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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

Vol. 52, No. 2—April 2024

THE ANNALS OF STATISTICS Vol. 52, No. 2, pp. 415–867 April 2024

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

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

The Annals of Statistics **[ISSN 0090-5364 (print); ISSN 2168-8966 (online)],** Volume 52, Number 2, April 2024. Published bimonthly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, OH 44094, USA. Periodicals postage paid at Cleveland, Ohio, and at additional mailing offices.

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

The Annals of Statistics
2024, Vol. 52, No. 2, 415–440
https://doi.org/10.1214/24-AOS2352

HIGH-DIMENSIONAL INFERENCE FOR DYNAMIC TREATMENT EFFECTS

BY JELENA BRADIC[1,a], WEIJIE JI[2,b] AND YUQIAN ZHANG[3,c]

[1]*Department of Mathematics and Halicioglu Data Science Institute, University of California San Diego,* [a]*jbradic@ucsd.edu*

[2]*School of Statistics and Management, Shanghai University of Finance and Economics,* [b]*w6ji@ucsd.edu*

[3]*Institute of Statistics and Big Data, Renmin University of China,* [c]*yuqianzhang@ruc.edu.cn*

Estimating dynamic treatment effects is a crucial endeavor in causal inference, particularly when confronted with high-dimensional confounders. Doubly robust (DR) approaches have emerged as promising tools for estimating treatment effects due to their flexibility. However, we showcase that the traditional DR approaches that only focus on the DR representation of the expected outcomes may fall short of delivering optimal results. In this paper, we propose a novel DR representation for intermediate conditional outcome models that leads to superior robustness guarantees. The proposed method achieves consistency even with high-dimensional confounders, as long as at least one nuisance function is appropriately parametrized for each exposure time and treatment path. Our results represent a significant step forward as they provide faster convergence rates and new robustness guarantees. The key to achieving these results lies in utilizing DR representations for intermediate conditional outcome models, which offer superior inferential performance while requiring weaker assumptions. Lastly, we examine finite sample behavior through simulations and a real data application.

REFERENCES

AVAGYAN, V. and VANSTEELANDT, S. (2021). High-dimensional inference for the average treatment effect under model misspecification using penalized bias-reduced double-robust estimation. *Biostat*. *Epidemiol*. 1–18.

BABINO, L., ROTNITZKY, A. and ROBINS, J. (2019). Multiple robust estimation of marginal structural mean models for unconstrained outcomes. *Biometrics* **75** 90–99. MR3953710 https://doi.org/10.1111/biom.12924

BANG, H. and ROBINS, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61** 962–973. MR2216189 https://doi.org/10.1111/j.1541-0420.2005.00377.x

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

BODORY, H., HUBER, M. and LAFFÉRS, L. (2022). Evaluating (weighted) dynamic treatment effects by double machine learning. *Econom*. *J*. **25** 628–648. MR4530103 https://doi.org/10.1093/ectj/utac018

BRADIC, J., JI, W. and ZHANG, Y. (2024). Supplement to “High-dimensional inference for dynamic treatment effects.” https://doi.org/10.1214/24-AOS2352SUPP

CHEN, X. and FLORES, C. A. (2015). Bounds on treatment effects in the presence of sample selection and noncompliance: The wage effects of Job Corps. *J*. *Bus*. *Econom*. *Statist*. **33** 523–540. MR3416598 https://doi.org/10.1080/07350015.2014.975229

CHERNOZHUKOV, V., CHETVERIKOV, D., DEMIRER, M., DUFLO, E., HANSEN, C., NEWEY, W. and ROBINS, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econom*. *J*. **21** C1–C68. MR3769544 https://doi.org/10.1111/ectj.12097

D’AMOUR, A., DING, P., FELLER, A., LEI, L. and SEKHON, J. (2021). Overlap in observational studies with high-dimensional covariates. *J*. *Econometrics* **221** 644–654. MR4215042 https://doi.org/10.1016/j.jeconom.2019.10.014

DANIEL, R. M., COUSENS, S. N., DE STAVOLA, B. L., KENWARD, M. G. and STERNE, J. A. C. (2013). Methods for dealing with time-dependent confounding. *Stat*. *Med*. **32** 1584–1618. MR3060620 https://doi.org/10.1002/sim.5686

DÍAZ, I., WILLIAMS, N., HOFFMAN, K. L. and SCHENCK, E. J. (2023). Nonparametric causal effects based on longitudinal modified treatment policies. *J*. *Amer*. *Statist*. *Assoc*. **118** 846–857. MR4595460 https://doi.org/10.1080/01621459.2021.1955691

MSC2020 subject classifications. Primary 62C99, 62G35; secondary 62J07.

Key words and phrases. Double-robustness, imputed Lasso, model misspecification, missing data.

DUKES, O., AVAGYAN, V. and VANSTEELANDT, S. (2020). Doubly robust tests of exposure effects under high-dimensional confounding. *Biometrics* **76** 1190–1200. MR4186835 https://doi.org/10.1111/biom.13231

DUKES, O. and VANSTEELANDT, S. (2021). Inference for treatment effect parameters in potentially misspecified high-dimensional models. *Biometrika* **108** 321–334. MR4259134 https://doi.org/10.1093/biomet/asaa071

FARRELL, M. H. (2015). Robust inference on average treatment effects with possibly more covariates than observations. *J. Econometrics* **189** 1–23. MR3397349 https://doi.org/10.1016/j.jeconom.2015.06.017

FLORES, C. A., FLORES-LAGUNES, A., GONZALEZ, A. and NEUMANN, T. C. (2012). Estimating the effects of length of exposure to instruction in a training program: The case of job corps. *Rev. Econ. Stat.* **94** 153–171.

HERNÁN, M. A., BRUMBACK, B. and ROBINS, J. M. (2001). Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *J. Amer. Statist. Assoc.* **96** 440–448. MR1939347 https://doi.org/10.1198/016214501753168154

HERNÁN, M. A., SAUER, B. C., HERNÁNDEZ-DÍAZ, S., PLATT, R. and SHRIER, I. (2016). Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J. Clin. Epidemiol.* **79** 70–75.

HUBER, M., HSU, Y.-C., LEE, Y.-Y. and LETTRY, L. (2020). Direct and indirect effects of continuous treatments based on generalized propensity score weighting. *J. Appl. Econometrics* **35** 814–840. MR4186786 https://doi.org/10.1002/jae.2765

KALLUS, N. and SANTACATTERINA, M. (2021). Optimal balancing of time-dependent confounders for marginal structural models. *J. Causal Inference* **9** 345–369. MR4362148 https://doi.org/10.1515/jci-2020-0033

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

LEE, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Rev. Econ. Stud.* **76** 1071–1102.

LEWIS, G. and SYRGKANIS, V. (2021). Double/debiased machine learning for dynamic treatment effects. *Adv. Neural Inf. Process. Syst.* **34** 22695–22707.

LUEDTKE, A. R., SOFRYGIN, O., VAN DER LAAN, M. J. and CARONE, M. (2017). Sequential double robustness in right-censored longitudinal models. arXiv preprint. Available at arXiv:1705.02459.

MOLINA, J., ROTNITZKY, A., SUED, M. and ROBINS, J. M. (2017). Multiple robustness in factorized likelihood models. *Biometrika* **104** 561–581. MR3694583 https://doi.org/10.1093/biomet/asx027

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

MURPHY, S. A., VAN DER LAAN, M. J., ROBINS, J. M. and CONDUCT PROBLEMS PREVENTION RESEARCH GROUP (2001). Marginal mean models for dynamic regimes. *J. Amer. Statist. Assoc.* **96** 1410–1423. MR1946586 https://doi.org/10.1198/016214501753382327

NEGAHBAN, S. N., RAVIKUMAR, P., WAINWRIGHT, M. J. and YU, B. (2012). A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers. *Statist. Sci.* **27** 538–557. MR3025133 https://doi.org/10.1214/12-STS400

NING, Y., SIDA, P. and IMAI, K. (2020). Robust estimation of causal effects via a high-dimensional covariate balancing propensity score. *Biometrika* **107** 533–554. MR4138975 https://doi.org/10.1093/biomet/asaa020

ORELLANA, L., ROTNITZKY, A. and ROBINS, J. M. (2010). Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, Part I: Main content. *Int. J. Biostat.* **6** Art. 8, 49. MR2602551 https://doi.org/10.2202/1557-4679.1200

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

ROBINS, J. M. (1987). Addendum to: "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** (1986), no. 9–12, 1393–1512; MR0877758 (87m:92078)]. *Comput. Math. Appl.* **14** 923–945. MR0922792 https://doi.org/10.1016/0898-1221(87)90238-0

ROBINS, J. M. (1997). Causal inference from complex longitudinal data. In *Latent Variable Modeling and Applications to Causality* (*Los Angeles*, *CA*, 1994). *Lect. Notes Stat.* **120** 69–117. Springer, New York. MR1601279 https://doi.org/10.1007/978-1-4612-1842-5_4

ROBINS, J. M. (2000a). Marginal structural models versus structural nested models as tools for causal inference. In *Statistical Models in Epidemiology*, *the Environment*, *and Clinical Trials* (*Minneapolis*, *MN*, 1997). *IMA Vol. Math. Appl.* **116** 95–133. Springer, New York. MR1731682 https://doi.org/10.1007/978-1-4612-1284-3_2

ROBINS, J. M. (2000b). Robust estimation in sequentially ignorable missing data and causal inference models. In *Proceedings of the American Statistical Association* **1999** 6–10, Indianapolis, IN.

ROBINS, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium in Biostatistics*. *Lect. Notes Stat.* **179** 189–326. Springer, New York. MR2129402 https://doi.org/10.1007/978-1-4419-9076-1_11

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

ROTNITZKY, A., ROBINS, J. and BABINO, L. (2017). On the multiply robust estimation of the mean of the g-functional. arXiv preprint. Available at arXiv:1705.08582.

SCHOCHET, P., BELLOTTI, J., RUO-JIAO, C., GLAZERMAN, S., GRADY, A., GRITZ, M., MCCONNELL, S., JOHNSON, T. and BURGHARDT, J. (2003). *National Job Corps Study*: *Data Documentation and Public Use Files*, *Vols*. *I–IV*. Mathematica Policy Research, Inc., Washington, DC.

SCHOCHET, P. Z. (2001). National Job Corps Study: The impacts of Job Corps on participants' employment and related outcomes. US Department of Labor, Employment and Training Administration. Office of Policy and Research.

SCHOCHET, P. Z., BURGHARDT, J. and MCCONNELL, S. (2008). Does job corps work? Impact findings from the national job corps study. *Amer*. *Econ*. *Rev*. **98** 1864–1886.

SINGH, R., XU, L. and GRETTON, A. (2021). Kernel methods for multistage causal inference: Mediation analysis and dynamic treatment effects. arXiv preprint. Available at arXiv:2111.03950.

SMUCLER, E., ROTNITZKY, A. and ROBINS, J. M. (2019). A unifying approach for doubly-robust l_1 regularized estimation of causal contrasts. arXiv preprint. Available at arXiv:1904.03737.

TAN, Z. (2020). Model-assisted inference for treatment effects using regularized calibrated estimation with high-dimensional data. *Ann*. *Statist*. **48** 811–837. MR4102677 https://doi.org/10.1214/19-AOS1824

VAN DER LAAN, M. J. and GRUBER, S. (2012). Targeted minimum loss based estimation of causal effects of multiple time point interventions. *Int*. *J*. *Biostat*. **8** Art. 9, 41. MR2923282 https://doi.org/10.1515/1557-4679.1370

VIVIANO, D. and BRADIC, J. (2021). Dynamic covariate balancing: Estimating treatment effects over time. arXiv preprint. Available at arXiv:2103.01280.

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

YIU, S. and SU, L. (2018). Covariate association eliminating weights: A unified weighting framework for causal effect estimation. *Biometrika* **105** 709–722. MR3842894 https://doi.org/10.1093/biomet/asy015

YU, Z. and VAN DER LAAN, M. (2006). Double robust estimation in longitudinal marginal structural models. *J*. *Statist*. *Plann*. *Inference* **136** 1061–1089. MR2181989 https://doi.org/10.1016/j.jspi.2004.08.011

ZHANG, J. L., RUBIN, D. B. and MEALLI, F. (2008). Evaluating the effects of job training programs on wages through principal stratification. In *Modelling and Evaluating Treatment Effects in Econometrics*. *Adv*. *Econom*. **21** 117–145. Elsevier/JAI, Amsterdam. MR2544066 https://doi.org/10.1016/S0731-9053(07)00005-9

ZHU, W., ZENG, D. and SONG, R. (2019). Proper inference for value function in high-dimensional Q-learning for dynamic treatment regimes. *J*. *Amer*. *Statist*. *Assoc*. **114** 1404–1417. MR4011788 https://doi.org/10.1080/01621459.2018.1506341

The Annals of Statistics
2024, Vol. 52, No. 2, 441–465
https://doi.org/10.1214/24-AOS2353

THE CURSE OF OVERPARAMETRIZATION IN ADVERSARIAL TRAINING: PRECISE ANALYSIS OF ROBUST GENERALIZATION FOR RANDOM FEATURES REGRESSION

BY HAMED HASSANI[1,a] AND ADEL JAVANMARD[2,b]

[1]*Department of Electrical and Systems Engineering, University of Pennsylvania,* [a]*hassani@seas.upenn.edu*
[2]*Data Sciences and Operations Department, University of Southern California,* [b]*ajavanma@usc.edu*

Successful deep learning models often involve training neural network architectures that contain more parameters than the number of training samples. Such overparametrized models have recently been extensively studied, and the virtues of overparametrization have been established from both the statistical perspective, via the double-descent phenomenon, and the computational perspective via the structural properties of the optimization landscape. Despite this success, it is also well known that these models are highly vulnerable to small adversarial perturbations in their inputs. Even when adversarially trained, their performance on perturbed inputs (robust generalization) is considerably worse than their best attainable performance on benign inputs (standard generalization). It is thus imperative to understand how overparametrization fundamentally affects robustness.

In this paper, we will provide a precise characterization of the role of overparametrization on robustness by focusing on random features regression models (two-layer neural networks with random first layer weights). We consider a regime where the sample size, the input dimension and the number of parameters grow proportionally, and derive an asymptotically exact formula for the robust generalization error when the model is adversarially trained. Our developed theory reveals the nontrivial effect of overparametrization on robustness and indicates that high overparametrization can hurt robust generalization.

REFERENCES

[1] ABBASI, E., SALEHI, F. and HASSIBI, B. (2019). Universality in learning from linear measurements. *Adv. Neural Inf. Process. Syst.* **32** 12372–12382.

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

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

[4] BELKIN, M., MA, S. and MANDAL, S. (2018). To understand deep learning we need to understand kernel learning. In *International Conference on Machine Learning* 541–549.

[5] BIGGIO, B., CORONA, I., MAIORCA, D., NELSON, B., ŠRNDIĆ, N., LASKOV, P., GIACINTO, G. and ROLI, F. (2013). Evasion attacks against machine learning at test time. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* 387–402. Springer, Berlin.

[6] BUBECK, S., LI, Y. and NAGARAJ, D. M. (2021). A law of robustness for two-layers neural networks. In *Conference on Learning Theory* 804–820. PMLR.

[7] BUBECK, S. and SELLKE, M. (2023). A universal law of robustness via isoperimetry. *J. ACM* **70** 1–18. MR4571363

[8] CARMON, Y., RAGHUNATHAN, A., SCHMIDT, L., LIANG, P. and DUCHI, J. C. (2019). Unlabeled data improves adversarial robustness. arXiv preprint. Available at arXiv:1905.13736.

[9] CHENG, X. and SINGER, A. (2013). The spectrum of random inner-product kernel matrices. *Random Matrices Theory Appl.* **2** 1350010, 47. MR3149440 https://doi.org/10.1142/S201032631350010X

MSC2020 subject classifications. Primary 62E20, 62F12; secondary 62F35.
Key words and phrases. Adversarial training, random features models, precise high-dimensional asymptotics, Gaussian equivalence property.

[10] COHEN, J., ROSENFELD, E. and KOLTER, Z. (2019). Certified adversarial robustness via randomized smoothing. In *International Conference on Machine Learning* 1310–1320. PMLR.

[11] DANIELY, A. (2017). SGD learns the conjugate kernel class of the network. In *Advances in Neural Information Processing Systems* 2422–2430.

[12] DANIELY, A., FROSTIG, R. and SINGER, Y. (2016). Toward deeper understanding of neural networks: The power of initialization and a dual view on expressivity. In *Proceedings of the* 30*th International Conference on Neural Information Processing Systems* 2261–2269.

[13] DENG, Z., KAMMOUN, A. and THRAMPOULIDIS, C. (2022). A model of double descent for high-dimensional binary linear classification. *Inf. Inference* **11** 435–495. MR4474343 https://doi.org/10.1093/imaiai/iaab002

[14] DENG, Z., ZHANG, L., GHORBANI, A. and ZOU, J. (2021). Improving adversarial robustness via unlabeled out-of-domain data. In *International Conference on Artificial Intelligence and Statistics* 2845–2853. PMLR.

[15] DEPERSIN, J. and LECUÉ, G. (2023). On the robustness to adversarial corruption and to heavy-tailed data of the Stahel–Donoho median of means. *Inf. Inference* **12** 814–850. MR4565752 https://doi.org/10.1093/imaiai/iaac026

[16] DHIFALLAH, O. and LU, Y. M. (2020). A precise performance analysis of learning with random features. arXiv preprint. Available at arXiv:2008.11904.

[17] DHIFALLAH, O., THRAMPOULIDIS, C. and LU, Y. M. (2018). Phase retrieval via polytope optimization: Geometry, phase transitions, and new algorithms. arXiv preprint. Available at arXiv:1805.09555.

[18] DOBRIBAN, E., HASSANI, H., HONG, D. and ROBEY, A. (2023). Provable tradeoffs in adversarially robust classification. *IEEE Trans. Inf. Theory* **69** 7793–7822. MR4692644 https://doi.org/10.1109/tit.2022.3205449

[19] DOHMATOB, E. (2021). Fundamental tradeoffs between memorization and robustness in random features and neural tangent regimes. arXiv preprint. Available at arXiv:2106.02630.

[20] DONHAUSER, K., TIFREA, A., AERNI, M., HECKEL, R. and YANG, F. (2021). Interpolation can hurt robust generalization even when there is no noise. *Adv. Neural Inf. Process. Syst.* **34**.

[21] DONOHO, D. L., MALEKI, A. and MONTANARI, A. (209). Message-passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci.* **106** 18914–18919.

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

[23] GERACE, F., LOUREIRO, B., KRZAKALA, F., MÉZARD, M. and ZDEBOROVÁ, L. (2020). Generalisation error in learning with random features and the hidden manifold model. In *International Conference on Machine Learning* 3452–3462. PMLR.

[24] GILMER, J., METZ, L., FAGHRI, F., SCHOENHOLZ, S. S., RAGHU, M., WATTENBERG, M. and GOODFELLOW, I. (2018). Adversarial spheres. arXiv preprint. Available at arXiv:1801.02774.

[25] GOLDT, S., LOUREIRO, B., REEVES, G., KRZAKALA, F., MÉZARD, M. and ZDEBOROVÁ, L. (2020). The Gaussian equivalence of generative models for learning with shallow neural networks. arXiv preprint. Available at arXiv:2006.14709.

[26] GOLDT, S., MÉZARD, M., KRZAKALA, F. and ZDEBOROVÁ, L. (2020). Modeling the influence of data structure on learning in neural networks: The hidden manifold model. *Phys. Rev. X* **10** 041044.

[27] GOODFELLOW, I. J., SHLENS, J. and SZEGEDY, C. (2015). Explaining and harnessing adversarial examples. In 3*rd International Conference on Learning Representations*, *ICLR* 2015, *San Diego*, *CA*, *USA*, *May* 7–9, 2015, *Conference Track Proceedings*.

[28] GORDON, Y. (1988). On Milman's inequality and random subspaces which escape through a mesh in $\mathbf{R}^n$. In *Geometric Aspects of Functional Analysis* (1986/87). *Lecture Notes in Math.* **1317** 84–106. Springer, Berlin. MR0950977 https://doi.org/10.1007/BFb0081737

[29] GOWAL, S., QIN, C., UESATO, J., MANN, T. and KOHLI, P. (2020). Uncovering the limits of adversarial training against norm-bounded adversarial examples. arXiv preprint. Available at arXiv:2010.03593.

[30] GUNASEKAR, S., LEE, J. D., SOUDRY, D. and SREBRO, N. (2018). Implicit bias of gradient descent on linear convolutional networks. In *Advances in Neural Information Processing Systems* (S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi and R. Garnett, eds.) **31** 9461–9471. Curran Associates, Red Hook, NY.

[31] HASSANI, H. and JAVANMARD, A. (2024). Supplement to "The curse of overparametrization in adversarial training: Precise analysis of robust generalization for random features regression." https://doi.org/10.1214/24-AOS2353SUPP

[32] HASTIE, T., MONTANARI, A., ROSSET, S. and TIBSHIRANI, R. J. (2022). Surprises in high-dimensional ridgeless least squares interpolation. *Ann. Statist.* **50** 949–986. MR4404925 https://doi.org/10.1214/21-aos2133

[33] HU, H. and LU, Y. M. (2019). Asymptotics and optimal designs of SLOPE for sparse linear regression. In 2019 *IEEE International Symposium on Information Theory* (*ISIT*) **68** 375–379. IEEE, New York City, U.S.

[34] HU, H. and LU, Y. M. (2023). Universality laws for high-dimensional learning with random features. *IEEE Trans*. *Inf*. *Theory* **69** 1932–1964. MR4564688

[35] HUANG, S.-T. and LEDERER, J. (2023). DeepMoM: Robust deep learning with median-of-means. *J*. *Comput*. *Graph*. *Statist*. **32** 181–195. MR4552946 https://doi.org/10.1080/10618600.2022.2090947

[36] JACOT, A., GABRIEL, F. and HONGLER, C. (2018). Neural tangent kernel: Convergence and generalization in neural networks. In *Advances in Neural Information Processing Systems* (S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi and R. Garnett, eds.) **31** 8571–8580. Curran Associates, Red Hook, NY.

[37] JALAL, A., ILYAS, A., DASKALAKIS, C. and DIMAKIS, A. G. (2017). The robust manifold defense: Adversarial training using generative models. arXiv preprint. Available at arXiv:1712.09196.

[38] JAVANMARD, A., MONDELLI, M. and MONTANARI, A. (2020). Analysis of a two-layer neural network via displacement convexity. *Ann*. *Statist*. **48** 3619–3642. MR4185822 https://doi.org/10.1214/20-AOS1945

[39] JAVANMARD, A. and SOLTANOLKOTABI, M. (2022). Precise statistical analysis of classification accuracies for adversarial training. *Ann*. *Statist*. **50** 2127–2156. MR4474485 https://doi.org/10.1214/22-aos2180

[40] JAVANMARD, A., SOLTANOLKOTABI, M. and HASSANI, H. (2020). Precise tradeoffs in adversarial training for linear regression. In *Conference on Learning Theory* 2034–2078. PMLR.

[41] KURAKIN, A., GOODFELLOW, I. and BENGIO, S. (2016). Adversarial machine learning at scale. arXiv preprint. Available at arXiv:1611.01236.

[42] LAI, L. and BAYRAKTAR, E. (2020). On the adversarial robustness of robust estimators. *IEEE Trans*. *Inf*. *Theory* **66** 5097–5109. MR4130664 https://doi.org/10.1109/TIT.2020.2985966

[43] LI, Y. and LIANG, Y. (2018). Learning overparameterized neural networks via stochastic gradient descent on structured data. In *NeurIPS*.

[44] LIANG, T. and SUR, P. (2022). A precise high-dimensional asymptotic theory for boosting and minimum-$\ell 1$-norm interpolated classifiers. *Ann*. *Statist*. **50** 1669–1695. MR4441136 https://doi.org/10.1214/22-aos2170

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

[46] LOUART, C., LIAO, Z. and COUILLET, R. (2018). A random matrix approach to neural networks. *Ann*. *Appl*. *Probab*. **28** 1190–1248. MR3784498 https://doi.org/10.1214/17-AAP1328

[47] MADRY, A., MAKELOV, A., SCHMIDT, L., TSIPRAS, D. and VLADU, A. (2018). Towards deep learning models resistant to adversarial attacks. In 6*th International Conference on Learning Representations*, *ICLR* 2018, *Vancouver*, *BC*, *Canada*, *April* 30–*May* 3, 2018, *Conference Track Proceedings*.

[48] MAHLOUJIFAR, S., DIOCHNOS, D. I. and MAHMOODY, M. (2019). The curse of concentration in robust learning: Evasion and poisoning attacks from concentration of measure. In *Proceedings of the AAAI Conference on Artificial Intelligence* **33** 4536–4543.

[49] MAHLOUJIFAR, S. and MAHMOODY, M. (2019). Can adversarially robust learning leverage computational hardness? In *Algorithmic Learning Theory* 2019. *Proc*. *Mach*. *Learn*. *Res*. (*PMLR*) **98** 581–609. PMLR. MR3932860

[50] MEI, S. and MONTANARI, A. (2022). The generalization error of random features regression: Precise asymptotics and the double descent curve. *Comm*. *Pure Appl*. *Math*. **75** 667–766. MR4400901 https://doi.org/10.1002/cpa.22008

[51] MEI, S., MONTANARI, A. and NGUYEN, P.-M. (2018). A mean field view of the landscape of two-layer neural networks. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **115** E7665–E7671. MR3845070 https://doi.org/10.1073/pnas.1806579115

[52] MIN, Y., CHEN, L. and KARBASI, A. (2021). The curious case of adversarially robust models: More data can help, double descend, or hurt generalization. In *Uncertainty in Artificial Intelligence* 129–139. PMLR.

[53] MONTANARI, A., RUAN, F., SOHN, Y. and YAN, J. (2019). The generalization error of max-margin linear classifiers: High-dimensional asymptotics in the overparametrized regime. arXiv preprint. Available at arXiv:1911.01544.

[54] MONTANARI, A., ZHONG, Y. and ZHOU, K. (2021). Tractability from overparametrization: The example of the negative perceptron. arXiv preprint. Available at arXiv:2110.15824.

[55] NAJAFI, A., MAEDA, S.-I., KOYAMA, M. and MIYATO, T. (2019). Robustness to adversarial perturbations in learning from incomplete data. arXiv preprint. Available at arXiv:1905.13021.

[56] PENNINGTON, J. and WORAH, P. (2019). Nonlinear random matrix theory for deep learning. *J*. *Stat*. *Mech*. *Theory Exp*. **2019** 124005. MR4063603 https://doi.org/10.1088/1742-5468/ab3bc3

[57] RAGHUNATHAN, A., XIE, S. M., YANG, F., DUCHI, J. C. and LIANG, P. (2019). Adversarial training can hurt generalization. arXiv preprint. Available at arXiv:1906.06032.

[58] RAHIMI, A. and RECHT, B. (2007). Random features for large-scale kernel machines. *Adv. Neural Inf. Process. Syst.* **20** 1177–1184.

[59] RAHIMI, A. and RECHT, B. (2008). Uniform approximation of functions with random bases. In 2008 46*th Annual Allerton Conference on Communication*, *Control*, *and Computing* 555–561. IEEE Press, New York.

[60] REBUFFI, S.-A., GOWAL, S., CALIAN, D. A., STIMBERG, F., WILES, O. and MANN, T. A. (2021). Data augmentation can improve robustness. *Adv. Neural Inf. Process. Syst.* **34**.

[61] RICHARDSON, T. and URBANKE, R. (2008). *Modern Coding Theory*. Cambridge Univ. Press, Cambridge. MR2494807 https://doi.org/10.1017/CBO9780511791338

[62] SALEHI, F., ABBASI, E. and HASSIBI, B. (2019). The impact of regularization on high-dimensional logistic regression. In *Advances in Neural Information Processing Systems* (H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox and R. Garnett, eds.). **32**. Curran Associates, Red Hook, NY.

[63] SEHWAG, V., MAHLOUJIFAR, S., HANDINA, T., DAI, S., XIANG, C., CHIANG, M. and MITTAL, P. (2021). Improving adversarial robustness using proxy distributions. arXiv preprint. Available at arXiv:2104.09425.

[64] SHAFAHI, A., HUANG, W. R., STUDER, C., FEIZI, S. and GOLDSTEIN, T. (2019). Are adversarial examples inevitable? In 7*th International Conference on Learning Representations*, *ICLR* 2019, *New Orleans*, *LA*, *USA*, *May* 6–9, 2019.

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

[66] SOUDRY, D., HOFFER, E., NACSON, M. S., GUNASEKAR, S. and SREBRO, N. (2018). The implicit bias of gradient descent on separable data. *J. Mach. Learn. Res.* **19** 2822–2878. MR3899772

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

[68] STOJNIC, M. (2013). Meshes that trap random subspaces. arXiv preprint. Available at arXiv:1304.0003.

[69] STOJNIC, M. (2013). Upper-bounding ℓ_1-optimization weak thresholds. arXiv preprint. Available at arXiv:1303.7289.

[70] SU, D., ZHANG, H., CHEN, H., YI, J., CHEN, P.-Y. and GAO, Y. (2018). Is robustness the cost of accuracy?—a comprehensive study on the robustness of 18 deep image classification models. In *Proceedings of the European Conference on Computer Vision* (*ECCV*) 631–648.

[71] SZEGEDY, C., ZAREMBA, W., SUTSKEVER, I., BRUNA, J., ERHAN, D., GOODFELLOW, I. J. and FERGUS, R. (2014). Intriguing properties of neural networks. ICLR. Available at arXiv:1312.6199.

[72] TAHERI, H., PEDARSANI, R. and THRAMPOULIDIS, C. (2020). Asymptotic behavior of adversarial training in binary classification. arXiv preprint. Available at arXiv:2010.13275.

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

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

[75] THRAMPOULIDIS, C., OYMAK, S. and SOLTANOLKOTABI, M. (2020). Theoretical insights into multiclass classification: A high-dimensional asymptotic view. arXiv preprint. Available at arXiv:2011.07729.

[76] TSIPRAS, D., SANTURKAR, S., ENGSTROM, L., TURNER, A. and MADRY, A. (2019). Robustness may be at odds with accuracy. In *International Conference on Learning Representations*.

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

[78] WONG, E. and KOLTER, J. Z. (2018). Provable defenses against adversarial examples via the convex outer adversarial polytope. In *Proceedings of the* 35*th International Conference on Machine Learning*, *ICML* 2018, *Stockholmsmässan*, *Stockholm*, *Sweden*, *July* 10–15, 2018 5283–5292.

[79] WU, B., CHEN, J., CAI, D., HE, X. and GU, Q. (2021). Do wider neural networks really help adversarial robustness? *Adv. Neural Inf. Process. Syst.* **34**.

[80] ZHAI, R., CAI, T., HE, D., DAN, C., HE, K., HOPCROFT, J. and WANG, L. (2019). Adversarially robust generalization just requires more unlabeled data. arXiv preprint. Available at arXiv:1906.00555.

[81] ZHANG, H., WU, Y. and HUANG, H. (2022). How many data are needed for robust learning? arXiv preprint. Available at arXiv:2202.11592.

[82] ZHANG, H., YU, Y., JIAO, J., XING, E. P., GHAOUI, L. E. and JORDAN, M. I. (2019). Theoretically principled trade-off between robustness and accuracy. In *Proceedings of the* 36*th International Conference on Machine Learning*, *ICML* 2019, 9–15 *June* 2019, *Long Beach*, *California*, *USA* 7472–7482.

The Annals of Statistics
2024, Vol. 52, No. 2, 466–490
https://doi.org/10.1214/24-AOS2354

CONVERGENCE RATES OF OBLIQUE REGRESSION TREES FOR FLEXIBLE FUNCTION LIBRARIES

BY MATIAS D. CATTANEO[a], RAJITA CHANDAK[b] AND JASON M. KLUSOWSKI[c]

Department of Operations Research and Financial Engineering, Princeton University,
[a]*cattaneo@princeton.edu,* [b]*rchandak@princeton.edu,* [c]*jason.klusowski@princeton.edu*

We develop a theoretical framework for the analysis of oblique decision trees, where the splits at each decision node occur at linear combinations of the covariates (as opposed to conventional tree constructions that force axis-aligned splits involving only a single covariate). While this methodology has garnered significant attention from the computer science and optimization communities since the mid-80s, the advantages they offer over their axis-aligned counterparts remain only empirically justified, and explanations for their success are largely based on heuristics. Filling this long-standing gap between theory and practice, we show that oblique regression trees (constructed by recursively minimizing squared error) satisfy a type of oracle inequality and can adapt to a rich library of regression models consisting of linear combinations of ridge functions and their limit points. This provides a quantitative baseline to compare and contrast decision trees with other less interpretable methods, such as projection pursuit regression and neural networks, which target similar model forms. Contrary to popular belief, one needs not always trade-off interpretability with accuracy. Specifically, we show that, under suitable conditions, oblique decision trees achieve similar predictive accuracy as neural networks for the same library of regression models. To address the combinatorial complexity of finding the optimal splitting hyperplane at each decision node, our proposed theoretical framework can accommodate many existing computational tools in the literature. Our results rely on (arguably surprising) connections between recursive adaptive partitioning and sequential greedy approximation algorithms for convex optimization problems (e.g., orthogonal greedy algorithms), which may be of independent theoretical interest. Using our theory and methods, we also study oblique random forests.

REFERENCES

[1] BARRON, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans. Inf. Theory* **39** 930–945. MR1237720 https://doi.org/10.1109/18.256500

[2] BARRON, A. R. (1994). Approximation and estimation bounds for artificial neural networks. *Mach. Learn.* **14** 115–133.

[3] 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

[4] BENNETT, K. P. (1994). Global tree optimization: A non-greedy decision tree algorithm. *J. Comput. Sci. Stat.* 156–156.

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

[6] BERTSIMAS, D. and DUNN, J. (2019). *Machine Learning Under a Modern Optimization Lens*. Dynamic Ideas LLC.

[7] BERTSIMAS, D., DUNN, J. and WANG, Y. (2021). Near-optimal nonlinear regression trees. *Oper. Res. Lett.* **49** 201–206. MR4204496 https://doi.org/10.1016/j.orl.2021.01.002

[8] BERTSIMAS, D., MAZUMDER, R. and SOBIESK, M. (2018). Optimal classification and regression trees with hyperplanes are as powerful as classification and regression neural networks. Unpublished manuscript.

MSC2020 subject classifications. Primary 62G08; secondary 62L12.
Key words and phrases. Decision trees, neural networks, projection pursuit regression, CART, random forest.

[9] BERTSIMAS, D. and STELLATO, B. (2021). The voice of optimization. *Mach. Learn.* **110** 249–277. MR4207500 https://doi.org/10.1007/s10994-020-05893-5

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

[11] 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

[12] BRODLEY, C. E. and UTGOFF, P. E. (1995). Multivariate decision trees. *Mach. Learn.* **19** 45–77.

[13] BUCILUUNDEFINED, C., CARUANA, R. and NICULESCU-MIZIL, A. (2006). Model compression. In *Proceedings of the* 12*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. *KDD'*06 535–541. Association for Computing Machinery, New York, NY, USA.

[14] CATTANEO, M. D., CHANDAK, R. and KLUSOWSKI, J. M. (2024). Supplement to "Convergence rates of oblique regression trees for flexible function libraries." https://doi.org/10.1214/24-AOS2354SUPP

[15] CATTANEO, M. D., FARRELL, M. H. and FENG, Y. (2020). Large sample properties of partitioning-based series estimators. *Ann. Statist.* **48** 1718–1741. MR4124341 https://doi.org/10.1214/19-AOS1865

[16] 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

[17] DEVORE, R., NOWAK, R. D., PARHI, R. and SIEGEL, J. W. (2023). Weighted variation spaces and approximation by shallow ReLU networks. ArXiv preprint. Available at arXiv:2307.15772.

[18] DUNN, J. W. (2018). Optimal trees for prediction and prescription. Ph.D. thesis, Massachusetts Institute of Technology.

[19] DURRETT, R. (2019). *Probability—Theory and Examples*, 5th ed. *Cambridge Series in Statistical and Probabilistic Mathematics* **49**. Cambridge Univ. Press, Cambridge. MR3930614 https://doi.org/10.1017/9781108591034

[20] FROSST, N. and HINTON, G. (2017). Distilling a neural network into a soft decision tree. Preprint. Available at arXiv:1711.09784.

[21] GHOSH, P., AZAM, S., JONKMAN, M., KARIM, A., SHAMRAT, F. M. J. M., IGNATIOUS, E., SHULTANA, S., BEERAVOLU, A. R. and DE BOER, F. (2021). Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. *IEEE Access* **9** 19304–19326.

[22] 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

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

[24] HEATH, D., KASIF, S. and SALZBERG, S. (1993). Induction of oblique decision trees. *J. Artificial Intelligence Res.* **2** 1–32.

[25] HUANG, J. Z. (2003). Local asymptotics for polynomial spline regression. *Ann. Statist.* **31** 1600–1635. MR2012827 https://doi.org/10.1214/aos/1065705120

[26] HÜLLERMEIER, E., MOHR, F., TORNEDE, A. and WEVER, M. (2021). Automated machine learning, bounded rationality, and rational metareasoning. Preprint. Available at arXiv:2109.04744.

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

[28] KLUSOWSKI, J. M. and TIAN, P. (2023). Large scale prediction with decision trees. *J. Amer. Statist. Assoc.*

[29] LEE, G.-H. and JAAKKOLA, T. S. (2020). Oblique decision trees from derivatives of ReLU networks. In *International Conference on Learning Representations*.

[30] LI, X.-B., SWEIGART, J. R., TENG, J. T. C., DONOHUE, J. M., THOMBS, L. A. and WANG, S. M. (2003). Multivariate decision trees using linear discriminants and tabu search. *IEEE Trans. Syst. Man Cybern., Part A, Syst. Humans* **33** 194–205.

[31] LOH, W.-Y. and SHIH, Y.-S. (1997). Split selection methods for classification trees. *Statist. Sinica* **7** 815–840. MR1488644

[32] LÓPEZ-CHAU, A., CERVANTES, J., LÓPEZ-GARCÍA, L. and LAMONT, F. G. (2013). Fisher's decision tree. *Expert Syst. Appl.* **40** 6283–6291.

[33] MENZE, B. H., KELM, B. M., SPLITTHOFF, D. N., KOETHE, U. and HAMPRECHT, F. A. (2011). On oblique random forests. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* 453–469. Springer, Berlin.

[34] MINGERS, J. (1989). An empirical comparison of pruning methods for decision tree induction. *Mach. Learn.* **4** 227–243.

[35] MURDOCH, W. J., SINGH, C., KUMBIER, K., ABBASI-ASL, R. and YU, B. (2019). Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci. USA* **116** 22071–22080. MR4030584 https://doi.org/10.1073/pnas.1900654116

[36] MURTHY, S. K., KASIF, S. and SALZBERG, S. (1994). A system for induction of oblique decision trees. *J. Artificial Intelligence Res.* **2** 1–32.

[37] PARHI, R. and NOWAK, R. D. (2023). Deep learning meets sparse regularization: A signal processing perspective. Preprint. Available at arXiv:2301.09554.

[38] QUINLAN, J. R. (1993). C4.5, programs for machine learning. In *Proc. of* 10*th International Conference on Machine Learning* 252–259.

[39] RAYMAEKERS, J., ROUSSEEUW, P. J., VERDONCK, T. and YAO, R. (2023). Fast linear model trees by PILOT. Preprint. Available at arXiv:2302.03931.

[40] RODRIGUEZ, J. J., KUNCHEVA, L. I. and ALONSO, C. J. (2006). Rotation forest: A new classifier ensemble method. *IEEE Trans. Pattern Anal. Mach. Intell.* **28** 1619–1630.

[41] RUDIN, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **1** 206–215. https://doi.org/10.1038/s42256-019-0048-x

[42] 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

[43] SYRGKANIS, V. and ZAMPETAKIS, M. (2020). Estimation and inference with trees and forests in high dimensions. In *Proceedings of Thirty Third Conference on Learning Theory* (J. Abernethy and S. Agarwal, eds.). *Proceedings of Machine Learning Research* **125** 3453–3454. PMLR.

[44] TOMITA, T. M., BROWNE, J., SHEN, C., CHUNG, J., PATSOLIC, J. L., FALK, B., PRIEBE, C. E., YIM, J., BURNS, R. et al. (2020). Sparse projection oblique randomer forests. *J. Mach. Learn. Res.* **21** 1–39.

[45] WAGER, S. and ATHEY, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *J. Amer. Statist. Assoc.* **113** 1228–1242. MR3862353 https://doi.org/10.1080/01621459.2017.1319839

[46] YANG, Y., MORILLO, I. G. and HOSPEDALES, T. M. (2018). Deep neural decision trees. In *ICML Workshop on Human Interpretability in Machine Learning* (*WHI*).

[47] ZHAN, H., LIU, Y. and XIA, Y. (2023). Consistency of the oblique decision tree and its random forest. Preprint. Available at arXiv:2211.12653.

[48] ZHANG, T. (2003). Sequential greedy approximation for certain convex optimization problems. *IEEE Trans. Inf. Theory* **49** 682–691. MR1967192 https://doi.org/10.1109/TIT.2002.808136

[49] ZHU, H., MURALI, P., PHAN, D., NGUYEN, L. and KALAGNANAM, J. (2020). A scalable MIP-based method for learning optimal multivariate decision trees. In *Advances in Neural Information Processing Systems* (H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan and H. Lin, eds.) **33** 1771–1781. Curran Associates, Red Hook, NY.

The Annals of Statistics
2024, Vol. 52, No. 2, 491–518
https://doi.org/10.1214/24-AOS2356

EARLY STOPPING FOR L^2-BOOSTING IN HIGH-DIMENSIONAL LINEAR MODELS

BY BERNHARD STANKEWITZ[a]

Bocconi Institute for Data Science and Analytics, Bocconi University,, [a]*bernhard.stankewitz@unibocconi.it*

Increasingly high-dimensional data sets require that estimation methods do not only satisfy statistical guarantees but also remain computationally feasible. In this context, we consider L^2-boosting via orthogonal matching pursuit in a high-dimensional linear model and analyze a data-driven early stopping time τ of the algorithm, which is sequential in the sense that its computation is based on the first τ iterations only. This approach is much less costly than established model selection criteria, that require the computation of the full boosting path, which may even be computationally infeasible in truly high-dimensional applications. We prove that sequential early stopping preserves statistical optimality in this setting in terms of a fully general oracle inequality for the empirical risk and recently established optimal convergence rates for the population risk. Finally, an extensive simulation study shows that at a significantly reduced computational cost, the performance of early stopping methods is on par with other state of the art algorithms such as the cross-validated Lasso or model selection via a high-dimensional Akaike criterion based on the full boosting path.

REFERENCES

[1] 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
[2] BAXTER, G. (1962). An asymptotic result for the finite predictor. *Math. Scand.* **10** 137–144. MR0149584 https://doi.org/10.7146/math.scand.a-10520
[3] BLANCHARD, G., HOFFMANN, M. and REISS, M. (2018). Early stopping for statistical inverse problems via truncated SVD estimation. *Electron. J. Stat.* **12** 3204–3231. MR3859376 https://doi.org/10.1214/18-EJS1482
[4] BLANCHARD, G., HOFFMANN, M. and REISS, M. (2018). Optimal adaptation for early stopping in statistical inverse problems. *SIAM/ASA J. Uncertain. Quantificat.* **6** 1043–1075. MR3829522 https://doi.org/10.1137/17M1154096
[5] BLANCHARD, G. and MATHÉ, P. (2012). Discrepancy principle for statistical inverse problems with application to conjugate gradient iteration. *Inverse Probl.* **28** 115011/1–115011/23. MR2992966 https://doi.org/10.1088/0266-5611/28/11/115011
[6] BÜHLMANN, P. (2006). Boosting for high-dimensional linear models. *Ann. Statist.* **34** 559–583. MR2281878 https://doi.org/10.1214/009053606000000092
[7] BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer Series in Statistics.* Springer, Heidelberg. MR2807761 https://doi.org/10.1007/978-3-642-20192-9
[8] BÜHLMANN, P. and YU, B. (2003). Boosting with the L_2 loss: Regression and classification. *J. Amer. Statist. Assoc.* **98** 324–339. MR1995709 https://doi.org/10.1198/016214503000125
[9] CELISSE, A. and WAHL, M. (2021). Analyzing the discrepancy principle for kernelized spectral filter learning algorithms. *J. Mach. Learn. Res.* **22** Paper No. 76, 59. MR4253769
[10] ENGL, H. W., HANKE, M. and NEUBAUER, A. (1996). *Regularization of Inverse Problems. Mathematics and Its Applications* **375**. Kluwer Academic, Dordrecht. MR1408680
[11] GILL, R. D. and LEVIT, B. Y. (1995). Applications of the Van Trees inequality: A Bayesian Cramér–Rao bound. *Bernoulli* **1** 59–79. MR1354456 https://doi.org/10.2307/3318681

MSC2020 subject classifications. Primary 62G05, 62J07; secondary 62F35.
Key words and phrases. Early stopping, discrepancy principle, adaptive estimation, oracle inequalities, L2-boosting, orthogonal matching pursuit.

[12] ING, C.-K. (2020). Model selection for high-dimensional linear regression with dependent observations. *Ann. Statist.* **48** 1959–1980. MR4134782 https://doi.org/10.1214/19-AOS1872

[13] ING, C.-K. and LAI, T. L. (2011). A stepwise regression method and consistent model selection for high-dimensional sparse linear models. *Statist. Sinica* **21** 1473–1513. MR2895106 https://doi.org/10.5705/ss.2010.081

[14] JAHN, T. (2022). Optimal convergence of the discrepancy principle for polynomially and exponentially ill-posed operators under white noise. *Numer. Funct. Anal. Optim.* **43** 145–167. MR4413866 https://doi.org/10.1080/01630563.2021.2013881

[15] KUECK, J., LUO, Y., SPINDLER, M. and WANG, Z. (2023). Estimation and inference of treatment effects with L_2-boosting in high-dimensional settings. *J. Econometrics* **234** 714–731. MR4574119 https://doi.org/10.1016/j.jeconom.2022.02.005

[16] MEYER, M., MCMURRY, T. and POLITIS, D. (2015). Baxter's inequality for triangular arrays. *Math. Methods Statist.* **24** 135–146. MR3366950 https://doi.org/10.3103/S1066530715020040

[17] MIKA, G. and SZKUTNIK, Z. (2021). Towards adaptivity via a new discrepancy principle for Poisson inverse problems. *Electron. J. Stat.* **15** 2029–2059. MR4255321 https://doi.org/10.1214/21-ejs1835

[18] NEEDELL, D. and VERSHYNIN, R. (2010). Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. *IEEE J. Sel. Top. Signal Process.* **4** 310–316.

[19] PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A. et al. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12** 2825–2830. MR2854348

[20] SCHAPIRE, R. E. and FREUND, Y. (2012). *Boosting: Foundations and Algorithms. Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR2920188

[21] SCHERZER, O. (1993). The use of Morozov's discrepancy principle for Tikhonov regularization for solving nonlinear ill-posed problems. *Computing* **51** 45–60. MR1242658 https://doi.org/10.1007/BF02243828

[22] STANKEWITZ, B. (2020). Smoothed residual stopping for statistical inverse problems via truncated SVD estimation. *Electron. J. Stat.* **14** 3396–3428. MR4152147 https://doi.org/10.1214/20-EJS1747

[23] STANKEWITZ, B. (2024). Supplement to "Early stopping for L^2-boosting in high-dimensional linear models." https://doi.org/10.1214/24-AOS2356SUPP

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

[25] TEMLYAKOV, V. N. (2000). Weak greedy algorithms. *Adv. Comput. Math.* **12** 213–227. MR1745113 https://doi.org/10.1023/A:1018917218956

[26] TROPP, J. A. (2004). Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inf. Theory* **50** 2231–2242. MR2097044 https://doi.org/10.1109/TIT.2004.834793

[27] TROPP, J. A. and GILBERT, A. C. (2007). Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inf. Theory* **53** 4655–4666. MR2446929 https://doi.org/10.1109/TIT.2007.909108

[28] 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

The Annals of Statistics
2024, Vol. 52, No. 2, 519–549
https://doi.org/10.1214/24-AOS2357
This research was funded, in whole or in part, by EPSRC (UKRI), EP/V026259. A CC BY 4.0 license is applied to this article arising from this submission, in accordance with the grant's open access conditions.

CONSISTENT INFERENCE FOR DIFFUSIONS FROM LOW FREQUENCY MEASUREMENTS

BY RICHARD NICKL[a]

Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, [a]*nickl@maths.cam.ac.uk*

Let (X_t) be a reflected diffusion process in a bounded convex domain in $\mathbb{R}^d$, solving the stochastic differential equation

$$dX_t = \nabla f(X_t)\,dt + \sqrt{2f(X_t)}\,dW_t, \quad t \geq 0,$$

with W_t a d-dimensional Brownian motion. The data $X_0, X_D, \dots, X_{ND}$ consist of discrete measurements and the time interval D between consecutive observations is fixed so that one cannot 'zoom' into the observed path of the process. The goal is to infer the diffusivity f and the associated transition operator $P_{t,f}$. We prove injectivity theorems and stability inequalities for the maps $f \mapsto P_{t,f} \mapsto P_{D,f}$, $t < D$. Using these estimates, we establish the statistical consistency of a class of Bayesian algorithms based on Gaussian process priors for the infinite-dimensional parameter f, and show optimality of some of the convergence rates obtained. We discuss an underlying relationship between the degree of ill-posedness of this inverse problem and the 'hot spots' conjecture from spectral geometry.

REFERENCES

[1] ABRAHAM, K. and NICKL, R. (2019). On statistical Calderón problems. *Math. Stat. Learn.* **2** 165–216. MR4130599

[2] AECKERLE-WILLEMS, C. and STRAUCH, C. (2022). Sup-norm adaptive drift estimation for multivariate nonreversible diffusions. *Ann. Statist.* **50** 3484–3509. MR4524505 https://doi.org/10.1214/22-aos2237

[3] ATAR, R. and BURDZY, K. (2004). On Neumann eigenfunctions in lip domains. *J. Amer. Math. Soc.* **17** 243–265. MR2051611 https://doi.org/10.1090/S0894-0347-04-00453-9

[4] BAKRY, D., GENTIL, I. and LEDOUX, M. (2014). *Analysis and Geometry of Markov Diffusion Operators. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **348**. Springer, Cham. MR3155209 https://doi.org/10.1007/978-3-319-00227-9

[5] BAÑUELOS, R. and BURDZY, K. (1999). On the "hot spots" conjecture of J. Rauch. *J. Funct. Anal.* **164** 1–33. MR1694534 https://doi.org/10.1006/jfan.1999.3397

[6] BASS, R. F. (1998). *Diffusions and Elliptic Operators. Probability and Its Applications* (*New York*). Springer, New York. MR1483890

[7] BASS, R. F. (2011). *Stochastic Processes. Cambridge Series in Statistical and Probabilistic Mathematics* **33**. Cambridge Univ. Press, Cambridge. MR2856623 https://doi.org/10.1017/CBO9780511997044

[8] 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

[9] BOHR, J. and NICKL, R. (2021). On log-concave approximations of high-dimensional posterior measures and stability properties in non-linear inverse problems. *Ann. Inst. Henri Poincaré Probab. Stat.*, to appear.

[10] BONITO, A., COHEN, A., DEVORE, R., PETROVA, G. and WELPER, G. (2017). Diffusion coefficients estimation for elliptic partial differential equations. *SIAM J. Math. Anal.* **49** 1570–1592. MR3639575 https://doi.org/10.1137/16M1094476

[11] BURDZY, K. (2006). Neumann eigenfunctions and Brownian couplings. In *Potential Theory in Matsue. Adv. Stud. Pure Math.* **44** 11–23. Math. Soc. Japan, Tokyo. MR2277819 https://doi.org/10.2969/aspm/04410011

MSC2020 subject classifications. 35R30, 62G20, 62F15.
Key words and phrases. Bayesian inverse problems, reflected diffusion process, spectral PCA.

[12] BURDZY, K. and WERNER, W. (1999). A counterexample to the "hot spots" conjecture. *Ann. of Math.* (2) **149** 309–317. MR1680567 https://doi.org/10.2307/121027

[13] CALDERÓN, A.-P. (1980). On an inverse boundary value problem. In *Seminar on Numerical Analysis and Its Applications to Continuum Physics* (*Rio de Janeiro*, 1980) 65–73. Soc. Brasil. Mat., Rio de Janeiro. MR0590275

[14] 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

[15] CHEN, M. F. and WANG, F. Y. (1995). Estimation of the first eigenvalue of second order elliptic operators. *J. Funct. Anal.* **131** 345–363. MR1345035 https://doi.org/10.1006/jfan.1995.1092

[16] 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

[17] CUI, T., LAW, K. J. H. and MARZOUK, Y. M. (2016). Dimension-independent likelihood-informed MCMC. *J. Comput. Phys.* **304** 109–137. MR3422405 https://doi.org/10.1016/j.jcp.2015.10.008

[18] DALALYAN, A. and REISS, M. (2007). Asymptotic statistical equivalence for ergodic diffusions: The multidimensional case. *Probab. Theory Related Fields* **137** 25–47. MR2278451 https://doi.org/10.1007/s00440-006-0502-7

[19] DAVIES, E. B. (1995). *Spectral Theory and Differential Operators*. *Cambridge Studies in Advanced Mathematics* **42**. Cambridge Univ. Press, Cambridge. MR1349825 https://doi.org/10.1017/CBO9780511623721

[20] ENGL, H. W., HANKE, M. and NEUBAUER, A. (1996). *Regularization of Inverse Problems*. *Mathematics and Its Applications* **375**. Kluwer Academic, Dordrecht. MR1408680

[21] EVANS, L. C. (2010). *Partial Differential Equations*, 2nd ed. *Graduate Studies in Mathematics* **19**. Amer. Math. Soc., Providence, RI. MR2597943 https://doi.org/10.1090/gsm/019

[22] 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

[23] GINÉ, E. and NICKL, R. (2011). Rates of contraction for posterior distributions in L^r-metrics, $1 \le r \le \infty$. *Ann. Statist.* **39** 2883–2911. MR3012395 https://doi.org/10.1214/11-AOS924

[24] 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

[25] GIORDANO, M. and NICKL, R. (2020). Consistency of Bayesian inference with Gaussian process priors in an elliptic inverse problem. *Inverse Probl.* **36** 085001. MR4151406 https://doi.org/10.1088/1361-6420/ab7d2a

[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] GOBET, E., HOFFMANN, M. and REISS, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.* **32** 2223–2253. MR2102509 https://doi.org/10.1214/009053604000000797

[28] GRAHAM, I. G., KUO, F. Y., NICHOLS, J. A., SCHEICHL, R., SCHWAB, C. and SLOAN, I. H. (2015). Quasi-Monte Carlo finite element methods for elliptic PDEs with lognormal random coefficients. *Numer. Math.* **131** 329–368. MR3385149 https://doi.org/10.1007/s00211-014-0689-y

[29] GUGUSHVILI, S. and SPREIJ, P. (2014). Nonparametric Bayesian drift estimation for multidimensional stochastic differential equations. *Lith. Math. J.* **54** 127–141. MR3212631 https://doi.org/10.1007/s10986-014-9232-1

[30] HECKERT, A., DAHAL, L., TJIAN, R. and DARZACQ, X. (2022). Recovering mixtures of fast-diffusing states from short single-particle trajectories. *eLife* **11**. https://doi.org/10.7554/eLife.70169

[31] HELTBERG, M., MINÉ-HATTAB, J., TADDEI, A., WALCZAK, A. and MORA, T. (2023). Physical observables to determine the nature of membrane-less cellular sub-compartments. *eLife* **10** e69181.

[32] HOFFMANN, M. and RAY, K. (2022). Bayesian estimation in a multidimensional diffusion model with high frequency data. ArXiv preprint.

[33] JERISON, D. and NADIRASHVILI, N. (2000). The "hot spots" conjecture for domains with two axes of symmetry. *J. Amer. Math. Soc.* **13** 741–772. MR1775736 https://doi.org/10.1090/S0894-0347-00-00346-5

[34] JUDGE, C. and MONDAL, S. (2020). Euclidean triangles have no hot spots. *Ann. of Math.* (2) **191** 167–211. MR4045963 https://doi.org/10.4007/annals.2020.191.1.3

[35] KALTENBACHER, B., NEUBAUER, A. and SCHERZER, O. (2008). *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*. *Radon Series on Computational and Applied Mathematics* **6**. de Gruyter, Berlin. MR2459012 https://doi.org/10.1515/9783110208276

[36] KAWOHL, B. (1985). *Rearrangements and Convexity of Level Sets in PDE*. *Lecture Notes in Math*. **1150**. Springer, Berlin. MR0810619 https://doi.org/10.1007/BFb0075060

[37] KOHN, R. and VOGELIUS, M. (1984). Determining conductivity by boundary measurements. *Comm*. *Pure Appl*. *Math*. **37** 289–298. MR0739921 https://doi.org/10.1002/cpa.3160370302

[38] KOLTCHINSKII, V. (2021). Asymptotically efficient estimation of smooth functionals of covariance operators. *J*. *Eur*. *Math*. *Soc*. (*JEMS*) **23** 765–843. MR4210724 https://doi.org/10.4171/jems/1023

[39] LAW, K., STUART, A. and ZYGALAKIS, K. (2015). *Data Assimilation*. *Texts in Applied Mathematics* **62**. Springer, Cham. A mathematical introduction. MR3363508 https://doi.org/10.1007/978-3-319-20325-6

[40] LIONS, J.-L. and MAGENES, E. (1972). *Non-homogeneous Boundary Value Problems and Applications*. *Vol*. *I*. *Die Grundlehren der Mathematischen Wissenschaften*, *Band* 181. Springer, New York. MR0350177

[41] LÖFFLER, M. and PICARD, A. (2021). Spectral thresholding for the estimation of Markov chain transition operators. *Electron*. *J*. *Stat*. **15** 6281–6310. MR4355708 https://doi.org/10.1214/21-ejs1935

[42] LUNARDI, A. (1995). *Analytic Semigroups and Optimal Regularity in Parabolic Problems*. *Modern Birkhäuser Classics*. Birkhäuser/Springer Basel AG, Basel. MR3012216

[43] MAJDA, A. J. and HARLIM, J. (2012). *Filtering Complex Turbulent Systems*. Cambridge Univ. Press, Cambridge. MR2934167 https://doi.org/10.1017/CBO9781139061308

[44] MILO, R. and PHILLIPS, R. (2015). *Cell Biology by the Numbers*. Garland, New York.

[45] 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

[46] MONARD, F., NICKL, R. and PATERNAIN, G. P. (2021). Statistical guarantees for Bayesian uncertainty quantification in nonlinear inverse problems with Gaussian process priors. *Ann*. *Statist*. **49** 3255–3298. MR4352530 https://doi.org/10.1214/21-aos2082

[47] NACHMAN, A. I. (1988). Reconstructions from boundary measurements. *Ann*. *of Math*. (2) **128** 531–576. MR0970610 https://doi.org/10.2307/1971435

[48] 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

[49] 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

[50] NICKL, R. (2024). Supplement to "Consistent inference for diffusions from low frequency measurements." https://doi.org/10.1214/24-AOS2357SUPP

[51] 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

[52] NICKL, R. and SÖHL, J. (2017). Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions. *Ann*. *Statist*. **45** 1664–1693. MR3670192 https://doi.org/10.1214/16-AOS1504

[53] NICKL, R. and TITI, E. S. (2023). On posterior consistency of data assimilation with Gaussian process priors: The 2D Navier-Stokes equations. ArXiv 2023.

[54] NICKL, R., VAN DE GEER, S. and WANG, S. (2020). Convergence rates for penalized least squares estimators in PDE constrained regression problems. *SIAM/ASA J*. *Uncertain*. *Quantificat*. **8** 374–413. MR4074017 https://doi.org/10.1137/18M1236137

[55] NICKL, R. and WANG, S. (2022). On polynomial-time computation of high-dimensional posterior measures by Langevin-type algorithms. *J*. *Eur*. *Math*. *Soc*. (*JEMS*) to appear.

[56] PAULIN, D. (2015). Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electron*. *J*. *Probab*. **20** 79. MR3383563 https://doi.org/10.1214/EJP.v20-4039

[57] REICH, S. and COTTER, C. (2015). *Probabilistic Forecasting and Bayesian Data Assimilation*. Cambridge Univ. Press, New York. MR3242790 https://doi.org/10.1017/CBO9781107706804

[58] RICHTER, G. R. (1981). An inverse problem for the steady state diffusion equation. *SIAM J*. *Appl*. *Math*. **41** 210–221. MR0628945 https://doi.org/10.1137/0141016

[59] STEINERBERGER, S. (2020). Hot spots in convex domains are in the tips (up to an inradius). *Comm*. *Partial Differential Equations* **45** 641–654. MR4107000 https://doi.org/10.1080/03605302.2020.1750427

[60] STRAUCH, C. (2016). Exact adaptive pointwise drift estimation for multidimensional ergodic diffusions. *Probab*. *Theory Related Fields* **164** 361–400. MR3449393 https://doi.org/10.1007/s00440-014-0614-4

[61] STRAUCH, C. (2018). Adaptive invariant density estimation for ergodic diffusions over anisotropic classes. *Ann*. *Statist*. **46** 3451–3480. MR3852658 https://doi.org/10.1214/17-AOS1664

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

[63] SYLVESTER, J. and UHLMANN, G. (1987). A global uniqueness theorem for an inverse boundary value problem. *Ann*. *of Math*. (2) **125** 153–169. MR0873380 https://doi.org/10.2307/1971291

[64] TANAKA, H. (1979). Stochastic differential equations with reflecting boundary condition in convex regions. *Hiroshima Math. J.* **9** 163–177. MR0529332

[65] TAYLOR, M. E. (2011). *Partial Differential Equations III. Nonlinear Equations*, 2nd ed. *Applied Mathematical Sciences* **117**. Springer, New York. MR2744149 https://doi.org/10.1007/978-1-4419-7049-7

[66] TAYLOR, M. E. (2011). *Partial Differential Equations III. Nonlinear Equations*, 2nd ed. *Applied Mathematical Sciences* **117**. Springer, New York. MR2744149 https://doi.org/10.1007/978-1-4419-7049-7

[67] TRIEBEL, H. (1983). *Theory of Function Spaces*. *Monographs in Mathematics* **78**. Birkhäuser, Basel. MR0781540 https://doi.org/10.1007/978-3-0346-0416-1

[68] UHLMANN, G. (2009). Electrical impedance tomography and Calderón's problem. *Inverse Probl.* **25** 123011. MR3460047 https://doi.org/10.1088/0266-5611/25/12/123011

[69] VAN DER MEULEN, F. and SCHAUER, M. (2017). Bayesian estimation of discretely observed multi-dimensional diffusion processes using guided proposals. *Electron. J. Stat.* **11** 2358–2396. MR3656495 https://doi.org/10.1214/17-EJS1290

[70] VAN DER MEULEN, F. and VAN ZANTEN, H. (2013). Consistent nonparametric Bayesian inference for discretely observed scalar diffusions. *Bernoulli* **19** 44–63. MR3019485 https://doi.org/10.3150/11-BEJ385

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

[72] WANG, S. (2019). The nonparametric LAN expansion for discretely observed diffusions. *Electron. J. Stat.* **13** 1329–1358. MR3935851 https://doi.org/10.1214/19-ejs1545

The Annals of Statistics
2024, Vol. 52, No. 2, 550–579
https://doi.org/10.1214/24-AOS2358

A GENERAL FRAMEWORK TO QUANTIFY DEVIATIONS FROM STRUCTURAL ASSUMPTIONS IN THE ANALYSIS OF NONSTATIONARY FUNCTION-VALUED PROCESSES

BY ANNE VAN DELFT[1,a] AND HOLGER DETTE[2,b]

[1]*Department of Statistics, Columbia University,* [a]*anne.vandelft@columbia.edu*
[2]*Fakultät für Mathematik, Ruhr-Universität Bochum,,* [b]*holger.dette@rub.de*

We present a general theory to quantify the uncertainty from imposing structural assumptions on the second-order structure of nonstationary Hilbert space-valued processes, which can be measured via functionals of time-dependent spectral density operators. The second-order dynamics are well known to be elements of the space of trace class operators, the latter is a Banach space of type 1 and of cotype 2, which makes the development of statistical inference tools more challenging. A part of our contribution is to obtain a weak invariance principle as well as concentration inequalities for (functionals of) the sequential time-varying spectral density operator. In addition, we introduce deviation measures in the nonstationary context, and derive corresponding estimators that are asymptotically pivotal. We then apply this framework and propose statistical methodology to investigate the validity of structural assumptions for nonstationary functional data, such as low-rank assumptions in the context of time-varying dynamic fPCA and principle separable component analysis, deviations from stationarity with respect to the square root distance, and deviations from zero functional canonical coherency.

REFERENCES

[1] ANDERSON, T. W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Stat.* **34** 122–148. MR0145620 https://doi.org/10.1214/aoms/1177704248

[2] ASTON, J. A. D., PIGOLI, D. and TAVAKOLI, S. (2017). Tests for separability in nonparametric covariance operators of random surfaces. *Ann. Statist.* **45** 1431–1461. MR3670184 https://doi.org/10.1214/16-AOS1495

[3] BAGCHI, P. and DETTE, H. (2020). A test for separability in covariance operators of random surfaces. *Ann. Statist.* **48** 2303–2322. MR4134796 https://doi.org/10.1214/19-AOS1888

[4] BRILLINGER, D. R. (1981). *Time Series: Data Analysis and Theory*, 2nd ed. *Holden-Day Series in Time Series Analysis*. Holden-Day, Oakland, CA. MR0595684

[5] CONSTANTINOU, P., KOKOSZKA, P. and REIMHERR, M. (2017). Testing separability of space-time functional processes. *Biometrika* **104** 425–437. MR3698263 https://doi.org/10.1093/biomet/asx013

[6] CUPIDON, J., GILLIAM, D. S., EUBANK, R. and RUYMGAART, F. (2007). The delta method for analytic functions of random operators with application to functional data. *Bernoulli* **13** 1179–1194. MR2364231 https://doi.org/10.3150/07-BEJ6180

[7] DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37. MR1429916 https://doi.org/10.1214/aos/1034276620

[8] DAHLHAUS, R., RICHTER, S. and WU, W. B. (2019). Towards a general theory for nonlinear locally stationary processes. *Bernoulli* **25** 1013–1044. MR3920364 https://doi.org/10.3150/17-bej1011

[9] DE ACOSTA, A. D. (1970). Existence and convergence of probability measures in Banach spaces. *Trans. Amer. Math. Soc.* **152** 273–298. MR0267614 https://doi.org/10.2307/1995651

[10] DETTE, H., KOKOT, K. and VOLGUSHEV, S. (2020). Testing relevant hypotheses in functional time series via self-normalization. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 629–660. MR4112779 https://doi.org/10.1111/rssb.12370

MSC2020 subject classifications. Primary 62M10, 60B12; secondary 62M15, 60F17, 60J05.
Key words and phrases. Functional data analysis, principle components, relevant hypotheses, nuclear operators, locally stationary processes, concentration inequalities, martingale theory.

[11] 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

[12] FREMDT, S., STEINEBACH, J. G., HORVÁTH, L. and KOKOSZKA, P. (2013). Testing the equality of covariance operators in functional samples. *Scand. J. Stat.* **40** 138–152. MR3024036 https://doi.org/10.1111/j.1467-9469.2012.00796.x

[13] GENTON, M. G. (2007). Separable approximations of space-time covariance matrices. *Environmetrics* **18** 681–695. MR2408938 https://doi.org/10.1002/env.854

[14] HÖRMANN, S., KIDZIŃSKI, Ł. and HALLIN, M. (2015). Dynamic functional principal components. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 319–348. MR3310529 https://doi.org/10.1111/rssb.12076

[15] JENTSCH, C. and SUBBA RAO, S. (2015). A test for second order stationarity of a multivariate time series. *J. Econometrics* **185** 124–161. MR3300340 https://doi.org/10.1016/j.jeconom.2014.09.010

[16] JOLLIFFE, I. T. (2002). *Principal Component Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2036084

[17] KREISS, J.-P. and PAPARODITIS, E. (2015). Bootstrapping locally stationary processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 267–290. MR3299408 https://doi.org/10.1111/rssb.12068

[18] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces*: *Isoperimetry and Processes*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete* (3) [*Results in Mathematics and Related Areas* (3)] **23**. Springer, Berlin. MR1102015 https://doi.org/10.1007/978-3-642-20212-4

[19] LIU, W. and WU, W. B. (2010). Asymptotics of spectral density estimates. *Econometric Theory* **26** 1218–1245. MR2660298 https://doi.org/10.1017/S026646660999051X

[20] MAS, A. (2006). A sufficient condition for the CLT in the space of nuclear operators—application to covariance of random functions. *Statist. Probab. Lett.* **76** 1503–1509. MR2245571 https://doi.org/10.1016/j.spl.2006.03.010

[21] MASAK, T., SARKAR, S. and PANARETOS, V. M. (2023). Separable expansions for covariance estimation via the partial inner product. *Biometrika* **110** 225–247. MR4565453 https://doi.org/10.1093/biomet/asac035

[22] NASON, G. P., VON SACHS, R. and KROISANDT, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 271–292. MR1749539 https://doi.org/10.1111/1467-9868.00231

[23] PANARETOS, V. M., KRAUS, D. and MADDOCKS, J. H. (2010). Second-order comparison of Gaussian random functions and the geometry of DNA minicircles. *J. Amer. Statist. Assoc.* **105** 670–682. Supplementary materials available online. MR2724851 https://doi.org/10.1198/jasa.2010.tm09239

[24] PANARETOS, V. M. and TAVAKOLI, S. (2013). Cramér–Karhunen–Loève representation and harmonic principal component analysis of functional time series. *Stochastic Process. Appl.* **123** 2779–2807. MR3054545 https://doi.org/10.1016/j.spa.2013.03.015

[25] PIGOLI, D., ASTON, J. A. D., DRYDEN, I. L. and SECCHI, P. (2014). Distances and inference for covariance operators. *Biometrika* **101** 409–422. MR3215356 https://doi.org/10.1093/biomet/asu008

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

[27] SHAO, X. (2015). Self-normalization for time series: A review of recent developments. *J. Amer. Statist. Assoc.* **110** 1797–1817. MR3449074 https://doi.org/10.1080/01621459.2015.1050493

[28] TOMCZAK-JAEGERMANN, N. (1974). The moduli of smoothness and convexity and the Rademacher averages of trace classes S_p $(1 \leq p < \infty)$. *Studia Math.* **50** 163–182. MR0355667

[29] VAN DELFT, A. (2020). A note on quadratic forms of stationary functional time series under mild conditions. *Stochastic Process. Appl.* **130** 4206–4251. MR4102264 https://doi.org/10.1016/j.spa.2019.12.002

[30] VAN DELFT, A., CHARACIEJUS, V. and DETTE, H. (2021). A nonparametric test for stationarity in functional time series. *Statist. Sinica* **31** 1375–1395. MR4297704 https://doi.org/10.5705/ss.202018.0320

[31] VAN DELFT, A. and DETTE, H. (2022). Pivotal tests for relevant differences in the second order dynamics of functional time series. *Bernoulli* **28** 2260–2293. MR4474543 https://doi.org/10.3150/21-bej1418

[32] VAN DELFT, A. and DETTE, H. (2024). Supplement to "A general framework to quantify deviations from structural assumptions in the analysis of nonstationary function-valued processes." https://doi.org/10.1214/24-AOS2358SUPP

[33] VAN DELFT, A. and EICHLER, M. (2018). Locally stationary functional time series. *Electron. J. Stat.* **12** 107–170. MR3746979 https://doi.org/10.1214/17-EJS1384

[34] VAN DELFT, A. and EICHLER, M. (2020). A note on Herglotz's theorem for time series on function spaces. *Stochastic Process. Appl.* **130** 3687–3710. MR4092417 https://doi.org/10.1016/j.spa.2019.10.006

[35] 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

[36] YANG, J. and ZHOU, Z. (2022). Spectral inference under complex temporal dynamics. *J. Amer. Statist. Assoc.* **117** 133–155. MR4399075 https://doi.org/10.1080/01621459.2020.1764365

[37] YANG, W., MÜLLER, H.-G. and STADTMÜLLER, U. (2011). Functional singular component analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 303–324. MR2815778 https://doi.org/10.1111/j.1467-9868.2010.00769.x

[38] 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

[39] ZHANG, X. and SHAO, X. (2015). Two sample inference for the second-order property of temporally dependent functional data. *Bernoulli* **21** 909–929. MR3338651 https://doi.org/10.3150/13-BEJ592

The Annals of Statistics
2024, Vol. 52, No. 2, 580–601
https://doi.org/10.1214/24-AOS2359

THE EDGE OF DISCOVERY: CONTROLLING THE LOCAL FALSE DISCOVERY RATE AT THE MARGIN

BY JAKE A. SOLOFF[1,a], DANIEL XIANG[1,b] AND WILLIAM FITHIAN[2,c]

[1]*Department of Statistics, University of Chicago,* [a]*soloff@uchicago.edu,* [b]*dxiang@uchicago.edu*
[2]*Department of Statistics, University of California, Berkeley,* [c]*wfithian@berkeley.edu*

Despite the popularity of the false discovery rate (FDR) as an error control metric for large-scale multiple testing, its close Bayesian counterpart the local false discovery rate (lfdr), defined as the posterior probability that a particular null hypothesis is false, is a more directly relevant standard for justifying and interpreting individual rejections. However, the lfdr is difficult to work with in small samples, as the prior distribution is typically unknown. We propose a simple multiple testing procedure and prove that it controls the expectation of the maximum lfdr across all rejections; equivalently, it controls the probability that the rejection with the largest p-value is a false discovery. Our method operates without knowledge of the prior, assuming only that the p-value density is uniform under the null and decreasing under the alternative. We also show that our method asymptotically implements the oracle Bayes procedure for a weighted classification risk, optimally trading off between false positives and false negatives. We derive the limiting distribution of the attained maximum lfdr over the rejections, and the limiting empirical Bayes regret relative to the oracle procedure.

REFERENCES

AUBERT, J., BAR-HEN, A., DAUDIN, J.-J. and ROBIN, S. (2004). Determination of the differentially expressed genes in microarray experiments using local FDR. *BMC Bioinform*. **5** 1–9.

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

BENJAMINI, Y. and HOCHBERG, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J. Educ. Behav. Stat.* **25** 60–83.

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

CHERNOFF, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.* **16** 31–41. MR0172382 https://doi.org/10.1007/BF02868560

DYKSTRA, R. and CAROLAN, C. (1999). The distribution of the argmax of two-sided Brownian motion with quadratic drift. *J. Stat. Comput. Simul.* **63** 47–58. MR1703044 https://doi.org/10.1080/00949659908811948

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

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

EFRON, B. (2019). Bayes, oracle Bayes and empirical Bayes. *Statist. Sci.* **34** 177–201. MR3983318 https://doi.org/10.1214/18-STS674

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

FINNER, H. and ROTERS, M. (2001). On the false discovery rate and expected type I errors. *Biom. J.* **43** 985–1005. MR1878272 https://doi.org/10.1002/1521-4036(200112)43:8<985::AID-BIMJ985>3.0.CO;2-4

FINNER, H. and ROTERS, M. (2002). Multiple hypotheses testing and expected number of type I errors. *Ann. Statist.* **30** 220–238. MR1892662 https://doi.org/10.1214/aos/1015362191

GENOVESE, C. and WASSERMAN, L. (2004). A stochastic process approach to false discovery control. *Ann. Statist.* **32** 1035–1061. MR2065197 https://doi.org/10.1214/009053604000000283

MSC2020 subject classifications. 62J15, 62C12.
Key words and phrases. Local false discovery rates, multiple testing, empirical Bayes, Grenander estimator.

GRENANDER, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39** 125–153. MR0093415 https://doi.org/10.1080/03461238.1956.10414944

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

GROENEBOOM, P. and WELLNER, J. A. (2001). Computing Chernoff's distribution. *J. Comput. Graph. Statist.* **10** 388–400. MR1939706 https://doi.org/10.1198/10618600152627997

LANGAAS, M., LINDQVIST, B. H. and FERKINGSTAD, E. (2005). Estimating the proportion of true null hypotheses, with application to DNA microarray data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 555–572. MR2168204 https://doi.org/10.1111/j.1467-9868.2005.00515.x

LIAO, J., LIN, Y., SELVANAYAGAM, Z. E. and SHIH, W. J. (2004). A mixture model for estimating the local false discovery rate in DNA microarray analysis. *Bioinformatics* **20** 2694–2701.

MURALIDHARAN, O. (2010). An empirical Bayes mixture method for effect size and false discovery rate estimation. *Ann. Appl. Stat.* **4** 422–438. MR2758178 https://doi.org/10.1214/09-AOAS276

NEUVIAL, P. and ROQUAIN, E. (2012). On false discovery rate thresholding for classification under sparsity. *Ann. Statist.* **40** 2572–2600. MR3097613 https://doi.org/10.1214/12-AOS1042

PATRA, R. K. and SEN, B. (2016). Estimation of a two-component mixture model with applications to multiple testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 869–893. MR3534354 https://doi.org/10.1111/rssb.12148

POUNDS, S. and CHENG, C. (2004). Improving false discovery rate estimation. *Bioinformatics* **20** 1737–1745. https://doi.org/10.1093/bioinformatics/bth160

POUNDS, S. and MORRIS, S. W. (2003). Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values. *Bioinformatics* **19** 1236–1242.

PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36. MR0267677

REINER, A., YEKUTIELI, D. and BENJAMINI, Y. (2003). Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* **19** 368–375.

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, Calif. MR0044803

ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. Wiley, Chichester. MR0961262

ROBIN, S., BAR-HEN, A., DAUDIN, J.-J. and PIERRE, L. (2007). A semi-parametric approach for mixture models: Application to local false discovery rate estimation. *Comput. Statist. Data Anal.* **51** 5483–5493. MR2407654 https://doi.org/10.1016/j.csda.2007.02.028

SCHEID, S. and SPANG, R. (2004). A stochastic downhill search algorithm for estimating the local false discovery rate. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **1** 98–108. https://doi.org/10.1109/TCBB.2004.24

SCHWEDER, T. and SPJØTVOLL, E. (1982). Plots of p-values to evaluate many tests simultaneously. *Biometrika* **69** 493–502.

SEEGER, P. (1968). A note on a method for the analysis of significances en masse. *Technometrics* **10** 586–593.

SHORACK, G. R. and WELLNER, J. A. (2009). *Empirical Processes with Applications to Statistics*. *Classics in Applied Mathematics* **59**. SIAM, Philadelphia, PA. Reprint of the 1986 original [MR0838963]. MR3396731 https://doi.org/10.1137/1.9780898719017.ch1

SOLOFF, J. A., XIANG, D. and FITHIAN, W. (2024). Supplement to "The edge of discovery: Controlling the local false discovery rate at the margin." https://doi.org/10.1214/24-AOS2359SUPPA, https://doi.org/10.1214/24-AOS2359SUPPB

STEPHENS, M. (2017). False discovery rates: A new deal. *Biostatistics* **18** 275–294. MR3824755 https://doi.org/10.1093/biostatistics/kxw041

STOREY, J. D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 479–498. MR1924302 https://doi.org/10.1111/1467-9868.00346

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

STRIMMER, K. (2008). A unified approach to false discovery rate estimation. *BMC Bioinform.* **9** 1–14.

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

TAKÁCS, L. (1967). On combinatorial methods in the theory of stochastic processes. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability* (*Berkeley, Calif.*, 1965/66), *Vol. II: Contributions to Probability Theory, Part* 1 431–447. Univ. California Press, Berkeley, CA. MR0214129

TUCKER, H. G. (1959). A generalization of the Glivenko–Cantelli theorem. *Ann. Math. Stat.* **30** 828–830. MR0107891 https://doi.org/10.1214/aoms/1177706212

The Annals of Statistics
2024, Vol. 52, No. 2, 602–627
https://doi.org/10.1214/24-AOS2360

ℓ^2 INFERENCE FOR CHANGE POINTS IN HIGH-DIMENSIONAL TIME SERIES VIA A TWO-WAY MOSUM

BY JIAQI LI[1,a], LIKAI CHEN[1,b], WEINING WANG[2,c] AND WEI BIAO WU[3,d]

[1]*Department of Statistics and Data Science, Washington University in St. Louis,* [a]*lijiaqi@wustl.edu,* [b]*likai.chen@wustl.edu*
[2]*Faculty of Economics and Business, University of Groningen,* [c]*weining.wang@rug.nl*
[3]*Department of Statistics, University of Chicago,* [d]*wbwu@galton.uchicago.edu*

We propose an inference method for detecting multiple change points in high-dimensional time series, targeting dense or spatially clustered signals. Our method aggregates moving sum (MOSUM) statistics cross-sectionally by an ℓ^2-norm and maximizes them over time. We further introduce a novel Two-Way MOSUM, which utilizes spatial-temporal moving regions to search for breaks, with the added advantage of enhancing testing power when breaks occur in only a few groups. The limiting distribution of an ℓ^2-aggregated statistic is established for testing break existence by extending a high-dimensional Gaussian approximation theorem to spatial-temporal nonstationary processes. Simulation studies exhibit promising performance of our test in detecting nonsparse weak signals. Two applications on equity returns and COVID-19 cases in the United States show the real-world relevance of our algorithms. The R package "L2hdchange" is available on CRAN.

REFERENCES

[1] ADDARIO-BERRY, L., BROUTIN, N., DEVROYE, L. and LUGOSI, G. (2010). On combinatorial testing problems. *Ann. Statist.* **38** 3063–3092. MR2722464 https://doi.org/10.1214/10-AOS817

[2] ARIAS-CASTRO, E., CANDÈS, E. J. and DURAND, A. (2011). Detection of an anomalous cluster in a network. *Ann. Statist.* **39** 278–304. MR2797847 https://doi.org/10.1214/10-AOS839

[3] ARIAS-CASTRO, E., CANDÈS, E. J., HELGASON, H. and ZEITOUNI, O. (2008). Searching for a trail of evidence in a maze. *Ann. Statist.* **36** 1726–1757. MR2435454 https://doi.org/10.1214/07-AOS526

[4] BAI, J. (2010). Common breaks in means and variances for panel data. *J. Econometrics* **157** 78–92. MR2652280 https://doi.org/10.1016/j.jeconom.2009.10.020

[5] BAI, J., HAN, X. and SHI, Y. (2020). Estimation and inference of change points in high-dimensional factor models. *J. Econometrics* **219** 66–100. MR4152786 https://doi.org/10.1016/j.jeconom.2019.08.013

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

[7] BARNETT, I. and ONNELA, J.-P. (2016). Change point detection in correlation networks. *Sci. Rep.* **6** 18893. https://doi.org/10.1038/srep18893

[8] CHAN, J., HORVÁTH, L. and HUŠKOVÁ, M. (2013). Darling–Erdős limit results for change-point detection in panel data. *J. Statist. Plann. Inference* **143** 955–970. MR3011306 https://doi.org/10.1016/j.jspi.2012.11.004

[9] CHEN, C. Y.-H., OKHRIN, Y. and WANG, T. (2022). Monitoring network changes in social media. *J. Bus. Econom. Statist.* To appear. https://doi.org/10.1080/07350015.2021.2016425

[10] CHEN, L., WANG, W. and WU, W. B. (2021). Dynamic semiparametric factor model with structural breaks. *J. Bus. Econom. Statist.* **39** 757–771. MR4272933 https://doi.org/10.1080/07350015.2020.1730857

[11] CHEN, L., WANG, W. and WU, W. B. (2022). Inference of breakpoints in high-dimensional time series. *J. Amer. Statist. Assoc.* **117** 1951–1963. MR4528482 https://doi.org/10.1080/01621459.2021.1893178

[12] CHEN, Y., WANG, T. and SAMWORTH, R. J. (2022). High-dimensional, multiscale online changepoint detection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 234–266. MR4400396 https://doi.org/10.1111/rssb.12447

MSC2020 subject classifications. Primary 62E20, 62M10; secondary 62G20.
Key words and phrases. Nonlinear time series, multiple change-point detection, ℓ^2 inference, Two-Way MOSUM, Gaussian approximation, temporal and spatial dependence.

[13] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. MR3693963 https://doi.org/10.1214/16-AOP1113

[14] CHERNOZHUKOV, V., CHETVERIKOV, D. and KOIKE, Y. (2023). Nearly optimal central limit theorem and bootstrap approximations in high dimensions. *Ann. Appl. Probab.* **33** 2374–2425. MR4583674 https://doi.org/10.1214/22-aap1870

[15] CHO, H. (2016). Change-point detection in panel data via double CUSUM statistic. *Electron. J. Stat.* **10** 2000–2038. MR3522667 https://doi.org/10.1214/16-EJS1155

[16] CHO, H. and FRYZLEWICZ, P. (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 475–507. MR3310536 https://doi.org/10.1111/rssb.12079

[17] CHO, H. and KIRCH, C. (2022). Two-stage data segmentation permitting multiscale change points, heavy tails and dependence. *Ann. Inst. Statist. Math.* **74** 653–684. MR4444107 https://doi.org/10.1007/s10463-021-00811-5

[18] CRESSIE, N. A. C. (2015). *Statistics for Spatial Data. Wiley Classics Library*. Wiley, New York. MR3559472

[19] EICHINGER, B. and KIRCH, C. (2018). A MOSUM procedure for the estimation of multiple random change points. *Bernoulli* **24** 526–564. MR3706768 https://doi.org/10.3150/16-BEJ887

[20] ENIKEEVA, F. and HARCHAOUI, Z. (2019). High-dimensional change-point detection under sparse alternatives. *Ann. Statist.* **47** 2051–2079. MR3953444 https://doi.org/10.1214/18-AOS1740

[21] ESFAHLANI, F. Z., JO, Y., FASKOWITZ, J., BYRGE, L., KENNEDY, D. P., SPORNS, O. and BETZEL, R. F. (2020). High-amplitude cofluctuations in cortical activity drive functional connectivity. *Proc. Natl. Acad. Sci. USA* **117** 28393–28401.

[22] FASKOWITZ, J., ESFAHLANI, F. Z., JO, Y., SPORNS, O. and BETZEL, R. F. (2020). Edge-centric functional network representations of human cerebral cortex reveal overlapping system-level architecture. *Nat. Neurosci.* **23** 1644–1654.

[23] FRYZLEWICZ, P. (2014). Wild binary segmentation for multiple change-point detection. *Ann. Statist.* **42** 2243–2281. MR3269979 https://doi.org/10.1214/14-AOS1245

[24] HORVÁTH, L. and HUŠKOVÁ, M. (2012). Change-point detection in panel data. *J. Time Series Anal.* **33** 631–648. MR2944843 https://doi.org/10.1111/j.1467-9892.2012.00796.x

[25] HORVÁTH, L., HUŠKOVÁ, M., RICE, G. and WANG, J. (2017). Asymptotic properties of the CUSUM estimator for the time of change in linear panel data models. *Econometric Theory* **33** 366–412. MR3600047 https://doi.org/10.1017/S0266466615000468

[26] HUŠKOVÁ, M. and SLABÝ, A. (2001). Permutation tests for multiple changes. *Kybernetika* (*Prague*) **37** 605–622. MR1877077

[27] JIRAK, M. (2015). Uniform change point tests in high dimension. *Ann. Statist.* **43** 2451–2483. MR3405600 https://doi.org/10.1214/15-AOS1347

[28] KILLICK, R., FEARNHEAD, P. and ECKLEY, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *J. Amer. Statist. Assoc.* **107** 1590–1598. MR3036418 https://doi.org/10.1080/01621459.2012.737745

[29] KIRCH, C. and KLEIN, P. (2023). Moving sum data segmentation for stochastic processes based on invariance. *Statist. Sinica* **33** 873–892. MR4575326 https://doi.org/10.5705/ss.202021.0048

[30] KUCHIBHOTLA, A. K., BROWN, L. D., BUJA, A., GEORGE, E. I. and ZHAO, L. (2023). Uniform-in-submodel bounds for linear regression in a model-free framework. *Econometric Theory* **39** 1202–1248. MR4678102 https://doi.org/10.1017/s0266466621000219

[31] LEE, S., SEO, M. H. and SHIN, Y. (2016). The lasso for high dimensional regression with a possible change point. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 193–210. MR3453652 https://doi.org/10.1111/rssb.12108

[32] LÉVY-LEDUC, C. and ROUEFF, F. (2009). Detection and localization of change-points in high-dimensional network traffic data. *Ann. Appl. Stat.* **3** 637–662. MR2750676 https://doi.org/10.1214/08-AOAS232

[33] LI, D., QIAN, J. and SU, L. (2016). Panel data models with interactive fixed effects and multiple structural breaks. *J. Amer. Statist. Assoc.* **111** 1804–1819. MR3601737 https://doi.org/10.1080/01621459.2015.1119696

[34] LI, J., CHEN, L., WANG, W. and WU, W. B. (2024). Supplement to "ℓ^2 inference for change points in high-dimensional time series via a Two-Way MOSUM." https://doi.org/10.1214/24-AOS2360SUPP

[35] LIU, B., QI, Z., ZHANG, X. and LIU, Y. (2022). Change point detection for high-dimensional linear models: A general tail-adaptive approach. Preprint. Available at arXiv:2207.11532.

[36] MADRID PADILLA, O. H., YU, Y. and RINALDO, A. (2021). Lattice partition recovery with dyadic CART. *Adv. Neural Inf. Process. Syst.* **34** 26143–26155.

[37] MATSUDA, Y. and YAJIMA, Y. (2009). Fourier analysis of irregularly spaced data on $\mathbb{R}^d$. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 191–217. MR2655530 https://doi.org/10.1111/j.1467-9868.2008.00685.x

[38] OLSHEN, A. B., VENKATRAMAN, E. S., LUCITO, R. and WIGLER, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5** 557–572. https://doi.org/10.1093/biostatistics/kxh008

[39] ONNELA, J. P., CHAKRABORTI, A., KASKI, K., KERTÉSZ, J. and KANTO, A. (2003). Dynamics of market correlations: Taxonomy and portfolio analysis. *Phys. Rev. E* **68** 056110.

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

[41] SCOTT, A. J. and KNOTT, M. (1974). A cluster analysis method for grouping means in the analysis of variance. *Biometrics* **30** 507–512.

[42] SHAO, X. (2010). A self-normalized approach to confidence interval construction in time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 343–366. MR2758116 https://doi.org/10.1111/j.1467-9868.2009.00737.x

[43] 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

[44] TIBSHIRANI, R. and WANG, P. (2008). Spatial smoothing and hot spot detection for CGH data using the fused lasso. *Biostatistics* **9** 18–29.

[45] WANG, D. and ZHAO, Z. (2022). Optimal change-point testing for high-dimensional linear models with temporal dependence. Preprint. Available at arXiv:2205.03880.

[46] WANG, R. and SHAO, X. (2020). Hypothesis testing for high-dimensional time series via self-normalization. *Ann. Statist.* **48** 2728–2758. MR4152119 https://doi.org/10.1214/19-AOS1904

[47] WANG, R. and SHAO, X. (2023). Dating the break in high-dimensional data. *Bernoulli* **29** 2879–2901. MR4632124 https://doi.org/10.3150/22-bej1567

[48] WANG, R., ZHU, C., VOLGUSHEV, S. and SHAO, X. (2022). Inference for change points in high-dimensional data via selfnormalization. *Ann. Statist.* **50** 781–806. MR4405366 https://doi.org/10.1214/21-aos2127

[49] WANG, T. and SAMWORTH, R. J. (2018). High dimensional change point estimation via sparse projection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 57–83. MR3744712 https://doi.org/10.1111/rssb.12243

[50] 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

[51] WU, W. B. and ZHAO, Z. (2007). Inference of trends in time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 391–410. MR2323759 https://doi.org/10.1111/j.1467-9868.2007.00594.x

[52] XIE, Y. and SIEGMUND, D. (2013). Sequential multi-sensor change-point detection. *Ann. Statist.* **41** 670–692. MR3099117 https://doi.org/10.1214/13-AOS1094

[53] XU, H., WANG, D., ZHAO, Z. and YU, Y. (2022). Change point inference in high-dimensional regression models under temporal dependence. Preprint. Available at arXiv:2207.12453.

[54] YEO, B. T. T., KRIENEN, F. M., SEPULCRE, J., SABUNCU, M. R., LASHKARI, D., HOLLINSHEAD, M., ROFFMAN, J. L., SMOLLER, J. W., ZÖLLEI, L. et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106** 1125–1165.

[55] YU, M. and CHEN, X. (2021). Finite sample change point inference and identification for high-dimensional mean vectors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 247–270. MR4250275 https://doi.org/10.1111/rssb.12406

[56] YU, Y. (2020). A review on minimax rates in change point detection and localisation. Preprint. Available at arXiv:2011.01857.

[57] YU, Y., MADRID, O. and RINALDO, A. (2022). Optimal partition recovery in general graphs. In *International Conference on Artificial Intelligence and Statistics* 4339–4358. PMLR.

[58] ZHANG, N. R., SIEGMUND, D. O., JI, H. and LI, J. Z. (2010). Detecting simultaneous changepoints in multiple sequences. *Biometrika* **97** 631–645. MR2672488 https://doi.org/10.1093/biomet/asq025

The Annals of Statistics
2024, Vol. 52, No. 2, 628–653
https://doi.org/10.1214/24-AOS2361

TESTING FOR PRACTICALLY SIGNIFICANT DEPENDENCIES IN HIGH DIMENSIONS VIA BOOTSTRAPPING MAXIMA OF U-STATISTICS

BY PATRICK BASTIAN[1,a], HOLGER DETTE[1,b] AND JOHANNES HEINY[2,c]

[1]*Fakultät für Mathematik, Ruhr-Universität Bochum,* [a]*patrick.bastian@rub.de,* [b]*holger.dette@rub.de*
[2]*Department of Mathematics, Stockholm University,* [c]*johannes.heiny@math.su.se*

This paper takes a different look on the problem of testing the mutual independence of the components of a high-dimensional vector. Instead of testing if all pairwise associations (e.g., all pairwise Kendall's τ) between the components vanish, we are interested in the (null) hypothesis that all pairwise associations do not exceed a certain threshold in absolute value. The consideration of these hypotheses is motivated by the observation that in the high-dimensional regime, it is rare, and perhaps impossible, to have a null hypothesis that can be exactly modeled by assuming that all pairwise associations are precisely equal to zero.

The formulation of the null hypothesis as a composite hypothesis makes the problem of constructing tests nonstandard and in this paper we provide a solution for a broad class of dependence measures, which can be estimated by U-statistics. In particular, we develop an asymptotic and a bootstrap level α-test for the new hypotheses in the high-dimensional regime. We also prove that the new tests are minimax-optimal and investigate their finite sample properties by means of a small simulation study and a data example.

REFERENCES

ADAM, B.-L., QU, Y., DAVIS, J. W., WARD, M. D., CLEMENTS, M. A., CAZARES, L. H., SEMMES, O. J., SCHELLHAMMER, P. F., YASUI, Y. et al. (2002). Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy Men1. *Cancer Res*. **62** 3609–3614.

ALBERT, M., BOURET, Y., FROMONT, M. and REYNAUD-BOURET, P. (2015). Bootstrap and permutation tests of independence for point processes. *Ann. Statist*. **43** 2537–2564. MR3405603 https://doi.org/10.1214/15-AOS1351

ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York.

ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1989). Two moments suffice for Poisson approximations: The Chen–Stein method. *Ann. Probab.* **17** 9–25. MR0972770

BAI, Z., JIANG, D., YAO, J.-F. and ZHENG, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist*. **37** 3822–3840. MR2572444 https://doi.org/10.1214/09-AOS694

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

BASTIAN, P., DETTE, H. and HEINY, J. (2024). Supplement to "Testing for practically significant dependencies in high dimensions via bootstrapping maxima of U-statistics." https://doi.org/10.1214/24-AOS2361SUPP

BERGER, J. O. and DELAMPADY, M. (1987). Testing precise hypotheses. *Statist. Sci.* **2** 317–352. MR0920141

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

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist*. **9** 1196–1217. MR0630103

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

MSC2020 subject classifications. 62G10, 62F40, 62C20.
Key words and phrases. Independence testing, relevant association, U-statistics, gaussian approximation, bootstrap, minimax optimality.

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

BOSCO, F. A., AGUINIS, H., SINGH, K., FIELD, J. G. and PIERCE, C. A. (2015). Correlational effect size benchmarks. *J. Appl. Psychol.* **100** 431–449. https://doi.org/10.1037/a0038047

BRYDGES, C. R. (2019). Effect size guidelines, sample size calculations, and statistical power in gerontology. *Innov. Aging* **3** igz036.

CAI, T. T. and JIANG, T. (2012). Phase transition in limiting distributions of coherence of high-dimensional random matrices. *J. Multivariate Anal.* **107** 24–39. MR2890430 https://doi.org/10.1016/j.jmva.2011.11.008

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

CHEN, H. and JIANG, T. (2018). A study of two high-dimensional likelihood ratio tests under alternative hypotheses. *Random Matrices Theory Appl.* **7** 1750016. MR3756424 https://doi.org/10.1142/S2010326317500162

CHEN, X. (2018). Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications. *Ann. Statist.* **46** 642–678. MR3782380 https://doi.org/10.1214/17-AOS1563

CHEN, X. and KATO, K. (2019). Randomized incomplete U-statistics in high dimensions. *Ann. Statist.* **47** 3127–3156. MR4025737 https://doi.org/10.1214/18-AOS1773

CHENG, G., LIU, Z. and PENG, L. (2022). Gaussian approximations for high-dimensional non-degenerate U-statistics via exchangeable pairs. *Statist. Probab. Lett.* **182** 109295. MR4340819 https://doi.org/10.1016/j.spl.2021.109295

CHERNOZHUKOV, V., CHETVERIKOV, D., KATO, K. and KOIKE, Y. (2019). Improved Central Limit Theorem and bootstrap approximations in high dimensions. Papers, arXiv.org. Available at arXiv:1912.10529.

CHETVERIKOV, D., WILHELM, D. and KIM, D. (2021). An adaptive test of stochastic monotonicity. *Econometric Theory* **37** 495–536. MR4271925 https://doi.org/10.1017/S0266466620000225

CHOW, S.-C. and LIU, P.-J. (1992). *Design and Analysis of Bioavailability and Bioequivalence Studies*. Dekker, New York.

COHEN, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Erlbaum, Hillsdale.

DETTE, H. and DÖRNEMANN, N. (2020). Likelihood ratio tests for many groups in high dimensions. *J. Multivariate Anal.* **178** 104605. MR4079037 https://doi.org/10.1016/j.jmva.2020.104605

DETTE, H., SIBURG, K. F. and STOIMENOV, P. A. (2013). A copula-based non-parametric measure of regression dependence. *Scand. J. Stat.* **40** 21–41. MR3024030 https://doi.org/10.1111/j.1467-9469.2011.00767.x

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

EDELMANN, D., TERZER, T. and RICHARDS, D. (2021). A basic treatment of the distance covariance. *Sankhya B* **83** S12–S25. MR4258084 https://doi.org/10.1007/s13571-021-00248-z

EVEN-ZOHAR, C. (2020). Independence: Fast rank tests. Arxiv preprint. Available at arXiv:2010.09712.

FANG, Z. and SANTOS, A. (2019). Inference on directionally differentiable functions. *Rev. Econ. Stud.* **86** 377–412. MR3936869 https://doi.org/10.1093/restud/rdy049

GEENENS, G. and LAFAYE DE MICHEAUX, P. (2022). The Hellinger correlation. *J. Amer. Statist. Assoc.* **117** 639–653. MR4436302 https://doi.org/10.1080/01621459.2020.1791132

GRETTON, A., FUKUMIZU, K., TEO, C., SONG, L., SCHÖLKOPF, B. and SMOLA, A. (2008). A kernel statistical test of independence. In *Advances in Neural Information Processing Systems* (J. Platt, D. Koller, Y. Singer and S. Roweis, eds.) **20** 585–592. Curran Associates, Red Hook.

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

HE, Y., XU, G., WU, C. and PAN, W. (2021). Asymptotically independent U-statistics in high-dimensional testing. *Ann. Statist.* **49** 154–181. MR4206673 https://doi.org/10.1214/20-AOS1951

HEINY, J., MIKOSCH, T. and YSLAS, J. (2021). Point process convergence for the off-diagonal entries of sample covariance matrices. *Ann. Appl. Probab.* **31** 538–560. MR4254488 https://doi.org/10.1214/20-aap1597

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

HEMPHILL, J. (2003). Interpreting the magnitude of correlation coefficients. *Amer. Psychol.* **58** 78–9.

HOEFFDING, W. (1948a). A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* **19** 293–325. MR0026294 https://doi.org/10.1214/aoms/1177730196

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

HUANG, W., TANG, M., CHEN, Y.-L., ZHANG, T.-L., HONG, T., LI, J., LV, G.-H., YAN, Y., OUYANG, Z.-H. et al. (2022). Incidence and risk factors for cerebrovascular-specific mortality in patients with colorectal cancer: A registry-based cohort study involving 563298 patients. *Cancers* **14** 2053.

JIANG, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **14** 865–880. MR2052906 https://doi.org/10.1214/105051604000000143

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

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

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

KIM, I. (2020). Multinomial goodness-of-fit based on U-statistics: High-dimensional asymptotic and minimax optimality. *J. Statist. Plann. Inference* **205** 74–91. MR4011624 https://doi.org/10.1016/j.jspi.2019.06.005

LEE, Y. (1971). Distribution of the canonical correlations and asymptotic expansions for distributions of certain independence test statistics. *Ann. Math. Stat.* **42** 526–537. MR0286215 https://doi.org/10.1214/aoms/1177693403

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

LEVINA, E., ROTHMAN, A. and ZHU, J. (2008). Sparse estimation of large covariance matrices via a nested Lasso penalty. *Ann. Appl. Stat.* **2** 245–263. MR2415602 https://doi.org/10.1214/07-AOAS139

LI, D., LIU, W.-D. and ROSALSKY, A. (2010). Necessary and sufficient conditions for the asymptotic distribution of the largest entry of a sample correlation matrix. *Probab. Theory Related Fields* **148** 5–35. MR2653220 https://doi.org/10.1007/s00440-009-0220-z

LI, Z., WANG, Q. and LI, R. (2021). Central limit theorem for linear spectral statistics of large dimensional Kendall's rank correlation matrices and its applications. *Ann. Statist.* **49** 1569–1593. MR4298873 https://doi.org/10.1214/20-aos2013

LINDSKOG, F., MCNEIL, A. and SCHMOCK, U. (2003). Kendall's tau for elliptical distributions. In *Credit Risk. Contributions to Economics. Physica-Verlag* (G. Bol, G. Nakhaeizadeh, S. T. Rachev, T. Ridder and K. H. Vollmer, eds.) 149–156.

LIU, W.-D., LIN, Z. and SHAO, Q.-M. (2008). The asymptotic distribution and Berry–Esseen bound of a new test for independence in high dimension with an application to stochastic optimization. *Ann. Appl. Probab.* **18** 2337–2366. MR2474539 https://doi.org/10.1214/08-AAP527

LOVAKOV, A. and AGADULLINA, E. (2021). Empirically derived guidelines for effect size interpretation in social psychology. *Eur. J. Soc. Psychol.* **51** 485–504.

NAGAO, H. (1973). On some test criteria for covariance matrix. *Ann. Statist.* **1** 700–709. MR0339405

NARAIN, R. D. (1950). On the completely unbiassed character of tests of independence in multivariate normal systems. *Ann. Math. Stat.* **21** 293–298. MR0035957 https://doi.org/10.1214/aoms/1177729848

PEARSON, K. (1920). Notes on the history of correlation. *Biometrika* **13** 25–45.

QIU, Y. and CHEN, S. X. (2012). Test for bandedness of high-dimensional covariance matrices and bandwidth estimation. *Ann. Statist.* **40** 1285–1314. MR3015026 https://doi.org/10.1214/12-AOS1002

QUINTANA, D. S. (2017). Statistical considerations for reporting and planning heart rate variability case-control studies. *Psychophysiology* **54** 344–349. https://doi.org/10.1111/psyp.12798

ROY, S. N. (1957). *Some Aspects of Multivariate Analysis*. Wiley, New York. MR0092296

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

SHAO, Q.-M. and ZHOU, W.-X. (2014). Necessary and sufficient conditions for the asymptotic distributions of coherence of ultra-high dimensional random matrices. *Ann. Probab.* **42** 623–648. MR3178469 https://doi.org/10.1214/13-AOP837

SONG, Y., CHEN, X. and KATO, K. (2019). Approximating high-dimensional infinite-order U-statistics: Statistical and computational guarantees. *Electron. J. Stat.* **13** 4794–4848. MR4038726 https://doi.org/10.1214/19-EJS1643

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

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

TUKEY, J. W. (1991). The philosophy of multiple comparisons. *Statist. Sci.* **6** 100–116.

WANG, R., ZHU, C., VOLGUSHEV, S. and SHAO, X. (2022). Inference for change points in high-dimensional data via selfnormalization. *Ann. Statist.* **50** 781–806. MR4405366 https://doi.org/10.1214/21-aos2127

WELLEK, S. (2010). *Testing Statistical Hypotheses of Equivalence and Noninferiority*, 2nd ed. CRC Press, Boca Raton, FL. MR2676002 https://doi.org/10.1201/EBK1439808184

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

ZHOU, C., HAN, F., ZHANG, X.-S. and LIU, H. (2019). An extreme-value approach for testing the equality of large U-statistic based correlation matrices. *Bernoulli* **25** 1472–1503. MR3920379 https://doi.org/10.3150/18-bej1027

ZHOU, W. (2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Trans. Amer. Math. Soc.* **359** 5345–5363. MR2327033 https://doi.org/10.1090/S0002-9947-07-04192-X

The Annals of Statistics
2024, Vol. 52, No. 2, 654–678
https://doi.org/10.1214/24-AOS2362

TRANSFER LEARNING FOR FUNCTIONAL MEAN ESTIMATION: PHASE TRANSITION AND ADAPTIVE ALGORITHMS

BY T. TONY CAI[a], DONGWOO KIM[b] AND HONGMING PU[c]

Department of Statistics and Data Science, The Wharton School, University of Pennsylvania, [a]*tcai@wharton.upenn.edu,* [b]*dongwooo@wharton.upenn.edu,* [c]*hpu@wharton.upenn.edu*

This paper studies transfer learning for estimating the mean of random functions based on discretely sampled data, where in addition to observations from the target distribution, auxiliary samples from similar but distinct source distributions are available. The paper considers both common and independent designs and establishes the minimax rates of convergence for both designs. The results reveal an interesting phase transition phenomenon under the two designs and demonstrate the benefits of utilizing the source samples in the low sampling frequency regime.

For practical applications, this paper proposes novel data-driven adaptive algorithms that attain the optimal rates of convergence within a logarithmic factor simultaneously over a large collection of parameter spaces. The theoretical findings are complemented by a simulation study that further supports the effectiveness of the proposed algorithms.

REFERENCES

[1] ASSOUAD, P. (1983). Densité et dimension. *Ann. Inst. Fourier* (*Grenoble*) **33** 233–282. MR0723955
[2] BREIMAN, L. (1996). Bagging predictors. *Mach. Learn.* **24** 123–140. https://doi.org/10.1007/BF00058655
[3] CAI, C., CAI, T. T. and LI, H. (2024). Transfer learning for contextual multi-armed bandits. *Ann. Statist.* **52** 207–232. MR4718413 https://doi.org/10.1214/23-aos2341
[4] LI, S., ZHANG, L., CAI, T. T. and LI, H. (2023). Estimation and inference for high-dimensional generalized linear models with knowledge transfer. *J. Amer. Statist. Assoc.* https://doi.org/10.1080/01621459.2023.2184373
[5] CAI, T. T. and HALL, P. (2006). Prediction in functional linear regression. *Ann. Statist.* **34** 2159–2179. MR2291496 https://doi.org/10.1214/009053606000000830
[6] CAI, T. T, KIM, D. and PU, H. (2024). Supplement to "Transfer learning for functional mean estimation: Phase transition and adaptive algorithms." https://doi.org/10.1214/24-AOS2362SUPP
[7] CAI, T. T. and PU, H. (2022). Transfer Learning for Nonparametric Regression: Non-Asymptotic Minimax Analysis and Adaptive Procedure Technical report.
[8] CAI, T. T. and WEI, H. (2021). Transfer learning for nonparametric classification: Minimax rate and adaptive classifier. *Ann. Statist.* **49** 100–128. MR4206671 https://doi.org/10.1214/20-AOS1949
[9] CAI, T. T. and YUAN, M. (2010). Nonparametric Covariance Function Estimation for Functional and Longitudinal Data Technical report.
[10] 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
[11] CAI, T. T. and YUAN, M. (2012). Minimax and adaptive prediction for functional linear regression. *J. Amer. Statist. Assoc.* **107** 1201–1216. MR3010906 https://doi.org/10.1080/01621459.2012.716337
[12] CHOI, K., FAZEKAS, G., SANDLER, M. and CHO, K. (2017). Transfer Learning for Music Classification and Regression Tasks Technical report. https://doi.org/10.48550/arXiv.1703.09179
[13] DEGRAS, D. (2017). Simultaneous confidence bands for the mean of functional data. *Wiley Interdiscip. Rev.: Comput. Stat.* **9** e1397, 15. MR3648600 https://doi.org/10.1002/wics.1397
[14] GONG, B., SHI, Y., SHA, F. and GRAUMAN, K. (2012). Geodesic flow kernel for unsupervised domain adaptation. In 2012 *IEEE Conference on Computer Vision and Pattern*. Recognition 2066–2073. https://doi.org/10.1109/CVPR.2012.6247911

MSC2020 subject classifications. Primary 62J05; secondary 62G20.
Key words and phrases. Adaptivity, common design, functional data analysis, independent design, mean function, minimax rate of convergence, phase transition, transfer learning.

[15] HUANG, J.-T., LI, J., YU, D., DENG, L. and GONG, Y. (2013). Cross-language knowledge transfer using multilingual deep neural network with shared hidden layers. In 2013 *IEEE International Conference on Acoustics, Speech and Signal Processing* 7304–7308. https://doi.org/10.1109/ICASSP.2013.6639081

[16] JAMES, N. and MENZIES, M. (2021). Trends in COVID-19 prevalence and mortality: A year in review. *Phys. D* **425** Paper No. 132968, 12. MR4275048 https://doi.org/10.1016/j.physd.2021.132968

[17] JIANG, C.-R., ASTON, J. A. D. and WANG, J.-L. (2009). Smoothing dynamic positron emission tomography time courses using functional principal components. *NeuroImage* **47** 184–193. https://doi.org/10.1016/j.neuroimage.2009.03.051

[18] KALOGRIDIS, I. and VAN AELST, S. (2023). Robust optimal estimation of location from discretely sampled functional data. *Scand. J. Stat.* **50** 411–451. MR4599920

[19] KOZLOFF, N., MULSANT, B. H., STERGIOPOULOS, V. and VOINESKOS, A. N. (2020). The Covid-19 global pandemic: Implications for people with schizophrenia and related disorders. *Schizophr. Bull.* **46** 752–757. https://doi.org/10.1093/schbul/sbaa051

[20] KPOTUFE, S. and MARTINET, G. (2018). Marginal singularity, and the benefits of labels in covariate-shift. In *Proceedings of the* 31*st Conference on Learning Theory* 1882–1886. *PMLR*.

[21] LENG, X. and MÜLLER, H.-G. (2006). Classification using functional data analysis for temporal gene expression data. *Bioinformatics* **22** 68–76. https://doi.org/10.1093/bioinformatics/bti742

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

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

[24] MANTÉ, C., DURBEC, J. P. and DAUVIN, J. C. (2005). A functional data-analytic approach to the classification of species according to their spatial dispersion. Application to a marine macrobenthic community from the Bay of Morlaix (Western English Channel). *J. Appl. Stat.* **32** 831–840. MR2214313 https://doi.org/10.1080/02664760500080124

[25] PAGE, A., AYALA, G., LEÓN, M. T., PEYDRO, M. F. and PRAT, J. M. (2006). Normalizing temporal patterns to analyze sit-to-stand movements by using registration of functional data. *J. Biomech.* **39** 2526–2534. https://doi.org/10.1016/j.jbiomech.2005.07.032

[26] PAN, S. J. and YANG, Q. (2010). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **22** 1345–1359. https://doi.org/10.1109/TKDE.2009.191

[27] PARK, C., KOO, J.-Y., KIM, S., SOHN, I. and LEE, J. W. (2008). Classification of gene functions using support vector machine for time-course gene expression data. *Comput. Statist. Data Anal.* **52** 2578–2587. MR2411960 https://doi.org/10.1016/j.csda.2007.09.002

[28] POMANN, G.-M., STAICU, A.-M. and GHOSH, S. (2016). A two-sample distribution-free test for functional data with application to a diffusion tensor imaging study of multiple sclerosis. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **65** 395–414. MR3470583 https://doi.org/10.1111/rssc.12130

[29] RAMSAY, J. O. and SILVERMAN, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies. Springer Series in Statistics*. Springer, New York. MR1910407 https://doi.org/10.1007/b98886

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

[31] RICE, J. A. and SILVERMAN, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B* **53** 233–243. MR1094283

[32] SONG, J. J., DENG, W., LEE, H.-J. and KWON, D. (2008). Optimal classification for time-course gene expression data using functional data analysis. *Comput. Biol. Chem.* **32** 426–432. MR2474550 https://doi.org/10.1016/j.compbiolchem.2008.07.007

[33] STAICU, A.-M., CRAINICEANU, C. M., REICH, D. S. and RUPPERT, D. (2012). Modeling functional data with spatially heterogeneous shape characteristics. *Biometrics* **68** 331–343. MR2959599 https://doi.org/10.1111/j.1541-0420.2011.01669.x

[34] TIAN, Y. and FENG, Y. (2023). Transfer learning under high-dimensional generalized linear models. *J. Amer. Statist. Assoc.* **118** 2684–2697. MR4681613 https://doi.org/10.1080/01621459.2022.2071278

[35] TZENG, E., HOFFMAN, J., SAENKO, K. and DARRELL, T. (2017). Adversarial discriminative domain adaptation. In 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*) 2962–2971. https://doi.org/10.1109/CVPR.2017.316

[36] 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

[37] WANG, J.-L., CHIOU, J.-M. and MÜLLER, H.-G. (2016). Functional data analysis. *Annu. Rev. Stat. Appl.* **3** 257–295. https://doi.org/10.1146/annurev-statistics-041715-033624

[38] WEAVER, C., XIAO, L. and LU, W. (2023). Functional data analysis for longitudinal data with informative observation times. *Biometrics* **79** 722–733. MR4606310

[39] WEISS, K., KHOSHGOFTAAR, T. M. and WANG, D. (2016). A survey of transfer learning. *J. Big Data* **3**. https://doi.org/10.1186/s40537-016-0043-6

[40] WU, P.-S. and MÜLLER, H.-G. (2010). Functional embedding for the classification of gene expression profiles. *Bioinformatics* **26** 509–517. https://doi.org/10.1093/bioinformatics/btp711

[41] YUAN, M. and CAI, T. T. (2010). A reproducing kernel Hilbert space approach to functional linear regression. *Ann. Statist.* **38** 3412–3444. MR2766857 https://doi.org/10.1214/09-AOS772

[42] ZHOU, L., LIN, H. and LIANG, H. (2018). Efficient estimation of the nonparametric mean and covariance functions for longitudinal and sparse functional data. *J. Amer. Statist. Assoc.* **113** 1550–1564. MR3902229 https://doi.org/10.1080/01621459.2017.1356317

The Annals of Statistics
2024, Vol. 52, No. 2, 679–707
https://doi.org/10.1214/24-AOS2363

FINDING THE OPTIMAL DYNAMIC TREATMENT REGIMES USING SMOOTH FISHER CONSISTENT SURROGATE LOSS

BY NILANJANA LAHA[1,a], AARON SONABEND-W[2,b], RAJARSHI MUKHERJEE[2,c] AND TIANXI CAI[2,d]

[1]*Department of Statistics, Texas A&M University,* [a]*nlaha@tamu.edu*
[2]*Department of Biostatistics, Harvard University,* [b]*asonabend@gmail.com,* [c]*ram521@mail.harvard.edu,* [d]*tcai@hsph.harvard.edu*

Large health care data repositories such as electronic health records (EHR) open new opportunities to derive individualized treatment strategies for complicated diseases such as sepsis. In this paper, we consider the problem of estimating sequential treatment rules tailored to a patient's individual characteristics, often referred to as dynamic treatment regimes (DTRs). Our main objective is to find the optimal DTR that maximizes a discontinuous value function through direct maximization of Fisher consistent surrogate loss functions. In this regard, we demonstrate that a large class of concave surrogates fails to be Fisher consistent—a behavior that differs from the classical binary classification problems. We further characterize a nonconcave family of Fisher consistent smooth surrogate functions, which is amenable to gradient-descent type optimization algorithms. Compared to the existing direct search approach under the support vector machine framework (*J. Amer. Statist. Assoc.* **110** (2015) 583–598), our proposed DTR estimation via surrogate loss optimization (DTRESLO) method is more computationally scalable to large sample sizes and allows for broader functional classes for treatment policies. We establish theoretical properties for our proposed DTR estimator and obtain a sharp upper bound on the regret corresponding to our DTRESLO method. The finite sample performance of our proposed estimator is evaluated through extensive simulations. We also illustrate the working principles and benefits of our method for estimating an optimal DTR for treating sepsis using EHR data from sepsis patients admitted to intensive care units.

REFERENCES

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

BARTLETT, P. L., JORDAN, M. I. and MCAULIFFE, J. D. (2006). Convexity, classification, and risk bounds. *J. Amer. Statist. Assoc.* **101** 138–156. MR2268032 https://doi.org/10.1198/016214505000000907

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

BLANCHARD, G., BOUSQUET, O. and MASSART, P. (2008). Statistical performance of support vector machines. *Ann. Statist.* **36** 489–531. MR2396805 https://doi.org/10.1214/009053607000000839

CALAUZENES, C., USUNIER, N. and GALLINARI, P. (2012). On the (non-) existence of convex, calibrated surrogate losses for ranking. *Adv. Neural Inf. Process. Syst.* **25** 197–205.

CHAKRABORTY, B. and MOODIE, E. E. M. (2013). *Statistical Methods for Dynamic Treatment Regimes: Reinforcement Learning, Causal Inference, and Personalized Medicine. Statistics for Biology and Health*. Springer, New York. MR3112454 https://doi.org/10.1007/978-1-4614-7428-9

CHEN, G., ZENG, D. and KOSOROK, M. R. (2016). Personalized dose finding using outcome weighted learning. *J. Amer. Statist. Assoc.* **111** 1509–1521. MR3601705 https://doi.org/10.1080/01621459.2016.1148611

CHEN, S., TIAN, L., CAI, T. and YU, M. (2017). A general statistical framework for subgroup identification and comparative treatment scoring. *Biometrics* **73** 1199–1209. MR3744534 https://doi.org/10.1111/biom.12676

Key words and phrases. Dynamic treatment regimes, classification, empirical risk minimization, nonconvex optimization.

CUI, Y. and TCHETGEN TCHETGEN, E. (2021). A semiparametric instrumental variable approach to optimal treatment regimes under endogeneity. *J. Amer. Statist. Assoc.* **116** 162–173. MR4227683 https://doi.org/10.1080/01621459.2020.1783272

DEMBCZYŃSKI, K., WAEGEMAN, W., CHENG, W. and HÜLLERMEIER, E. (2012). On label dependence and loss minimization in multi-label classification. *Mach. Learn.* **88** 5–45. MR2942603 https://doi.org/10.1007/s10994-012-5285-8

DUCHI, J., KHOSRAVI, K. and RUAN, F. (2018). Multiclass classification, information, divergence and surrogate risk. *Ann. Statist.* **46** 3246–3275. MR3852651 https://doi.org/10.1214/17-AOS1657

DUCHI, J. C., MACKEY, L. W. and JORDAN, M. I. (2010). On the consistency of ranking algorithms. In *ICML*.

FENG, H., NING, Y. and ZHAO, J. (2022). Nonregular and minimax estimation of individualized thresholds in high dimension with binary responses. *Ann. Statist.* **50** 2284–2305. MR4474491 https://doi.org/10.1214/22-aos2188

GAO, W. and ZHOU, Z.-H. (2011). On the consistency of multi-label learning. In *Proceedings of the* 24*th Annual Conference on Learning Theory* 341–358.

HIRIART-URRUTY, J.-B. and LEMARÉCHAL, C. (2001). *Fundamentals of Convex Analysis*. *Grundlehren Text Editions*. Springer, Berlin. Abridged version of ıt Convex analysis and minimization algorithms. I [Springer, Berlin, 1993; MR1261420 (95m:90001)] and ıt II [ibid.; MR1295240 (95m:90002)]. MR1865628 https://doi.org/10.1007/978-3-642-56468-0

HOROWITZ, J. L. (1992). A smoothed maximum score estimator for the binary response model. *Econometrica* **60** 505–531. MR1162997 https://doi.org/10.2307/2951582

JIANG, B., SONG, R., LI, J. and ZENG, D. (2019). Entropy learning for dynamic treatment regimes. *Statist. Sinica* **29** 1633–1656. MR3970323

KALLUS, N. (2018). Balanced policy evaluation and learning. *Adv. Neural Inf. Process. Syst.* **31**.

KALLUS, N. (2019). Discussion: "Entropy learning for dynamic treatment regimes" [MR3970323]. *Statist. Sinica* **29** 1697–1705. MR3970330

KARP, R. M. (1972). Reducibility among combinatorial problems. In *Complexity of Computer Computations* (*Proc. Sympos.*, *IBM Thomas J. Watson Res. Center, Yorktown Heights, N.Y.*, 1972). *The IBM Research Symposia Series* 85–103. Plenum, New York. MR0378476

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

KOSOROK, M. R. and LABER, E. B. (2019). Precision medicine. *Annu. Rev. Stat. Appl.* **6** 263–286. MR3939521 https://doi.org/10.1146/annurev-statistics-030718-105251

LABER, E. B. and DAVIDIAN, M. (2017). Dynamic treatment regimes, past, present, and future: A conversation with experts. *Stat. Methods Med. Res.* **26** 1605–1610. MR3687166 https://doi.org/10.1177/0962280217708661

LABER, E. B., LIZOTTE, D. J., QIAN, M., PELHAM, W. E. and MURPHY, S. A. (2014). Dynamic treatment regimes: Technical challenges and applications. *Electron. J. Stat.* **8** 1225–1272. MR3263118 https://doi.org/10.1214/14-EJS920

LABER, E. B. and ZHAO, Y. Q. (2015). Tree-based methods for individualized treatment regimes. *Biometrika* **102** 501–514. MR3394271 https://doi.org/10.1093/biomet/asv028

LAHA, N., SONABEND-W, A., MUKHERJEE, R. and CAI, T. (2024). Supplement to "Finding the optimal dynamic treatment regimes using smooth Fisher consistent surrogate loss." https://doi.org/10.1214/24-AOS2363SUPP

LIN, Y. (2004). A note on margin-based loss functions in classification. *Statist. Probab. Lett.* **68** 73–82. MR2064687 https://doi.org/10.1016/j.spl.2004.03.002

LIU, Y. (2007). Fisher consistency of multicategory support vector machines. In *Artificial Intelligence and Statistics* 291–298.

LIU, Y. and SHEN, X. (2006). Multicategory ψ-learning. *J. Amer. Statist. Assoc.* **101** 500–509. MR2256170 https://doi.org/10.1198/016214505000000781

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

MOODIE, E. E., DEAN, N. and SUN, Y. R. (2014). Q-learning: Flexible learning about useful utilities. *Stat. Biosci.* **6** 223–243.

MUKHERJEE, D., BANERJEE, M. and RITOV, Y. (2021). Optimal linear discriminators for the discrete choice model in growing dimensions. *Ann. Statist.* **49** 3324–3357. MR4352532 https://doi.org/10.1214/21-aos2085

MURPHY, S. A. (2003). Optimal dynamic treatment regimes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 331–366. MR1983752 https://doi.org/10.1111/1467-9868.00389

MURPHY, S. A. (2005). A generalization error for Q-learning. *J. Mach. Learn. Res.* **6** 1073–1097. MR2249849

MURPHY, S. A., VAN DER LAAN, M. J. and ROBINS, J. M. (2001). Marginal mean models for dynamic regimes. *J. Amer. Statist. Assoc.* **96** 1410–1423. MR1946586 https://doi.org/10.1198/016214501753382327

NEYKOV, M., LIU, J. S. and CAI, T. (2016). On the characterization of a class of Fisher-consistent loss functions and its application to boosting. *J. Mach. Learn. Res.* **17** Paper No. 70, 32. MR3517093

ORELLANA, L., ROTNITZKY, A. and ROBINS, J. M. (2010). Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, Part I: Main content. *Int. J. Biostat.* **6** Art. 8, 49. MR2602551 https://doi.org/10.2202/1557-4679.1200

PEDREGOSA, F., BACH, F. and GRAMFORT, A. (2017). On the consistency of ordinal regression methods. *J. Mach. Learn. Res.* **18** Paper No. 55, 35. MR3687598

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

ROBINS, J. M. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Comm. Statist. Theory Methods* **23** 2379–2412. MR1293185 https://doi.org/10.1080/03610929408831393

ROBINS, J. M. (1997). Causal inference from complex longitudinal data. In *Latent Variable Modeling and Applications to Causality* (*Los Angeles*, *CA*, 1994). *Lect. Notes Stat.* **120** 69–117. Springer, New York. MR1601279 https://doi.org/10.1007/978-1-4612-1842-5_4

ROBINS, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium in Biostatistics*. *Lect. Notes Stat.* **179** 189–326. Springer, New York. MR2129402 https://doi.org/10.1007/978-1-4419-9076-1_11

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

SCHULTE, P. J., TSIATIS, A. A., LABER, E. B. and DAVIDIAN, M. (2014). Q- and A-learning methods for estimating optimal dynamic treatment regimes. *Statist. Sci.* **29** 640–661. MR3300363 https://doi.org/10.1214/13-STS450

SONABEND-W, A. (2022). DTR estimation via surrogate loss. Available at https://github.com/asonabend?tab=repositories. Github code for "Finding the optimal dynamic treatment regimes using smooth Fisher consistent surrogate loss".

SONABEND-W, A., LAHA, N., ANANTHAKRISHNAN, A. N., CAI, T. and MUKHERJEE, R. (2023). Semi-supervised off-policy reinforcement learning and value estimation for dynamic treatment regimes. *J. Mach. Learn. Res.* **24** 86. MR4690272

SONG, R., KOSOROK, M., ZENG, D., ZHAO, Y., LABER, E. and YUAN, M. (2015). On sparse representation for optimal individualized treatment selection with penalized outcome weighted learning. *Stat* **4** 59–68. MR3405390 https://doi.org/10.1002/sta4.78

STEINWART, I. and SCOVEL, C. (2007). Fast rates for support vector machines using Gaussian kernels. *Ann. Statist.* **35** 575–607. MR2336860 https://doi.org/10.1214/009053606000001226

SUN, Y. and WANG, L. (2021). Stochastic tree search for estimating optimal dynamic treatment regimes. *J. Amer. Statist. Assoc.* **116** 421–432. MR4227704 https://doi.org/10.1080/01621459.2020.1819294

TEWARI, A. and BARTLETT, P. L. (2007). On the consistency of multiclass classification methods. *J. Mach. Learn. Res.* **8** 1007–1025. MR2320680 https://doi.org/10.1007/11503415_10

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

WANG, R., FOSTER, D. P. and KAKADE, S. M. (2020). What are the statistical limits of offline RL with linear function approximation? arXiv preprint. Available at arXiv:2010.11895.

WATKINS, C. J. C. H. (1989). Learning from delayed rewards.

XU, T., WANG, J. and FANG, Y. (2014). A model-free estimation for the covariate-adjusted Youden index and its associated cut-point. *Stat. Med.* **33** 4963–4974. MR3276512 https://doi.org/10.1002/sim.6290

XU, Y., MÜLLER, P., WAHED, A. S. and THALL, P. F. (2016). Bayesian nonparametric estimation for dynamic treatment regimes with sequential transition times. *J. Amer. Statist. Assoc.* **111** 921–950. MR3561917 https://doi.org/10.1080/01621459.2015.1086353

ZAJONC, T. (2012). Bayesian inference for dynamic treatment regimes: Mobility, equity, and efficiency in student tracking. *J. Amer. Statist. Assoc.* **107** 80–92. MR2949343 https://doi.org/10.1080/01621459.2011.643747

ZHANG, J., LIU, T. and TAO, D. (2022). On the rates of convergence from surrogate risk minimizers to the Bayes optimal classifier. *IEEE Trans. Neural Netw. Learn. Syst.* **33** 5766–5774. MR4497102

ZHANG, T. (2010). Analysis of multi-stage convex relaxation for sparse regularization. *J. Mach. Learn. Res.* **11** 1081–1107. MR2629825

ZHANG, Y., LABER, E. B., DAVIDIAN, M. and TSIATIS, A. A. (2018). Estimation of optimal treatment regimes using lists. *J. Amer. Statist. Assoc.* **113** 1541–1549. MR3902228 https://doi.org/10.1080/01621459.2017.1345743

ZHAO, Y., ZENG, D., RUSH, A. J. and KOSOROK, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *J. Amer. Statist. Assoc.* **107** 1106–1118. MR3010898 https://doi.org/10.1080/01621459.2012.695674

ZHAO, Y.-Q., ZENG, D., LABER, E. B. and KOSOROK, M. R. (2015). New statistical learning methods for estimating optimal dynamic treatment regimes. *J. Amer. Statist. Assoc.* **110** 583–598. MR3367249 https://doi.org/10.1080/01621459.2014.937488

ZHOU, X. and KOSOROK, M. R. (2017). Augmented outcome-weighted learning for optimal treatment regimes. arXiv preprint. Available at arXiv:1711.10654.

The Annals of Statistics
2024, Vol. 52, No. 2, 708–728
https://doi.org/10.1214/24-AOS2365

EDGE DIFFERENTIALLY PRIVATE ESTIMATION IN THE β-MODEL VIA JITTERING AND METHOD OF MOMENTS

BY JINYUAN CHANG[1,a], QIAO HU[1,b], ERIC D. KOLACZYK[2,c], QIWEI YAO[3,d] AND FENGTING YI[4,e]

[1]*Joint Laboratory of Data Science and Business Intelligence, Southwestern University of Finance and Economics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences,* [a]*changjinyuan@swufe.edu.cn,* [b]*huqiao@smail.swufe.edu.cn*

[2]*Department of Mathematics and Statistics, McGill University,* [c]*eric.kolaczyk@mcgill.ca*

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

[4]*Yunnan Key Laboratory of Statistical Modeling and Data Analysis, Yunnan University,* [e]*yifengting@ynu.edu.cn*

A standing challenge in data privacy is the trade-off between the level of privacy and the efficiency of statistical inference. Here, we conduct an in-depth study of this trade-off for parameter estimation in the β-model (*Ann. Appl. Probab.* **21** (2011) 1400–1435) for edge differentially private network data released via jittering (*J. R. Stat. Soc. Ser. C. Appl. Stat.* **66** (2017) 481–500). Unlike most previous approaches based on maximum likelihood estimation for this network model, we proceed via the method of moments. This choice facilitates our exploration of a substantially broader range of privacy levels—corresponding to stricter privacy—than has been to date. Over this new range, we discover our proposed estimator for the parameters exhibits an interesting phase transition, with both its convergence rate and asymptotic variance following one of three different regimes of behavior depending on the level of privacy. Because identification of the operable regime is difficult, if not impossible in practice, we devise a novel adaptive bootstrap procedure to construct uniform inference across different phases. In fact, leveraging this bootstrap we are able to provide for simultaneous inference of all parameters in the β-model (i.e., equal to the number of nodes), which, to our best knowledge, is the first result of its kind. Numerical experiments confirm the competitive and reliable finite sample performance of the proposed inference methods, next to a comparable maximum likelihood method, as well as significant advantages in terms of computational speed and memory.

REFERENCES

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

BLOCKI, J., BLUM, A., DATTA, A. and SHEFFET, O. (2013). Differentially private data analysis of social networks via restricted sensitivity. In *ITCS'*13*—Proceedings of the* 2013 *ACM Conference on Innovations in Theoretical Computer Science* 87–96. ACM, New York. MR3385388

CHANG, J., CHEN, X. and WU, M. (2024). Central limit theorems for high dimensional dependent data. *Bernoulli* **30** 712–742. MR4665595 https://doi.org/10.3150/23-bej1614

CHANG, J., HU, Q., KOLACZYK, E. D., YAO, Q. and YI, F. (2024). Supplement to "Edge differentially private estimation in the β-model via jittering and method of moments." https://doi.org/10.1214/24-AOS2365SUPP

CHANG, J., KOLACZYK, E. D. and YAO, Q. (2020). Discussion of 'Network cross-validation by edge sampling' [4108931]. *Biometrika* **107** 277–280. MR4496202 https://doi.org/10.1093/biomet/asaa017

CHANG, J., KOLACZYK, E. D. and YAO, Q. (2022). Estimation of subgraph densities in noisy networks. *J. Amer. Statist. Assoc.* **117** 361–374. MR4399091 https://doi.org/10.1080/01621459.2020.1778482

CHANG, J., QIU, Y., YAO, Q. and ZOU, T. (2018). Confidence regions for entries of a large precision matrix. *J. Econometrics* **206** 57–82. MR3840783 https://doi.org/10.1016/j.jeconom.2018.03.020

MSC2020 subject classifications. Primary 62F12, 68P27; secondary 91D30.

Key words and phrases. Adaptive inference, β-model, bootstrap inference, data privacy, data release mechanism, edge differential privacy, phase transition.

CHANG, J., ZHENG, C., ZHOU, W.-X. and ZHOU, W. (2017). Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity. *Biometrics* **73** 1300–1310. MR3744543 https://doi.org/10.1111/biom.12695

CHATTERJEE, S., DIACONIS, P. and SLY, A. (2011). Random graphs with a given degree sequence. *Ann. Appl. Probab.* **21** 1400–1435. MR2857452 https://doi.org/10.1214/10-AAP728

CHEN, M., KATO, K. and LENG, C. (2021). Analysis of networks via the sparse β-model. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 887–910. MR4349121 https://doi.org/10.1111/rssb.12444

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

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. MR3693963 https://doi.org/10.1214/16-AOP1113

DE LA PEÑA, V. H. and MONTGOMERY-SMITH, S. J. (1995). Decoupling inequalities for the tail probabilities of multivariate U-statistics. *Ann. Probab.* **23** 806–816. MR1334173

DWORK, C. (2006). Differential privacy. In *33rd International Colloquium on Automata, Languages and Programming. Lecture Notes. Comput. Sci.* **4052** 1–12. Springer, Berlin. MR2307219 https://doi.org/10.1007/11787006_1

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

FAN, Y., ZHANG, H. and YAN, T. (2020). Asymptotic theory for differentially private generalized β-models with parameters increasing. *Stat. Interface* **13** 385–398. MR4091804 https://doi.org/10.4310/SII.2020.v13.n3.a8

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

HENNIG, C. (2007). Cluster-wise assessment of cluster stability. *Comput. Statist. Data Anal.* **52** 258–271. MR2409980 https://doi.org/10.1016/j.csda.2006.11.025

JIANG, H., PEI, J., YU, D., YU, J., GONG, B. and CHENG, X. (2020). Applications of differential privacy in social network analysis: A survey. *IEEE Trans. Knowl. Data Eng.* **35** 108–127.

KARWA, V., KRIVITSKY, P. N. and SLAVKOVIĆ, A. B. (2017). Sharing social network data: Differentially private estimation of exponential family random-graph models. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **66** 481–500. MR3632338 https://doi.org/10.1111/rssc.12185

KARWA, V. and SLAVKOVIĆ, A. (2016). Inference using noisy degrees: Differentially private β-model and synthetic graphs. *Ann. Statist.* **44** 87–112. MR3449763 https://doi.org/10.1214/15-AOS1358

MUKHERJEE, R., MUKHERJEE, S. and SEN, S. (2018). Detection thresholds for the β-model on sparse graphs. *Ann. Statist.* **46** 1288–1317. MR3798004 https://doi.org/10.1214/17-AOS1585

NISSIM, K., RASKHODNIKOVA, S. and SMITH, A. (2007). Smooth sensitivity and sampling in private data analysis. In *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing* 75–84. ACM, New York. MR2402430 https://doi.org/10.1145/1250790.1250803

RINALDO, A., PETROVIĆ, S. and FIENBERG, S. E. (2013). Maximum likelihood estimation in the β-model. *Ann. Statist.* **41** 1085–1110. MR3113804 https://doi.org/10.1214/12-AOS1078

STEIN, S. and LENG, C. (2020). A sparse β-model with covariates for networks. Preprint. Available at arXiv:2010.13604.

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

YAN, T. and XU, J. (2013). A central limit theorem in the β-model for undirected random graphs with a diverging number of vertices. *Biometrika* **100** 519–524. MR3068452 https://doi.org/10.1093/biomet/ass084

ZHANG, Y., WANG, Q., ZHANG, Y., YAN, T. and LUO, J. (2021). L-2 regularized maximum likelihood for β-model in large and sparse networks. Preprint. Available at arXiv:2110.11856.

The Annals of Statistics
2024, Vol. 52, No. 2, 729–756
https://doi.org/10.1214/24-AOS2366

INFERENCE FOR HETEROSKEDASTIC PCA WITH MISSING DATA

BY YULING YAN[1,a], YUXIN CHEN[2,b] AND JIANQING FAN[3,c]

[1]*Institute for Data, Systems, and Society, Massachusetts Institute of Technology,* [a]*yulingy@mit.edu*
[2]*Department of Statistics and Data Science, Wharton School, University of Pennsylvania,* [b]*yuxinc@wharton.upenn.edu*
[3]*Department of Operations Research and Financial Engineering, Princeton University,* [c]*jqfan@princeton.edu*

This paper studies how to construct confidence regions for principal component analysis (PCA) in high dimension, a problem that has been vastly underexplored. While computing measures of uncertainty for nonlinear/nonconvex estimators is in general difficult in high dimension, the challenge is further compounded by the prevalent presence of missing data and heteroskedastic noise. We propose a novel approach to performing valid inference on the principal subspace, on the basis of an estimator called HeteroPCA (*Ann. Statist.* **50** (2022b) 53–80). We develop nonasymptotic distributional guarantees for HeteroPCA, and demonstrate how these can be invoked to compute both confidence regions for the principal subspace and entrywise confidence intervals for the spiked covariance matrix. Our inference procedures are fully data-driven and adaptive to heteroskedastic random noise, without requiring prior knowledge about the noise levels.

REFERENCES

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

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

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

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

BAO, Z., DING, X., WANG, J. and WANG, K. (2022a). 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., WANG, J. and WANG, K. (2022b). 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

BLOEMENDAL, A., KNOWLES, A., YAU, H.-T. and YIN, J. (2016). On the principal components of sample covariance matrices. *Probab. Theory Related Fields* **164** 459–552. MR3449395 https://doi.org/10.1007/s00440-015-0616-x

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, C., POOR, H. V. and CHEN, Y. (2023). Uncertainty quantification for nonconvex tensor completion: Confidence intervals, heteroscedasticity and optimality. *IEEE Trans. Inf. Theory* **69** 407–452. MR4544966

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

MSC2020 subject classifications. Primary 62H25; secondary 62E17.
Key words and phrases. Principal component analysis, confidence regions, missing data, uncertainty quantification, heteroskedastic data, subspace estimation.

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

CANDÈS, E. J. (2014). Mathematics of sparsity (and a few other things). In *Proceedings of the International Congress of Mathematicians—Seoul* 2014, *Vol.* 1 235–258. Kyung Moon Sa, Seoul. MR3728471

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

CAPE, J., TANG, M. and PRIEBE, C. E. (2019). The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *Ann. Statist.* **47** 2405–2439. MR3988761 https://doi.org/10.1214/18-AOS1752

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

CHEN, J., LIU, D. and LI, X. (2020). Nonconvex rectangular matrix completion via gradient descent without $\ell_{2,\infty}$ regularization. *IEEE Trans. Inf. Theory* **66** 5806–5841. MR4158648 https://doi.org/10.1109/TIT.2020.2992234

CHEN, P., GAO, C. and ZHANG, A. Y. (2022). Partial recovery for top-k ranking: Optimality of MLE and suboptimality of the spectral method. *Ann. Statist.* **50** 1618–1652. MR4441134 https://doi.org/10.1214/21-aos2166

CHEN, Y., CHENG, C. and FAN, J. (2021). Asymmetry helps: Eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices. *Ann. Statist.* **49** 435–458. MR4206685 https://doi.org/10.1214/20-AOS1963

CHEN, Y., CHI, Y., FAN, J. and MA, C. (2019). Gradient descent with random initialization: Fast global convergence for nonconvex phase retrieval. *Math. Program.* **176** 5–37. MR3960803 https://doi.org/10.1007/s10107-019-01363-6

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

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 WANG, K. (2019a). Spectral method and regularized MLE are both optimal for top-K ranking. *Ann. Statist.* **47** 2204–2235. MR3953449 https://doi.org/10.1214/18-AOS1745

CHEN, Y., FAN, J., MA, C. and YAN, Y. (2019b). 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

CHEN, Y. and WAINWRIGHT, M. J. (2015). Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees. Available at arXiv:1509.03025.

CHENG, C., WEI, Y. and CHEN, Y. (2021). Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise. *IEEE Trans. Inf. Theory* **67** 7380–7419. MR4345128 https://doi.org/10.1109/TIT.2021.3111828

CHERNOZHUKOV, V., HANSEN, C., LIAO, Y. and ZHU, Y. (2023). Inference for low-rank models. *Ann. Statist.* **51** 1309–1330. MR4630950 https://doi.org/10.1214/23-aos2293

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

CHO, J., KIM, D. and ROHE, K. (2017). Asymptotic theory for estimating the singular vectors and values of a partially-observed low rank matrix with noise. *Statist. Sinica* **27** 1921–1948. MR3726772

DAVIS, C. and KAHAN, W. M. (1970). The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.* **7** 1–46. MR0264450 https://doi.org/10.1137/0707001

DING, X. (2020). High dimensional deformed rectangular matrices with applications in matrix denoising. *Bernoulli* **26** 387–417. MR4036038 https://doi.org/10.3150/19-BEJ1129

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

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

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

ELDRIDGE, J., BELKIN, M. and WANG, Y. (2018). Unperturbed: Spectral analysis beyond Davis–Kahan. In *Proceedings of Algorithmic Learning Theory. Proc. Mach. Learn. Res.* (*PMLR*) **83** 38. PMLR. MR3857310

FAN, J., FAN, Y., HAN, X. and LV, J. (2022). Asymptotic theory of eigenvectors for random matrices with diverging spikes. *J. Amer. Statist. Assoc.* **117** 996–1009. MR4436328 https://doi.org/10.1080/01621459.2020.1840990

FAN, J., LI, K. and LIAO, Y. (2021). Recent developments on factor models and its applications in econometric learning. *Annu. Rev. Financ. Econ.* **13** 401–430.

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

FAN, J., WANG, W. and ZHONG, Y. (2017). An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. *J. Mach. Learn. Res.* **18** Paper No. 207, 42. MR3827095

FAN, J. and YAO, Q. (2017). *The Elements of Financial Econometrics*. Cambridge Univ. Press, Cambridge.

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

GAGLIARDINI, P., OSSOLA, E. and SCAILLET, O. (2020). Estimation of large dimensional conditional factor models in finance. In *Handbook of Econometrics, Vol. 7A*. *Handbooks in Econom.* 219–282. Elsevier, Amsterdam. MR4254181 https://doi.org/10.1016/bs.hoe.2020.10.001

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

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* **106** 1277–1292. https://doi.org/10.1109/JPROC.2018.2846730

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

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

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

KOLTCHINSKII, V., LÖFFLER, M. and NICKL, R. (2020). Efficient estimation of linear functionals of principal components. *Ann. Statist.* **48** 464–490. MR4065170 https://doi.org/10.1214/19-AOS1816

LEI, L. (2019). Unified $\ell_{2,\infty}$ eigenspace perturbation theory for symmetric random matrices. arXiv preprint. Available at arXiv:1909.04798.

LI, G., CAI, C., GU, Y.,, POOR, H. V. and CHEN, Y. (2021). Minimax Estimation of Linear Functions of Eigenvectors in the Face of Small Eigen-Gaps. arXiv preprint. Available at arXiv:2104.03298.

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

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., RUAN, F. and YAN, J. (2018). Adapting to unknown noise distribution in matrix denoising. arXiv preprint. Available at arXiv:1810.02954.

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

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

NEGAHBAN, S. and WAINWRIGHT, M. J. (2012). Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *J. Mach. Learn. Res.* **13** 1665–1697. MR2930649

NING, Y. and LIU, H. (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *Ann. Statist.* **45** 158–195. MR3611489 https://doi.org/10.1214/16-AOS1448

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

PORTER, M. E. et al. (1998). *Clusters and the New Economics of Competition* 76. Harvard Business Review, Boston, MA.

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

SCHÖNEMANN, P. H. (1966). A generalized solution of the orthogonal Procrustes problem. *Psychometrika* **31** 1–10. MR0215870 https://doi.org/10.1007/BF02289451

SUN, R. and LUO, Z.-Q. (2016). Guaranteed matrix completion via non-convex factorization. *IEEE Trans. Inf. Theory* **62** 6535–6579. MR3565131 https://doi.org/10.1109/TIT.2016.2598574

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

VASWANI, N., CHI, Y. and BOUWMANS, T. (2018). Rethinking PCA for modern data sets: Theory, algorithms, and applications. *Proc. IEEE* **106** 1274–1276.

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

WAHBA, G. (1965). A least squares estimate of satellite attitude. *SIAM Rev.* **7** 409–409.

XIA, D. (2019). Confidence region of singular subspaces for low-rank matrix regression. *IEEE Trans. Inf. Theory* **65** 7437–7459. MR4030894 https://doi.org/10.1109/TIT.2019.2924900

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. 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

XIE, F. (2021). Entrywise limit theorems of eigenvectors and their one-step refinement for sparse random graphs. arXiv preprint. Available at arXiv:2106.09840.

YAN, Y., CHEN, Y. and FAN, J. (2021). Inference for heteroskedastic PCA with missing data (full version). arXiv preprint. Available at arXiv:2107.12365.

YAN, Y., CHEN, Y. and FAN, J. (2024). Supplement to "Inference for heteroskedastic PCA with missing data." https://doi.org/10.1214/24-AOS2366SUPP

YAN, Y. and WAINWRIGHT, M. J. (2024). Entrywise inference for causal panel data: A simple and instance-optimal approach. arXiv preprint. Available at arXiv:2401.13665.

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

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

ZHENG, Q. and LAFFERTY, J. (2016). Convergence analysis for rectangular matrix completion using Burer–Monteiro factorization and gradient descent. Available at arXiv:1605.07051.

ZHONG, Y. and BOUMAL, N. (2018). Near-optimal bound for phase synchronization. *SIAM J. Optim.* MR3566919 https://doi.org/10.1137/16M105808X

ZHOU, Y. and CHEN, Y. (2023a). Deflated HeteroPCA: Overcoming the curse of ill-conditioning in heteroskedastic PCA. arXiv preprint. Available at arXiv:2303.06198.

ZHOU, Y. and CHEN, Y. (2023b). Heteroskedastic tensor clustering. arXiv preprint. Available at arXiv:2311.02306/3.

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
2024, Vol. 52, No. 2, 757–792
https://doi.org/10.1214/24-AOS2368

METRIC STATISTICS: EXPLORATION AND INFERENCE FOR RANDOM OBJECTS WITH DISTANCE PROFILES

BY PAROMITA DUBEY[1,a], YAQING CHEN[2,b] AND HANS-GEORG MÜLLER[3,c]

[1]*Department of Data Sciences and Operations, Marshall School of Business, University of Southern California,* [a]*paromita@marshall.usc.edu*

[2]*Department of Statistics, Rutgers University,* [b]*yqchen@stat.rutgers.edu*

[3]*Department of Statistics, University of California, Davis,* [c]*hgmueller@ucdavis.edu*

This article provides an overview on the statistical modeling of complex data as increasingly encountered in modern data analysis. It is argued that such data can often be described as elements of a metric space that satisfies certain structural conditions and features a probability measure. We refer to the random elements of such spaces as random objects and to the emerging field that deals with their statistical analysis as metric statistics. Metric statistics provides methodology, theory and visualization tools for the statistical description, quantification of variation, centrality and quantiles, regression and inference for populations of random objects, inferring these quantities from available data and samples. In addition to a brief review of current concepts, we focus on distance profiles as a major tool for object data in conjunction with the pairwise Wasserstein transports of the underlying one-dimensional distance distributions. These pairwise transports lead to the definition of intuitive and interpretable notions of transport ranks and transport quantiles as well as two-sample inference. An associated profile metric complements the original metric of the object space and may reveal important features of the object data in data analysis. We demonstrate these tools for the analysis of complex data through various examples and visualizations.

REFERENCES

AHIDAR-COUTRIX, A., LE GOUIC, T. and PARIS, Q. (2020). Convergence rates for empirical barycenters in metric spaces: Curvature, convexity and extendable geodesics. *Probab. Theory Related Fields* **177** 323–368. MR4095017 https://doi.org/10.1007/s00440-019-00950-0

AITCHISON, J. (1986). *The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability*. CRC Press, London. MR0865647 https://doi.org/10.1007/978-94-009-4109-0

AMBROSIO, L., GIGLI, N. and SAVARÉ, G. (2008). *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed. *Lectures in Mathematics ETH Zürich*. Birkhäuser, Basel. MR2401600

BARABÁSI, A.-L. and ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. MR2091634 https://doi.org/10.1126/science.286.5439.509

BARDEN, D., LE, H. and OWEN, M. (2018). Limiting behaviour of Fréchet means in the space of phylogenetic trees. *Ann. Inst. Statist. Math.* **70** 99–129. MR3742820 https://doi.org/10.1007/s10463-016-0582-9

BHATTACHARJEE, S., LI, B. and XUE, L. (2023). Nonlinear global Fréchet regression for random objects via weak conditional expectation. arXiv preprint. Available at arXiv:2310.07817.

BHATTACHARJEE, S. and MÜLLER, H.-G. (2023). Single index Fréchet regression. *Ann. Statist.* **51** 1770–1798. MR4658576 https://doi.org/10.1214/23-aos2307

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

BILLARD, L. and DIDAY, E. (2003). From the statistics of data to the statistics of knowledge: Symbolic data analysis. *J. Amer. Statist. Assoc.* **98** 470–487. MR1982575 https://doi.org/10.1198/016214503000242

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

Key words and phrases. Distributional data, Fréchet mean, Fréchet regression, functional data analysis, metric variance, profile metric, transport rank, transport quantile, visualization, Wasserstein metric.

BISWAL, B., YETKIN, F. Z., HAUGHTON, V. M. and HYDE, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* **34** 537–541. https://doi.org/10.1002/mrm.1910340409

BLEI, R., GAO, F. and LI, W. V. (2007). Metric entropy of high dimensional distributions. *Proc. Amer. Math. Soc.* **135** 4009–4018. MR2341952 https://doi.org/10.1090/S0002-9939-07-08935-6

BOLSTAD, B. M., IRIZARRY, R. A., ÅSTRAND, M. and SPEED, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19** 185–193.

BURAGO, D., BURAGO, Y. and IVANOV, S. (2001). *A Course in Metric Geometry*. *Graduate Studies in Mathematics* **33**. Amer. Math. Soc., Providence, RI. MR1835418 https://doi.org/10.1090/gsm/033

CHAVEL, I. (2006). *Riemannian Geometry*: *A Modern Introduction*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **98**. Cambridge Univ. Press, Cambridge. MR2229062 https://doi.org/10.1017/CBO9780511616822

CHEN, H. and FRIEDMAN, J. H. (2017). A new graph-based two-sample test for multivariate and object data. *J. Amer. Statist. Assoc.* **112** 397–409. MR3646580 https://doi.org/10.1080/01621459.2016.1147356

CHEN, K., DELICADO, P. and MÜLLER, H.-G. (2017). Modelling function-valued stochastic processes, with applications to fertility dynamics. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 177–196. MR3597969 https://doi.org/10.1111/rssb.12160

CHEN, Y., DUBEY, P. and MÜLLER, H.-G. (2024). ODP: Exploration for random objects using distance profiles R package version 0.1.0. Available at https://github.com/yqgchen/ODP.

CHEN, Y., GAJARDO, A., FAN, J., ZHONG, Q., DUBEY, P., HAN, K., BHATTACHARJEE, S. and MÜLLER, H.-G. (2020). frechet: Statistical analysis for random objects and non-Euclidean data. R package version 0.2.0. Available at https://CRAN.R-project.org/package=frechet.

CHEN, H., and MÜLLER, H.-G. (2023). Sliced Wasserstein regression. arXiv preprint. Available at arXiv:2306.10601.

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

CHEN, Y. and MÜLLER, H.-G. (2022). Uniform convergence of local Fréchet regression with applications to locating extrema and time warping for metric space valued trajectories. *Ann. Statist.* **50** 1573–1592. MR4441132 https://doi.org/10.1214/21-aos2163

CHENG, M.-Y. and WU, H.-T. (2013). Local linear regression on manifolds and its geometric interpretation. *J. Amer. Statist. Assoc.* **108** 1421–1434. MR3174718 https://doi.org/10.1080/01621459.2013.827984

CHOLAQUIDIS, A., FRAIMAN, R. and MORENO, L. (2023). Level sets of depth measures in abstract spaces. *TEST* **32** 942–957. MR4656906 https://doi.org/10.1007/s11749-023-00858-x

CORNEA, E., ZHU, H., KIM, P. and IBRAHIM, J. G. (2017). Regression models on Riemannian symmetric spaces. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 463–482. MR3611755 https://doi.org/10.1111/rssb.12169

CUTURI, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems* 2292–2300.

DABO-NIANG, S. (2002). Estimation de la densité dans un espace de dimension infinie: Application aux diffusions. *C. R. Math. Acad. Sci. Paris* **334** 213–216. MR1891061 https://doi.org/10.1016/S1631-073X(02)02247-1

DAI, X. (2022). Statistical inference on the Hilbert sphere with application to random densities. *Electron. J. Stat.* **16** 700–736. MR4366819 https://doi.org/10.1214/21-ejs1942

DAI, X., LIN, Z. and MÜLLER, H.-G. (2021). Modeling sparse longitudinal data on Riemannian manifolds. *Biometrics* **77** 1328–1341. MR4357841 https://doi.org/10.1111/biom.13385

DAI, X. and LOPEZ-PINTADO, S. (2023). Tukey's depth for object data. *J. Amer. Statist. Assoc.* **118** 1760–1772. Authors writing for the Alzheimer's Disease Neuroimaging Initiative. MR4646604 https://doi.org/10.1080/01621459.2021.2011298

DONG, Y. and WU, Y. (2022). Fréchet kernel sliced inverse regression. *J. Multivariate Anal.* **191** Paper No. 105032, 14. MR4432318 https://doi.org/10.1016/j.jmva.2022.105032

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

DRYDEN, I. L. and MARDIA, K. V. (2016). *Statistical Shape Analysis with Applications in R*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR3559734 https://doi.org/10.1002/9781119072492

DUBEY, P., CHEN, Y. and MÜLLER, H.-G. (2024). Supplement to "Metric statistics: Exploration and inference for random objects With distance profiles." https://doi.org/10.1214/24-AOS2368SUPP

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

DUBEY, P. and MÜLLER, H.-G. (2020a). Functional models for time-varying random objects. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 275–327. MR4084166

DUBEY, P. and MÜLLER, H.-G. (2020b). Fréchet change-point detection. *Ann. Statist.* **48** 3312–3335. MR4185810 https://doi.org/10.1214/19-AOS1930

ELTZNER, B. and HUCKEMANN, S. F. (2019). A smeary central limit theorem for manifolds with application to high-dimensional spheres. *Ann. Statist.* **47** 3360–3381. MR4025745 https://doi.org/10.1214/18-AOS1781

FERAGEN, A., LAUZE, F. and HAUBERG, S. (2015). Geodesic exponential kernels: When curvature and linearity conflict. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 3032–3042.

FILZMOSER, P., HRON, K. and TEMPL, M. (2019). *Applied Compositional Data Analysis*: *With Worked Examples in R*. Springer.

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

FRISTON, K. J., FRITH, C. D., LIDDLE, P. F. and FRACKOWIAK, R. S. J. (1993). Functional connectivity: The principal-component analysis of large (PET) data sets. *J. Cereb. Blood Flow Metab.* **13** 5–14.

GAO, F. and WELLNER, J. A. (2009). On the rate of convergence of the maximum likelihood estimator of a k-monotone density. *Sci. China Ser. A* **52** 1525–1538. MR2520591 https://doi.org/10.1007/s11425-009-0102-y

GARBA, M. K., NYE, T. M. W., LUEG, J. and HUCKEMANN, S. F. (2021). Information geometry for phylogenetic trees. *J. Math. Biol.* **82** Paper No. 19, 39. MR4218000 https://doi.org/10.1007/s00285-021-01553-x

GEENENS, G., NIETO-REYES, A. and FRANCISCI, G. (2023). Statistical depth in abstract metric spaces. *Stat. Comput.* **33** Paper No. 46, 15. MR4554146 https://doi.org/10.1007/s11222-023-10216-4

GHODRATI, L. and PANARETOS, V. M. (2023). On distributional autoregression and iterated transportation. arXiv preprint. Available at arXiv:2303.09469.

GHOSAL, A., MEIRING, W. and PETERSEN, A. (2023). Fréchet single index models for object response regression. *Electron. J. Stat.* **17** 1074–1112. MR4575027 https://doi.org/10.1214/23-ejs2120

GHOSAL, R., VARMA, V. R., VOLFSON, D., HILLEL, I., URBANEK, J., HAUSDORFF, J. M., WATTS, A. and ZIPUNNIKOV, V. (2023). Distributional data analysis via quantile functions and its application to modeling digital biomarkers of gait in Alzheimer's Disease. *Biostatistics* **24** 539–561. MR4615240 https://doi.org/10.1093/biostatistics/kxab041

GINESTET, C. E., LI, J., BALACHANDRAN, P., ROSENBERG, S. and KOLACZYK, E. D. (2017). Hypothesis testing for network data in functional neuroimaging. *Ann. Appl. Stat.* **11** 725–750. MR3693544 https://doi.org/10.1214/16-AOAS1015

HRON, K., MENAFOGLIO, A., TEMPL, M., HRŮZOVÁ, K. and FILZMOSER, P. (2016). Simplicial principal component analysis for density functions in Bayes spaces. *Comput. Statist. Data Anal.* **94** 330–350. MR3412829 https://doi.org/10.1016/j.csda.2015.07.007

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

HUCKEMANN, S. F. and ELTZNER, B. (2021). Data analysis on nonstandard spaces. *Wiley Interdiscip. Rev.*: *Comput. Stat.* **13** Paper No. e1526, 19. MR4242812 https://doi.org/10.1002/wics.1526

JEON, J. M. and PARK, B. U. (2020). Additive regression with Hilbertian responses. *Ann. Statist.* **48** 2671–2697. MR4152117 https://doi.org/10.1214/19-AOS1902

JUNG, S., DRYDEN, I. L. and MARRON, J. S. (2012). Analysis of principal nested spheres. *Biometrika* **99** 551–568. MR2966769 https://doi.org/10.1093/biomet/ass022

JUNG, S., SCHWARTZMAN, A. and GROISSER, D. (2015). Scaling-rotation distance and interpolation of symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **36** 1180–1201. MR3379023 https://doi.org/10.1137/140967040

KANTOROVITCH, L. (1958). On the translocation of masses. *Manage. Sci.* **5** 1–4. MR0096552 https://doi.org/10.1287/mnsc.5.1.1

KIM, J., ROSENBERG, N. A. and PALACIOS, J. A. (2020). Distance metrics for ranked evolutionary trees. *Proc. Natl. Acad. Sci. USA* **117** 28876–28886.

KLEBANOV, L. B. (2005). *N-Distances and Their Applications*. Karolinum Press, Charles Univ. Prague, Czech Republic.

KNEIP, A. and UTIKAL, K. J. (2001). Inference for density families using functional principal component analysis. *J. Amer. Statist. Assoc.* **96** 519–542. With comments and a rejoinder by the authors. MR1946423 https://doi.org/10.1198/016214501753168235

KOLACZYK, E. D., LIN, L., ROSENBERG, S., WALTERS, J. and XU, J. (2020). Averages of unlabeled networks: Geometric characterization and asymptotic behavior. *Ann. Statist.* **48** 514–538. MR4065172 https://doi.org/10.1214/19-AOS1820

KOLOURI, S., NADJAHI, K., SIMSEKLI, U., BADEAU, R. and ROHDE, G. (2019). Generalized sliced Wasserstein distances. *Adv. Neural Inf. Process. Syst.* **32** 261–272.

KOLOURI, S., ZOU, Y. and ROHDE, G. K. (2016). Sliced Wasserstein kernels for probability distributions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 5258–5267.

LIN, Z. (2019). Riemannian geometry of symmetric positive definite matrices via Cholesky decomposition. *SIAM J. Matrix Anal. Appl.* **40** 1353–1370. MR4032859 https://doi.org/10.1137/18M1221084

LIN, Z. and MÜLLER, H.-G. (2021). Total variation regularized Fréchet regression for metric-space valued data. *Ann. Statist.* **49** 3510–3533. MR4352539 https://doi.org/10.1214/21-aos2095

LINDQUIST, M. A. (2008). The statistical analysis of fMRI data. *Statist. Sci.* **23** 439–464. MR2530545 https://doi.org/10.1214/09-STS282

LIU, R. Y. and SINGH, K. (1992). Ordering directional data: Concepts of data depth on circles and spheres. *Ann. Statist.* **20** 1468–1484. MR1186260 https://doi.org/10.1214/aos/1176348779

LUEG, J., GARBA, M. K., NYE, T. M. W. and HUCKEMANN, S. F. (2022). Foundations of the Wald space for phylogenetic trees. arXiv preprint. Available ar arXiv:2209.05332.

LUNAGÓMEZ, S., OLHEDE, S. C. and WOLFE, P. J. (2021). Modeling network populations via graph distances. *J. Amer. Statist. Assoc.* **116** 2023–2040. MR4353730 https://doi.org/10.1080/01621459.2020.1763803

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

MARDIA, K. V. (1978). Some properties of classical multi-dimensional scaling. *Comm. Statist. Theory Methods* **7** 1233–1241. MR0514645 https://doi.org/10.1080/03610927808827707

MARRON, J. S. and DRYDEN, I. L. (2021). *Object Oriented Data Analysis*. CRC Press, Boca Raton.

MATABUENA, M., PETERSEN, A., VIDAL, J. C. and GUDE, F. (2021). Glucodensities: A new representation of glucose profiles using distributional data analysis. *Stat. Methods Med. Res.* **30** 1445–1464. MR4269959 https://doi.org/10.1177/0962280221998064

MÜLLER, H.-G. (2016). Peter Hall, functional data analysis and random objects. *Ann. Statist.* **44** 1867–1887. MR3546436 https://doi.org/10.1214/16-AOS1492

PANARETOS, V. M. and ZEMEL, Y. (2020). *An Invitation to Statistics in Wasserstein Space*. *SpringerBriefs in Probability and Mathematical Statistics*. Springer, Cham. MR4350694 https://doi.org/10.1007/978-3-030-38438-8

PEGORARO, M. and BERAHA, M. (2022). Projected statistical methods for distributional data on the real line with the Wasserstein metric. *J. Mach. Learn. Res.* **23** Paper No. [37], 59. MR4420762

PETERSEN, A. and MÜLLER, H.-G. (2016a). 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

PETERSEN, A. and MÜLLER, H.-G. (2016b). Fréchet integration and adaptive metric selection for interpretable covariances of multivariate functional data. *Biometrika* **103** 103–120. MR3465824 https://doi.org/10.1093/biomet/asv054

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

PETERSEN, A., ZHANG, C. and KOKOSZKA, P. (2022). Modeling probability density functions as data objects. *Econom. Stat.* **21** 159–178. MR4366852 https://doi.org/10.1016/j.ecosta.2021.04.004

PIGOLI, D., ASTON, J. A. D., DRYDEN, I. L. and SECCHI, P. (2014). Distances and inference for covariance operators. *Biometrika* **101** 409–422. MR3215356 https://doi.org/10.1093/biomet/asu008

POWER, J. D., COHEN, A. L., NELSON, S. M., WIG, G. S., BARNES, K. A., CHURCH, J. A., VOGEL, A. C., LAUMANN, T. O., MIEZIN, F. M. et al. (2011). Functional network organization of the human brain. *Neuron* **72** 665–678. https://doi.org/10.1016/j.neuron.2011.09.006

R CORE TEAM (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

SCEALY, J. L. and WELSH, A. H. (2011). Regression for compositional data by using distributions defined on the hypersphere. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 351–375. MR2815780 https://doi.org/10.1111/j.1467-9868.2010.00766.x

SCEALY, J. L. and WELSH, A. H. (2014). Colours and cocktails: Compositional data analysis 2013 Lancaster lecture. *Aust. N. Z. J. Stat.* **56** 145–169. MR3226434 https://doi.org/10.1111/anzs.12073

SCHOENBERG, I. J. (1937). On certain metric spaces arising from Euclidean spaces by a change of metric and their imbedding in Hilbert space. *Ann. of Math.* (2) **38** 787–793. MR1503370 https://doi.org/10.2307/1968835

SCHOENBERG, I. J. (1938). Metric spaces and positive definite functions. *Trans. Amer. Math. Soc.* **44** 522–536. MR1501980 https://doi.org/10.2307/1989894

SCHÖTZ, C. (2019). Convergence rates for the generalized Fréchet mean via the quadruple inequality. *Electron. J. Stat.* **13** 4280–4345. MR4023955 https://doi.org/10.1214/19-EJS1618

SCHÖTZ, C. (2022). Nonparametric regression in nonstandard spaces. *Electron. J. Stat.* **16** 4679–4741. MR4489238 https://doi.org/10.1214/22-ejs2056

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

SEVERN, K. E., DRYDEN, I. L. and PRESTON, S. P. (2022). Manifold valued data analysis of samples of networks, with applications in corpus linguistics. *Ann*. *Appl*. *Stat*. **16** 368–390. MR4400514 https://doi.org/10.1214/21-aoas1480

STEINKE, F. and HEIN, M. (2009). Non-parametric regression between manifolds. *Adv*. *Neural Inf*. *Process*. *Syst*. 1561–1568.

STEINKE, F., HEIN, M. and SCHÖLKOPF, B. (2010). Nonparametric regression between general Riemannian manifolds. *SIAM J*. *Imaging Sci*. **3** 527–563. MR2736019 https://doi.org/10.1137/080744189

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

SZÉKELY, G. J. and RIZZO, M. L. (2004). Testing for equal distributions in high dimension. *Interstate* **5** 1–6.

SZÉKELY, G. J. and RIZZO, M. L. (2017). The energy of data. *Annu*. *Rev*. *Stat*. *Appl*. **4** 447–479.

TUCKER, D. C., WU, Y. and MÜLLER, H.-G. (2023). Variable selection for global Fréchet regression. *J*. *Amer*. *Statist*. *Assoc*. **118** 1023–1037. MR4595474 https://doi.org/10.1080/01621459.2021.1969240

VAKHANIA, N. N., TARIELADZE, V. I. and CHOBANYAN, S. A. (1987). *Probability Distributions on Banach Spaces*. *Mathematics and Its Applications* (*Soviet Series*) **14**. Reidel, Dordrecht. Translated from the Russian and with a preface by Wojbor A. Woyczynski. MR1435288 https://doi.org/10.1007/978-94-009-3873-1

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

VENET, N. (2019). Nonexistence of fractional Brownian fields indexed by cylinders. *Electron*. *J*. *Probab*. **24** Paper No. 75, 26. MR3978225 https://doi.org/10.1214/18-EJP256

VIRTA, J., LEE, K.-Y. and LI, L. (2022). Sliced inverse regression in metric spaces. *Statist*. *Sinica* **32** 2315–2337. MR4485085

WANG, H. and MARRON, J. S. (2007). Object oriented data analysis: Sets of trees. *Ann*. *Statist*. **35** 1849–1873. MR2363955 https://doi.org/10.1214/009053607000000217

WANG, J.-L., CHIOU, J.-M. and MÜLLER, H.-G. (2016). Functional data analysis. *Annu*. *Rev*. *Stat*. *Appl*. **3** 257–295.

WANG, X., ZHU, J., PAN, W., ZHU, J. and ZHANG, H. (2023). Nonparametric statistical inference via metric distribution function in metric spaces. *J*. *Amer*. *Statist*. *Assoc*. (to appear). https://doi.org/10.1080/01621459.2023.2277417.

YUAN, Y., ZHU, H., LIN, W. and MARRON, J. S. (2012). Local polynomial regression for symmetric positive definite matrices. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **74** 697–719. MR2965956 https://doi.org/10.1111/j.1467-9868.2011.01022.x

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

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

ZHANG, Q., XUE, L. and LI, B. (2021). Dimension reduction and data visualization for Fréchet regression. arXiv preprint. Available at arXiv:2110.00467.

ZHOU, H. and MÜLLER, H.-G. (2023). Optimal transport representations and functional principal components for distribution-valued processes. arXiv preprint. Available at arXiv:2310.20088.

ZHOU, Y. and MÜLLER, H.-G. (2022). Network regression with graph Laplacians. *J*. *Mach*. *Learn*. *Res*. **23** Paper No. [320], 41. MR4577759 https://doi.org/10.22405/2226-8383-2022-23-5-320-336

ZHU, C. and MÜLLER, H.-G. (2023a). Autoregressive optimal transport models. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **85** 1012–1033.

ZHU, C. and MÜLLER, H.-G. (2023b). Geodesic optimal transport regression. arXiv preprint. Available at arXiv:2312.15376.

ZHU, C. and MÜLLER, H.-G. (2024). Spherical autoregressive models, with application to distributional and compositional time series. *J*. *Econometrics* **239** Paper No. 105389, 16. MR4708615 https://doi.org/10.1016/j.jeconom.2022.12.008

The Annals of Statistics
2024, Vol. 52, No. 2, 793–816
https://doi.org/10.1214/24-AOS2369

MINIMAX RATES FOR HETEROGENEOUS CAUSAL EFFECT ESTIMATION

BY EDWARD H. KENNEDY[a], SIVARAMAN BALAKRISHNAN[b], JAMES M. ROBINS[c] AND LARRY WASSERMAN[d]

Department of Statistics and Data Science, Carnegie Mellon University, [a]*edward@stat.cmu.edu,* [b]*siva@stat.cmu.edu,* [c]*robins@hsph.harvard.edu,* [d]*larry@stat.cmu.edu*

Estimation of heterogeneous causal effects—that is, how effects of policies and treatments vary across subjects—is a fundamental task in causal inference. Many methods for estimating conditional average treatment effects (CATEs) have been proposed in recent years, but questions surrounding optimality have remained largely unanswered. In particular, a minimax theory of optimality has yet to be developed, with the minimax rate of convergence and construction of rate-optimal estimators remaining open problems. In this paper, we derive the minimax rate for CATE estimation, in a Hölder-smooth nonparametric model, and present a new local polynomial estimator, giving high-level conditions under which it is minimax optimal. Our minimax lower bound is derived via a localized version of the method of fuzzy hypotheses, combining lower bound constructions for nonparametric regression and functional estimation. Our proposed estimator can be viewed as a local polynomial R-Learner, based on a localized modification of higher-order influence function methods. The minimax rate we find exhibits several interesting features, including a nonstandard elbow phenomenon and an unusual interpolation between nonparametric regression and functional estimation rates. The latter quantifies how the CATE, as an estimand, can be viewed as a regression/functional hybrid.

REFERENCES

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

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

[3] BICKEL, P. J. and RITOV, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhyā Ser. A* **50** 381–393. MR1065550

[4] BIRGÉ, L. and MASSART, P. (1995). Estimation of integral functionals of a density. *Ann. Statist.* **23** 11–29. MR1331653 https://doi.org/10.1214/aos/1176324452

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

[6] CHERNOZHUKOV, V., GOLDMAN, M., SEMENOVA, V. and TADDY, M. (2017). Orthogonal machine learning for demand estimation: High dimensional causal inference in dynamic panels. Preprint. Available at arXiv:1712.09988.

[7] FAN, Q., HSU, Y.-C., LIELI, R. P. and ZHANG, Y. (2022). Estimation of conditional average treatment effects with high-dimensional data. *J. Bus. Econom. Statist.* **40** 313–327. MR4356575 https://doi.org/10.1080/07350015.2020.1811102

[8] FOSTER, D. J. and SYRGKANIS, V. (2023). Orthogonal statistical learning. *Ann. Statist.* **51** 879–908. MR4630373 https://doi.org/10.1214/23-AOS2258

[9] FOSTER, J. C., TAYLOR, J. M. G. and RUBERG, S. J. (2011). Subgroup identification from randomized clinical trial data. *Stat. Med.* **30** 2867–2880. MR2844689 https://doi.org/10.1002/sim.4322

[10] GAO, Z. and HAN, Y. (2020). Minimax optimal nonparametric estimation of heterogeneous treatment effects. Preprint. Available at arXiv:2002.06471.

MSC2020 subject classifications. 62G08, 62H12.
Key words and phrases. Causal inference, functional estimation, higher-order influence functions, nonparametric regression, optimal rates of convergence.

[11] 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

[12] HAHN, P. R., MURRAY, J. S. and CARVALHO, C. M. (2020). Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects (with discussion). *Bayesian Anal.* **15** 965–1056. MR4154846 https://doi.org/10.1214/19-BA1195

[13] HERNÁN, M. A. and ROBINS, J. M. (2020). *Causal Inference*: *What If*. CRC, Boca Raton, FL.

[14] IBRAGIMOV, I. A., NEMIROVSKIĬ, A. S. and KHAS'MINSKIĬ, R. Z. (1986). Some problems of nonparametric estimation in Gaussian white noise. *Teor. Veroyatn. Primen.* **31** 451–466. MR0866866

[15] IMAI, K. and RATKOVIC, M. (2013). Estimating treatment effect heterogeneity in randomized program evaluation. *Ann. Appl. Stat.* **7** 443–470. MR3086426 https://doi.org/10.1214/12-AOAS593

[16] INGSTER, YU. 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

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

[18] KENNEDY, E. H, BALAKRISHNAN, S., ROBINS, J. M and WASSERMAN, L. (2024). Supplement to "Minimax rates for heterogeneous causal effect estimation." https://doi.org/10.1214/24-AOS2369SUPP

[19] KUENZEL, S. R. (2019). *Heterogeneous Treatment Effect Estimation Using Machine Learning*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—University of California, Berkeley. MR4051203

[20] LEE, S., OKUI, R. and WHANG, Y.-J. (2017). Doubly robust uniform confidence band for the conditional average treatment effect function. *J. Appl. Econometrics* **32** 1207–1225. MR3734484 https://doi.org/10.1002/jae.2574

[21] LIU, L., MUKHERJEE, R., ROBINS, J. M. and TCHETGEN TCHETGEN, E. (2021). Adaptive estimation of nonparametric functionals. *J. Mach. Learn. Res.* **22** Paper No. 99, 66. MR4279750

[22] LUEDTKE, A. R. and VAN DER LAAN, M. J. (2016). Super-learning of an optimal dynamic treatment rule. *Int. J. Biostat.* **12** 305–332. MR3505699 https://doi.org/10.1515/ijb-2015-0052

[23] MUKHERJEE, R., NEWEY, W. K. and ROBINS, J. M. (2017). Semiparametric efficient empirical higher order influence function estimators. Preprint. Available at arXiv:1705.07577.

[24] MUKHERJEE, R., TCHETGEN, E. J. T. and ROBINS, J. M. (2015). Lepski's method and adaptive estimation of nonlinear integral functionals of density. Preprint. Available at arXiv:1508.00249.

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

[26] NIE, X. and WAGER, S. (2021). Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika* **108** 299–319. MR4259133 https://doi.org/10.1093/biomet/asaa076

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

[28] ROBINS, J., LI, L., TCHETGEN, E. and VAN DER VAART, A. W. (2009). Quadratic semiparametric von Mises calculus. *Metrika* **69** 227–247. MR2481922 https://doi.org/10.1007/s00184-008-0214-3

[29] ROBINS, J., TCHETGEN TCHETGEN, E., LI, L. and VAN DER VAART, A. (2009). Semiparametric minimax rates. *Electron. J. Stat.* **3** 1305–1321. MR2566189 https://doi.org/10.1214/09-EJS479

[30] ROBINS, J. M. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Comm. Statist. Theory Methods* **23** 2379–2412. MR1293185 https://doi.org/10.1080/03610929408831393

[31] ROBINS, J. M., LI, L., MUKHERJEE, R., TCHETGEN, E. T. and VAN DER VAART, A. (2017). Minimax estimation of a functional on a structured high-dimensional model. *Ann. Statist.* **45** 1951–1987. MR3718158 https://doi.org/10.1214/16-AOS1515

[32] ROBINS, J. M., MARK, S. D. and NEWEY, W. K. (1992). Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics* **48** 479–495. MR1173493 https://doi.org/10.2307/2532304

[33] ROBINSON, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica* **56** 931–954. MR0951762 https://doi.org/10.2307/1912705

[34] SEMENOVA, V. and CHERNOZHUKOV, V. (2017). Estimation and inference about conditional average treatment effect and other structural functions. arXiv–1702, arXiv.

[35] SHALIT, U., JOHANSSON, F. D. and SONTAG, D. (2017). Estimating individual treatment effect: Generalization bounds and algorithms. In *Proceedings of the* 34*th International Conference on Machine Learning—Volume* 70 3076–3085. JMLR.org.

[36] 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

[37] 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

[38] VAN DER LAAN, M. J. (2006). Statistical inference for variable importance. *Int. J. Biostat.* **2** Art. 2, 33. MR2275897 https://doi.org/10.2202/1557-4679.1008

[39] VAN DER LAAN, M. J. and ROBINS, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. *Springer Series in Statistics*. Springer, New York. MR1958123 https://doi.org/10.1007/978-0-387-21700-0

[40] VANSTEELANDT, S. and JOFFE, M. (2014). Structural nested models and G-estimation: The partially realized promise. *Statist. Sci.* **29** 707–731. MR3300367 https://doi.org/10.1214/14-STS493

[41] WAGER, S. and ATHEY, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *J. Amer. Statist. Assoc.* **113** 1228–1242. MR3862353 https://doi.org/10.1080/01621459.2017.1319839

[42] 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

[43] ZHAO, Q., SMALL, D. S. and ERTEFAIE, A. (2022). Selective inference for effect modification via the lasso. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 382–413. MR4412991 https://doi.org/10.1111/rssb.12483

[44] ZIMMERT, M. and LECHNER, M. (2019). Nonparametric estimation of causal heterogeneity under high-dimensional confounding. Preprint. Available at arXiv:1908.08779.

The Annals of Statistics
2024, Vol. 52, No. 2, 817–841
https://doi.org/10.1214/24-AOS2370

THE ONLINE CLOSURE PRINCIPLE

BY LASSE FISCHER[1,a], MARTA BOFILL ROIG[2,c] AND WERNER BRANNATH[1,b]

[1]*Competence Center for Clinical Trials Bremen, University of Bremen,* [a]*fischer1@uni-bremen.de,* [b]*brannath@uni-bremen.de*
[2]*Center for Medical Data Science, Medical University of Vienna,* [c]*marta.bofillroig@meduniwien.ac.at*

The closure principle is fundamental in multiple testing and has been used to derive many efficient procedures with familywise error rate control. However, it is often unsuitable for modern research, which involves flexible multiple testing settings where not all hypotheses are known at the beginning of the evaluation. In this paper, we focus on online multiple testing where a possibly infinite sequence of hypotheses is tested over time. At each step, it must be decided on the current hypothesis without having any information about the hypotheses that have not been tested yet. Our main contribution is a general and stringent mathematical definition of online multiple testing and a new online closure principle, which ensures that the resulting closed procedure can be applied in the online setting. We prove that any familywise error rate controlling online procedure can be derived by this online closure principle and provide admissibility results. In addition, we demonstrate how shortcuts of these online closed procedures can be obtained under a suitable consonance property.

REFERENCES

[1] 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

[2] BITTMAN, R. M., ROMANO, J. P., VALLARINO, C. and WOLF, M. (2009). Optimal testing of multiple hypotheses with common effect direction. *Biometrika* **96** 399–410. MR2507151 https://doi.org/10.1093/biomet/asp006

[3] BRETZ, F., MAURER, W., BRANNATH, W. and POSCH, M. (2009). A graphical approach to sequentially rejective multiple test procedures. *Stat. Med.* **28** 586–604. MR2655732 https://doi.org/10.1002/sim.3495

[4] DMITRIENKO, A., OFFEN, W. W. and WESTFALL, P. H. (2003). Gatekeeping strategies for clinical trials that do not require all primary effects to be significant. *Stat. Med.* **22** 2387–2400.

[5] DÖHLER, S., MEAH, I. and ROQUAIN, E. (2024). Online multiple testing with super-uniformity reward. *Electron. J. Stat.* **18** 1293–1354. MR4718473 https://doi.org/10.1214/24-ejs2230

[6] FENG, J., EMERSON, S. and SIMON, N. (2021). Approval policies for modifications to machine learning-based software as a medical device: A study of bio-creep. *Biometrics* **77** 31–44. MR4229719 https://doi.org/10.1111/biom.13379

[7] FENG, J., PENNLLO, G., PETRICK, N., SAHINER, B., PIRRACCHIO, R. and GOSSMANN, A. (2022). Sequential algorithmic modification with test data reuse. In *Uncertainty in Artificial Intelligence* 674–684. PMLR.

[8] FISCHER, L., BOFILL ROIG, M. and BRANNATH, W. (2024). Supplement to "The online closure principle." https://doi.org/10.1214/24-AOS2370SUPP

[9] FOSTER, D. P. and STINE, R. A. (2008). α-investing: A procedure for sequential control of expected false discoveries. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 429–444. MR2424761 https://doi.org/10.1111/j.1467-9868.2007.00643.x

[10] GABRIEL, K. R. (1969). Simultaneous test procedures—some theory of multiple comparisons. *Ann. Math. Stat.* **40** 224–250. MR0240931 https://doi.org/10.1214/aoms/1177697819

[11] GENOVESE, C. R. and WASSERMAN, L. (2006). Exceedance control of the false discovery proportion. *J. Amer. Statist. Assoc.* **101** 1408–1417. MR2279468 https://doi.org/10.1198/016214506000000339

MSC2020 subject classifications. 62L10.
Key words and phrases. Online multiple testing, closure principle, familywise error rate.

[12] GOEMAN, J. J., HEMERIK, J. and SOLARI, A. (2021). Only closed testing procedures are admissible for controlling false discovery proportions. *Ann. Statist.* **49** 1218–1238. MR4255125 https://doi.org/10.1214/20-aos1999

[13] GOEMAN, J. J., MEIJER, R. J., KREBS, T. J. P. and SOLARI, A. (2019). Simultaneous control of all false discovery proportions in large-scale multiple hypothesis testing. *Biometrika* **106** 841–856. MR4046036 https://doi.org/10.1093/biomet/asz041

[14] GOEMAN, J. J. and SOLARI, A. (2011). Multiple testing for exploratory research. *Statist. Sci.* **26** 584–597. MR2951390 https://doi.org/10.1214/11-STS356

[15] GRECHANOVSKY, E. and HOCHBERG, Y. (1999). Closed procedures are better and often admit a shortcut. *J. Statist. Plann. Inference* **76** 79–91. MR1673341 https://doi.org/10.1016/S0378-3758(98)00125-6

[16] HOMMEL, G., BRETZ, F. and MAURER, W. (2007). Powerful short-cuts for multiple testing procedures with special reference to gatekeeping strategies. *Stat. Med.* **26** 4063–4073. MR2405792 https://doi.org/10.1002/sim.2873

[17] IOANNIDIS, J. P. (2005). Why most published research findings are false. *PLoS Med.* **2** e124.

[18] JAVANMARD, A. and MONTANARI, A. (2018). Online rules for control of false discovery rate and false discovery exceedance. *Ann. Statist.* **46** 526–554. MR3782376 https://doi.org/10.1214/17-AOS1559

[19] KARP, N. A., MASON, J., BEAUDET, A. L., BENJAMINI, Y., BOWER, L., BRAUN, R. E., BROWN, S. D., CHESLER, E. J., DICKINSON, M. E. et al. (2017). Prevalence of sexual dimorphism in mammalian phenotypic traits. *Nat. Commun.* **8** 1–12.

[20] KOHAVI, R., DENG, A., FRASCA, B., WALKER, T., XU, Y. and POHLMANN, N. (2013). Online controlled experiments at large scale. In *Proceedings of the* 19*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 1168–1176.

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

[22] MARCUS, R., PERITZ, E. and GABRIEL, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* **63** 655–660. MR0468056 https://doi.org/10.1093/biomet/63.3.655

[23] MAURER, W., HOTHORN, L. and LEHRMACHER, W. (1995). Multiple comparisons in drug clinical trials and preclinical assays: A-priori ordered hypotheses. In *Biometrie in der Chemisch-Pharmazeutischen Industrie* (V. Joachim, ed.) 3–18. Fischer Verlag, Stuttgart.

[24] MUÑOZ-FUENTES, V., CACHEIRO, P., MEEHAN, T. F., AGUILAR-PIMENTEL, J. A., BROWN, S. D., FLENNIKEN, A. M., FLICEK, P., GALLI, A., MASHHADI, H. H. et al. (2018). The International Mouse Phenotyping Consortium (IMPC): A functional catalogue of the mammalian genome that informs conservation. *Conserv. Genet.* **19** 995–1005.

[25] RAMDAS, A., YANG, F., WAINWRIGHT, M. J. and JORDAN, M. I. (2017). Online control of the false discovery rate with decaying memory. In *Advances in Neural Information Processing Systems* **30**. Curran Associates, Red Hook.

[26] RAMDAS, A., ZRNIC, T., WAINWRIGHT, M. and JORDAN, M. (2018). SAFFRON: An adaptive algorithm for online control of the false discovery rate. In *International Conference on Machine Learning* 4286–4294. PMLR.

[27] ROBERTSON, D. S., WASON, J. M. S. and BRETZ, F. (2020). Graphical approaches for the control of generalized error rates. *Stat. Med.* **39** 3135–3155. MR4151924 https://doi.org/10.1002/sim.8595

[28] ROBERTSON, D. S., WASON, J. M. S., KÖNIG, F., POSCH, M. and JAKI, T. (2023). Online error rate control for platform trials. *Stat. Med.* **42** 2475–2495. MR4596806 https://doi.org/10.1002/sim.9733

[29] ROBERTSON, D. S., WASON, J. M. S. and RAMDAS, A. (2023). Online multiple hypothesis testing. *Statist. Sci.* **38** 557–575. MR4665026 https://doi.org/10.1214/23-sts901

[30] ROBERTSON, D. S., WILDENHAIN, J., JAVANMARD, A. and KARP, N. A. (2019). onlineFDR: An R package to control the false discovery rate for growing data repositories. *Bioinformatics* **35** 4196–4199.

[31] ROMANO, J. P., SHAIKH, A. and WOLF, M. (2011). Consonance and the closure method in multiple testing. *Int. J. Biostat.* **7** 12. MR2775079 https://doi.org/10.2202/1557-4679.1300

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

[33] SANDERCOCK, P. A., DARBYSHIRE, J., DEMETS, D., FOWLER, R., LALLOO, D. G., MUNAVVAR, M., STAPLIN, N., WARRIS, A., WITTES, J. et al. (2022). Experiences of the data monitoring committee for the RECOVERY trial, a large-scale adaptive platform randomised trial of treatments for patients hospitalised with COVID-19. *Trials* **23** 881.

[34] SCHWEDER, T. and SPJØTVOLL, E. (1982). Plots of p-values to evaluate many tests simultaneously. *Biometrika* **69** 493–502.

[35] SONNEMANN, E. and FINNER, H. (1988). Vollständigkeitssätze für multiple Testprobleme. In *Multiple Hypothesenprüfung/Multiple Hypotheses Testing* (P. Bauer, G. Hommel and E. Sonnemann, eds.) 121–135. Springer, Berlin.

[36] TIAN, J. and RAMDAS, A. (2019). ADDIS: An adaptive discarding algorithm for online FDR control with conservative nulls. In *Advances in Neural Information Processing Systems* **32**. Curran Associates, Red Hook.

[37] TIAN, J. and RAMDAS, A. (2021). Online control of the familywise error rate. *Stat. Methods Med. Res.* **30** 976–993. MR4259882 https://doi.org/10.1177/0962280220983381

[38] ZEHETMAYER, S., POSCH, M. and KOENIG, F. (2022). Online control of the False Discovery Rate in group-sequential platform trials. *Stat. Methods Med. Res.* **31** 2470–2485. MR4513312 https://doi.org/10.1177/09622802221129051

[39] ZHAO, Q., SMALL, D. S. and SU, W. (2019). Multiple testing when many p-values are uniformly conservative, with application to testing qualitative interaction in educational interventions. *J. Amer. Statist. Assoc.* **114** 1291–1304. MR4011780 https://doi.org/10.1080/01621459.2018.1497499

[40] ZRNIC, T., JIANG, D., RAMDAS, A. and JORDAN, M. (2020). The power of batching in multiple hypothesis testing. In *International Conference on Artificial Intelligence and Statistics* **108** 3806–3815. PMLR.

[41] ZRNIC, T., RAMDAS, A. and JORDAN, M. I. (2021). Asynchronous online testing of multiple hypotheses. *J. Mach. Learn. Res.* **22** 33. MR4253726 https://doi.org/10.1515/ijnsns-2019-0210

The Annals of Statistics
2024, Vol. 52, No. 2, 842–867
https://doi.org/10.1214/24-AOS2371

PARAMETER ESTIMATION IN NONLINEAR MULTIVARIATE STOCHASTIC DIFFERENTIAL EQUATIONS BASED ON SPLITTING SCHEMES

BY PREDRAG PILIPOVIC[1,a], ADELINE SAMSON[2,b] AND SUSANNE DITLEVSEN[1,c]

[1]*Department of Mathematical Sciences, University of Copenhagen,* [a]*predrag@math.ku.dk*
[2]*Université Grenoble Alpes, CNRS, Grenoble INP, LJK,* [b]*adeline.leclercq-samson@univ-grenoble-alpes.fr,*
[c]*susanne@math.ku.dk*

The likelihood functions for discretely observed nonlinear continuous time models based on stochastic differential equations are not available except for a few cases. Various parameter estimation techniques have been proposed, each with advantages, disadvantages and limitations depending on the application. Most applications still use the Euler–Maruyama discretization, despite many proofs of its bias. More sophisticated methods, such as Kessler's Gaussian approximation, Ozaki's local linearization, Aït–Sahalia's Hermite expansions or MCMC methods, might be complex to implement, do not scale well with increasing model dimension or can be numerically unstable. We propose two efficient and easy-to-implement likelihood-based estimators based on the Lie–Trotter (LT) and the Strang (S) splitting schemes. We prove that S has L^p convergence rate of order 1, a property already known for LT. We show that the estimators are consistent and asymptotically efficient under the less restrictive one-sided Lipschitz assumption. A numerical study on the 3-dimensional stochastic Lorenz system complements our theoretical findings. The simulation shows that the S estimator performs the best when measured on precision and computational speed compared to the state-of-the-art.

REFERENCES

ABDULLE, A., VILMART, G. and ZYGALAKIS, K. C. (2015). Long time accuracy of Lie–Trotter splitting methods for Langevin dynamics. *SIAM J. Numer. Anal.* **53** 1–16. MR3296612 https://doi.org/10.1137/140962644

ABLEIDINGER, M. and BUCKWAR, E. (2016). Splitting integrators for the stochastic Landau–Lifshitz equation. *SIAM J. Sci. Comput.* **38** A1788–A1806. MR3511359 https://doi.org/10.1137/15M103529X

ABLEIDINGER, M., BUCKWAR, E. and HINTERLEITNER, H. (2017). A stochastic version of the Jansen and Rit neural mass model: Analysis and numerics. *J. Math. Neurosci.* **7** Paper No. 8, 35. MR3683994 https://doi.org/10.1186/s13408-017-0046-4

AÏT-SAHALIA, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica* **70** 223–262. MR1926260 https://doi.org/10.1111/1468-0262.00274

AÏT-SAHALIA, Y. (2008). Closed-form likelihood expansions for multivariate diffusions. *Ann. Statist.* **36** 906–937. MR2396819 https://doi.org/10.1214/009053607000000622

ALAMO, A. and SANZ-SERNA, J. M. (2016). A technique for studying strong and weak local errors of splitting stochastic integrators. *SIAM J. Numer. Anal.* **54** 3239–3257. MR3570281 https://doi.org/10.1137/16M1058765

ALYUSHINA, L. A. (1987). Euler polygonal lines for Itô equations with monotone coefficients. *Teor. Veroyatn. Primen.* **32** 367–373. MR0902767

ANN, N., PEBRIANTI, D., ABAS, M. and BAYUAJI, L. (2022). Parameter estimation of Lorenz attractor: A combined deep neural network and K-means clustering approach. In *Recent Trends in Mechatronics Towards Industry* 4.0. *Lecture Notes in Electrical Engineering* **730** 321–331. Springer, Singapore.

ARNST, M., LOUPPE, G., VAN HULLE, R., GILLET, L., BUREAU, F. and DENOËL, V. (2022). A hybrid stochastic model and its Bayesian identification for infectious disease screening in a university campus with application to massive COVID-19 screening at the University of Liège. *Math. Biosci.* **347** Paper No. 108805, 14. MR4403088 https://doi.org/10.1016/j.mbs.2022.108805

MSC2020 subject classifications. Primary 62F12, 62H12, 62M99; secondary 37M15, 60G65.
Key words and phrases. Asymptotic normality, consistency, L^p convergence, splitting schemes, stochastic differential equations, stochastic Lorenz system.

BARBU, V. (1988). A product formula approach to nonlinear optimal control problems. *SIAM J. Control Optim.* **26** 497–520. MR0937669 https://doi.org/10.1137/0326030

BAYDIN, A. M. G., PEARLMUTTER, B. A., RADUL, A. A. and SISKIND, J. M. (2017). Automatic differentiation in machine learning: A survey. *J. Mach. Learn. Res.* **18** Paper No. 153, 43. MR3800512

BENSOUSSAN, A., GLOWINSKI, R. and RĂŞCANU, A. (1992). Approximation of some stochastic differential equations by the splitting up method. *Appl. Math. Optim.* **25** 81–106. MR1133253 https://doi.org/10.1007/BF01184157

BIBBY, B. M. and SØRENSEN, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli* **1** 17–39. MR1354454 https://doi.org/10.2307/3318679

BLANES, S., CASAS, F. and MURUA, A. (2008). Splitting and composition methods in the numerical integration of differential equations. *Bol. Soc. Esp. Mat. Apl.* 45, 89–145. MR2477860

BOU-RABEE, N. and OWHADI, H. (2010). Long-run accuracy of variational integrators in the stochastic context. *SIAM J. Numer. Anal.* **48** 278–297. MR2608370 https://doi.org/10.1137/090758842

BRÉHIER, C-E. and GOUDENÈGE, L. (2019). Analysis of some splitting schemes for the stochastic Allen-Cahn equation. *Discrete Contin. Dyn. Syst. Ser. B* **24** 4169–4190. MR3986273 https://doi.org/10.3934/dcdsb.2019077

BUCKWAR, E., SAMSON, A., TAMBORRINO, M. and TUBIKANEC, I. (2022). A splitting method for SDEs with locally Lipschitz drift: Illustration on the FitzHugh–Nagumo model. *Appl. Numer. Math.* **179** 191–220. MR4422320 https://doi.org/10.1016/j.apnum.2022.04.018

BUCKWAR, E., TAMBORRINO, M. and TUBIKANEC, I. (2020). Spectral density-based and measure-preserving ABC for partially observed diffusion processes. An illustration on Hamiltonian SDEs. *Stat. Comput.* **30** 627–648. MR4065223 https://doi.org/10.1007/s11222-019-09909-6

CHANG, J. and CHEN, S. X. (2011). On the approximate maximum likelihood estimation for diffusion processes. *Ann. Statist.* **39** 2820–2851. MR3012393 https://doi.org/10.1214/11-AOS922

CHOI, S. (2013). Closed-form likelihood expansions for multivariate time-inhomogeneous diffusions. *J. Econometrics* **174** 45–65. MR3045019 https://doi.org/10.1016/j.jeconom.2011.12.007

CHOI, S. (2015). Explicit form of approximate transition probability density functions of diffusion processes. *J. Econometrics* **187** 57–73. MR3347294 https://doi.org/10.1016/j.jeconom.2015.02.003

CHOPIN, N. and PAPASPILIOPOULOS, O. (2020). *An Introduction to Sequential Monte Carlo*. *Springer Series in Statistics*. Springer, Cham. MR4215639 https://doi.org/10.1007/978-3-030-47845-2

DACUNHA-CASTELLE, D. and FLORENS-ZMIROU, D. (1986). Estimation of the coefficients of a diffusion from discrete observations. *Stochastics* **19** 263–284. MR0872464 https://doi.org/10.1080/17442508608833428

DIPPLE, S., CHOUDHARY, A., FLAMINO, J., SZYMANSKI, B. and KORNISS, G. (2020). Using correlated stochastic differential equations to forecast cryptocurrency rates and social media activities. *Appl. Netw. Sci.* **5**. https://doi.org/10.1007/s41109-020-00259-1

DITLEVSEN, P. and DITLEVSEN, S. (2023). Warning of a forthcoming collapse of the Atlantic meridional overturning circulation. *Nat. Commun.* **14** 4254. https://doi.org/10.1038/s41467-023-39810-w

DITLEVSEN, S. and SAMSON, A. (2019). Hypoelliptic diffusions: Filtering and inference from complete and partial observations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 361–384. MR3928146

DITLEVSEN, S. and SØRENSEN, M. (2004). Inference for observations of integrated diffusion processes. *Scand. J. Stat.* **31** 417–429. MR2087834 https://doi.org/10.1111/j.1467-9469.2004.02_023.x

DITLEVSEN, S., TAMBORRINO, M. and TUBIKANEC, I. (2023). Network inference in a stochastic multi-population neural mass model via approximate Bayesian computation. Available at arXiv:2306.15787.

DOHNAL, G. (1987). On estimating the diffusion coefficient. *J. Appl. Probab.* **24** 105–114. MR0876173 https://doi.org/10.2307/3214063

DUBOIS, P., GOMEZ, T., PLANCKAERT, L. and PERRET, L. (2020). Data-driven predictions of the Lorenz system. *Phys. D* **408** 132495, 10. MR4087348 https://doi.org/10.1016/j.physd.2020.132495

FALBEL, D. and LURASCHI, J. (2022). torch: Tensors and neural networks with 'GPU' acceleration. Available at https://torch.mlverse.org/docs, https://github.com/mlverse/torch.

FLORENS-ZMIROU, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. *Statistics* **20** 547–557. MR1047222 https://doi.org/10.1080/02331888908802205

FORMAN, J. L. and SØRENSEN, M. (2008). The Pearson diffusions: A class of statistically tractable diffusion processes. *Scand. J. Stat.* **35** 438–465. MR2446729 https://doi.org/10.1111/j.1467-9469.2007.00592.x

FUCHS, C. (2013). *Inference for Diffusion Processes*: *With Applications in Life Sciences*. Springer, Heidelberg. With a foreword by Ludwig Fahrmeir. MR3015023 https://doi.org/10.1007/978-3-642-25969-2

GENON-CATALOT, V. and JACOD, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **29** 119–151. MR1204521

GLOAGUEN, P., ETIENNE, M.-P. and LE CORFF, S. (2018). Stochastic differential equation based on a multimodal potential to model movement data in ecology. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **67** 599–619. MR3787968 https://doi.org/10.1111/rssc.12251

GLOTER, A. (2006). Parameter estimation for a discretely observed integrated diffusion process. *Scand. J. Stat.* **33** 83–104. MR2255111 https://doi.org/10.1111/j.1467-9469.2006.00465.x

GLOTER, A. and YOSHIDA, N. (2021a). Adaptive estimation for degenerate diffusion processes. *Electron. J. Stat.* **15** 1424–1472. MR4255288 https://doi.org/10.1214/20-ejs1777

GLOTER, A. and YOSHIDA, N. (2021b). Adaptive estimation for degenerate diffusion processes. *Electron. J. Stat.* **15** 1424–1472. MR4255288 https://doi.org/10.1214/20-ejs1777

GOBET, E. (2002). LAN property for ergodic diffusions with discrete observations. *Ann. Inst. Henri Poincaré Probab. Stat.* **38** 711–737. MR1931584 https://doi.org/10.1016/S0246-0203(02)01107-X

GU, W., WU, H. and XUE, H. (2020). Parameter estimation for multivariate nonlinear stochastic differential equation models: A comparison study. In *Statistical Modeling for Biological Systems: In Memory of Andrei Yakovlev* 245–258. Springer, Cham. https://doi.org/10.1007/978-3-030-34675-1_13

HAIRER, E., NØRSETT, S. P. and WANNER, G. (1993). *Solving Ordinary Differential Equations. I : Nonstiff Problems*, 2nd ed. *Springer Series in Computational Mathematics* **8**. Springer, Berlin. MR1227985

HILBORN, R. C. (1994). *Chaos and Nonlinear Dynamics: An Introduction for Scientists and Engineers: An Introduction for Scientists and Engineers*. Oxford Univ. Press, New York. MR1263025

HOPKINS, W. E. JR. and WONG, W. S. (1986). Lie–Trotter product formulas for nonlinear filtering. *Stochastics* **17** 313–337. MR0854651 https://doi.org/10.1080/17442508608833395

HUMPHRIES, A. R. and STUART, A. M. (1994). Runge–Kutta methods for dissipative and gradient dynamical systems. *SIAM J. Numer. Anal.* **31** 1452–1485. MR1293524 https://doi.org/10.1137/0731075

HUMPHRIES, A. R. and STUART, A. M. (2002). Deterministic and random dynamical systems: Theory and numerics. In *Modern Methods in Scientific Computing and Applications* (*Montréal, QC*, 2001). *NATO Sci. Ser. II Math. Phys. Chem.* **75** 211–254. Kluwer Academic, Dordrecht. MR2004356

HURN, A. S., JEISMAN, J. I. and LINDSAY, K. A. (2007). Seeing the wood for the trees: A critical evaluation of methods to estimate the parameters of stochastic differential equations. *J. Financ. Econ.* **5** 390–455.

HUTZENTHALER, M., JENTZEN, A. and KLOEDEN, P. E. (2011). Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **467** 1563–1576. MR2795791 https://doi.org/10.1098/rspa.2010.0348

IGUCHI, Y., BESKOS, A. and GRAHAM, M. M. (2022). Parameter estimation with increased precision for elliptic and hypo-elliptic diffusions. Available at arXiv:2211.16384.

JENSEN, B. and POULSEN, R. (2002). Transition densities of diffusion processes: Numerical comparison of approximation techniques. *J. Deriv.* **9** 18–32.

JIMENEZ, J. C., MORA, C. and SELVA, M. (2017). A weak local linearization scheme for stochastic differential equations with multiplicative noise. *J. Comput. Appl. Math.* **313** 202–217. MR3573236 https://doi.org/10.1016/j.cam.2016.09.013

JIMENEZ, J. C., SHOJI, I. and OZAKI, T. (1999). Simulation of stochastic differential equations through the local linearization method. A comparative study. *J. Stat. Phys.* **94** 587–602. MR1675365 https://doi.org/10.1023/A:1004504506041

KAREEM, A. M. and AL-AZZAWI, S. N. (2021). A stochastic differential equations model for internal COVID-19 dynamics. *J. Phys., Conf. Ser.* **1818** 012121. https://doi.org/10.1088/1742-6596/1818/1/012121

KELLER, H. (1996). Attractors and bifurcations of the stochastic Lorenz system Technical report. Institut für Dynamische syteme, Universität Bremen.

KESSLER, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Stat.* **24** 211–229. MR1455868 https://doi.org/10.1111/1467-9469.00059

KLOEDEN, P. E. and PLATEN, E. (1992). *Numerical Solution of Stochastic Differential Equations*. *Applications of Mathematics* (*New York*) **23**. Springer, Berlin. MR1214374 https://doi.org/10.1007/978-3-662-12616-5

KRYLOV, N. V. (1990). A simple proof of the existence of a solution to the Itô equation with monotone coefficients. *Teor. Veroyatn. Primen.* **35** 576–580. MR1091217 https://doi.org/10.1137/1135082

LAZZÚS, J. A., RIVERA, M. and LÓPEZ-CARABALLO, C. H. (2016). Parameter estimation of Lorenz chaotic system using a hybrid swarm intelligence algorithm. *Phys. Lett. A* **380** 1164–1171. MR3457318 https://doi.org/10.1016/j.physleta.2016.01.040

LEIMKUHLER, B. and MATTHEWS, C. (2015). *Molecular Dynamics: With Deterministic and Stochastic Numerical Methods*. *Interdisciplinary Applied Mathematics* **39**. Springer, Cham. MR3362507

LI, C. (2013). Maximum-likelihood estimation for diffusion processes via closed-form density expansions. *Ann. Statist.* **41** 1350–1380. MR3113814 https://doi.org/10.1214/13-AOS1118

LÓPEZ-PÉREZ, A., FEBRERO-BANDE, A. and GONZÁLEZ-MANTEIGAV, W. (2021). Parametric estimation of diffusion processes: A review and comparative study. *Mathematics* **9** 859. https://doi.org/10.3390/math9080859

LORENZ, E. N. (1963). Deterministic nonperiodic flow. *J. Atmos. Sci.* **20** 130–141. MR4021434 https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2

MAO, X. (2007). *Stochastic Differential Equations and Applications*. Elsevier, Amsterdam.

MCLACHLAN, R. I. and QUISPEL, G. R. W. (2002). Splitting methods. *Acta Numer.* **11** 341–434. MR2009376 https://doi.org/10.1017/S0962492902000053

MICHELOT, T., GLENNIE, R., HARRIS, C. and THOMAS, L. (2021). Varying-coefficient stochastic differential equations with applications in ecology. *J. Agric. Biol. Environ. Stat.* **26** 446–463. MR4292797 https://doi.org/10.1007/s13253-021-00450-6

MICHELOT, T., GLOAGUEN, P., BLACKWELL, P. and ETIENNE, M.-P. (2019). The Langevin diffusion as a continuous-time model of animal movement and habitat selection. *Methods Ecol. Evol.* **10**.

MILSTEIN, G. N. (1987). A theorem on the order of convergence of mean-square approximations of solutions of systems of stochastic differential equations. *Teor. Veroyatn. Primen.* **32** 809–811. MR0927268

MILSTEIN, G. N. and TRETYAKOV, M. V. (2003). Quasi-symplectic methods for Langevin-type equations. *IMA J. Numer. Anal.* **23** 593–626. MR2011342 https://doi.org/10.1093/imanum/23.4.593

MISAWA, T. (2001). A Lie algebraic approach to numerical integration of stochastic differential equations. *SIAM J. Sci. Comput.* **23** 866–890. MR1860968 https://doi.org/10.1137/S106482750037024X

OZAKI, T. (1985). Statistical identification of storage models with application to stochastic hydrology. *J. Amer. Water Resour. Assoc.* **21** 663–675.

OZAKI, T. (1992). A bridge between nonlinear time series models and nonlinear stochastic dynamical systems: A local linearization approach. *Statist. Sinica* **2** 113–135. MR1152300

OZAKI, T., JIMENEZ, J. C. and HAGGAN-OZAKI, V. (2000). The role of the likelihood function in the estimation of chaos models. *J. Time Series Anal.* **21** 363–387. MR1787661 https://doi.org/10.1111/1467-9892.00189

PICCHINI, U. and DITLEVSEN, S. (2011). Practical estimation of high dimensional stochastic differential mixed-effects models. *Comput. Statist. Data Anal.* **55** 1426–1444. MR2741425 https://doi.org/10.1016/j.csda.2010.10.003

PILIPOVIC, P., SAMSON, A. and DITLEVSEN, S. (2024). Supplement to "Parameter estimation in nonlinear multivariate stochastic differential equations based on splitting schemes." https://doi.org/10.1214/24-AOS2371SUPPA, https://doi.org/10.1214/24-AOS2371SUPPB

RIEDMILLER, M. and BRAUN, H. (1992). RPROP—a fast adaptive learning algorithm. Technical report. Proc. of ISCIS VII, Universitat.

SHOJI, I. (1998). Approximation of continuous time stochastic processes by a local linearization method. *Math. Comp.* **67** 287–298. MR1432134 https://doi.org/10.1090/S0025-5718-98-00888-6

SHOJI, I. (2011). A note on convergence rate of a linearization method for the discretization of stochastic differential equations. *Commun. Nonlinear Sci. Numer. Simul.* **16** 2667–2671. MR2772282 https://doi.org/10.1016/j.cnsns.2010.09.008

SHOJI, I. and OZAKI, T. (1998). Estimation for nonlinear stochastic differential equations by a local linearization method. *Stoch. Anal. Appl.* **16** 733–752. MR1632562 https://doi.org/10.1080/07362999808809559

SØRENSEN, M. (2012). Estimating functions for diffusion-type processes. In *Statistical Methods for Stochastic Differential Equations* 1 1–97. CRC Press, Boca Raton. https://doi.org/10.1201/b12126-2

SØRENSEN, M. and UCHIDA, M. (2003). Small-diffusion asymptotics for discretely sampled stochastic differential equations. *Bernoulli* **9** 1051–1069. MR2046817 https://doi.org/10.3150/bj/1072215200

TABOR, M. (1989). *Chaos and Integrability in Nonlinear Dynamics*: *An Introduction*. *A Wiley-Interscience Publication*. Wiley, New York. MR1007309

TIAN, Y. and FAN, M. (2020). Nonlinear integral inequality with power and its application in delay integro-differential equations. *Adv. Difference Equ.* Paper No. 142, 11. MR4085951 https://doi.org/10.1186/s13662-020-02596-y

TRETYAKOV, M. V. and ZHANG, Z. (2013). A fundamental mean-square convergence theorem for SDEs with locally Lipschitz coefficients and its applications. *SIAM J. Numer. Anal.* **51** 3135–3162. MR3129758 https://doi.org/10.1137/120902318

UCHIDA, M. and YOSHIDA, N. (2012). Adaptive estimation of an ergodic diffusion process based on sampled data. *Stochastic Process. Appl.* **122** 2885–2924. MR2931346 https://doi.org/10.1016/j.spa.2012.04.001

VAN LOAN, C. F. (1978). Computing integrals involving the matrix exponential. *IEEE Trans. Automat. Control* **23** 395–404. MR0494865 https://doi.org/10.1109/TAC.1978.1101743

VATIWUTIPONG, P. and PHEWCHEAN, N. (2019). Alternative way to derive the distribution of the multivariate Ornstein–Uhlenbeck process. *Adv. Difference Equ.* Paper No. 276, 7. MR3978552 https://doi.org/10.1186/s13662-019-2214-1

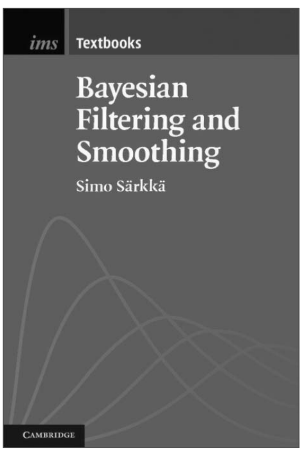
YANG, N., CHEN, N. and WAN, X. (2019). A new delta expansion for multivariate diffusions via the Itô–Taylor expansion. *J. Econometrics* **209** 256–288. MR3944752 https://doi.org/10.1016/j.jeconom.2019.01.003

ZHUANG, L., CAO, L., WU, Y., ZHONG, Y., ZHANGZHONG, L., ZHENG, W. and WANG, L. (2020). Parameter estimation of Lorenz chaotic system based on a hybrid Jaya–Powell algorithm. *IEEE Access* **8** 20514–20522.

R CORE TEAM (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

The Institute of Mathematical Statistics presents

IMS TEXTBOOKS



Bayesian Filtering and Smoothing

Simo Särkkä

Filtering and smoothing methods are used to produce an accurate estimate of the state of a time-varying system based on multiple observational inputs (data). Interest in these methods has exploded in recent years, with numerous applications emerging in fields such as navigation, aerospace engineering, telecommunications and medicine. This compact, informal introduction for graduate students and advanced undergraduates presents the current state-of-the-art filtering and smoothing methods in a unified Bayesian framework. Readers learn what non-linear Kalman filters and particle filters are, how they are related, and their relative advantages and disadvantages. They also discover how state-of-the-art Bayesian parameter estimation methods can be combined with state-of-the-art filtering and smoothing algorithms. The book's practical and algorithmic approach assumes only modest mathematical prerequisites. Examples include MATLAB computations, and the numerous end-of-chapter exercises include computational assignments. MATLAB/GNU Octave source code is available for download at www.cambridge.org/sarkka, promoting hands-on work with the methods.

IMS member? Claim your 40% discount:
www.cambridge.org/ims

Hardback US$59.40
Paperback US$22.19

Cambridge University Press, in conjunction with the Institute of Mathematical Statistics, established the IMS Monographs and IMS Textbooks series of high-quality books. The Series Editors are Xiao-Li Meng, Susan Holmes, Ben Hambly, D. R. Cox and Alan Agresti.