov diffusion models. *Electron. J. Stat.* **15** 5811–5854. MR4355698 https://doi.org/10.1214/21-ejs1922

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

[24] GIESECKE, K., SCHWENKLER, G. and SIRIGNANO, J. A. (2020). Inference for large financial systems. *Math. Finance* **30** 3–46. MR4067069 https://doi.org/10.1111/mafi.12222

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

[26] GIORDANO, M. and RAY, K. (2022). Nonparametric Bayesian inference for reversible multidimensional diffusions. *Ann. Statist.* **50** 2872–2898. MR4500628 https://doi.org/10.1214/22-aos2213

[27] GOLSE, F. (2016). On the dynamics of large particle systems in the mean field limit. In *Macroscopic and Large Scale Phenomena: Coarse Graining, Mean Field Limits and Ergodicity*. *Lect. Notes Appl. Math. Mech.* **3** 1–144. Springer, Cham. MR3468297 https://doi.org/10.1007/978-3-319-26883-5_1

[28] GOMES, S. N., PAVLIOTIS, G. A. and VAES, U. (2020). Mean field limits for interacting diffusions with colored noise: Phase transitions and spectral numerical methods. *Multiscale Model. Simul.* **18** 1343–1370. MR4143413 https://doi.org/10.1137/19M1258116

[29] GOODWIN, T., EVENHUIS, C., WOODCOCK, S. and QUIROZ, M. (2020). Bayesian inference on the Keller–Segel model. In *Proceedings of the 14th Biennial Engineering Mathematics and Applications Conference, EMAC-2019*. *ANZIAM J.* **61** C181–C196.

[30] GUILLIN, A., BRIS, P. L. and MONMARCHÉ, P. (2024). Uniform in time propagation of chaos for the 2D vortex model and other singular stochastic systems. J. Eur. Math. Soc. Published Online First.

[31] GUPTA, S., CAMPA, A. and RUFFO, S. (2018). *Statistical Physics of Synchronization*. *SpringerBriefs in Complexity*. Springer, Cham. With a foreword by Steven Strogatz. MR3837140 https://doi.org/10.1007/978-3-319-96664-9

[32] GVALANI, R. S. (2020). Many-particle systems: Phase transitions, distinguished limits, and gradient flows Ph.D. thesis Imperial College London.

[33] HEGSELMANN, R. and KRAUSE, U. (2002). Opinion dynamics and bounded confidence: Models, analysis and simulation. *J. Artif. Soc. Soc. Simul.* **5**.

[34] HELLMUTH, K., KLINGENBERG, C., LI, Q. and TANG, M. (2021). Multiscale convergence of the inverse problem for Chemotaxis in the Bayesian setting. *Computation* **9**. https://doi.org/10.3390/computation9110119

[35] HOFFMANN, M. and RAY, K. (2024). Nonparametric Bayesian estimation in a multidimensional diffusion model with high frequency data. *Probab. Theory Related Fields*. To appear.
[36] KASONGA, R. A. (1990). Maximum likelihood theory for large interacting systems. *SIAM J. Appl. Math.* **50** 865–875. MR1050917 https://doi.org/10.1137/0150050
[37] KOTZ, S., KOZUBOWSKI, T. J. and PODGÓRSKI, K. (2001). *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*. Birkhäuser, Boston, MA. MR1935481 https://doi.org/10.1007/978-1-4612-0173-1
[38] LACKER, D. and LE FLEM, L. (2023). Sharp uniform-in-time propagation of chaos. *Probab. Theory Related Fields* **187** 443–480. MR4634344 https://doi.org/10.1007/s00440-023-01192-x
[39] LEE, S., PSARELLIS, Y. M., SIETTOS, C. I. and KEVREKIDIS, I. G. (2023). Learning black- and gray-box chemotactic PDEs/closures from agent based Monte Carlo simulation data. *J. Math. Biol.* **87** Paper No. 15, 28. MR4606241 https://doi.org/10.1007/s00285-023-01946-0
[40] LIONS, J.-L. and MAGENES, E. (1972). *Non-homogeneous Boundary Value Problems and Applications. Vol. II. Die Grundlehren der Mathematischen Wissenschaften, Band* 182. Springer, New York. Translated from the French by P. Kenneth. MR0350178
[41] MALRIEU, F. (2001). Logarithmic Sobolev inequalities for some nonlinear PDE's. *Stochastic Process. Appl.* **95** 109–132. MR1847094 https://doi.org/10.1016/S0304-4149(01)00095-3
[42] MILANI, A. (1999). A remark on the Sobolev regularity of classical solutions to uniformly parabolic equations. *Math. Nachr.* **199** 115–144. MR1676322 https://doi.org/10.1002/mana.19991990107
[43] MISCHLER, S. and MOUHOT, C. (2013). Kac's program in kinetic theory. *Invent. Math.* **193** 1–147. MR3069113 https://doi.org/10.1007/s00222-012-0422-3
[44] MONARD, F., NICKL, R. and PATERNAIN, G. P. (2021). Consistent inversion of noisy non-Abelian X-ray transforms. *Comm. Pure Appl. Math.* **74** 1045–1099. MR4230066 https://doi.org/10.1002/cpa.21942
[45] NICKL, R. (2020). Bernstein–von Mises theorems for statistical inverse problems I: Schrödinger equation. *J. Eur. Math. Soc. (JEMS)* **22** 2697–2750. MR4118619 https://doi.org/10.4171/JEMS/975
[46] NICKL, R. (2024). Consistent inference for diffusions from low frequency measurements. *Ann. Statist.* **52** 519–549. MR4744186 https://doi.org/10.1214/24-aos2357
[47] NICKL, R. (2023). *Bayesian Non-linear Statistical Inverse Problems. Zurich Lectures in Advanced Mathematics*. EMS Press, Berlin. MR4604099 https://doi.org/10.4171/zlam/30
[48] NICKL, R. and RAY, K. (2020). Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions. *Ann. Statist.* **48** 1383–1408. MR4124327 https://doi.org/10.1214/19-AOS1851
[49] NICKL, R. and TITI, E. S. (2024). On posterior consistency of data assimilation with Gaussian process priors: The 2D-Navier–Stokes equations. *Ann. Statist.* **52** 1825–1844. MR4804829 https://doi.org/10.1214/24-aos2427
[50] NICKL, R. and WANG, S. (2024). On polynomial-time computation of high-dimensional posterior measures by Langevin-type algorithms. *J. Eur. Math. Soc. (JEMS)* **26** 1031–1112. MR4721029 https://doi.org/10.4171/jems/1304
[51] OELSCHLÄGER, K. (1984). A martingale approach to the law of large numbers for weakly interacting stochastic processes. *Ann. Probab.* **12** 458–479. MR0735849
[52] PAINTER, K. J., HILLEN, T. and POTTS, J. R. (2024). Biological modeling with nonlocal advection-diffusion equations. *Math. Models Methods Appl. Sci.* **34** 57–107. MR4683281 https://doi.org/10.1142/S0218202524400025
[53] PARESCHI, L. and TOSCANI, G. (2013). *Interacting Multiagent Systems: Kinetic Equations and Monte Carlo Methods*. OUP Oxford.
[54] PAVLIOTIS, G. A., STUART, A. M. and VAES, U. (2022). Derivative-free Bayesian inversion using multiscale dynamics. *SIAM J. Appl. Dyn. Syst.* **21** 284–326. MR4368987 https://doi.org/10.1137/21M1397416
[55] PAVLIOTIS, G. A. and ZANONI, A. (2022). Eigenfunction martingale estimators for interacting particle systems and their mean field limit. *SIAM J. Appl. Dyn. Syst.* **21** 2338–2370. MR4513316 https://doi.org/10.1137/21M1464348
[56] PAVLIOTIS, G. A. and ZANONI, A. (2024). A method of moments estimator for interacting particle systems and their mean field limit. *SIAM/ASA J. Uncertain. Quantificat.* **12** 262–288. MR4727693 https://doi.org/10.1137/22M153848X
[57] PINNAU, R., TOTZECK, C., TSE, O. and MARTIN, S. (2017). A consensus-based model for global optimization and its mean-field limit. *Math. Models Methods Appl. Sci.* **27** 183–204. MR3597012 https://doi.org/10.1142/S0218202517400061
[58] RASMUSSEN, A. K., SEIZILLES, F., GIROLAMI, M. and KAZLAUSKAITE, I. (2024). The Bayesian approach to inverse Robin problems. *SIAM/ASA J. Uncertain. Quantificat.* **12** 1050–1084. MR4796198 https://doi.org/10.1137/23M1620624
[59] RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning. Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR2514435

[60] REHMEIER, M. and RÖCKNER, M. (2022). On nonlinear Markov processes in the sense of McKean. Preprint. Available at arXiv:2212.12424.

[61] REISS, M. (2008). Asymptotic equivalence for nonparametric regression with multivariate and random design. *Ann. Statist.* **36** 1957–1982. MR2435461 https://doi.org/10.1214/07-AOS525

[62] SCHLICHTING, A. and SEIS, C. (2022). The Scharfetter–Gummel scheme for aggregation-diffusion equations. *IMA J. Numer. Anal.* **42** 2361–2402. MR4454925 https://doi.org/10.1093/imanum/drab039

[63] SHARROCK, L., KANTAS, N., PARPAS, P. and PAVLIOTIS, G. A. (2023). Online parameter estimation for the McKean–Vlasov stochastic differential equation. *Stochastic Process. Appl.* **162** 481–546. MR4597535 https://doi.org/10.1016/j.spa.2023.05.002

[64] SPRUNGK, B., WEISSMANN, S. and ZECH, J. (2023). Metropolis-adjusted interacting particle sampling.

[65] STUART, A. M. (2010). Inverse problems: A Bayesian perspective. *Acta Numer.* **19** 451–559. MR2652785 https://doi.org/10.1017/S0962492910000061

[66] SZNITMAN, A.-S. (1991). Topics in propagation of chaos. In *École D'Été de Probabilités de Saint-Flour XIX*—1989. *Lecture Notes in Math.* **1464** 165–251. Springer, Berlin. MR1108185 https://doi.org/10.1007/BFb0085169

[67] WLOKA, J. (1987). *Partial Differential Equations*. Cambridge Univ. Press, Cambridge. Translated from the German by C. B. Thomas and M. J. Thomas. MR0895589 https://doi.org/10.1017/CBO9781139171755

The Annals of Statistics
2025, Vol. 53, No. 1, 194–218
https://doi.org/10.1214/24-AOS2460

ENSEMBLE PROJECTION PURSUIT FOR GENERAL NONPARAMETRIC REGRESSION

BY HAORAN ZHAN[1,a], MINGKE ZHANG[1,b] AND YINGCUN XIA[1,2,c]

[1]*Department of Statistics and Data Science, National University of Singapore,* [a]*haoran.zhan@u.nus.edu,* [b]*mingke.zhang@u.nus.edu,* [c]*yingcun.xia@nus.edu.sg*

[2]*School of Mathematics, University of Electronic Science and Technology of China*

Projection pursuit regression (PPR) and artificial neural networks (ANNs), both introduced around the same time, share a similar approach to approximating complex structures in data. However, PPR has received far less attention than ANNs and other popular statistical learning methods such as random forests (RF) and support vector machines (SVM). In this paper, we revisit the estimation of PPR and propose an *optimal* greedy algorithm and an ensemble approach via feature bagging, hereafter referred to as ePPR, to improve its effectiveness. Compared to RF, ePPR has two main advantages. First, its theoretical consistency can be proved for more general regression functions, as long as they are L^2 integrable, and higher consistency rates can be achieved. Second, ePPR does not split the samples, so each term of the PPR is estimated using the entire data, making minimisation more efficient and ensuring the smoothness of the estimator. Extensive comparisons on real data sets show that ePPR is more efficient in regression and classification than RF and other competitors.

REFERENCES

[1] ATHEY, S., TIBSHIRANI, J. and WAGER, S. (2019). Generalized random forests. *Ann. Statist.* **47** 1148–1178. MR3909963 https://doi.org/10.1214/18-AOS1709

[2] AU, T. C. (2018). Random forests, decision trees, and categorical predictors: The "absent levels" problem. *J. Mach. Learn. Res.* **19** Paper No. 45, 30 pp. MR3874153

[3] BAGIROV, A. M., CLAUSEN, C. and KOHLER, M. (2009). Estimation of a regression function by maxima of minima of linear functions. *IEEE Trans. Inf. Theory* **55** 833–845. MR2597271 https://doi.org/10.1109/TIT.2008.2009835

[4] BARRON, A. R., COHEN, A., DAHMEN, W. and DEVORE, R. A. (2008). Approximation and learning by greedy algorithms. *Ann. Statist.* **36** 64–94. MR2387964 https://doi.org/10.1214/009053607000000631

[5] BARTLETT, P., MAIOROV, V. and MEIR, R. (1998). Almost linear VC dimension bounds for piecewise polynomial networks. *Adv. Neural Inf. Process. Syst.* **11**.

[6] BREIMAN, L. (2001). Statistical modeling: The two cultures. *Statist. Sci.* **16** 199–231. MR1874152 https://doi.org/10.1214/ss/1009213726

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

[8] CAI, T. T. and WANG, L. (2011). Orthogonal matching pursuit for sparse signal recovery with noise. *IEEE Trans. Inf. Theory* **57** 4680–4688. MR2840484 https://doi.org/10.1109/TIT.2011.2146090

[9] CHEN, H. (1991). Estimation of a projection-pursuit type regression model. *Ann. Statist.* **19** 142–157. MR1091843 https://doi.org/10.1214/aos/1176347974

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

[11] CORTES, C. and VAPNIK, V. (1995). Support-vector networks. *Mach. Learn.* **20** 273–297.

[12] DEVORE, R. A. and TEMLYAKOV, V. N. (1996). Some remarks on greedy algorithms. *Adv. Comput. Math.* **5** 173–187. MR1399379 https://doi.org/10.1007/BF02124742

MSC2020 subject classifications. Primary 62G08; secondary 62G20.

Key words and phrases. Convergence of algorithm, consistency of estimation, greedy algorithm, nonparametric regression and classification, projection pursuit regression, random forest.

[13] FRIEDBERG, R., TIBSHIRANI, J., ATHEY, S. and WAGER, S. (2021). Local linear forests. *J. Comput. Graph. Statist.* **30** 503–517. MR4270520 https://doi.org/10.1080/10618600.2020.1831930

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

[15] FRIEDMAN, J. H. and TUKEY, J. W. (1974). A projection pursuit algorithm for exploratory data analysis. *IEEE Trans. Comput.* **100** 881–890.

[16] 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

[17] 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

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

[19] HO, T. K. (1995). Random decision forests. In *Document Analysis and Recognition*, 1995, *Proceedings of the Third International Conference on* **1** 278–282. IEEE.

[20] JONES, L. K. (1987). On a conjecture of Huber concerning the convergence of projection pursuit regression. *Ann. Statist.* **15** 880–882. MR0888447 https://doi.org/10.1214/aos/1176350382

[21] JONES, L. K. (1992). A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training. *Ann. Statist.* **20** 608–613. MR1150368 https://doi.org/10.1214/aos/1176348546

[22] KLUSOWSKI, J. M. and TIAN, P. M. (2024). Large scale prediction with decision trees. *J. Amer. Statist. Assoc.* **119** 525–537. MR4713911 https://doi.org/10.1080/01621459.2022.2126782

[23] 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

[24] KRUSKAL, J. B. (1969). Toward a practical method which helps uncover the structure of a set of multivariate observations by finding the linear transformation which optimizes a new index of condensation. In *Statistical Computation* Academic Press, San Diego.

[25] LESHNO, M., LIN, V. Y., PINKUS, A. and SCHOCKEN, S. (1993). Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Netw.* **6** 861–867.

[26] PATI, Y. C., REZAIIFAR, R. and KRISHNAPRASAD, P. S. (1993). Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Proceedings of* 27*th Asilomar Conference on Signals*, *Systems and Computers* 40–44. IEEE.

[27] PETRUSHEV, P. P. (1999). Approximation by ridge functions and neural networks. *SIAM J. Math. Anal.* **30** 155–189. MR1646689 https://doi.org/10.1137/S0036141097322959

[28] RASHID, A. A. and VUGAR, E. I. (2020). A representation problem for smooth sums of ridge functions. *J. Approx. Theory* **257** 105448.

[29] 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

[30] SCORNET, E., BIAU, G. and VERT, J.-P. (2015). Consistency of random forests. *Ann. Statist.* **43** 1716–1741. MR3357876 https://doi.org/10.1214/15-AOS1321

[31] SHALEV-SHWARTZ, S. and BEN-DAVID, S. (2014). *Understanding Machine Learning*: *From Theory to Algorithms*. Cambridge Univ. Press, Cambridge.

[32] SIEGEL, J. W. and XU, J. (2022). Optimal convergence rates for the orthogonal greedy algorithm. *IEEE Trans. Inf. Theory* **68** 3354–3361. MR4433225 https://doi.org/10.1109/tit.2022.3147984

[33] SIEGEL, J. W. and XU, J. (2024). Sharp bounds on the approximation rates, metric entropy, and n-widths of shallow neural networks. *Found. Comput. Math.* **24** 481–537. MR4733356 https://doi.org/10.1007/s10208-022-09595-3

[34] TEMLYAKOV, V. N. (2015). Greedy approximation in convex optimization. *Constr. Approx.* **41** 269–296. MR3315679 https://doi.org/10.1007/s00365-014-9272-0

[35] VAN DER VAART, A. W. and WELLNER, J. A. (2023). *Weak Convergence and Empirical Processes—with Applications to Statistics*. *Springer Series in Statistics*. Springer, Cham. MR4628026 https://doi.org/10.1007/978-3-031-29040-4

[36] WAGER, S. and WALTHER, G. (2015). Adaptive concentration of regression trees, with application to random forests. Preprint arXiv:1503.06388.

[37] XIA, Y., TONG, H. and LI, W. K. (1999). On extended partially linear single-index models. *Biometrika* **86** 831–842. MR1741980 https://doi.org/10.1093/biomet/86.4.831

[38] 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

[39] ZHAN, H., ZHANG, M. and XIA, Y. (2025). Supplement to "Ensemble projection pursuit for general nonparametric regression." https://doi.org/10.1214/24-AOS2460SUPP

[40] ZHANG, H. and WANG, M. (2009). Search for the smallest random forest. *Stat*. *Interface* **2** 381–388. MR2540095 https://doi.org/10.4310/SII.2009.v2.n3.a11

[41] ZHANG, T. and YU, B. (2005). Boosting with early stopping: Convergence and consistency. *Ann*. *Statist*. **33** 1538–1579. MR2166555 https://doi.org/10.1214/009053605000000255

[42] ZHU, R., ZENG, D. and KOSOROK, M. R. (2015). Reinforcement learning trees. *J*. *Amer. Statist*. *Assoc*. **110** 1770–1784. MR3449072 https://doi.org/10.1080/01621459.2015.1036994

The Annals of Statistics
2025, Vol. 53, No. 1, 219–244
https://doi.org/10.1214/24-AOS2461

RATE-OPTIMAL ESTIMATION OF MIXED SEMIMARTINGALES

BY CARSTEN H. CHONG[1,a], THOMAS DELERUE[2,b] AND FABIAN MIES[3,c]

[1]*Department of Information Systems, Business Statistics and Operations Management, The Hong Kong University of Science and Technology,* [a]*carstenchong@ust.hk*

[2]*Institute of Epidemiology, Helmholtz Munich,* [b]*thomas.delerue@gmail.com*

[3]*Delft Institute of Applied Mathematics, Delft University of Technology,* [c]*f.mies@tudelft.nl*

Consider the sum $Y = B + B(H)$ of a Brownian motion B and an independent fractional Brownian motion $B(H)$ with Hurst parameter $H \in (0, 1)$. Even though $B(H)$ is not a semimartingale, it was shown by Cheridito (*Bernoulli* **7** (2001) 913–934) that Y is a semimartingale if $H > 3/4$. Moreover, Y is locally equivalent to B in this case, so H cannot be consistently estimated from local observations of Y. This paper pivots on another unexpected feature in this model: if B and $B(H)$ become correlated, then Y will never be a semimartingale, and H can be identified, regardless of its value. This and other results will follow from a detailed statistical analysis of a more general class of processes called *mixed semimartingales*, which are semiparametric extensions of Y with stochastic volatility in both the martingale and the fractional component. In particular, we derive consistent estimators and feasible central limit theorems for all parameters and processes that can be identified from high-frequency observations. We further show that our estimators achieve optimal rates in a minimax sense.

REFERENCES

[1] AÏT-SAHALIA, Y. and JACOD, J. (2009). Testing for jumps in a discretely observed process. *Ann. Statist.* **37** 184–222. MR2488349 https://doi.org/10.1214/07-AOS568

[2] AÏT-SAHALIA, Y. and JACOD, J. (2014). *High-Frequency Financial Econometrics*. Princeton Univ. Press, Princeton, NJ.

[3] AÏT-SAHALIA, Y. and YU, J. (2009). High frequency market microstructure noise estimates and liquidity measures. *Ann. Appl. Stat.* **3** 422–457. MR2668714 https://doi.org/10.1214/08-AOAS200

[4] AMBLARD, P.-O., COEURJOLLY, J.-F., LAVANCIER, F. and PHILIPPE, A. (2013). Basic properties of the multivariate fractional Brownian motion. In *Self-Similar Processes and Their Applications. Sémin. Congr.* **28** 63–84. Soc. Math. France, Paris. MR3203519

[5] BARNDORFF-NIELSEN, O. E., CORCUERA, J. M. and PODOLSKIJ, M. (2011). Multipower variation for Brownian semistationary processes. *Bernoulli* **17** 1159–1194. MR2854768 https://doi.org/10.3150/10-BEJ316

[6] BARNDORFF-NIELSEN, O. E. and SHEPHARD, N. (2005). Power variation and time change. *Teor. Veroyatn. Primen.* **50** 115–130. MR2222740 https://doi.org/10.1137/S0040585X97981482

[7] BROUSTE, A. and FUKASAWA, M. (2018). Local asymptotic normality property for fractional Gaussian noise under high-frequency observations. *Ann. Statist.* **46** 2045–2061. MR3845010 https://doi.org/10.1214/17-AOS1611

[8] CAI, C., CHIGANSKY, P. and KLEPTSYNA, M. (2016). Mixed Gaussian processes: A filtering approach. *Ann. Probab.* **44** 3032–3075. MR3531685 https://doi.org/10.1214/15-AOP1041

[9] CHEN, D., CHENG, Y., CHONG, C. H., GENTINE, P., JIA, W., MONIER, B. and SHEN, S. (2022). Pre-averaging fractional processes contaminated by noise, with an application to turbulence. Available at arXiv:2212.00867.

[10] CHEN, D., MYKLAND, P. A. and ZHANG, L. (2020). The five trolls under the bridge: Principal component analysis with asynchronous and noisy high frequency data. *J. Amer. Statist. Assoc.* **115** 1960–1977. MR4189770 https://doi.org/10.1080/01621459.2019.1672555

MSC2020 subject classifications. Primary 60G22, 62G20, 62M09; secondary 60F05, 62P20.

Key words and phrases. Central limit theorem, high-frequency observations, Hurst parameter, KL divergence, minimax rate, mixed fractional Brownian motion, rough noise.

[11] CHERIDITO, P. (2001). Mixed fractional Brownian motion. *Bernoulli* **7** 913–934. MR1873835 https://doi.org/10.2307/3318626

[12] CHIGANSKY, P. and KLEPTSYNA, M. (2023). Estimation of the Hurst parameter from continuous noisy data. *Electron. J. Stat.* **17** 2343–2385. MR4649984 https://doi.org/10.1214/23-ejs2156

[13] CHONG, C. H., DELERUE, T. and LI, G. (2024). When frictions are fractional: Rough noise in high-frequency data. *J. Amer. Statist. Assoc.* Forthcoming. https://doi.org/10.1080/01621459.2024.2428466

[14] CHONG, C. H., DELERUE, T. and MIES, F. (2025). Supplement to “Rate-optimal estimation of mixed semimartingales.” https://doi.org/10.1214/24-AOS2461SUPP

[15] CHONG, C. H., HOFFMANN, M., LIU, Y., ROSENBAUM, M. and SZYMANSKI, G. (2024). Statistical inference for rough volatility: Central limit theorems. *Ann. Appl. Probab.* **34** 2600–2649. MR4757488 https://doi.org/10.1214/23-aap2002

[16] COMTE, F. and RENAULT, E. (1998). Long memory in continuous-time stochastic volatility models. *Math. Finance* **8** 291–323. MR1645101 https://doi.org/10.1111/1467-9965.00057

[17] CORCUERA, J. M., HEDEVANG, E., PAKKANEN, M. S. and PODOLSKIJ, M. (2013). Asymptotic theory for Brownian semi-stationary processes with application to turbulence. *Stochastic Process. Appl.* **123** 2552–2574. MR3054536 https://doi.org/10.1016/j.spa.2013.03.011

[18] DELATTRE, S. and JACOD, J. (1997). A central limit theorem for normalized functions of the increments of a diffusion process, in the presence of round-off errors. *Bernoulli* **3** 1–28. MR1466543 https://doi.org/10.2307/3318650

[19] DIEBOLD, F. X. and STRASSER, G. (2013). On the correlation structure of microstructure noise: A financial economic approach. *Rev. Econ. Stud.* **80** 1304–1337. MR3129949 https://doi.org/10.1093/restud/rdt008

[20] DOZZI, M., MISHURA, Y. and SHEVCHENKO, G. (2015). Asymptotic behavior of mixed power variations and statistical estimation in mixed models. *Stat. Inference Stoch. Process.* **18** 151–175. MR3348583 https://doi.org/10.1007/s11203-014-9106-5

[21] GATHERAL, J., JAISSON, T. and ROSENBAUM, M. (2018). Volatility is rough. *Quant. Finance* **18** 933–949. MR3805308 https://doi.org/10.1080/14697688.2017.1393551

[22] HANSEN, P. R. and LUNDE, A. (2006). Realized variance and market microstructure noise. *J. Bus. Econom. Statist.* **24** 127–218. MR2234447 https://doi.org/10.1198/073500106000000071

[23] HURST, H. E. (1951). Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civ. Eng.* **116** 770–799.

[24] JACOD, J., LI, Y. and ZHENG, X. (2017). Statistical properties of microstructure noise. *Econometrica* **85** 1133–1174. MR3681766 https://doi.org/10.3982/ECTA13085

[25] JACOD, J. and PROTTER, P. (2012). *Discretization of Processes. Stochastic Modelling and Applied Probability* **67**. Springer, Heidelberg. MR2859096 https://doi.org/10.1007/978-3-642-24127-7

[26] JACOD, J. and SØRENSEN, M. (2018). A review of asymptotic theory of estimating functions. *Stat. Inference Stoch. Process.* **21** 415–434. MR3824976 https://doi.org/10.1007/s11203-018-9178-8

[27] JACOD, J. and TODOROV, V. (2014). Efficient estimation of integrated volatility in presence of infinite variation jumps. *Ann. Statist.* **42** 1029–1069. MR3210995 https://doi.org/10.1214/14-AOS1213

[28] LELAND, W., TAQQU, M., WILLINGER, W. and WILSON, D. (1994). On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Trans. Netw.* **2** 1–15.

[29] LI, J. and XIU, D. (2016). Generalized method of integrated moments for high-frequency data. *Econometrica* **84** 1613–1633. MR3537165 https://doi.org/10.3982/ECTA12306

[30] LI, Y. and MYKLAND, P. A. (2015). Rounding errors and volatility estimation. *J. Financ. Econom.* **13** 478–504.

[31] LIEBERMAN, O., ROSEMARIN, R. and ROUSSEAU, J. (2012). Asymptotic theory for maximum likelihood estimation of the memory parameter in stationary Gaussian processes. *Econometric Theory* **28** 457–470. MR2913638 https://doi.org/10.1017/S0266466611000399

[32] MIES, F. and PODOLSKIJ, M. (2023). Estimation of mixed fractional stable processes using high-frequency data. *Ann. Statist.* **51** 1946–1964. MR4678791 https://doi.org/10.1214/23-aos2312

[33] MIKOSCH, T., RESNICK, S., ROOTZÉN, H. and STEGEMAN, A. (2002). Is network traffic approximated by stable Lévy motion or fractional Brownian motion? *Ann. Appl. Probab.* **12** 23–68. MR1890056 https://doi.org/10.1214/aoap/1015961155

[34] MISHURA, Y. S. (2008). *Stochastic Calculus for Fractional Brownian Motion and Related Processes. Lecture Notes in Math.* **1929**. Springer, Berlin. MR2378138 https://doi.org/10.1007/978-3-540-75873-0

[35] MOLZ, F. J., LIU, H. H. and SZULGA, J. (1997). Fractional Brownian motion and fractional Gaussian noise in subsurface hydrology: A review, presentation of fundamental properties, and extensions. *Water Resour. Res.* **33** 2273–2286.

[36] NEWEY, W. K. and WEST, K. D. (1994). Automatic lag selection in covariance matrix estimation. *Rev. Econ. Stud.* **61** 631–653. MR1299308 https://doi.org/10.2307/2297912

[37] PANDEY, G., LOVEJOY, S. and SCHERTZER, D. (1998). Multifractal analysis of daily river flows including extremes for basins of five to two million square kilometres, one day to 75 years. *J. Hydrol.* **208** 62–81.

[38] ROSENBAUM, M. (2009). Integrated volatility and round-off error. *Bernoulli* **15** 687–720. MR2555195 https://doi.org/10.3150/08-BEJ170

[39] SAMORODNITSKY, G. and TAQQU, M. S. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance. Stochastic Modeling*. CRC Press, New York. MR1280932

[40] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation. Springer Series in Statistics*. Springer, New York. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats. MR2724359 https://doi.org/10.1007/b13794

[41] VAN ZANTEN, H. (2007). When is a linear combination of independent fBm's equivalent to a single fBm? *Stochastic Process. Appl.* **117** 57–70. MR2287103 https://doi.org/10.1016/j.spa.2006.05.013

[42] WILLIAMS, S. D., BOCK, Y., FANG, P., JAMASON, P., NIKOLAIDIS, R. M., PRAWIRODIRDJO, L., MILLER, M. and JOHNSON, D. J. (2004). Error analysis of continuous GPS position time series. *J. Geophys. Res.* **109** B03412.

[43] XU, C. (2017). Detecting periodic oscillations in astronomy data: Revisiting wavelet analysis with coloured and white noise. *Mon. Not. R. Astron. Soc.* **466** 3827–3833.

The Annals of Statistics
2025, Vol. 53, No. 1, 245–267
https://doi.org/10.1214/24-AOS2463

SOME THEORY ABOUT EFFICIENT DIMENSION REDUCTION REGARDING THE INTERACTION BETWEEN TWO RESPONSES

BY WEI LUO[1,a]

[1]*Center for Data Science, Zhejiang University,* [a]*weiluo@zju.edu.cn*

Efficient dimension reduction regarding the interaction between two response variables, which facilitates statistical analysis in multiple important application scenarios, was initially discussed by Luo (*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** (2022) 269–294). The efficient dimension reduction subspaces were introduced, and, under mild conditions on the predictor, they were equated with the family of dual inverse regression subspaces. Besides the general framework, however, limited theory has been proposed to uncover the mystery of these spaces. In this paper, we propose a thorough characterization of the family of dual inverse regression subspaces, including their uniform lower and upper bounds, their explicit forms, their consistent and exhaustive estimation, some interesting special cases, and certain subfamilies that have desired sparsity. In addition, we extend some of these results to provide more insights into the efficient dimension reduction subspaces under the general settings, including their uniform lower and upper bounds, and the sufficient and necessary conditions for the uniqueness of the space that are assessable in practice. These results largely complete the theoretical foundation for the new type of dimension reduction, and, as such, enhance its applicability in statistical problems such as missing data analysis and causal inference. The latter is illustrated by simulation studies and a real data example at the end.

REFERENCES

[1] CHEN, X., ZOU, C. and COOK, R. D. (2010). Coordinate-independent sparse sufficient dimension reduction and variable selection. *Ann. Statist.* **38** 3696–3723. MR2766865 https://doi.org/10.1214/10-AOS826

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

[3] COOK, R. D. and WEISBERG, S. (1991). Discussion of "Sliced inverse regression for dimension reduction". *J. Amer. Statist. Assoc.* **86** 316–342.

[4] COOK, R. D. and ZHANG, X. (2015). Foundations for envelope models and methods. *J. Amer. Statist. Assoc.* **110** 599–611. MR3367250 https://doi.org/10.1080/01621459.2014.983235

[5] DE LUNA, X., WAERNBAUM, I. and RICHARDSON, T. S. (2011). Covariate selection for the nonparametric estimation of an average treatment effect. *Biometrika* **98** 861–875. MR2860329 https://doi.org/10.1093/biomet/asr041

[6] DEHEJIA, R. H. and WAHBA, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *J. Amer. Statist. Assoc.* **94** 1053–1062.

[7] DIACONIS, P. and FREEDMAN, D. (1984). Asymptotics of graphical projection pursuit. *Ann. Statist.* **12** 793–815. MR0751274 https://doi.org/10.1214/aos/1176346703

[8] FAN, J., XUE, L. and YAO, J. (2017). Sufficient forecasting using factor models. *J. Econometrics* **201** 292–306. MR3717565 https://doi.org/10.1016/j.jeconom.2017.08.009

[9] GHOSH, D. (2011). Propensity score modelling in observational studies using dimension reduction methods. *Statist. Probab. Lett.* **81** 813–820. MR2793748 https://doi.org/10.1016/j.spl.2011.03.002

[10] GREENLAND, S. (2003). Quantifying biases in causal models: Classical confounding vs collider-stratification bias. *Epidemiology* **14** 300–306.

MSC2020 subject classifications. Primary 62B05; secondary 62H12, 62H20.

Key words and phrases. Causal inference, dual inverse regression, efficient dimension reduction, exhaustive estimation, missing data analysis.

[11] Greenland, S., Pearl, J. and Robins, J. M. (2007). Causal diagrams for epidemiologic research. *Epidemiology* **10** 37–48.

[12] Hall, P. and Li, K.-C. (1993). On almost linearity of low-dimensional projections from high-dimensional data. *Ann. Statist.* **21** 867–889. MR1232523 https://doi.org/10.1214/aos/1176349155

[13] LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *Amer. Econ. Rev.* 604–620.

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

[15] Li, B. and Dong, Y. (2009). Dimension reduction for nonelliptically distributed predictors. *Ann. Statist.* **37** 1272–1298. MR2509074 https://doi.org/10.1214/08-AOS598

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

[17] Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.* **86** 316–342. With discussion and a rejoinder by the author. MR1137117

[18] Lin, Q., Zhao, Z. and Liu, J. S. (2019). Sparse sliced inverse regression via Lasso. *J. Amer. Statist. Assoc.* **114** 1726–1739. MR4047295 https://doi.org/10.1080/01621459.2018.1520115

[19] Luo, W. (2022). On efficient dimension reduction with respect to the interaction between two response variables. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 269–294. MR4412987

[20] Luo, W. (2025). Supplement to "Some theory about efficient dimension reduction regarding the interaction between two responses." https://doi.org/10.1214/24-AOS2463SUPP

[21] Luo, W. and Li, B. (2021). On order determination by predictor augmentation. *Biometrika* **108** 557–574. MR4298764 https://doi.org/10.1093/biomet/asaa077

[22] Luo, W., Xue, L., Yao, J. and Yu, X. (2022). Inverse moment methods for sufficient forecasting using high-dimensional predictors. *Biometrika* **109** 473–487. MR4430969 https://doi.org/10.1093/biomet/asab037

[23] Luo, W. and Zhu, Y. (2020). Matching using sufficient dimension reduction for causal inference. *J. Bus. Econom. Statist.* **38** 888–900. MR4154895 https://doi.org/10.1080/07350015.2019.1609974

[24] Luo, W., Zhu, Y. and Ghosh, D. (2017). On estimating regression-based causal effects using sufficient dimension reduction. *Biometrika* **104** 51–65. MR3626479 https://doi.org/10.1093/biomet/asw068

[25] Ma, Y. and Zhu, L. (2012). A semiparametric approach to dimension reduction. *J. Amer. Statist. Assoc.* **107** 168–179. MR2949349 https://doi.org/10.1080/01621459.2011.646925

[26] Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math. Model.* **7** 1393–1512. MR0877758 https://doi.org/10.1016/0270-0255(86)90088-6

[27] Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55. MR0742974 https://doi.org/10.1093/biomet/70.1.41

[28] 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

[29] Shin, S. J., Wu, Y., Zhang, H. H. and Liu, Y. (2017). Principal weighted support vector machines for sufficient dimension reduction in binary classification. *Biometrika* **104** 67–81. MR3626475 https://doi.org/10.1093/biomet/asw057

[30] Wang, Q., Zhang, T. and Härdle, W. K. (2016). An extended single-index model with missing response at random. *Scand. J. Stat.* **43** 1140–1152. MR3573680 https://doi.org/10.1111/sjos.12233

[31] Xia, Y., Tong, H., Li, W. K. and Zhu, L.-X. (2002). An adaptive estimation of dimension reduction space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 363–410. MR1924297 https://doi.org/10.1111/1467-9868.03411

[32] Zeng, J., Mai, Q. and Zhang, X. (2024). Subspace estimation with automatic dimension and variable selection in sufficient dimension reduction. *J. Amer. Statist. Assoc.* **119** 343–355. MR4713897 https://doi.org/10.1080/01621459.2022.2118601

The Annals of Statistics
2025, Vol. 53, No. 1, 268–294
https://doi.org/10.1214/24-AOS2466

GENERAL SPATIO-TEMPORAL FACTOR MODELS FOR HIGH-DIMENSIONAL RANDOM FIELDS ON A LATTICE

BY MATTEO BARIGOZZI[1,2,a], DAVIDE LA VECCHIA[3,b] AND HANG LIU[4,c]

[1]*Department of Economics, University of Bologna*
[2]*CREST, ENSAE, Paris,* [a]*matteo.barigozzi@unibo.it*
[3]*Geneva School of Economics and Management, University of Geneva,* [b]*Davide.LaVecchia@unige.ch*
[4]*Faculty of Business in SciTech, School of Management, University of Science and Technology of China,* [c]*hliu01@ustc.edu.cn*

Motivated by the need for analysing large spatio-temporal panel data, we introduce a novel nonparametric methodology for n-dimensional random fields observed across S spatial locations and T time periods. We call it general spatio-temporal factor model (GSTFM). First, we provide the probabilistic and mathematical underpinning needed for the representation of a random field as the sum of two components: the common component (driven by a small number q of latent factors) and the idiosyncratic component (mildly cross-correlated). We show that the two components are identified as $n \to \infty$. Second, we propose an estimator of the common component and derive its statistical guarantees (consistency and rate of convergence) as $\min(n, S, T) \to \infty$. Third, we propose an information criterion to determine the number of factors. Estimation makes use of Fourier analysis in the frequency domain and thus it fully exploits the information on the spatio-temporal covariance structure of the whole panel. Synthetic data examples illustrate the applicability of GSTFM and its advantages over the extant generalized dynamic factor model that ignores the spatial correlations.

REFERENCES

ALTISSIMO, F., CRISTADORO, R., FORNI, M., LIPPI, M. and VERONESE, G. (2010). New eurocoin: Tracking economic growth in real time. *Rev. Econ. Stat.* **92** 1024–1034.

ANDERSON, B. D. O. and DEISTLER, M. (2008). Properties of zero-free transfer function matrices. *SICE J. Control Meas. Syst. Integr.* **1** 284–292.

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

BALTAGI, B. (2008). *Econometric Analysis of Panel Data*. Wiley, New York.

BARIGOZZI, M., CHO, H. and OWENS, D. (2023). FNETS: Factor-adjusted network estimation and forecasting for high-dimensional time series. *J. Bus. Econom. Statist.* **42** 890–902. MR4757113 https://doi.org/10.1080/07350015.2023.2257270

BARIGOZZI, M. and FARNÈ, M. (2022). An algebraic estimator for large spectral density matrices. *J. Amer. Statist. Assoc.* **119** 498–510. MR4713909 https://doi.org/10.1080/01621459.2022.2126780

BARIGOZZI, M., LA VECCHIA, D. and LIU, H. (2025). Supplement to "General spatio-temporal factor models for high-dimensional random fields on a lattice." https://doi.org/10.1214/24-AOS2466SUPPA, https://doi.org/10.1214/24-AOS2466SUPPB

BODELET, J. and LA VECCHIA, D. (2022). Robust sieve M-estimation with an application to dimensionality reduction. *Electron. J. Stat.* **16** 3996–4030. MR4456782 https://doi.org/10.1214/22-ejs2038

BRILLINGER, D. R. (1970). The frequency analysis of relations between stationary spatial series. In *Proc. Twelfth Biennial Sem. Canad. Math. Congr. on Time Series and Stochastic Processes*; *Convexity and Combinatorics* (*Vancouver, B.C.*, 1969) (R. Pyke, ed.) 39–81. Canad. Math. Congr., Montreal, QC. MR0273763

BROCKWELL, P. J. and DAVIS, R. A. (2006). *Time Series*: *Theory and Methods*. *Springer Series in Statistics*. Springer, New York. Reprint of the second (1991) edition. MR2839251

MSC2020 subject classifications. 62H25, 62M15, 62M40, 62H20, 60G35.
Key words and phrases. Factor models, high-dimensional random fields, spectral analysis, functional/phsyical dependence.

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

CHRISTENSEN, W. F. and AMEMIYA, Y. (2002). Latent variable analysis of multivariate spatial data. *J. Amer. Statist. Assoc.* **97** 302–317. MR1947288 https://doi.org/10.1198/016214502753479437

CRESSIE, N. and WIKLE, C. K. (2011). *Statistics for Spatio-Temporal Data. Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2848400

CRISTADORO, R., FORNI, M., REICHLIN, L. and VERONESE, G. (2005). A core inflation indicator for the euro area. *J. Money Credit Bank.* **37** 539–560.

DAHLHAUS, R. and KÜNSCH, H. (1987). Edge effects and efficient parameter estimation for stationary random fields. *Biometrika* **74** 877–882. MR0919857 https://doi.org/10.1093/biomet/74.4.877

DEB, S., POURAHMADI, M. and WU, W. B. (2017). An asymptotic theory for spectral analysis of random fields. *Electron. J. Stat.* **11** 4297–4322. MR3724221 https://doi.org/10.1214/17-EJS1326

EL MACHKOURI, M., VOLNÝ, D. and WU, W. B. (2013). A central limit theorem for stationary random fields. *Stochastic Process. Appl.* **123** 1–14. MR2988107 https://doi.org/10.1016/j.spa.2012.08.014

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. With 33 discussions by 57 authors and a reply by Fan, Liao and Mincheva. MR3091653 https://doi.org/10.1111/rssb.12016

FORNI, M., GIANNONE, D., LIPPI, M. and REICHLIN, L. (2009). Opening the black box: Structural factor models with large cross sections. *Econometric Theory* **25** 1319–1347. MR2540502 https://doi.org/10.1017/S026646660809052X

FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Rev. Econ. Stat.* **82** 540–554.

FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *J. Amer. Statist. Assoc.* **100** 830–840. MR2201012 https://doi.org/10.1198/016214504000002050

FORNI, M., HALLIN, M., LIPPI, M. and ZAFFARONI, P. (2015). Dynamic factor models with infinite-dimensional factor spaces: One-sided representations. *J. Econometrics* **185** 359–371. MR3311826 https://doi.org/10.1016/j.jeconom.2013.10.017

FORNI, M., HALLIN, M., LIPPI, M. and ZAFFARONI, P. (2017). Dynamic factor models with infinite-dimensional factor space: Asymptotic analysis. *J. Econometrics* **199** 74–92. MR3652406 https://doi.org/10.1016/j.jeconom.2017.04.002

FORNI, M. and LIPPI, M. (2001). The generalized dynamic factor model: Representation theory. *Econometric Theory* **17** 1113–1141. MR1867540 https://doi.org/10.1017/S0266466601176048

GERSING, P., RUST, C. and DEISTLER, M. (2024). Reconciling the theory of factor sequences. Available at arXiv:2307.10067.

GEWEKE, J. (1977). The dynamic factor analysis of economic time series. In *Latent Variables in Socio-Economic Models*, *Vol.* 1 (D. J. Aigner and A. S. Goldberger, eds.) 365–383. North-Holland, Amsterdam.

HALLIN, M., HÖRMANN, S. and LIPPI, M. (2018). Optimal dimension reduction for high-dimensional and functional time series. *Stat. Inference Stoch. Process.* **21** 385–398. MR3824974 https://doi.org/10.1007/s11203-018-9172-1

HALLIN, M. and LIPPI, M. (2013). Factor models in high-dimensional time series—a time-domain approach. *Stochastic Process. Appl.* **123** 2678–2695. MR3054541 https://doi.org/10.1016/j.spa.2013.04.001

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

HALLIN, M. and LIŠKA, R. (2011). Dynamic factors in the presence of blocks. *J. Econometrics* **163** 29–41. MR2803664 https://doi.org/10.1016/j.jeconom.2010.11.004

HALLIN, M. and TRUCÍOS, C. (2023). Forecasting value-at-risk and expected shortfall in large portfolios: A general dynamic factor model approach. *Econom. Stat.* **27** 1–15. MR4608335 https://doi.org/10.1016/j.ecosta.2021.04.006

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

LAZAR, N. (2008). *The Statistical Analysis of Functional MRI Data*. Springer, Berlin.

LEONENKO, N. (1999). *Limit Theorems for Random Fields with Singular Spectrum. Mathematics and Its Applications* **465**. Kluwer Academic, Dordrecht. MR1687092 https://doi.org/10.1007/978-94-011-4607-4

MANDREKAR, V. S. and REDETT, D. A. (2018). *Weakly Stationary Random Fields, Invariant Subspaces and Applications*. CRC Press, Boca Raton, FL. MR3751383

MOTTA, G., WU, W. B. and POURAHMADI, M. (2023). $\sqrt{2}$-estimation for smooth eigenvectors of matrix-valued functions. *Biometrika* **110** 1077–1098. MR4667441 https://doi.org/10.1093/biomet/asad018

PARK, B. U., MAMMEN, E., HÄRDLE, W. and BORAK, S. (2009). Time series modelling with semiparametric factor dynamics. *J. Amer. Statist. Assoc.* **104** 284–298. MR2504378 https://doi.org/10.1198/jasa.2009.0105

ROYDEN, H. L. and FITZPATRICK, P. (1988). *Real Analysis*, 3rd ed. Macmillan Co., New York. MR1013117

ROZANOV, Y. A. (1967). *Stationary Random Processes*. Holden-Day, San Francisco. Translated from the Russian by A. Feinstein. MR0214134

RUDIN, W. (1987). *Real and Complex Analysis*, 3rd ed. McGraw-Hill, New York. MR0924157

SARGENT, T. J. and SIMS, C. A. (1977). Business cycle modeling without pretending to have too much a priori economic theory. *New Method. Bus. Cycle Res.* **1** 145–168.

STEIN, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging. Springer Series in Statistics*. Springer, New York. MR1697409 https://doi.org/10.1007/978-1-4612-1494-6

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

TRUCÍOS, C., MAZZEU, J. H. G., HALLIN, M., HOTTA, L. K., PEREIRA, P. L. V. and ZEVALLOS, M. (2023). Forecasting conditional covariance matrices in high-dimensional time series: A general dynamic factor approach. *J. Bus. Econom. Statist.* **41** 40–52. MR4522151 https://doi.org/10.1080/07350015.2021.1996380

VANMARCKE, E. (2010). *Random Fields: Analysis and Synthesis*. World Scientific, Hackensack, NJ. MR2676223 https://doi.org/10.1142/5807

WANG, F. and WALL, M. M. (2003). Generalized common spatial factor model. *Biostatistics* **4** 569–582. https://doi.org/10.1093/biostatistics/4.4.569

WHITTLE, P. (1954). On stationary processes in the plane. *Biometrika* **41** 434–449. MR0067450 https://doi.org/10.1093/biomet/41.3-4.434

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. and ZAFFARONI, P. (2018). Asymptotic theory for spectral density estimates of general multivariate time series. *Econometric Theory* **34** 1–22. MR3739234 https://doi.org/10.1017/S0266466617000068

ZHANG, D. and WU, W. B. (2021). Convergence of covariance and spectral density estimates for high-dimensional locally stationary processes. *Ann. Statist.* **49** 233–254. MR4206676 https://doi.org/10.1214/20-AOS1954

The Annals of Statistics
2025, Vol. 53, No. 1, 295–321
https://doi.org/10.1214/24-AOS2467

OPTIMAL HETEROSKEDASTICITY TESTING IN NONPARAMETRIC REGRESSION

BY SUBHODH KOTEKAL[a] AND SOUMYABRATA KUNDU[b]

Department of Statistics, University of Chicago, [a]*skotekal@uchicago.edu,* [b]*soumyabratakundu@uchicago.edu*

Heteroskedasticity testing in nonparametric regression is a classic statistical problem with important practical applications, yet fundamental limits are unknown. Adopting a minimax perspective, this article considers the testing problem in the context of an α-Hölder mean and a β-Hölder variance function. For $\alpha > 0$ and $\beta \in (0, 1/2)$, the sharp minimax separation rate $n^{-4\alpha} + n^{-4\beta/(4\beta+1)} + n^{-2\beta}$ is established. To achieve the minimax separation rate, a kernel-based statistic using first-order squared differences is developed. Notably, the statistic estimates a proxy rather than a natural quadratic functional (the squared distance between the variance function and its best L^2 approximation by a constant) suggested in previous work.

The setting where no smoothness is assumed on the variance function is also studied; the variance profile across the design points can be arbitrary. Despite the lack of structure, consistent testing turns out to still be possible by using the Gaussian character of the noise, and the minimax rate is shown to be $n^{-4\alpha} + n^{-1/2}$. Exploiting noise information happens to be a fundamental necessity, as consistent testing is impossible if nothing more than zero mean and unit variance is known about the noise distribution. Furthermore, in the setting where the variance function is β-Hölder but heteroskedasticity is measured only with respect to the design points, the minimax separation rate is shown to be $n^{-4\alpha} + n^{-((1/2)\vee(4\beta/(4\beta+1)))}$ when the noise is Gaussian and $n^{-4\alpha} + n^{-4\beta/(4\beta+1)} + n^{-2\beta}$ when the noise distribution is unknown.

REFERENCES

[1] 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

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

[3] BICKEL, P. J. (1978). Using residuals robustly. I. Tests for heteroscedasticity, nonlinearity. *Ann. Statist.* **6** 266–291. MR0474625

[4] BREUSCH, T. S. and PAGAN, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica* **47** 1287–1294. MR0545960 https://doi.org/10.2307/1911963

[5] BROWN, M. B. and FORSYTHE, A. B. (1974). Robust tests for the equality of variances. *J. Amer. Statist. Assoc.* **69** 364–367. https://doi.org/10.1080/01621459.1974.10482955

[6] CAI, T. T. and LOW, M. G. (2006). Optimal adaptive estimation of a quadratic functional. *Ann. Statist.* **34** 2298–2325. MR2291501 https://doi.org/10.1214/009053606000000849

[7] CAI, T. T. and YUAN, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *Ann. Statist.* **39** 2330–2355. MR2906870 https://doi.org/10.1214/11-AOS898

[8] CHHOR, J., MUKHERJEE, R. and SEN, S. (2024). Sparse signal detection in heteroscedastic Gaussian sequence models: Sharp minimax rates. *Bernoulli* **30** 2127–2153. MR4746602 https://doi.org/10.3150/23-bej1667

[9] CHOWN, J. and MÜLLER, U. U. (2018). Detecting heteroscedasticity in non-parametric regression using weighted empirical processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 951–974. MR3874305 https://doi.org/10.1111/rssb.12282

MSC2020 subject classifications. 62C20, 62G08, 62G10.
Key words and phrases. Heteroskedasticity, nonparametric regression, minimax testing.

[10] COMMINGES, L. and DALALYAN, A. S. (2013). Minimax testing of a composite null hypothesis defined via a quadratic functional in the model of regression. *Electron. J. Stat.* **7** 146–190. MR3020417 https://doi.org/10.1214/13-EJS766

[11] COOK, R. D. and WEISBERG, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika* **70** 1–10. MR0742970 https://doi.org/10.1093/biomet/70.1.1

[12] DETTE, H. (2002). A consistent test for heteroscedasticity in nonparametric regression based on the kernel method. *J. Statist. Plann. Inference* **103** 311–329. MR1896998 https://doi.org/10.1016/S0378-3758(01)00229-4

[13] DETTE, H. and HETZLER, B. (2009). A simple test for the parametric form of the variance function in nonparametric regression. *Ann. Inst. Statist. Math.* **61** 861–886. MR2556768 https://doi.org/10.1007/s10463-008-0169-1

[14] DETTE, H. and MUNK, A. (1998). Testing heteroscedasticity in nonparametric regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 693–708. MR1649535 https://doi.org/10.1111/1467-9868.00149

[15] DETTE, H., NEUMEYER, N. and VAN KEILEGOM, I. (2007). A new test for the parametric form of the variance function in non-parametric regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 903–917. MR2368576 https://doi.org/10.1111/j.1467-9868.2007.00616.x

[16] DEVORE, R. A. and LORENTZ, G. G. (1993). *Constructive Approximation. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **303**. Springer, Berlin. MR1261635

[17] EFROMOVICH, S. and LOW, M. (1996). On optimal adaptive estimation of a quadratic functional. *Ann. Statist.* **24** 1106–1125. MR1401840 https://doi.org/10.1214/aos/1032526959

[18] ERMAKOV, M. S. (2004). Minimax detection of a signal in weighted Gaussian white noise. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov.* (*POMI*) **320** 54–68, 226. MR2115865 https://doi.org/10.1007/s10958-006-0244-1

[19] EUBANK, R. L. and THOMAS, W. (1993). Detecting heteroscedasticity in nonparametric regression. *J. Roy. Statist. Soc. Ser. B* **55** 145–155. MR1210427

[20] GASSER, T., SROKA, L. and JENNEN-STEINMETZ, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika* **73** 625–633. MR0897854 https://doi.org/10.1093/biomet/73.3.625

[21] GASTWIRTH, J. L., GEL, Y. R. and MIAO, W. (2009). The impact of Levene's test of equality of variances on statistical theory and practice. *Statist. Sci.* **24** 343–360. MR2757435 https://doi.org/10.1214/09-STS301

[22] GINÉ, E. and NICKL, R. (2008). A simple adaptive estimator of the integrated square of a density. *Bernoulli* **14** 47–61. MR2401653 https://doi.org/10.3150/07-BEJ110

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

[24] GOLDFELD, S. M. and QUANDT, R. E. (1965). Some tests for homoscedasticity. *J. Amer. Statist. Assoc.* **60** 539–547. MR0184358

[25] HALL, P., KAY, J. W. and TITTERINGTON, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* **77** 521–528. MR1087842 https://doi.org/10.1093/biomet/77.3.521

[26] HARRISON, M. J. and MCCABE, B. P. M. (1979). A test for heteroscedasticity based on ordinary least squares residuals. *J. Amer. Statist. Assoc.* **74** 494–499. MR0548049

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

[28] KOENKER, R. and BASSETT, G. JR. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* **50** 43–61. MR0640165 https://doi.org/10.2307/1912528

[29] KOTEKAL, S. and KUNDU, S. (2025). Supplement to "Optimal heteroskedasticity testing in nonparametric regression." https://doi.org/10.1214/24-AOS2467SUPP

[30] LAURENT, B., LOUBES, J.-M. and MARTEAU, C. (2012). Non asymptotic minimax rates of testing in signal detection with heterogeneous variances. *Electron. J. Stat.* **6** 91–122. MR2879673 https://doi.org/10.1214/12-EJS667

[31] LEPSKIĬ, O. V. (1990). A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatn. Primen.* **35** 459–470. MR1091202 https://doi.org/10.1137/1135065

[32] LEVENE, H. (1960). Robust tests for equality of variances. In *Contributions to Probability and Statistics. Stanford Studies in Mathematics and Statistics* **2** 278–292. Stanford Univ. Press, Stanford, CA. MR0120709

[33] SHEN, Y., GAO, C., WITTEN, D. and HAN, F. (2020). Optimal estimation of variance in nonparametric regression with random design. *Ann. Statist.* **48** 3589–3618. MR4185821 https://doi.org/10.1214/20-AOS1944

[34] SPOKOINY, V. G. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.* **24** 2477–2498. MR1425962 https://doi.org/10.1214/aos/1032181163

[35] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. Revised and extended from the 2004 French original. Translated by Vladimir Zaiats. MR2724359 https://doi.org/10.1007/b13794

[36] WANG, L., BROWN, L. D., CAI, T. T. and LEVINE, M. (2008). Effect of mean on variance function estimation in nonparametric regression. *Ann. Statist.* **36** 646–664. MR2396810 https://doi.org/10.1214/009053607000000901

[37] WHITE, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48** 817–838. MR0575027 https://doi.org/10.2307/1912934

[38] WU, Y. and YANG, P. (2020). Polynomial methods in statistical inference: Theory and practice. *Found. Trends Commun. Inf. Theory* **17** 402–586. https://doi.org/10.1561/0100000095

[39] ZHENG, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics* **75** 263–289. MR1413644 https://doi.org/10.1016/0304-4076(95)01760-7

[40] ZHENG, X. (2009). Testing heteroscedasticity in nonlinear and nonparametric regressions. *Canad. J. Statist.* **37** 282–300. MR2531832 https://doi.org/10.1002/cjs.10020

The Annals of Statistics
2025, Vol. 53, No. 1, 322–351
https://doi.org/10.1214/24-AOS2468

A STATISTICAL FRAMEWORK OF WATERMARKS FOR LARGE LANGUAGE MODELS: PIVOT, DETECTION EFFICIENCY AND OPTIMAL RULES

BY XIANG LI[1,a], FENG RUAN[2,e], HUIYUAN WANG[1,b], QI LONG[1,c] AND WEIJIE J. SU[1,d]

[1]*University of Pennsylvania,* [a]*lx10077@upenn.edu,* [b]*huiyuanw@upenn.edu,* [c]*qlong@upenn.edu,* [d]*suw@wharton.upenn.edu*
[2]*Northwestern University,* [e]*fengruan@northwestern.edu*

Since ChatGPT was introduced in November 2022, embedding (nearly) unnoticeable statistical signals into text generated by large language models (LLMs), also known as watermarking, has been used as a principled approach to provable detection of LLM-generated text from its human-written counterpart. In this paper, we introduce a general and flexible framework for reasoning about the statistical efficiency of watermarks and designing powerful detection rules. Inspired by the hypothesis testing formulation of watermark detection, our framework starts by selecting a pivotal statistic of the text and a secret key—provided by the LLM to the verifier—to control the false positive rate (the error of mistakenly detecting human-written text as LLM-generated). Next, this framework allows one to evaluate the power of watermark detection rules by obtaining a closed-form expression of the asymptotic false negative rate (the error of incorrectly classifying LLM-generated text as human-written). Our framework further reduces the problem of determining the optimal detection rule to solving a minimax optimization program. We apply this framework to two representative watermarks—one of which has been internally implemented at OpenAI—and obtain several findings that can be instrumental in guiding the practice of implementing watermarks. In particular, we derive optimal detection rules for these watermarks under our framework. These theoretically derived detection rules are demonstrated to be competitive and sometimes enjoy a higher power than existing detection approaches through numerical experiments.

REFERENCES

[1] AARONSON, S. (2023). Watermarking of large language models. https://simons.berkeley.edu/talks/scott-aaronson-ut-austin-openai-2023-08-17.

[2] ACHIAM, J., ADLER, S., AGARWAL, S., AHMAD, L., AKKAYA, I., ALEMAN, F. L., ALMEIDA, D., ALTENSCHMIDT, J., ALTMAN, S. et al. (2023). GPT-4 technical report. arXiv preprint arXiv:2303.08774.

[3] ACKLEY, D. H., HINTON, G. E. and SEJNOWSKI, T. J. (1985). A learning algorithm for Boltzmann machines. *Cogn. Sci.* **9** 147–169.

[4] ATALLAH, M. J., RASKIN, V., CROGAN, M., HEMPELMANN, C., KERSCHBAUM, F., MOHAMED, D. and NAIK, S. (2001). Natural language watermarking: Design, analysis, and a proof-of-concept implementation. In *Information Hiding* 185–200.

[5] BAHADUR, R. R. (1960). On the asymptotic efficiency of tests and estimates. *Sankhyā* **22** 229–252. MR0293767

[6] BAO, G., ZHAO, Y., TENG, Z., YANG, L. and ZHANG, Y. (2023). Fast-Detectgpt: Efficient zero-shot detection of machine-generated text via conditional probability curvature. arXiv preprint arXiv:2310.05130.

[7] BARAK, B. (2021). An intensive introduction to cryptography, lectures notes for Harvard CS 127. https://intensecrypto.org/public/index.html.

[8] BARTZ, D. and HU, K. (2023). OpenAI, Google, others pledge to watermark AI content for safety, White House says. https://www.reuters.com/technology/openai-google-others-pledge-watermark-ai-content-safety-white-house-2023-07-21/. Accessed: 2023-10-03.

MSC2020 subject classifications. Primary 62C05; secondary 62F03.
Key words and phrases. Large language model, watermark, pivot, detection efficiency, large deviation theory.

[9] BIDEN, J. (2023). Fact sheet: President Biden issues executive order on safe, secure, and trustworthy artificial intelligence. https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/. Accessed: 2023-11-01.

[10] BROWN, T., MANN, B., RYDER, N., SUBBIAH, M., KAPLAN, J. D., DHARIWAL, P., NEELAKANTAN, A., SHYAM, P., SASTRY, G. et al. (2020). Language models are few-shot learners. In *Advances in Neural Information Processing Systems* **33** 1877–1901.

[11] CAI, Z., LIU, S., WANG, H., ZHONG, H. and LI, X. (2024). Towards better statistical understanding of watermarking LLMs. arXiv preprint arXiv:2403.13027.

[12] CAYRE, F., FONTAINE, C. and FURON, T. (2005). Watermarking security: Theory and practice. *IEEE Trans. Signal Process.* **53** 3976–3987. MR2239857 https://doi.org/10.1109/TSP.2005.855418

[13] CHAKRABORTY, S., BEDI, A. S., ZHU, S., AN, B., MANOCHA, D. and HUANG, F. (2023). On the possibilities of AI-generated text detection. arXiv preprint arXiv:2304.04736.

[14] CHRIST, M., GUNN, S. and ZAMIR, O. (2024). Undetectable watermarks for language models. In *Conference on Learning Theory* 1125–1139. PMLR.

[15] COX, I., MILLER, M., BLOOM, J., FRIDRICH, J. and KALKER, T. (2007). *Digital Watermarking and Steganography*. Morgan Kaufmann.

[16] DAS, D., DE LANGIS, K., MARTIN, A., KIM, J., LEE, M., KIM, Z. M., HAYATI, S., OWAN, R., HU, B. et al. (2024). Under the surface: Tracking the artifactuality of LLM-generated data. arXiv preprint arXiv:2401.14698.

[17] DEMBO, A. and ZEITOUNI, O. (2010). *Large Deviations Techniques and Applications*. *Stochastic Modelling and Applied Probability* **38**. Springer, Berlin. Corrected reprint of the second (1998) edition. MR2571413 https://doi.org/10.1007/978-3-642-03311-7

[18] DEVROYE, L. (2006). Nonuniform random variate generation. *Handbooks Oper. Res. Management Sci.* **13** 83–121.

[19] DONSKER, M. D. and VARADHAN, S. R. S. (1983). Asymptotic evaluation of certain Markov process expectations for large time. IV. *Comm. Pure Appl. Math.* **36** 183–212. MR0690656 https://doi.org/10.1002/cpa.3160360204

[20] FERNANDEZ, P., CHAFFIN, A., TIT, K., CHAPPELIER, V. and FURON, T. (2023). Three bricks to consolidate watermarks for large language models. arXiv preprint arXiv:2308.00113.

[21] GEHRMANN, S., STROBELT, H. and RUSH, A. (2019). GLTR: Statistical detection and visualization of generated text. In *Annual Meeting of the Association for Computational Linguistics: System Demonstrations*.

[22] GIBOULOT, E. and TEDDY, F. (2024). WaterMax: Breaking the LLM watermark detectability-robustness-quality trade-off. arXiv preprint arXiv:2403.04808.

[23] GPTZERO (2023). GPTZero: More than an AI detector preserve what's human. https://gptzero.me/.

[24] GUMBEL, E. J. (1954). *Statistical Theory of Extreme Values and Some Practical Applications. A Series of Lectures*. *National Bureau of Standards Applied Mathematics Series* **33**. U. S. Government Printing Office, Washington, DC. MR0061342

[25] HAZAN, T. and JAAKKOLA, T. (2012). On the partition function and random maximum a-posteriori perturbations. In *International Conference on Machine Learning* 1667–1674.

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

[27] HU, Z., CHEN, L., WU, X., WU, Y., ZHANG, H. and HUANG, H. (2024). Unbiased watermark for large language models. In *International Conference on Learning Representations*.

[28] HUANG, B., ZHU, H., ZHU, B., RAMCHANDRAN, K., JORDAN, M. I., LEE, J. D. and JIAO, J. (2023). Towards optimal statistical watermarking. arXiv preprint arXiv:2312.07930.

[29] JANG, E., GU, S. and POOLE, B. (2016). Categorical reparameterization with Gumbel-softmax. In *International Conference on Learning Representations*.

[30] KATZ, J. and LINDELL, Y. (2008). *Introduction to Modern Cryptography*. *Chapman & Hall/CRC Cryptography and Network Security*. CRC Press/CRC, Boca Raton, FL. MR2371431

[31] KIRCHENBAUER, J., GEIPING, J., WEN, Y., KATZ, J., MIERS, I. and GOLDSTEIN, T. (2023). A watermark for large language models. In *International Conference on Machine Learning* **202** 17061–17084.

[32] KIRCHENBAUER, J., GEIPING, J., WEN, Y., SHU, M., SAIFULLAH, K., KONG, K., FERNANDO, K., SAHA, A., GOLDBLUM, M. et al. (2024). On the reliability of watermarks for large language models. In *International Conference on Learning Representations*.

[33] KRISHNA, K., SONG, Y., KARPINSKA, M., WIETING, J. and IYYER, M. (2024). Paraphrasing evades detectors of AI-generated text, but retrieval is an effective defense. In *Advances in Neural Information Processing Systems* **36**.

[34] KUDITIPUDI, R., THICKSTUN, J., HASHIMOTO, T. and LIANG, P. (2024). Robust distortion-free watermarks for language models. *Trans. Mach. Learn. Res.*.
[35] LI, X., RUAN, F., WANG, H., LONG, Q. and SU, W. J. (2024). Robust detection of watermarks for large language models under human edits. arXiv preprint arXiv:2411.13868.
[36] LI, X., RUAN, F., WANG, H., LONG, Q. and SU, W. J. (2025). Supplement to "A statistical framework of watermarks for large language models: Pivot, detection efficiency and optimal rules." https://doi.org/10.1214/24-AOS2468SUPPA
[37] LI, X., RUAN, F., WANG, H., LONG, Q. and SU, W. J. (2025). Source codes to "A statistical framework of watermarks for large language models: Pivot, detection efficiency and optimal rules." https://doi.org/10.1214/24-AOS2468SUPPB
[38] LIANG, W., YUKSEKGONUL, M., MAO, Y., WU, E. and ZOU, J. (2023). GPT detectors are biased against non-native English writers. In *ICLR* 2023 *Workshop on Trustworthy and Reliable Large-Scale Machine Learning Models*.
[39] LIN, T., JIN, C. and JORDAN, M. I. (2020). Near-optimal algorithms for minimax optimization. In *Conference on Learning Theory* 2738–2779. PMLR.
[40] LIU, Y. and BU, Y. (2024). Adaptive text watermark for large language models. In *International Conference on Machine Learning*.
[41] MADDISON, C. J., TARLOW, D. and MINKA, T. (2014). A* sampling. In *Advances in Neural Information Processing Systems* **27**.
[42] MILANO, S., MCGRANE, J. A. and LEONELLI, S. (2023). Large language models challenge the future of higher education. *Nat. Mach. Intell.* **5** 333–334.
[43] MITCHELL, E., LEE, Y., KHAZATSKY, A., MANNING, C. D. and FINN, C. (2023). Detectgpt: Zero-shot machine-generated text detection using probability curvature. arXiv preprint arXiv:2301.11305.
[44] OPENAI (2023). ChatGPT: Optimizing language models for dialogue.
[45] PAAR, C. and PELZL, J. (2009). *Understanding Cryptography: A Textbook for Students and Practitioners*. Springer.
[46] PAPANDREOU, G. and YUILLE, A. L. (2011). Perturb-and-map random fields: Using discrete optimization to learn and sample from energy models. In 2011 *International Conference on Computer Vision* 193–200. IEEE.
[47] PIET, J., SITAWARIN, C., FANG, V., MU, N. and WAGNER, D. (2023). Mark my words: Analyzing and evaluating language model watermarks. arXiv preprint arXiv:2312.00273.
[48] RADFORD, A., KIM, J. W., XU, T., BROCKMAN, G., MCLEAVEY, C. and SUTSKEVER, I. (2023). Robust speech recognition via large-scale weak supervision. In *International Conference on Machine Learning* 28492–28518. PMLR.
[49] RADFORD, A., WU, J., CHILD, R., LUAN, D., AMODEI, D., SUTSKEVER, I. et al. (2019). Language models are unsupervised multitask learners. *OpenAI Blog* **1** 9.
[50] RAFFEL, C., SHAZEER, N., ROBERTS, A., LEE, K., NARANG, S., MATENA, M., ZHOU, Y., LI, W. and LIU, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.* **21** 5485–5551.
[51] RAHIMIAN, H. and MEHROTRA, S. (2022). Frameworks and results in distributionally robust optimization. *Open J. Math. Optim.* **3** Art. No. 4, 85. MR4459982 https://doi.org/10.5802/ojmo.15
[52] SADASIVAN, V. S., KUMAR, A., BALASUBRAMANIAN, S., WANG, W. and FEIZI, S. (2023). Can AI-generated text be reliably detected? arXiv preprint arXiv:2303.11156.
[53] SCHNEIER, B. (1996). *Applied Cryptography*. Wiley.
[54] SHUMAILOV, I., SHUMAYLOV, Z., ZHAO, Y., GAL, Y., PAPERNOT, N. and ANDERSON, R. (2023). The curse of recursion: Training on generated data makes models forget. arXiv preprint arXiv:2305.17493.
[55] SOUFIANI, H. A., PARKES, D. C. and XIA, L. (2012). Random utility theory for social choice. In *Neural Information Processing Systems* 126–134.
[56] STARBIRD, K. (2019). Disinformation's spread: Bots, trolls and all of us. *Nature* **571** 449–450.
[57] STINSON, D. R. (2006). *Cryptography: Theory and Practice*, 3rd ed. *Discrete Mathematics and Its Applications* (*Boca Raton*). CRC Press/CRC, Boca Raton, FL. MR2182472
[58] STOKEL-WALKER, C. (2022). AI bot ChatGPT writes smart essays—Should professors worry? *Nat. News*.
[59] TOPKARA, U., TOPKARA, M. and ATALLAH, M. J. (2006). The hiding virtues of ambiguity: Quantifiably resilient watermarking of natural language text through synonym substitutions. In *Proceedings of the* 8*th Workshop on Multimedia and Security* 164–174.
[60] TOUVRON, H., LAVRIL, T., IZACARD, G., MARTINET, X., LACHAUX, M.-A., LACROIX, T., ROZIÈRE, B., GOYAL, N., HAMBRO, E. et al. (2023). LLaMA: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971.

[61] TULCHINSKII, E., KUZNETSOV, K., KUSHNAREVA, L., CHERNIAVSKII, D., NIKOLENKO, S., BURNAEV, E., BARANNIKOV, S. and PIONTKOVSKAYA, I. (2024). Intrinsic dimension estimation for robust detection of AI-generated texts. In *Advances in Neural Information Processing Systems* **36**.

[62] 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

[63] VASWANI, A., SHAZEER, N., PARMAR, N., USZKOREIT, J., JONES, L., GOMEZ, A. N., KAISER, Ł. and POLOSUKHIN, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems* **30**.

[64] WALKER, A. J. (1974). New fast method for generating discrete random numbers with arbitrary frequency distributions. *Electron*. *Lett*. **8** 127–128.

[65] WALKER, A. J. (1977). An efficient method for generating discrete random variables with general distributions. *ACM Trans*. *Math*. *Software* **3** 253–256.

[66] WEBER-WULFF, D., ANOHINA-NAUMECA, A., BJELOBABA, S., FOLTÝNEK, T., GUERRERO-DIB, J., POPOOLA, O., ŠIGUT, P. and WADDINGTON, L. (2023). Testing of detection tools for AI-generated text. *Int*. *J*. *Educ*. *Integr*. **19** 26.

[67] WEIDINGER, L., MELLOR, J., RAUH, M., GRIFFIN, C., UESATO, J., HUANG, P.-S., CHENG, M., GLAESE, M., BALLE, B. et al. (2021). Ethical and social risks of harm from language models. arXiv preprint arXiv:2112.04359.

[68] WU, Y., HU, Z., ZHANG, H. and HUANG, H. (2023). DiPmark: A stealthy, efficient and resilient watermark for large language models. arXiv preprint arXiv:2310.07710.

[69] XIA, M., GAO, T., ZENG, Z. and CHEN, D. (2023). Sheared LLaMA: Accelerating language model pre-training via structured pruning. In *International Conference on Learning Representations*.

[70] XIE, Y., LI, X., MALLICK, T., SU, W. J. and ZHANG, R. (2024). Debiasing watermarks for large language models via maximal coupling. arXiv preprint arXiv:2411.11203.

[71] YANG, X., CHENG, W., PETZOLD, L., WANG, W. Y. and CHEN, H. (2023). DNA-GPT: Divergent N-gram analysis for training-free detection of GPT-generated text. arXiv preprint arXiv:2305.17359.

[72] ZELLERS, R., HOLTZMAN, A., RASHKIN, H., BISK, Y., FARHADI, A., ROESNER, F. and CHOI, Y. (2019). Defending against neural fake news. In *Advances in Neural Information Processing Systems* **32**.

[73] ZEROGPT (2023). ZeroGPT: Trusted GPT-4, ChatGPT and AI detector tool by ZeroGPT. https://www.zerogpt.com/.

[74] ZHANG, H., EDELMAN, B. L., FRANCATI, D., VENTURI, D., ATENIESE, G. and BARAK, B. (2023). Watermarks in the sand: Impossibility of strong watermarking for generative models. arXiv preprint arXiv:2311.04378.

[75] ZHANG, S., ROLLER, S., GOYAL, N., ARTETXE, M., CHEN, M., CHEN, S., DEWAN, C., DIAB, M., LI, X. et al. (2022). OPT: Open pre-trained transformer language models. arXiv preprint arXiv:2205.01068.

[76] ZHAO, X., ANANTH, P. V., LI, L. and WANG, Y.-X. (2024). Provable robust watermarking for AI-generated text. In *International Conference on Learning Representations*.

[77] ZHAO, X., LI, L. and WANG, Y.-X. (2024). Permute-and-flip: An optimally robust and watermarkable decoder for LLMs. arXiv preprint arXiv:2402.05864.

[78] ZHOU, X., ZHAO, W., WANG, Z. and PAN, L. (2009). Security theory and attack analysis for text watermarking. In *International Conference on E-Business and Information System Security* 1–6.

[79] ZIPF, G. K. (2016). *Human Behavior and the Principle of Least Effort*: *An Introduction to Human Ecology*. Ravenio Books.

The Annals of Statistics
2025, Vol. 53, No. 1, 352–373
https://doi.org/10.1214/24-AOS2472

EMPIRICAL LIKELIHOOD BASED TESTING FOR MULTIVARIATE REGULAR VARIATION

BY JOHN H. J. EINMAHL[1,a], ANDREA KRAJINA[2,b] AND JUAN JUAN CAI[3,c]

[1]*Department of Econometrics and OR and CentER, Tilburg University,* [a]*j.h.j.einmahl@uvt.nl*

[2]*Institute for Mathematical Stochastics, University of Göttingen,* [b]*akrajina@gmail.com*

[3]*Department of Econometrics and Data Science, Vrije Universiteit Amsterdam,* [c]*j.cai@vu.nl*

Multivariate regular variation is a common assumption in the statistics literature and needs to be verified in real-data applications. We develop a novel hypothesis test for multivariate regular variation, employing localized empirical likelihood. We establish the weak convergence of the test statistic to a nonstandard, distribution-free limit and hence can provide universal critical values for the test. We show the very good finite-sample behavior of the procedure through simulations and apply the test to several real-data examples.

REFERENCES

ALEXANDER, K. S. (1987). Rates of growth and sample moduli for weighted empirical processes indexed by sets. *Probab. Theory Related Fields* **75** 379–423. MR0890285 https://doi.org/10.1007/BF00318708

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

BASRAK, B. and SEGERS, J. (2009). Regularly varying multivariate time series. *Stochastic Process. Appl.* **119** 1055–1080. MR2508565 https://doi.org/10.1016/j.spa.2008.05.004

BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1987). *Regular Variation. Encyclopedia of Mathematics and Its Applications* **27**. Cambridge Univ. Press, Cambridge. MR0898871 https://doi.org/10.1017/CBO9780511721434

CAI, J.-J., EINMAHL, J. H. J. and DE HAAN, L. (2011). Estimation of extreme risk regions under multivariate regular variation. *Ann. Statist.* **39** 1803–1826. MR2850221 https://doi.org/10.1214/11-AOS891

DAS, B. and FASEN-HARTMANN, V. (2018). Risk contagion under regular variation and asymptotic tail independence. *J. Multivariate Anal.* **165** 194–215. MR3768761 https://doi.org/10.1016/j.jmva.2017.12.004

DAVIS, R. A. and HSING, T. (1995). Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Ann. Probab.* **23** 879–917. MR1334176

DAVIS, R. A. and MIKOSCH, T. (1998). The sample autocorrelations of heavy-tailed processes with applications to ARCH. *Ann. Statist.* **26** 2049–2080. MR1673289 https://doi.org/10.1214/aos/1024691368

EINMAHL, J. H. J. (1997). Poisson and Gaussian approximation of weighted local empirical processes. *Stochastic Process. Appl.* **70** 31–58. MR1472958 https://doi.org/10.1016/S0304-4149(97)00055-0

EINMAHL, J. H. J. and MCKEAGUE, I. W. (2003). Empirical likelihood based hypothesis testing. *Bernoulli* **9** 267–290. MR1997030 https://doi.org/10.3150/bj/1068128978

EINMAHL, J. H. J., YANG, F. and ZHOU, C. (2021). Testing the multivariate regular variation model. *J. Bus. Econom. Statist.* **39** 907–919. MR4319681 https://doi.org/10.1080/07350015.2020.1737533

GIRARD, S. and STUPFLER, G. (2015). Extreme geometric quantiles in a multivariate regular variation framework. *Extremes* **18** 629–663. MR3418771 https://doi.org/10.1007/s10687-015-0226-0

HE, Y. and EINMAHL, J. H. J. (2017). Estimation of extreme depth-based quantile regions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 449–461. MR3611754 https://doi.org/10.1111/rssb.12163

JESSEN, A. H. and MIKOSCH, T. (2006). Regularly varying functions. *Publ. Inst. Math.* (*Beograd*) (*N.S.*) **80** 171–192. MR2281913 https://doi.org/10.2298/PIM0694171J

JOE, H. and LI, H. (2011). Tail risk of multivariate regular variation. *Methodol. Comput. Appl. Probab.* **13** 671–693. MR2851832 https://doi.org/10.1007/s11009-010-9183-x

MSC2020 subject classifications. Primary 62G10, 62G20, 62G30, 62G32, 62H15; secondary 60F05, 60G70.

Key words and phrases. Asymptotic theory, distribution-free, empirical likelihood, empirical process, multivariate tail, regular variation, tail index.

KLEY, O., KLÜPPELBERG, C. and REINERT, G. (2016). Risk in a large claims insurance market with bipartite graph structure. *Oper. Res.* **64** 1159–1176. MR3558442 https://doi.org/10.1287/opre.2016.1502

KLEY, O., KLÜPPELBERG, C. and REINERT, G. (2018). Conditional risk measures in a bipartite market structure. *Scand. Actuar. J.* **2018** 328–355. MR3782703 https://doi.org/10.1080/03461238.2017.1350203

KULIK, R. and SOULIER, P. (2015). Heavy tailed time series with extremal independence. *Extremes* **18** 273–299. MR3351817 https://doi.org/10.1007/s10687-014-0213-x

LI, H. and SUN, Y. (2009). Tail dependence for heavy-tailed scale mixtures of multivariate distributions. *J. Appl. Probab.* **46** 925–937. MR2582698 https://doi.org/10.1239/jap/1261670680

MAINIK, G. and EMBRECHTS, P. (2013). Diversification in heavy-tailed portfolios: Properties and pitfalls. *Ann. Actuar. Sci.* **7** 26–45.

MAINIK, G., MITOV, G. and RÜSCHENDORF, L. (2015). Portfolio optimization for heavy-tailed assets: Extreme risk index vs. Markowitz. *J. Empir. Finance* **32** 115–134.

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

MAINIK, G. and RÜSCHENDORF, L. (2012). Ordering of multivariate risk models with respect to extreme portfolio losses. *Stat. Risk Model.* **29** 73–105. MR2901804 https://doi.org/10.1524/strm.2012.1103

OWEN, A. (2001). *Empirical Likelihood*. Chapman & Hall/CRC, Boca Raton, FL.

RESNICK, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. *Applied Probability. A Series of the Applied Probability Trust* **4**. Springer, New York. MR0900810 https://doi.org/10.1007/978-0-387-75953-1

RVAČEVA, E. L. (1962). On domains of attraction of multi-dimensional distributions. In *Select. Transl. Math. Statist. and Probability, Vol.* 2 183–205. Amer. Math. Soc., Providence, RI. MR0150795

TORRES, R., DI BERNARDINO, E., LANIADO, H. and LILLO, R. E. (2020). On the estimation of extreme directional multivariate quantiles. *Comm. Statist. Theory Methods* **49** 5504–5534. MR4156014 https://doi.org/10.1080/03610926.2019.1619770

ZHU, L. and LI, H. (2012). Asymptotic analysis of multivariate tail conditional expectations. *N. Am. Actuar. J.* **16** 350–363. MR2972977 https://doi.org/10.1080/10920277.2012.10590646

The Annals of Statistics
2025, Vol. 53, No. 1, 374–399
https://doi.org/10.1214/24-AOS2473

COMPUTATIONALLY EFFICIENT AND STATISTICALLY OPTIMAL ROBUST HIGH-DIMENSIONAL LINEAR REGRESSION

BY YINAN SHEN[a], JINGYANG LI[b], JIAN-FENG CAI[c] AND DONG XIA[d]

Department of Mathematics, Hong Kong University of Science and Technology, [a]*yshenay@connect.ust.hk,* [b]*jlieb@connect.ust.hk,* [c]*jfcai@ust.hk,* [d]*madxia@ust.hk*

High-dimensional linear regression under heavy-tailed noise or outlier corruption is challenging, both computationally and statistically. Convex approaches have been proven statistically optimal but suffer from high computational costs, especially since the robust loss functions are usually nonsmooth. More recently, computationally fast nonconvex approaches via subgradient descent are proposed, which, unfortunately, fail to deliver a statistically consistent estimator even under sub-Gaussian noise. In this paper, we introduce a projected subgradient descent algorithm for both the sparse linear regression and low-rank linear regression problems. The algorithm is not only computationally efficient with linear convergence but also statistically optimal, be the noise Gaussian or heavy-tailed with a finite $1 + \varepsilon$ moment. The convergence theory is established for a general framework and its specific applications to absolute loss, Huber loss and quantile loss are investigated. Compared with existing nonconvex methods, ours reveals a surprising phenomenon of *two-phase convergence*. In phase one, the algorithm behaves as in typical nonsmooth optimization that requires gradually decaying stepsizes. However, phase one only delivers a statistically suboptimal estimator, which is already observed in the existing literature. Interestingly, during phase two, the algorithm converges linearly *as if* minimizing a smooth and strongly convex objective function, and thus a constant stepsize suffices. Underlying the phase-two convergence is the *smoothing effect* of random noise to the nonsmooth robust losses in an area *close but not too close* to the truth. Numerical simulations confirm our theoretical discovery and showcase the superiority of our algorithm over prior methods.

REFERENCES

ABE, A., YAMAMOTO, Y., IBA, S., KANIE, T., OKAMOTO, A., TOKUDA, M., INAGUMA, Y., YANADA, M., MORISHIMA, S. et al. (2016). ETV6-LPXN fusion transcript generated by t (11; 12)(q12. 1; p13) in a patient with relapsing acute myeloid leukemia with NUP98-HOXA9. *Genes Chromosomes Cancer* **55** 242–250.

ADAMCZAK, R. (2008). A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.* **13** 1000–1034. MR2424985 https://doi.org/10.1214/EJP.v13-521

AGARWAL, A., NEGAHBAN, S. and WAINWRIGHT, M. J. (2010). Fast global convergence rates of gradient methods for high-dimensional statistical recovery. *Adv. Neural Inf. Process. Syst.* **23**.

ALQUIER, P., COTTET, V. and LECUÉ, G. (2019). Estimation bounds and sharp oracle inequalities of regularized procedures with Lipschitz loss functions. *Ann. Statist.* **47** 2117–2144. MR3953446 https://doi.org/10.1214/18-AOS1742

ANNAN, D. A., MAISHI, N., SOGA, T., DAWOOD, R., LI, C., KIKUCHI, H., HOJO, T., MORIMOTO, M., KITAMURA, T. et al. (2019). Carbonic anhydrase 2 (CAII) supports tumor blood endothelial cell survival under lactic acidosis in the tumor microenvironment. *Cell Commun. Signal.* **17** 1–15.

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

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

MSC2020 subject classifications. Primary 62J05; secondary 62F10.
Key words and phrases. Robust estimation, heavy-tailed noise, linear regression, low-rank matrix, sparse vector.

BLUMENSATH, T. and DAVIES, M. E. (2010). Normalized iterative hard thresholding: Guaranteed stability and performance. *IEEE J. Sel. Top. Signal Process.* **4** 298–309.

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

BURER, S. and MONTEIRO, R. D. (2003). A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Math. Program.* **95** 329–357.

CAI, J.-F., LI, J. and XIA, D. (2022). Provable tensor-train format tensor completion by Riemannian optimization. *J. Mach. Learn. Res.* **23** 123. 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

CAI, T. T. and ZHANG, A. (2015). ROP: Matrix recovery via rank-one projections. *Ann. Statist.* **43** 102–138. MR3285602 https://doi.org/10.1214/14-AOS1267

CAMBIER, L. and ABSIL, P.-A. (2016). Robust low-rank matrix completion by Riemannian optimization. *SIAM J. Sci. Comput.* **38** S440–S460. MR3565571 https://doi.org/10.1137/15M1025153

CANDÈS, E. J., LI, X., MA, Y. and WRIGHT, J. (2011). Robust principal component analysis? *J. ACM* **58** 11. MR2811000 https://doi.org/10.1145/1970392.1970395

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

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

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

CHANG, L.-C., JONES, D. K. and PIERPAOLI, C. (2005). RESTORE: Robust estimation of tensors by outlier rejection. *Magn. Reson. Med.* **53** 1088–1095.

CHARISOPOULOS, V., CHEN, Y., DAVIS, D., DÍAZ, M., DING, L. and DRUSVYATSKIY, D. (2021). Low-rank matrix recovery with composite optimization: Good conditioning and rapid convergence. *Found. Comput. Math.* **21** 1505–1593. MR4343018 https://doi.org/10.1007/s10208-020-09490-9

CHEN, J., ZHOU, J. and YE, J. (2011). Integrating low-rank and group-sparse structures for robust multi-task learning. In *Proceedings of the* 17*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 42–50.

CHEN, M., GAO, C. and REN, Z. (2016). A general decision theory for Huber's ϵ-contamination model. *Electron. J. Stat.* **10** 3752–3774. MR3579675 https://doi.org/10.1214/16-EJS1216

CHEN, M., GAO, C. and REN, Z. (2018). Robust covariance and scatter matrix estimation under Huber's contamination model. *Ann. Statist.* **46** 1932–1960. MR3845006 https://doi.org/10.1214/17-AOS1607

CHEN, X., LIU, W. and MAO, X. (2022). Robust reduced rank regression in a distributed setting. *Sci. China Math.* **65** 1707–1730. MR4456270 https://doi.org/10.1007/s11425-020-1785-0

CHEN, Y. and WAINWRIGHT, M. J. (2015). Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees.

CHINOT, G., LECUÉ, G. and LERASLE, M. (2020). Robust statistical learning with Lipschitz and convex loss functions. *Probab. Theory Related Fields* **176** 897–940. MR4087486 https://doi.org/10.1007/s00440-019-00931-3

CHIU, J., DENG, Y. and RUSH, A. (2021). Low-rank constraints for fast inference in structured models. *Adv. Neural Inf. Process. Syst.* **34**.

COLAS, E., PEREZ, C., CABRERA, S., PEDROLA, N., MONGE, M., CASTELLVI, J., EYZAGUIRRE, F., GREGORIO, J., RUIZ, A. et al. (2011). Molecular markers of endometrial carcinoma detected in uterine aspirates. *Int. J. Cancer* **129** 2435–2444.

DAVENPORT, M. A. and ROMBERG, J. (2016). An overview of low-rank matrix recovery from incomplete observations. *IEEE J. Sel. Top. Signal Process.* **10** 608–622.

DEPERSIN, J. (2024). Robust subgaussian estimation with VC-dimension. *Ann. Inst. Henri Poincaré Probab. Stat.* **60** 971–989. MR4757514 https://doi.org/10.1214/22-aihp1347

DEPERSIN, J. and LECUÉ, G. (2022). Robust sub-Gaussian estimation of a mean vector in nearly linear time. *Ann. Statist.* **50** 511–536. MR4382026 https://doi.org/10.1214/21-aos2118

DIAKONIKOLAS, I., KAMATH, G., KANE, D., LI, J., MOITRA, A. and STEWART, A. (2019). Robust estimators in high-dimensions without the computational intractability. *SIAM J. Comput.* **48** 742–864. MR3945261 https://doi.org/10.1137/17M1126680

DIAKONIKOLAS, I. and KANE, D. M. (2019). Recent advances in algorithmic high-dimensional robust statistics. arXiv preprint. Available at arXiv:1911.05911.

DUCHI, J. C., BARTLETT, P. L. and WAINWRIGHT, M. J. (2012). Randomized smoothing for stochastic optimization. *SIAM J. Optim.* **22** 674–701. MR2968871 https://doi.org/10.1137/110831659

EFRON, B., HASTIE, T., JOHNSTONE, I. and TIBSHIRANI, R. (2004). Least angle regression. *Ann. Statist.* **32** 407–499. MR2060166 https://doi.org/10.1214/009053604000000067

ELSENER, A. and VAN DE GEER, S. (2018). Robust low-rank matrix estimation. *Ann. Statist.* **46** 3481–3509. MR3852659 https://doi.org/10.1214/17-AOS1666

FAN, J., WANG, W. and ZHU, Z. (2021). A shrinkage principle for heavy-tailed data: High-dimensional robust low-rank matrix recovery. *Ann. Statist.* **49** 1239–1266. MR4298863 https://doi.org/10.1214/20-aos1980

FOUCART, S. and LECUÉ, G. (2017). An IHT algorithm for sparse recovery from subexponential measurements. *IEEE Signal Process. Lett.* **24** 1280–1283.

GOFFIN, J. L. (1977). On convergence rates of subgradient optimization methods. *Math. Program.* **13** 329–347. MR0519665 https://doi.org/10.1007/BF01584346

GROSS, D., LIU, Y.-K., FLAMMIA, S. T., BECKER, S. and EISERT, J. (2010). Quantum state tomography via compressed sensing. *Phys. Rev. Lett.* **105** 150401.

HAN, W., CAO, F., GAO, X.-J., WANG, H.-B., CHEN, F., CAI, S.-J., ZHANG, C., HU, Y.-W., MA, J. et al. (2018). ZIC1 acts a tumor suppressor in breast cancer by targeting survivin. *Int. J. Oncol.* **53** 937–948.

HOPKINS, S. B. (2020). Mean estimation with sub-Gaussian rates in polynomial time. *Ann. Statist.* **48** 1193–1213. MR4102693 https://doi.org/10.1214/19-AOS1843

HUBER, P. J. (1965). A robust version of the probability ratio test. *Ann. Math. Stat.* **36** 1753–1758. MR0185747 https://doi.org/10.1214/aoms/1177699803

HUBER, P. J. and RONCHETTI, E. M. (2009). *Robust Statistics*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2488795 https://doi.org/10.1002/9780470434697

JAIN, P., TEWARI, A. and KAR, P. (2014). On iterative hard thresholding methods for high-dimensional m-estimation. *Adv. Neural Inf. Process. Syst.* **27**.

KIM, J. H., KANG, H.-S., KIM, T. W. and KIM, S. J. (2011). Differential methylation hybridization profiling identifies involvement of STAT1-mediated pathways in breast cancer. *Int. J. Oncol.* **39** 955–963.

KLOPP, O., LOUNICI, K. and TSYBAKOV, A. B. (2017). Robust matrix completion. *Probab. Theory Related Fields* **169** 523–564. MR3704775 https://doi.org/10.1007/s00440-016-0736-y

KOENKER, R. and BASSETT, G. JR. (1978). Regression quantiles. *Econometrica* **46** 33–50. MR0474644 https://doi.org/10.2307/1913643

KOENKER, R. and HALLOCK, K. F. (2001). Quantile regression. *J. Econ. Perspect.* **15** 143–156.

LECUÉ, G. and LERASLE, M. (2019). Learning from MOM's principles: Le Cam's approach. *Stochastic Process. Appl.* **129** 4385–4410. MR4013866 https://doi.org/10.1016/j.spa.2018.11.024

LECUÉ, G. and MENDELSON, S. (2017). Sparse recovery under weak moment assumptions. *J. Eur. Math. Soc.* (*JEMS*) **19** 881–904. MR3612870 https://doi.org/10.4171/JEMS/682

LI, B., WEN, G., ZHAO, Y., TONG, J. and HEI, T. K. (2012). The role of TGFBI in mesothelioma and breast cancer: Association with tumor suppression. *BMC Cancer* **12** 1–12.

LI, X., ZHU, Z., SO, A. M.-C. and VIDAL, R. (2020). Nonconvex robust low-rank matrix recovery. *SIAM J. Optim.* **30** 660–686. MR4068332 https://doi.org/10.1137/18M1224738

LOH, P.-L. and WAINWRIGHT, M. J. (2012). High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *Ann. Statist.* **40** 1637–1664. MR3015038 https://doi.org/10.1214/12-AOS1018

LOUNICI, K., PONTIL, M., VAN DE GEER, S. and TSYBAKOV, A. B. (2011). Oracle inequalities and optimal inference under group sparsity. *Ann. Statist.* **39** 2164–2204. MR2893865 https://doi.org/10.1214/11-AOS896

LUGOSI, G. and MENDELSON, S. (2019a). Regularization, sparse recovery, and median-of-means tournaments. *Bernoulli* **25** 2075–2106. MR3961241 https://doi.org/10.3150/18-BEJ1046

LUGOSI, G. and MENDELSON, S. (2019b). 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

LUGOSI, G. and MENDELSON, S. (2020). Risk minimization by median-of-means tournaments. *J. Eur. Math. Soc.* (*JEMS*) **22** 925–965. MR4055993 https://doi.org/10.4171/jems/937

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

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

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** 96. MR4582518

MA, Z. and WU, Y. (2015). Volume ratio, sparsity, and minimaxity under unitarily invariant norms. *IEEE Trans. Inf. Theory* **61** 6939–6956. MR3430731 https://doi.org/10.1109/TIT.2015.2487541

MINSKER, S. (2015). Geometric median and robust estimation in Banach spaces. *Bernoulli* **21** 2308–2335. MR3378468 https://doi.org/10.3150/14-BEJ645

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

MINSKER, S., NDAOUD, M. and WANG, L. (2024). Robust and tuning-free sparse linear regression via square-root slope. *SIAM J. Math. Data Sci.* **6** 428–453. MR4752929 https://doi.org/10.1137/23M1608690

NAKATA, B., TAKASHIMA, T., OGAWA, Y., ISHIKAWA, T. and HIRAKAWA, K. (2004). Serum CYFRA 21-1 (cytokeratin-19 fragments) is a useful tumour marker for detecting disease relapse and assessing treatment efficacy in breast cancer. *Br. J. Cancer* **91** 873–878. https://doi.org/10.1038/sj.bjc.6602074

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

NEGAHBAN, S. and WAINWRIGHT, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *Ann. Statist.* **39** 1069–1097. MR2816348 https://doi.org/10.1214/10-AOS850

NISHIZUKA, S., CHARBONEAU, L., YOUNG, L., MAJOR, S., REINHOLD, W. C., WALTHAM, M., KOUROS-MEHR, H., BUSSEY, K. J., LEE, J. K. et al. (2003). Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays. *Proc. Natl. Acad. Sci. USA* **100** 14229–14234.

NIU, J., LI, X.-M., WANG, X., LIANG, C., ZHANG, Y.-D., LI, H.-Y., LIU, F.-Y., SUN, H., XIE, S.-Q. et al. (2019). DKK1 inhibits breast cancer cell migration and invasion through suppression of β-catenin/MMP7 signaling pathway. *Cancer Cell Int.* **19** 1–13.

OSTA, W. A., CHEN, Y., MIKHITARIAN, K., MITAS, M., SALEM, M., HANNUN, Y. A., COLE, D. J. and GILLANDERS, W. E. (2004). EpCAM is overexpressed in breast cancer and is a potential target for breast cancer gene therapy. *Cancer Res.* **64** 5818–5824.

PAN, X., SUN, Q. and ZHOU, W.-X. (2021). Iteratively reweighted ℓ_1-penalized robust regression. *Electron. J. Stat.* **15** 3287–3348. MR4280173 https://doi.org/10.1214/21-ejs1862

RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2011). Minimax rates of estimation for high-dimensional linear regression over ℓ_q-balls. *IEEE Trans. Inf. Theory* **57** 6976–6994. MR2882274 https://doi.org/10.1109/TIT.2011.2165799

RIGOLLET, P. and TSYBAKOV, A. (2011). Exponential screening and optimal rates of sparse estimation. *Ann. Statist.* **39** 731–771. MR2816337 https://doi.org/10.1214/10-AOS854

ROHDE, A. and TSYBAKOV, A. B. (2011). Estimation of high-dimensional low-rank matrices. *Ann. Statist.* **39** 887–930. MR2816342 https://doi.org/10.1214/10-AOS860

SHANKAVARAM, U. T., REINHOLD, W. C., NISHIZUKA, S., MAJOR, S., MORITA, D., CHARY, K. K., REIMERS, M. A., SCHERF, U., KAHN, A. et al. (2007). Transcript and protein expression profiles of the NCI-60 cancer cell panel: An integromic microarray study. *Mol. Cancer Ther.* **6** 820–832.

SHEN, J. and LI, P. (2017). A tight bound of hard thresholding. *J. Mach. Learn. Res.* **18** 208. MR3827096

SHEN, Y., LI, J., CAI, J.-F and XIA, D. (2025). Supplement to "Computationally Efficient and Statistically Optimal Robust High-Dimensional Linear Regression." https://doi.org/10.1214/24-AOS2473SUPP

SIDDIQI, S., BOOTS, B. and GORDON, G. (2010). Reduced-rank hidden Markov models. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* 741–748. JMLR Workshop and Conference Proceedings.

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

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

TAN, K. M., WANG, L. and ZHOU, W.-X. (2022). High-dimensional quantile regression: Convolution smoothing and concave regularization. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 205–233. MR4400395 https://doi.org/10.1111/rssb.12485

THOMPSON, P. (2020). Outlier-robust sparse/low-rank least-squares regression and robust matrix completion. arXiv preprint. Available at arXiv:2012.06750.

THOMPSON, P. (2023). Outlier-robust additive matrix decomposition and robust matrix completion. arXiv preprint. Available at arXiv:2310.19136.

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

TONG, T., MA, C. and CHI, Y. (2021a). Low-rank matrix recovery with scaled subgradient methods: Fast and robust convergence without the condition number. *IEEE Trans. Signal Process.* **69** 2396–2409. MR4253619 https://doi.org/10.1109/TSP.2021.3071560

TONG, T., MA, C. and CHI, Y. (2021b). Accelerating ill-conditioned low-rank matrix estimation via scaled gradient descent. *J. Mach. Learn. Res.* **22** 150. MR4318506

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

VAN DE GEER, S. A. and VAN DE GEER, S. (2000). *Empirical Processes in M-Estimation* **6**. Cambridge University Press, Cambridge.

WANG, L., PENG, B., BRADIC, J., LI, R. and WU, Y. (2020). A tuning-free robust and efficient approach to high-dimensional regression. *J. Amer. Statist. Assoc.* **115** 1700–1714. MR4189748 https://doi.org/10.1080/01621459.2020.1840989

WANG, L., WU, Y. and LI, R. (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *J. Amer. Statist. Assoc.* **107** 214–222. MR2949353 https://doi.org/10.1080/01621459.2012.656014

WANG, L., ZHENG, C., ZHOU, W. and ZHOU, W.-X. (2021). A new principle for tuning-free Huber regression. *Statist. Sinica* **31** 2153–2177. MR4328856 https://doi.org/10.5705/ss.202019.0045

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

WELSH, A. H. (1989). On M-processes and M-estimation. *Ann. Statist.* **17** 337–361. MR0981455 https://doi.org/10.1214/aos/1176347021

XIA, D. (2014). Optimal Schatten-q and ky-fan-k norm rate of low rank matrix estimation. arXiv preprint. Available at arXiv:1403.6499.

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

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

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

XIE, W.-Y., HE, R.-H., ZHANG, J., HE, Y.-J., WAN, Z., ZHOU, C.-F., TANG, Y.-J., LI, Z., MCLEOD, H. L. et al. (2019). β-blockers inhibit the viability of breast cancer cells by regulating the ERK/COX-2 signaling pathway and the drug response is affected by ADRB2 single-nucleotide polymorphisms. *Oncol. Rep.* **41** 341–350.

XUE, J., ZHU, X., QIAO, X., WANG, Y., BU, J., ZHANG, X., MA, Q., LIANG, L., SUN, L. et al. (2022). CEMIP as a potential biomarker and therapeutic target for breast cancer patients. *Int. J. Med. Sci.* **19** 434.

YANG, L., CHEN, Y., CUI, T., KNÖSEL, T., ZHANG, Q., ALBRING, K. F., HUBER, O. and PETERSEN, I. (2012). Desmoplakin acts as a tumor suppressor by inhibition of the Wnt/β-catenin signaling pathway in human lung cancer. *Carcinogenesis* **33** 1863–1870.

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

YE, X., ZHOU, J., TONG, D., WANG, D., WANG, H., GUO, J. and KANG, X. (2022). E2F1 affects the therapeutic response to neoadjuvant therapy in breast cancer. *Dis. Mark.* **2022**.

YU, W. and CHAI, H. (2017). Inhibition of BAMBI reduces the viability and motility of colon cancer via activating TGF-β/Smad pathway in vitro and in vivo Retraction in/10.3892/ol. 2021.12505. *Oncol. Lett.* **14** 4793–4799.

ZHANG, S., WANG, Z., LIU, W., LEI, R., SHAN, J., LI, L. and WANG, X. (2017). Distinct prognostic values of S100 mRNA expression in breast cancer. *Sci. Rep.* **7** 1–11.

ZHANG, Y. and XIA, D. (2022). Edgeworth expansions for network moments. *Ann. Statist.* **50** 726–753. MR4404918 https://doi.org/10.1214/21-aos2125

ZHAO, P., YANG, Y. and HE, Q.-C. (2022). High-dimensional linear regression via implicit regularization. *Biometrika* **109** 1033–1046. MR4519114 https://doi.org/10.1093/biomet/asac010

ZHAO, T., WANG, Z. and LIU, H. (2015). A nonconvex optimization framework for low rank matrix estimation. *Adv. Neural Inf. Process. Syst.* **28**.

ZHENG, Q. and LAFFERTY, J. (2016). A convergent gradient descent algorithm for rank minimization and semidefinite programming from random linear measurements.

ZHENG, Q., PENG, L. and HE, X. (2015). Globally adaptive quantile regression with ultra-high dimensional data. *Ann. Statist.* **43** 2225–2258. MR3396984 https://doi.org/10.1214/15-AOS1340

ZHOU, T., LI, Y., YANG, L., LIU, L., JU, Y. and LI, C. (2017). Silencing of ANXA3 expression by RNA interference inhibits the proliferation and invasion of breast cancer cells. *Oncol. Rep.* **37** 388–398.

ZHU, J., WEN, C., ZHU, J., ZHANG, H. and WANG, X. (2020). A polynomial algorithm for best-subset selection problem. *Proc. Natl. Acad. Sci. USA* **117** 33117–33123. MR4269239 https://doi.org/10.1073/pnas.2014241117

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** 169. MR4777411

The Annals of Statistics
2025, Vol. 53, No. 1, 400–425
https://doi.org/10.1214/24-AOS2475

MULTIVARIATE DYNAMIC MEDIATION ANALYSIS UNDER A REINFORCEMENT LEARNING FRAMEWORK

BY LAN LUO[1,a], CHENGCHUN SHI[2,b], JITAO WANG[3,c], ZHENKE WU[3,d] AND LEXIN LI[4,e]

[1]*Department of Biostatistics and Epidemiology, Rutgers University,* [a]*l.luo@rutgers.edu*
[2]*Department of Statistics, London School of Economics and Political Science,* [b]*C.Shi7@lse.ac.uk*
[3]*Department of Biostatistics, University of Michigan,* [c]*jitwang@umich.edu,* [d]*zhenkewu@umich.edu*
[4]*Division of Biostatistics, University of California, Berkeley,* [e]*lexinli@berkeley.edu*

Mediation analysis is an important analytic tool commonly used in a broad range of scientific applications. In this article, we study the problem of mediation analysis when there are multivariate and conditionally dependent mediators, and when the variables are observed over multiple time points. The problem is challenging, because the effect of a mediator involves not only the path from the treatment to this mediator itself at the current time point, but also all possible paths pointed to this mediator from its upstream mediators, as well as the carryover effects from all previous time points. We propose a novel multivariate dynamic mediation analysis approach. Drawing inspiration from the Markov decision process model that is frequently employed in reinforcement learning, we introduce a Markov mediation process paired with a system of time-varying linear structural equation models to formulate the problem. We then formally define the individual mediation effect, built upon the idea of simultaneous interventions and intervention calculus. We next derive the closed-form expression, propose an iterative estimation procedure under the Markov mediation process model and develop a bootstrap method to infer the individual mediation effect. We study both the asymptotic property and the empirical performance of the proposed methodology, and further illustrate its usefulness with a mobile health application.

REFERENCES

BARON, R. M. and KENNY, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* **51** 1173–1182. https://doi.org/10.1037//0022-3514.51.6.1173

BELLO, K., ARAGAM, B. and RAVIKUMAR, P. (2022). DAGMA: Learning DAGs via M-matrices and a log-determinant acyclicity characterization. arXiv preprint. Available at arXiv:2209.08037.

BI, X., YANG, L., LI, T., WANG, B., ZHU, H. and ZHANG, H. (2017). Genome-wide mediation analysis of psychiatric and cognitive traits through imaging phenotypes. *Hum. Brain Mapp.* **38** 4088–4097.

BOCA, S. M., SINHA, R., CROSS, A. J., MOORE, S. C. and SAMPSON, J. N. (2014). Testing multiple biological mediators simultaneously. *Bioinformatics* **30** 214–220. https://doi.org/10.1093/bioinformatics/btt633

BRADLEY, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** 107–144. Update of, and a supplement to, the 1986 original. MR2178042 https://doi.org/10.1214/154957805100000104

CAI, H., SONG, R. and LU, W. (2021). ANOCE: Analysis of causal effects with multiple mediators via constrained structural learning. In *International Conference on Learning Representations*.

CAI, X., COFFMAN, D. L., PIPER, M. E. and LI, R. (2022). Estimation and inference for the mediation effect in a time-varying mediation model. *BMC Med. Res. Methodol.* **22** 113.

CELLI, V. (2022). Causal mediation analysis in economics: Objectives, assumptions, models. *J. Econ. Surv.* **36** 214–234.

CHAKRABORTTY, A., NANDY, P. and LI, H. (2018). Inference for individual mediation effects and interventional effects in sparse high-dimensional causal graphical models. arXiv preprint. Available at arXiv:1809.10652.

MSC2020 subject classifications. 62L10.
Key words and phrases. Longitudinal data, Markov process, mediation analysis, mobile health, reinforcement learning.

DÍAZ, I., WILLIAMS, N. and RUDOLPH, K. E. (2023). Efficient and flexible mediation analysis with time-varying mediators, treatments, and confounders. *J. Causal Inference* **11** Paper No. 20220077, 17. MR4610658 https://doi.org/10.1515/jci-2022-0077

DJORDJILOVIĆ, V., HEMERIK, J. and THORESEN, M. (2022). On optimal two-stage testing of multiple mediators. *Biom. J.* **64** 1090–1108. MR4476372 https://doi.org/10.1002/bimj.202100190

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

GE, L., WANG, J., SHI, C., WU, Z. and SONG, R. (2023). A reinforcement learning framework for dynamic mediation analysis. arXiv preprint. Available at arXiv:2301.13348.

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

GUO, X., LI, R., LIU, J. and ZENG, M. (2023). Statistical inference for linear mediation models with high-dimensional mediators and application to studying stock reaction to COVID-19 pandemic. *J. Econometrics* **235** 166–179. MR4580425 https://doi.org/10.1016/j.jeconom.2022.03.001

HAO, B., JI, X., DUAN, Y., LU, H., SZEPESVARI, C. and WANG, M. (2021). Bootstrapping fitted q-evaluation for off-policy inference. In *International Conference on Machine Learning* 4074–4084. PMLR.

HEJAZI, N. S., RUDOLPH, K. E., VAN DER LAAN, M. J. and DÍAZ, I. (2023). Nonparametric causal mediation analysis for stochastic interventional (in)direct effects. *Biostatistics* **24** 686–707. MR4615248 https://doi.org/10.1093/biostatistics/kxac002

HU, Y. and WAGER, S. (2022). Switchback experiments under geometric mixing. arXiv preprint. Available at arXiv:2209.00197.

HUANG, J. and YUAN, Y. (2017). Bayesian dynamic mediation analysis. *Psychol. Methods* **22** 667–686.

HUANG, Y.-T. and PAN, W.-C. (2016). Hypothesis test of mediation effect in causal mediation model with high-dimensional continuous mediators. *Biometrics* **72** 402–413. MR3515767 https://doi.org/10.1111/biom.12421

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

KALLUS, N. and UEHARA, M. (2022). Efficiently breaking the curse of horizon in off-policy evaluation with double reinforcement learning. *Oper. Res.* **70** 3282–3302. MR4538517

KAUFMAN, J. S. and KAUFMAN, S. (2001). Assessment of structured socioeconomic effects on health. *Epidemiology* **12** 157–167. https://doi.org/10.1097/00001648-200103000-00006

LI, M., SHI, C., WU, Z. and FRYZLEWICZ, P. (2022). Testing stationarity and change point detection in reinforcement learning. arXiv preprint. Available at arXiv:2203.01707.

LIAO, P., KLASNJA, P. and MURPHY, S. (2021). Off-policy estimation of long-term average outcomes with applications to mobile health. *J. Amer. Statist. Assoc.* **116** 382–391. MR4227701 https://doi.org/10.1080/01621459.2020.1807993

LIAO, P., QI, Z., WAN, R., KLASNJA, P. and MURPHY, S. A. (2022). Batch policy learning in average reward Markov decision processes. *Ann. Statist.* **50** 3364–3387. MR4524500 https://doi.org/10.1214/22-aos2231

LIN, S.-H., YOUNG, J. G., LOGAN, R. and VANDERWEELE, T. J. (2017). Mediation analysis for a survival outcome with time-varying exposures, mediators, and confounders. *Stat. Med.* **36** 4153–4166. MR3713656 https://doi.org/10.1002/sim.7426

LIU, W., TU, J., ZHANG, Y. and CHEN, X. (2023). Online Estimation and Inference for Robust Policy Evaluation in Reinforcement Learning. arXiv preprint. Available at arXiv:2310.02581.

LUCKETT, D. J., LABER, E. B., KAHKOSKA, A. R., MAAHS, D. M., MAYER-DAVIS, E. and KOSOROK, M. R. (2020). Estimating dynamic treatment regimes in mobile health using V-learning. *J. Amer. Statist. Assoc.* **115** 692–706. MR4107673 https://doi.org/10.1080/01621459.2018.1537919

LUO, L., SHI, C., WANG, J., WU, Z. and LI, L. (2025). Supplement to “Multivariate Dynamic Mediation Analysis under a Reinforcement Learning Framework.” https://doi.org/10.1214/24-AOS2475SUPP

MAATHUIS, M. H., KALISCH, M. and BÜHLMANN, P. (2009). Estimating high-dimensional intervention effects from observational data. *Ann. Statist.* **37** 3133–3164. MR2549555 https://doi.org/10.1214/09-AOS685

MACKINNON, D. P. (2008). *Introduction to Statistical Mediation Analysis*. Taylor & Francis, London.

MALINSKY, D. and SPIRTES, P. (2018). Causal structure learning from multivariate time series in settings with unmeasured confounding. In *Proceedings of* 2018 *ACM SIGKDD Workshop on Causal Discovery* 23–47. PMLR.

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

MURPHY, S. A. (2003). Optimal dynamic treatment regimes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 331–355. MR1983752 https://doi.org/10.1111/1467-9868.00389

NANDY, P., MAATHUIS, M. H. and RICHARDSON, T. S. (2017). Estimating the effect of joint interventions from observational data in sparse high-dimensional settings. *Ann. Statist.* **45** 647–674. MR3650396 https://doi.org/10.1214/16-AOS1462

NECAMP, T., SEN, S., FRANK, E., WALTON, M. A., IONIDES, E. L., FANG, Y., TEWARI, A. and WU, Z. (2020). Assessing real-time moderation for developing adaptive mobile health interventions for medical interns: Micro-randomized trial. *J. Med. Internet Res.* **22** e15033.

PAMFIL, R., SRIWATTANAWORACHAI, N., DESAI, S., PILGERSTORFER, P., GEORGATZIS, K., BEAUMONT, P. and ARAGAM, B. (2020). Dynotears: Structure learning from time-series data. In *International Conference on Artificial Intelligence and Statistics* 1595–1605. PMLR.

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

PEARL, J. (2012). The causal mediation formula—a guide to the assessment of pathways and mechanisms. *Prev. Sci.* **13** 426–436.

PETERS, J. and BÜHLMANN, P. (2014). Identifiability of Gaussian structural equation models with equal error variances. *Biometrika* **101** 219–228. MR3180667 https://doi.org/10.1093/biomet/ast043

PREACHER, K. J. (2015). Advances in mediation analysis: A survey and synthesis of new developments. *Annu. Rev. Psychol.* **66** 825–852. https://doi.org/10.1146/annurev-psych-010814-015258

PUTERMAN, M. L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. *Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Wiley, New York. A Wiley-Interscience Publication. MR1270015

QIN, Z. T., ZHU, H. and YE, J. (2021). Reinforcement learning for ridesharing: A survey. In 2021 *IEEE International Intelligent Transportation Systems Conference* (*ITSC*) 2447–2454. IEEE Press, New York.

RAMPRASAD, P., LI, Y., YANG, Z., WANG, Z., SUN, W. W. and CHENG, G. (2023). Online bootstrap inference for policy evaluation in reinforcement learning. *J. Amer. Statist. Assoc.* **118** 2901–2914. MR4681629 https://doi.org/10.1080/01621459.2022.2096620

RUCKER, D. D., PREACHER, K. J., TORMALA, Z. L. and PETTY, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Soc. Pers. Psychol. Compass* **5** 359–371.

SAMPSON, J. N., BOCA, S. M., MOORE, S. C. and HELLER, R. (2018). FWER and FDR control when testing multiple mediators. *Bioinformatics* **34** 2418–2424. https://doi.org/10.1093/bioinformatics/bty064

SELIG, J. P. and PREACHER, K. J. (2009). Mediation models for longitudinal data in developmental research. *Res. Hum. Dev.* **6** 144–164.

SHAFFER, F. and GINSBERG, J. P. (2017). An overview of heart rate variability metrics and norms. *Front. Public Health* **5** 258. https://doi.org/10.3389/fpubh.2017.00258

SHI, C. and LI, L. (2022). Testing mediation effects using logic of Boolean matrices. *J. Amer. Statist. Assoc.* **117** 2014–2027. MR4528486 https://doi.org/10.1080/01621459.2021.1895177

SHI, C., ZHANG, S., LU, W. and SONG, R. (2022). Statistical inference of the value function for reinforcement learning in infinite-horizon settings. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 765–793. MR4460575

SHUAI, K., LIU, L., HE, Y. and LI, W. (2023). Mediation pathway selection with unmeasured mediator-outcome confounding. arXiv preprint. Available at arXiv:2311.16793.

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

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

VANDERWEELE, T. J. (2016). Mediation analysis: A practitioner's guide. *Annu. Rev. Public Health* **37** 17–32. https://doi.org/10.1146/annurev-publhealth-032315-021402

VANDERWEELE, T. J. and TCHETGEN TCHETGEN, E. J. (2017). Mediation analysis with time varying exposures and mediators. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 917–938. MR3641414 https://doi.org/10.1111/rssb.12194

WANG, J., QI, Z. and WONG, R. K. W. (2023). Projected state-action balancing weights for offline reinforcement learning. *Ann. Statist.* **51** 1639–1665. MR4658571 https://doi.org/10.1214/23-aos2302

WANG, J., SHI, C. and WU, Z. (2023). A robust test for the stationarity assumption in sequential decision making. In *International Conference on Machine Learning* 36355–36379. PMLR.

WEI, H., CAI, H., SHI, C. and SONG, R. (2024). On efficient inference of causal effects with multiple mediators. arXiv preprint. Available at arXiv:2401.05517.

WRIGHT, S. (1921). Correlation and causation. *J. Agric. Res.* **20** 557–585.

YUAN, Y. and QU, A. (2024). De-confounding causal inference using latent multiple-mediator pathways. *J. Amer. Statist. Assoc.* **119** 2051–2065. MR4797922 https://doi.org/10.1080/01621459.2023.2240461

YUAN, Y., SHEN, X., PAN, W. and WANG, Z. (2019). Constrained likelihood for reconstructing a directed acyclic Gaussian graph. *Biometrika* **106** 109–125. MR3912386 https://doi.org/10.1093/biomet/asy057

ZHANG, H., CHEN, X. and YANG, L. F. (2023). Adaptive liquidity provision in uniswap V3 with deep reinforcement learning. arXiv preprint. Available at arXiv:2309.10129.

ZHANG, H., ZHENG, Y., ZHANG, Z., GAO, T., JOYCE, B., YOON, G., ZHANG, W., SCHWARTZ, J., JUST, A. et al. (2016). Estimating and testing high-dimensional mediation effects in epigenetic studies. *Bioinformatics* **32** 3150–3154.

ZHAO, Y., LI, L. and INITIATIVE, A. D. N. (2022). Multimodal data integration via mediation analysis with high-dimensional exposures and mediators. *Hum. Brain Mapp.* **43** 2519–2533.

ZHAO, Y. and LUO, X. (2022). Pathway Lasso: Pathway estimation and selection with high-dimensional mediators. *Stat. Interface* **15** 39–50. MR4305014 https://doi.org/10.4310/21-SII673

ZHAO, Y., LUO, X., LINDQUIST, M. and CAFFO, B. (2018). Functional mediation analysis with an application to functional magnetic resonance imaging data. arXiv preprint. Available at arXiv:1805.06923.

ZHENG, W. and VAN DER LAAN, M. (2017). Longitudinal mediation analysis with time-varying mediators and exposures, with application to survival outcomes. *J. Causal Inference* **5** Art. No. 20160006, 24. MR4328877 https://doi.org/10.1515/jci-2016-0006

ZHENG, X., ARAGAM, B., RAVIKUMAR, P. K. and XING, E. P. (2018). Reinforcement learning for ridesharing: A survey. DAGs with NO TEARS: Continuous optimization for structure learning. In *In Advances in Neural Information Processing Systems* 9472–9483.

The Annals of Statistics
2025, Vol. 53, No. 1, 426–456
https://doi.org/10.1214/24-AOS2483

UNIFIED ALGORITHMS FOR RL WITH DECISION-ESTIMATION COEFFICIENTS: PAC, REWARD-FREE, PREFERENCE-BASED LEARNING AND BEYOND

BY FAN CHEN[1,a], SONG MEI[2,b] AND YU BAI[3,c]

[1]*Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology,* [a]*fanchen@mit.edu*
[2]*Department of Statistics, University of California, Berkeley,* [b]*songmei@berkeley.edu*
[3]*Work done at Salesforce Research,* [c]*yubai.pku@gmail.com*

Modern Reinforcement Learning (RL) is more than just learning the optimal policy; alternative learning goals such as exploring the environment, estimating the underlying model and learning from preference feedback are all of practical importance. While provably sample-efficient algorithms for each specific goal have been proposed, these algorithms often depend strongly on the particular learning goal, and thus admit different structures correspondingly. It is an urging open question whether these learning goals can rather be tackled by a single unified algorithm.

We make progress on this question by developing a unified algorithm framework for a large class of learning goals, building on the Decision-Estimation Coefficient (DEC) framework. Our framework handles many learning goals such as no-regret RL, PAC RL, reward-free learning, model estimation and preference-based learning, all by simply instantiating the same generic complexity measure called "Generalized DEC," and a corresponding generic algorithm. The generalized DEC also yields a sample complexity lower bound for each specific learning goal. As applications, we propose "decouplable representation" as a natural sufficient condition for bounding generalized DECs, and use it to obtain many new sample-efficient results (and recover existing results) for a wide range of learning goals and problem classes as direct corollaries. Finally, as a connection, we reanalyze two existing optimistic model-based algorithms based on posterior sampling and maximum likelihood estimation, showing that they enjoy sample complexity bounds under similar structural conditions as the DEC.

REFERENCES

[1] ABBASI-YADKORI, Y. and NEU, G. (2014). Online learning in MDPs with side information. Preprint. Available at arXiv:1406.6812.

[2] 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.

[3] AGARWAL, A., KAKADE, S., KRISHNAMURTHY, A. and SUN, W. (2020). Flambe: Structural complexity and representation learning of low rank MDPs. *Adv. Neural Inf. Process. Syst.* **33** 20095–20107.

[4] AGARWAL, A. and ZHANG, T. (2022). Model-based RL with optimistic posterior sampling: Structural conditions and sample complexity. Preprint. Available at arXiv:2206.07659.

[5] AGARWAL, A. and ZHANG, T. (2022). Non-linear reinforcement learning in large action spaces: Structural conditions and sample-efficiency of posterior sampling. Preprint. Available at arXiv:2203.08248.

[6] AGRAWAL, S. and JIA, R. (2017). Optimistic posterior sampling for reinforcement learning: Worst-case regret bounds. *Adv. Neural Inf. Process. Syst.* **30**.

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

MSC2020 subject classifications. Primary 62C20; secondary 90C40.
Key words and phrases. Reinforcement learning, decision-estimation coefficients, reward-free learning, preference-based learning.

[8] AMORTILA, P., FOSTER, D. J. and KRISHNAMURTHY, A. (2024). Scalable online exploration via coverability. Preprint. Available at arXiv:2403.06571.

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

[10] AUER, P., ORTNER, R. and SZEPESVÁRI, C. (2007). Improved rates for the stochastic continuum-armed bandit problem. In *Learning Theory. Lecture Notes in Computer Science* **4539** 454–468. Springer, Berlin. MR2397605 https://doi.org/10.1007/978-3-540-72927-3_33

[11] AYOUB, A., JIA, Z., SZEPESVARI, C., WANG, M. and YANG, L. (2020). Model-based reinforcement learning with value-targeted regression. In *International Conference on Machine Learning* 463–474. PMLR.

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

[13] BENGS, V., BUSA-FEKETE, R., EL MESAOUDI-PAUL, A. and HÜLLERMEIER, E. (2021). Preference-based online learning with dueling bandits: A survey. *J. Mach. Learn. Res.* **22** Paper No. 7, 108 pp. MR4253700

[14] BEYGELZIMER, A., LANGFORD, J., LI, L., REYZIN, L. and SCHAPIRE, R. (2011). Contextual bandit algorithms with supervised learning guarantees. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics* 19–26. JMLR Workshop and Conference Proceedings.

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

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

[17] BRAFMAN, R. I. and TENNENHOLTZ, M. (2003). R-MAX—a general polynomial time algorithm for near-optimal reinforcement learning. *J. Mach. Learn. Res.* **3** 213–231. MR1971337 https://doi.org/10.1162/153244303765208377

[18] BUBECK, S., DEKEL, O., KOREN, T. and PERES, Y. (2015). Bandit convex optimization: \sqrtt regret in one dimension. In *Conference on Learning Theory* 266–278. PMLR.

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

[20] CAI, Q., YANG, Z., JIN, C. and WANG, Z. (2020). Provably efficient exploration in policy optimization. In *International Conference on Machine Learning* 1283–1294. PMLR.

[21] CHEN, F., BAI, Y. and MEI, S. (2022). Partially observable RL with B-stability: Unified structural condition and sharp sample-efficient algorithms. Preprint. Available at arXiv:2209.14990.

[22] CHEN, F., MEI, S. and BAI, Y. (2025). Supplement to "Unified Algorithms for RL with Decision-Estimation Coefficients: PAC, Reward-Free, Preference-Based Learning and Beyond." https://doi.org/10.1214/24-AOS2483SUPP

[23] CHEN, J., MODI, A., KRISHNAMURTHY, A., JIANG, N. and AGARWAL, A. (2022). On the statistical efficiency of reward-free exploration in non-linear RL. Preprint. Available at arXiv:2206.10770.

[24] CHEN, X., HU, J., YANG, L. F. and WANG, L. (2021). Near-optimal reward-free exploration for linear mixture MDPs with plug-in solver. Preprint. Available at arXiv:2110.03244.

[25] CHEN, X., ZHONG, H., YANG, Z., WANG, Z. and WANG, L. (2022). Human-in-the-loop: Provably efficient preference-based reinforcement learning with general function approximation. In *International Conference on Machine Learning* 3773–3793. PMLR.

[26] CHOWDHURY, S. R., GOPALAN, A. and MAILLARD, O.-A. (2021). Reinforcement learning in parametric MDPs with exponential families. In *International Conference on Artificial Intelligence and Statistics* 1855–1863. PMLR.

[27] CHRISTIANO, P. F., LEIKE, J., BROWN, T., MARTIC, M., LEGG, S. and AMODEI, D. (2017). Deep reinforcement learning from human preferences. *Adv. Neural Inf. Process. Syst.* **30**.

[28] 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.

[29] CUI, Q., ZHANG, K. and DU, S. (2023). Breaking the curse of multiagents in a large state space: Rl in Markov games with independent linear function approximation. In *The Thirty Sixth Annual Conference on Learning Theory* 2651–2652. PMLR.

[30] DANN, C. and BRUNSKILL, E. (2015). Sample complexity of episodic fixed-horizon reinforcement learning. *Adv. Neural Inf. Process. Syst.* **28**.

[31] DANN, C., LATTIMORE, T. and BRUNSKILL, E. (2017). Unifying PAC and regret: Uniform PAC bounds for episodic reinforcement learning. *Adv. Neural Inf. Process. Syst.* **30**.

[32] DANN, C., LI, L., WEI, W. and BRUNSKILL, E. (2019). Policy certificates: Towards accountable reinforcement learning. In *International Conference on Machine Learning* 1507–1516. PMLR.

[33] DOMINGUES, O. D., MÉNARD, P., KAUFMANN, E. and VALKO, M. (2021). Episodic reinforcement learning in finite MDPs: Minimax lower bounds revisited. In *Algorithmic Learning Theory*. *Proc*. *Mach*. *Learn*. *Res*. (*PMLR*) **132** 21. PMLR. MR4227335

[34] DONOHO, D. L. and LIU, R. C. (1987). Geometrizing rates of convergence. I. Technical report.

[35] DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence. II. *Ann*. *Statist*. **19** 633–667. MR1105839 https://doi.org/10.1214/aos/1176348114

[36] DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence. III. *Ann*. *Statist*. **19** 668–701. MR1105839 https://doi.org/10.1214/aos/1176348114

[37] 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.

[38] DU, S., KRISHNAMURTHY, A., JIANG, N., AGARWAL, A., DUDIK, M. and LANGFORD, J. (2019). Provably efficient RL with rich observations via latent state decoding. In *International Conference on Machine Learning* 1665–1674. PMLR.

[39] DU, S. S., KAKADE, S. M., WANG, R. and YANG, L. F. (2020). Is a good representation sufficient for sample efficient reinforcement learning? In *International Conference on Learning Representations*.

[40] DUDÍK, M., HOFMANN, K., SCHAPIRE, R. E., SLIVKINS, A. and ZOGHI, M. (2015). Contextual dueling bandits. In *Conference on Learning Theory* 563–587. PMLR.

[41] FOSTER, D., FOSTER, D. J., GOLOWICH, N. and RAKHLIN, A. (2023). On the complexity of multi-agent decision making: From learning in games to partial monitoring. In *The Thirty Sixth Annual Conference on Learning Theory* 2678–2792. PMLR.

[42] 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.

[43] FOSTER, D. J., GOLOWICH, N. and HAN, Y. (2023). Tight guarantees for interactive decision making with the decision-estimation coefficient. Preprint. Available at arXiv:2301.08215.

[44] 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. Preprint. Available at arXiv:2211.14250.

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

[46] FOSTER, D. J., RAKHLIN, A., SEKHARI, A. and SRIDHARAN, K. (2022). On the complexity of adversarial decision making. Preprint. Available at arXiv:2206.13063.

[47] FOSTER, D. J., RAKHLIN, A., SIMCHI-LEVI, D. and XU, Y. (2020). Instance-dependent complexity of contextual bandits and reinforcement learning: A disagreement-based perspective. Preprint. Available at arXiv:2010.03104.

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

[49] 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.

[50] JIN, C., ALLEN-ZHU, Z., BUBECK, S. and JORDAN, M. I. (2018). Is Q-learning provably efficient? *Adv. Neural Inf. Process*. *Syst*. **31**.

[51] JIN, C., KAKADE, S., KRISHNAMURTHY, A. and LIU, Q. (2020). Sample-efficient reinforcement learning of undercomplete POMDPs. *Adv. Neural Inf. Process*. *Syst*. **33** 18530–18539.

[52] JIN, C., KRISHNAMURTHY, A., SIMCHOWITZ, M. and YU, T. (2020). Reward-free exploration for reinforcement learning. In *International Conference on Machine Learning* 4870–4879. PMLR.

[53] 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.

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

[55] KAKADE, S., KRISHNAMURTHY, A., LOWREY, K., OHNISHI, M. and SUN, W. (2020). Information theoretic regret bounds for online nonlinear control. *Adv. Neural Inf. Process*. *Syst*. **33** 15312–15325.

[56] KAKADE, S. and LANGFORD, J. (2002). Approximately optimal approximate reinforcement learning. In *Proceedings of the Nineteenth International Conference on Machine Learning* 267–274.

[57] KAUFMANN, E., MÉNARD, P., DOMINGUES, O. D., JONSSON, A., LEURENT, E. and VALKO, M. (2021). Adaptive reward-free exploration. In *Algorithmic Learning Theory* 865–891. PMLR.

[58] KEARNS, M. and SINGH, S. (2002). Near-optimal reinforcement learning in polynomial time. *Mach*. *Learn*. **49** 209–232.

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

[60] KLEINBERG, R., SLIVKINS, A. and UPFAL, E. (2019). Bandits and experts in metric spaces. *J. ACM* **66** Art. 30, 77 pp. MR3962354 https://doi.org/10.1145/3299873

[61] KUMAR, P. R. and VARAIYA, P. (2016). *Stochastic Systems: Estimation, Identification, and Adaptive Control. Classics in Applied Mathematics* **75**. SIAM, Philadelphia, PA. MR3471643 https://doi.org/10.1137/1.9781611974263

[62] LANGFORD, J. and ZHANG, T. (2007). The epoch-greedy algorithm for multi-armed bandits with side information. *Adv. Neural Inf. Process. Syst.* **20**.

[63] 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

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

[65] LATTIMORE, T. and SZEPESVÁRI, C. (2020). Exploration by optimisation in partial monitoring. In *Conference on Learning Theory* 2488–2515. PMLR.

[66] LATTIMORE, T., SZEPESVARI, C. and WEISZ, G. (2020). Learning with good feature representations in bandits and in RL with a generative model. In *International Conference on Machine Learning* 5662–5670. PMLR.

[67] LI, G., LI, J., KABRA, A., SREBRO, N., WANG, Z. and YANG, Z. (2022). Exponential family model-based reinforcement learning via score matching. *Adv. Neural Inf. Process. Syst.* **35** 28474–28487.

[68] LIU, Q., CHUNG, A., SZEPESVÁRI, C. and JIN, C. (2022). When is partially observable reinforcement learning not scary? Preprint. Available at arXiv:2204.08967.

[69] LIU, Q., NETRAPALLI, P., SZEPESVARI, C. and JIN, C. (2023). Optimistic MLE: A generic model-based algorithm for partially observable sequential decision making. In *STOC'23—Proceedings of the 55th Annual ACM Symposium on Theory of Computing* 363–376. ACM, New York. MR4617392 https://doi.org/10.1145/3564246.3585161

[70] LIU, Q., SZEPESVÁRI, C. and JIN, C. (2022). Sample-efficient reinforcement learning of partially observable Markov games. Preprint. Available at arXiv:2206.01315.

[71] LIU, Q., YU, T., BAI, Y. and JIN, C. (2021). A sharp analysis of model-based reinforcement learning with self-play. In *International Conference on Machine Learning* 7001–7010. PMLR.

[72] MANIA, H., JORDAN, M. I. and RECHT, B. (2022). Active learning for nonlinear system identification with guarantees. *J. Mach. Learn. Res.* **23** Paper No. [32], 30 pp. MR4420757

[73] MÉNARD, P., DOMINGUES, O. D., JONSSON, A., KAUFMANN, E., LEURENT, E. and VALKO, M. (2021). Fast active learning for pure exploration in reinforcement learning. In *International Conference on Machine Learning* 7599–7608. PMLR.

[74] METE, A., SINGH, R., LIU, X. and KUMAR, P. (2021). Reward biased maximum likelihood estimation for reinforcement learning. In *Learning for Dynamics and Control* 815–827. PMLR.

[75] MISRA, D., HENAFF, M., KRISHNAMURTHY, A. and LANGFORD, J. (2020). Kinematic state abstraction and provably efficient rich-observation reinforcement learning. In *International Conference on Machine Learning* 6961–6971. PMLR.

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

[77] MODI, A., CHEN, J., KRISHNAMURTHY, A., JIANG, N. and AGARWAL, A. (2024). Model-free representation learning and exploration in low-rank MDPs. *J. Mach. Learn. Res.* **25** Paper No. [6], 76 pp. MR4723856

[78] MODI, A., JIANG, N., SINGH, S. and TEWARI, A. (2018). Markov decision processes with continuous side information. In *Algorithmic Learning Theory* 2018 597–618. PMLR. MR3857322

[79] MODI, A., JIANG, N., TEWARI, A. and SINGH, S. (2020). Sample complexity of reinforcement learning using linearly combined model ensembles. In *International Conference on Artificial Intelligence and Statistics* 2010–2020. PMLR.

[80] MODI, A. and TEWARI, A. (2020). No-regret exploration in contextual reinforcement learning. In *Conference on Uncertainty in Artificial Intelligence* 829–838. PMLR.

[81] NOVOSELLER, E., WEI, Y., SUI, Y., YUE, Y. and BURDICK, J. (2020). Dueling posterior sampling for preference-based reinforcement learning. In *Conference on Uncertainty in Artificial Intelligence* 1029–1038. PMLR.

[82] OUYANG, L., WU, J., JIANG, X., ALMEIDA, D., WAINWRIGHT, C., MISHKIN, P., ZHANG, C., AGARWAL, S., SLAMA, K. et al. (2022). Training language models to follow instructions with human feedback. *Adv. Neural Inf. Process. Syst.* **35** 27730–27744.

[83] PACCHIANO, A., SAHA, A. and LEE, J. (2021). Dueling RL: Reinforcement learning with trajectory preferences. Preprint. Available at arXiv:2111.04850.
[84] QIU, S., YE, J., WANG, Z. and YANG, Z. (2021). On reward-free RL with kernel and neural function approximations: Single-agent MDP and Markov game. In *International Conference on Machine Learning* 8737–8747. PMLR.
[85] REN, T., ZHANG, T., SZEPESVÁRI, C. and DAI, B. (2022). A free lunch from the noise: Provable and practical exploration for representation learning. In *Uncertainty in Artificial Intelligence* 1686–1696. PMLR.
[86] RUSSO, D. (2019). Worst-case regret bounds for exploration via randomized value functions. *Adv. Neural Inf. Process. Syst.* **32**.
[87] RUSSO, D. and VAN ROY, B. (2016). An information-theoretic analysis of Thompson sampling. *J. Mach. Learn. Res.* **17** Paper No. 68, 30 pp. MR3517091
[88] RUSSO, D. and VAN ROY, B. (2018). Learning to optimize via information-directed sampling. *Oper. Res.* **66** 230–252. MR3781663 https://doi.org/10.1287/opre.2017.1663
[89] RUSSO, D. and VAN ROY, B. (2013). Eluder dimension and the sample complexity of optimistic exploration. *Adv. Neural Inf. Process. Syst.* **26**.
[90] SCHOUKENS, J. and LJUNG, L. (2019). Nonlinear system identification: A user-oriented road map. *IEEE Control Syst. Mag.* **39** 28–99. MR3971097
[91] SILVER, D., HUANG, A., MADDISON, C. J., GUEZ, A., SIFRE, L., VAN DEN DRIESSCHE, G., SCHRITTWIESER, J., ANTONOGLOU, I., PANNEERSHELVAM, V. et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature* **529** 484–489.
[92] 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.
[93] THOMPSON, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* **25** 285–294.
[94] UEHARA, M., ZHANG, X. and SUN, W. (2021). Representation learning for online and offline RL in low-rank MDPs. Preprint. Available at arXiv:2110.04652.
[95] VOVK, V. (1995). A game of prediction with expert advice. In *Proceedings of the Eighth Annual Conference on Computational Learning Theory* 51–60.
[96] 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.
[97] WANG, R., DU, S. S., YANG, L. and SALAKHUTDINOV, R. R. (2020). On reward-free reinforcement learning with linear function approximation. *Adv. Neural Inf. Process. Syst.* **33** 17816–17826.
[98] 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.
[99] WANG, Y., LIU, Q., BAI, Y. and JIN, C. (2023). Breaking the curse of multiagency: Provably efficient decentralized multi-agent RL with function approximation. In *The Thirty Sixth Annual Conference on Learning Theory* 2793–2848. PMLR.
[100] WANG, Y., LIU, Q. and JIN, C. (2023). Is RLHF more difficult than standard RL? Preprint. Available at arXiv:2306.14111.
[101] WANG, Y., WANG, R., DU, S. S. and KRISHNAMURTHY, A. (2021). Optimism in reinforcement learning with generalized linear function approximation. In *International Conference on Learning Representations*.
[102] WIRTH, C., AKROUR, R., NEUMANN, G. and FÜRNKRANZ, J. (2017). A survey of preference-based reinforcement learning methods. *J. Mach. Learn. Res.* **18** Paper No. 136, 46 pp. MR3763770
[103] XIE, T., FOSTER, D. J., BAI, Y., JIANG, N. and KAKADE, S. M. (2022). The role of coverage in online reinforcement learning. Preprint. Available at arXiv:2210.04157.
[104] XU, Y., WANG, R., YANG, L., SINGH, A. and DUBRAWSKI, A. (2020). Preference-based reinforcement learning with finite-time guarantees. *Adv. Neural Inf. Process. Syst.* **33** 18784–18794.
[105] YANG, L. and WANG, M. (2019). Sample-optimal parametric q-learning using linearly additive features. In *International Conference on Machine Learning* 6995–7004. PMLR.
[106] ZANETTE, A., BRANDFONBRENER, D., BRUNSKILL, E., PIROTTA, M. and LAZARIC, A. (2020). Frequentist regret bounds for randomized least-squares value iteration. In *International Conference on Artificial Intelligence and Statistics* 1954–1964. PMLR.
[107] ZANETTE, A. and BRUNSKILL, E. (2019). Tighter problem-dependent regret bounds in reinforcement learning without domain knowledge using value function bounds. In *International Conference on Machine Learning* 7304–7312. PMLR.

[108] ZANETTE, A., LAZARIC, A., KOCHENDERFER, M. and BRUNSKILL, E. (2020). Learning near optimal policies with low inherent Bellman error. In *International Conference on Machine Learning* 10978–10989. PMLR.

[109] ZANETTE, A., LAZARIC, A., KOCHENDERFER, M. J. and BRUNSKILL, E. (2020). Provably efficient reward-agnostic navigation with linear value iteration. *Adv. Neural Inf. Process. Syst.* **33** 11756–11766.

[110] ZHANG, T. (2022). Feel-good Thompson sampling for contextual bandits and reinforcement learning. *SIAM J. Math. Data Sci.* **4** 834–857. MR4443669 https://doi.org/10.1137/21M140924X

[111] ZHANG, W., ZHOU, D. and GU, Q. (2021). Reward-free model-based reinforcement learning with linear function approximation. *Adv. Neural Inf. Process. Syst.* **34** 1582–1593.

[112] ZHANG, X., MA, Y. and SINGLA, A. (2020). Task-agnostic exploration in reinforcement learning. *Adv. Neural Inf. Process. Syst.* **33** 11734–11743.

[113] ZHANG, Z., DU, S. and JI, X. (2021). Near optimal reward-free reinforcement learning. In *International Conference on Machine Learning* 12402–12412. PMLR.

[114] ZHANG, Z., ZHOU, Y. and JI, X. (2020). Almost optimal model-free reinforcement learningvia reference-advantage decomposition. *Adv. Neural Inf. Process. Syst.* **33** 15198–15207.

[115] ZHONG, H., XIONG, W., ZHENG, S., WANG, L., WANG, Z., YANG, Z. and ZHANG, T. (2022). GEC: A unified framework for interactive decision making in MDP, POMDP, and beyond. Preprint. Available at arXiv:2211.01962.

[116] ZHOU, D., GU, Q. and SZEPESVARI, C. (2021). Nearly minimax optimal reinforcement learning for linear mixture Markov decision processes. In *Conference on Learning Theory* 4532–4576. PMLR.

[117] ZHU, B., JIAO, J. and JORDAN, M. I. (2023). Principled reinforcement learning with human feedback from pairwise or K-wise comparisons. Preprint. Available at arXiv:2301.11270.