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

THE ANNALS *of* STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

Vol. 53, No. 2—April 2025

THE ANNALS OF STATISTICS Vol. 53, No. 2, pp. 457–906 April 2025

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

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

IMS OFFICERS

President: Tony Cai, Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, PA 19104-6304, USA

President-Elect: Kavita Ramanan, Division of Applied Mathematics, Brown University, Providence, RI 02912, USA

Past President: Michael Kosorok, Department of Biostatistics and Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, Chapel Hill, NC 27599, USA

Executive Secretary: Peter Hoff, Department of Statistical Science, Duke University, Durham, NC 27708-0251, USA

Treasurer: Jiashun Jin, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA

Program Secretary: Annie Qu, Department of Statistics, University of California, Irvine, Irvine, CA 92697-3425, USA

IMS EDITORS

The Annals of Statistics. *Editors*: Hans-Georg Müller, Department of Statistics, University of California, Davis, Davis, CA 95616, USA. Harrison Zhou, Department of Statistics and Data Science, Yale University, New Haven, CT 06520, USA

The Annals of Applied Statistics. *Editor-in-Chief*: Lexin Li, Department of Biostatistics and Epidemiology, University of California, Berkeley, Berkeley, CA 94720-7360, USA

The Annals of Probability. *Editors*: Paul Bourgade, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012-1185, USA. Julien Dubedat, Department of Mathematics, Columbia University, New York, NY 10027, USA

The Annals of Applied Probability. *Editors*: Jian Ding, School of Mathematical Sciences, Peking University, 100871, Bejing, China. Claudio Landim, IMPA, 22461-320, Rio de Janeiro, Brazil

Statistical Science. *Editor*: Moulinath Banerjee, Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA

The IMS Bulletin. *Editor*: Tati Howell, bulletin@imstat.org

The Annals of Statistics **[ISSN 0090-5364 (print); ISSN 2168-8966 (online)],** Volume 53, Number 2, April 2025. Published bimonthly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, OH 44094, USA. Periodicals postage paid at Cleveland, Ohio, and at additional mailing offices.

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

Copyright © 2025 by the Institute of Mathematical Statistics
Printed in the United States of America

The Annals of Statistics
2025, Vol. 53, No. 2, 457–476
https://doi.org/10.1214/24-AOS2398
© Institute of Mathematical Statistics, 2025

WALD TESTS WHEN RESTRICTIONS ARE LOCALLY SINGULAR

BY JEAN-MARIE DUFOUR[1,a], ERIC RENAULT[2,c] AND VICTORIA ZINDE-WALSH[1,b]

[1]*Department of Economics, McGill University,* [a]*jean-marie.dufour@mcgill.ca,* [b]*victoria.zinde-walsh@mcgill.ca*
[2]*Department of Economics, University of Warwick,* [c]*Eric.Renault@warwick.ac.uk*

This paper provides an exhaustive characterization of the asymptotic null distribution of Wald-type statistics for testing restrictions given by polynomial functions—which may involve singularities—when the limiting distribution of the parameter estimator is absolutely continuous (e.g., Gaussian). In addition to the well-known finite-sample noninvariance, there is also an asymptotic noninvariance (nonpivotality): with standard critical values, the test may either under-reject or over-reject, and even diverge under the null hypothesis. The asymptotic distribution of the test statistic can vary under the null hypothesis and depends on the true unknown parameter value. All these situations are possible in testing restrictions which arise in the statistical and econometric literatures, for example, for examining the specification of ARMA models, causality at different horizons, indirect effects, zero determinant hypotheses on matrices of coefficients, and other situations where singularities cannot be excluded. We provide the limit distribution and general bounds for the single restriction case. For multiple restrictions, we give a necessary and sufficient condition for the existence of a limit distribution, and its form if it exists.

REFERENCES

[1] AL-SADOON, M. M. (2017). A unifying theory of tests of rank. *J. Econometrics* **199** 49–62. MR3652404 https://doi.org/10.1016/j.jeconom.2017.03.002

[2] ALISON, P. D. (1995). The impact of random predictors on comparisons of coefficients between models: Comment on Clogg, Petkova, and Haritou. *Am. J. Sociol.* **100** 1294–1305.

[3] ANDREWS, D. W. K. (1987). Asymptotic results for generalized Wald tests. *Econometric Theory* **3** 348–358.

[4] AVRACHENKOV, K. E., HAVIV, M. and HOWLETT, P. G. (2001). Inversion of analytic matrix functions that are singular at the origin. *SIAM J. Matrix Anal. Appl.* **22** 1175–1189. MR1824064 https://doi.org/10.1137/S0895479898337555

[5] AZAÏS, J.-M., GASSIAT, É. and MERCADIER, C. (2006). Asymptotic distribution and local power of the log-likelihood ratio test for mixtures: Bounded and unbounded cases. *Bernoulli* **12** 775–799. MR2265342 https://doi.org/10.3150/bj/1161614946

[6] BARON, R. M. and KENNY, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* **51** 1173–1182.

[7] BOLLEN, K. A., LENNOX, R. D. and DAHLY, D. L. (2009). Practical application of the vanishing tetrad test for causal indicator measurement models: An example from health-related quality of life. *Stat. Med.* **28** 1524–1536. MR2649710 https://doi.org/10.1002/sim.3560

[8] BOLLEN, K. A. and TING, K. F. (2000). A tetrad test for causal indicators. *Psychol. Methods* **5** 3–22. https://doi.org/10.1037/1082-989x.5.1.3

[9] BOUDJELLABA, H., DUFOUR, J.-M. and ROY, R. (1992). Testing causality between two vectors in multivariate autoregressive moving average models. *J. Amer. Statist. Assoc.* **87** 1082–1090. MR1209566

[10] BOUDJELLABA, H., DUFOUR, J.-M. and ROY, R. (1994). Simplified conditions for noncausality between vectors in multivariate ARMA models. *J. Econometrics* **63** 271–287. MR1309985 https://doi.org/10.1016/0304-4076(93)01568-7

[11] BREUSCH, T. S. and SCHMIDT, P. (1988). Alternative forms of the Wald test: How long is a piece of string? *Comm. Statist. Theory Methods* **17** 2789–2795. MR0955361 https://doi.org/10.1080/03610928808829771

MSC2020 subject classifications. Primary 62E20, 62F03, 62F05; secondary 62D20.
Key words and phrases. Nonlinear restriction, Wald test, deficient rank, singular covariance matrix, nonstandard asymptotic theory, bound, indirect effect, mediation.

[12] CHERNOFF, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Stat.* **25** 573–578. MR0065087 https://doi.org/10.1214/aoms/1177728725

[13] CLOGG, C. C., PETKOVA, E. and HARITOU, A. (1995). Statistical methods for comparing regression coefficients between models. *Am. J. Sociol.* **100** 1261–1293.

[14] CLOGG, C. C., PETKOVA, E. and SHIHADEH, E. S. (1992). Statistical methods for analyzing collapsibility in regression models. *J. Educ. Stat.* **17** 51–74.

[15] CRITCHLEY, F., MARRIOTT, P. and SALMON, M. (1996). On the differential geometry of the Wald test with nonlinear restrictions. *Econometrica* **64** 1213–1222. MR1403235 https://doi.org/10.2307/2171963

[16] DAGENAIS, M. G. and DUFOUR, J.-M. (1991). Invariance, nonlinear models, and asymptotic tests. *Econometrica* **59** 1601–1615. MR1135644 https://doi.org/10.2307/2938281

[17] DAGENAIS, M. G. and DUFOUR, J.-M. (1994). Pitfalls of rescaling regression models with Box-Cox transformations. *Rev. Econ. Stat.* **76** 571–575.

[18] DAVIS, L. J. (1989). Intersection union tests for strict collapsibility in three-dimensional contingency tables. *Ann. Statist.* **17** 1693–1708.

[19] DRTON, M. (2009). Likelihood ratio tests and singularities. *Ann. Statist.* **37** 979–1012. MR2502658 https://doi.org/10.1214/07-AOS571

[20] DRTON, M., MASSAM, H. and OLKIN, I. (2008). Moments of minors of Wishart matrices. *Ann. Statist.* **36** 2261–2283. MR2458187 https://doi.org/10.1214/07-AOS522

[21] DRTON, M., STURMFELS, B. and SULLIVANT, S. (2007). Algebraic factor analysis: Tetrads, pentads and beyond. *Probab. Theory Related Fields* **138** 463–493. MR2299716 https://doi.org/10.1007/s00440-006-0033-2

[22] DRTON, M., STURMFELS, B. and SULLIVANT, S. (2009). *Lectures on Algebraic Statistics*. *Oberwolfach Seminars* **39**. Birkhäuser, Basel. MR2723140 https://doi.org/10.1007/978-3-7643-8905-5

[23] DRTON, M. and XIAO, H. (2016). Wald tests of singular hypotheses. *Bernoulli* **22** 38–59. MR3449776 https://doi.org/10.3150/14-BEJ620

[24] DUCHARME, G. R. and LEPAGE, Y. (1986). Testing collapsibility in contingency tables. *J. Roy. Statist. Soc. Ser. B* **48** 197–205. MR0867997

[25] DUFOUR, J.-M. (1989). Nonlinear hypotheses, inequality restrictions, and nonnested hypotheses: Exact simultaneous tests in linear regressions. *Econometrica* **57** 335–355. MR0996940 https://doi.org/10.2307/1912558

[26] DUFOUR, J.-M. (1997). Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica* **65** 1365–1387. MR1604305 https://doi.org/10.2307/2171740

[27] DUFOUR, J.-M. (2003). Identification, weak instruments and statistical inference in econometrics. *Can. J. Econ.* **36** 767–808.

[28] DUFOUR, J.-M. and DAGENAIS, M. G. (1992). Nonlinear models, rescaling and test invariance. *J. Statist. Plann. Inference* **32** 111–135. MR1177362 https://doi.org/10.1016/0378-3758(92)90155-L

[29] DUFOUR, J.-M., PELLETIER, D. and RENAULT, É. (2006). Short run and long run causality in time series: Inference. *J. Econometrics* **132** 337–362. MR2323984 https://doi.org/10.1016/j.jeconom.2005.02.003

[30] DUFOUR, J.-M. and RENAULT, E. (1998). Short run and long run causality in time series: Theory. *Econometrica* **66** 1099–1125. MR1639415 https://doi.org/10.2307/2999631

[31] DUFOUR, J.-M, RENAULT, E. and ZINDE-WALSH, V. (2025). Supplement to "Wald tests when restrictions are locally singular." https://doi.org/10.1214/24-AOS2398SUPP

[32] DUFOUR, J.-M., TROGNON, A. and TUVAANDORJ, P. (2017). Invariant tests based on M-estimators, estimating functions, and the generalized method of moments. *Econometric Rev.* **36** 182–204. MR3555472 https://doi.org/10.1080/07474938.2015.1114285

[33] EMSLEY, R., DUNN, G. and WHITE, I. R. (2010). Mediation and moderation of treatment effects in randomised controlled trials of complex interventions. *Stat. Methods Med. Res.* **19** 237–270. MR2757115 https://doi.org/10.1177/0962280209105014

[34] GAFFKE, N., HEILIGERS, B. and OFFINGER, R. (2002). On the asymptotic null-distribution of the Wald statistic at singular parameter points. *Statist. Decisions* **20** 379–398. MR1960710

[35] GAFFKE, N., STEYER, R. and VON DAVIER, A. A. (1999). On the asymptotic null-distribution of the Wald statistic at singular parameter points. *Statist. Decisions* **17** 339–358. MR1745064

[36] GALBRAITH, J. W. and ZINDE-WALSH, V. (1997). On some simple, autoregression-based estimation and identification techniques for ARMA models. *Biometrika* **84** 685–696. MR1603948 https://doi.org/10.1093/biomet/84.3.685

[37] GLONEK, G. F. V. (1993). On the behaviour of Wald statistics for the disjunction of two regular hypotheses. *J. Roy. Statist. Soc. Ser. B* **55** 749–755. MR1223941

[38] GOURIÉROUX, C. and JASIAK, J. (2013). Size distortion in the analysis of volatility and covolatility effects. In *Uncertainty Analysis in Econometrics with Applications* (V. N. Huynh, V. Kreinovich, S. Sriboonchita and K. Suriya, eds.) 91–118. Springer, Berlin.

[39] GOURIÉROUX, C., MONFORT, A. and RENAULT, E. (1989). Testing for common roots. *Econometrica* **57** 171–185. MR0988252 https://doi.org/10.2307/1912578
[40] GOURIÉROUX, C., MONFORT, A. and RENAULT, É. (1990). Bilinear constraints: Estimation and test. In *Essays in Honor of Edmond Malinvaud*, *Volume* 3: *Empirical Economics* (P. Champsaur, M. Deleau, J.-M. Grandmont, R. Guesnerie, C. Henry, J.-J. Laffont, G. Laroque, J. Mairesse, A. Monfort et al., eds.) 166–191. MIT Press, Cambridge, MA.
[41] GOURIÉROUX, C., MONFORT, A. and RENAULT, E. (1993). Tests sur le noyau, l'image et le rang de la matrice des coefficients d'un modèle linéaire multivarié. *Ann*. *Écon*. *Stat*. **32** 81–111. MR1252035 https://doi.org/10.2307/20075927
[42] GREGORY, A. and VEALL, M. (1985). Formulating Wald tests of nonlinear restrictions. *Econometrica* **53** 1465–1468.
[43] HAYES, A. F. (2013). *Introduction to Mediation*, *Moderation*, *and Conditional Process Analysis*: *A Regression-Based Approach*. Guilford, New York.
[44] HIPP, J. R. and BOLLEN, K. A. (2003). Model fit in structural equation models with censored, ordinal, and dichotomous variables: Testing vanishing tetrads. *Sociol*. *Method*. **3** 267–305.
[45] HÖRMANDER, L. (1990). *An Introduction to Complex Analysis in Several Variables*, 3rd ed. *North-Holland Mathematical Library* **7**. North-Holland, Amsterdam. MR1045639
[46] JOHNSON, T. R. and BODNER, T. E. (2007). A note on the use of bootstrap tetrad tests for covariance structures. *Struct*. *Equ*. *Model*. **14** 113–124. MR2339072 https://doi.org/10.1207/s15328007sem1401_6
[47] KATO, N. and KURIKI, S. (2013). Likelihood ratio tests for positivity in polynomial regressions. *J*. *Multivariate Anal*. **115** 334–346. MR3004562 https://doi.org/10.1016/j.jmva.2012.10.016
[48] KOO, N., LEITE, W. L. and ALGINA, J. (2016). Medicated effects with the parallel process latent growth model: An evaluation of methods for testing mediation in the presence of nonnormal data. *Struct*. *Equ*. *Model*. **23** 32–44. MR3455492 https://doi.org/10.1080/10705511.2014.959419
[49] LANGENHOP, C. E. (1977). The inverse of a matrix polynomial. *Linear Algebra Appl*. **16** 267–284. MR0469935 https://doi.org/10.1016/0024-3795(77)90009-x
[50] LÜTKEPOHL, H. and BURDA, M. M. (1997). Modified Wald tests under nonregular conditions. *J*. *Econometrics* **78** 315–332. MR1453483 https://doi.org/10.1016/S0304-4076(96)00015-2
[51] MACKINNON, D. P. (2008). *Introduction to Statistical Mediation Analysis*. Lawrence Erlbaum Associates, New York.
[52] MACKINNON, D. P., LOCKWOOD, C. M., HOFFMAN, J. M., WEST, S. G. and SHEETS, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *J*. *Pers*. *Soc*. *Psychol*. **7** 83–104.
[53] MACKINNON, D. P., LOCKWOOD, C. M. and WILLIAMS, J. (2006). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivar. Behav. Res*. **39** 99–128.
[54] MITYAGIN, B. S. (2020). The zero set of a real analytic function. *Mat*. *Zametki* **107** 473–475. MR4070868 https://doi.org/10.4213/mzm12620
[55] OKAMOTO, M. (1973). Distinctness of the eigenvalues of a quadratic form in a multivariate sample. *Ann*. *Statist*. **1** 763–765. MR0331643
[56] PHILLIPS, P. C. B. and PARK, J. Y. (1988). On the formulation of Wald tests of nonlinear restrictions. *Econometrica* **56** 1065–1083. MR0964149 https://doi.org/10.2307/1911359
[57] PILLAI, N. S. and MENG, X.-L. (2016). An unexpected encounter with Cauchy and Lévy. *Ann*. *Statist*. **44** 2089–2097. MR3546444 https://doi.org/10.1214/15-AOS1407
[58] RITZ, C. and SKOVGAARD, I. M. (2005). Likelihood ratio tests in curved exponential families with nuisance parameters present only under the alternative. *Biometrika* **92** 507–517. MR2202642 https://doi.org/10.1093/biomet/92.3.507
[59] ROBIN, J.-M. and SMITH, R. J. (2000). Tests of rank. *Econometric Theory* **16** 151–175. MR1763430 https://doi.org/10.1017/S0266466600162012
[60] SAIN, M. K. and MASSEY, J. L. (1969). Invertibility of linear time-invariant dynamical systems. *IEEE Trans*. *Automat*. *Control* **AC-14** 141–149. MR0246664 https://doi.org/10.1109/tac.1969.1099133
[61] SARGAN, J. D. (1980). Some tests of dynamic specification for a single equation. *Econometrica* **48** 879–897. MR0575031 https://doi.org/10.2307/1912938
[62] SCHATZMAN, M. (2002). *Numerical Analaysis*: *A Mathematical Introduction*. Clarendon Press, Oxford.
[63] SILVA, R., SCHEINES, R., GLYMOUR, C. and SPIRTES, P. (2006). Learning the structure of linear latent variable models. *J*. *Mach*. *Learn*. *Res*. **7** 191–246. MR2274367
[64] SIMPSON, E. H. (1951). The interpretation of interaction in contingency tables. *J*. *Roy*. *Statist*. *Soc*. *Ser*. *B* **13** 238–241. MR0051472
[65] SOBEL, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociol*. *Method*. **13** 290–312.

[66] SOBEL, M. E. (1986). Some new results on indirect effects and their standard errors in covariance structure models. *Sociol. Method.* **16** 159–186.

[67] SPIRTES, P., GLYMOUR, C. and SCHEINES, R. (2000). *Causation, Prediction, and Search. Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA.

[68] STEYER, R., GABLER, S. and RUCAI, A. A. (1996). Individual causal effects, average causal effects, and unconfoundedness in regression models. In *Softstat'95: Advances in Statistical Software* 5 (F. Faulbaum and W. Bandilla, eds.) 203–210. Lucius & Lucius, Heidelberg.

[69] STEYER, R., MAYER, A. and FIEGE, C. (2014). Causal inference on total, direct, and indirect effects. In *Encyclopedia of Quality of Life Research* (A. C. Michalos, ed.) 606–630. Springer, Dordrecht.

[70] SULLIVANT, S., TALASKA, K. and DRAISMA, J. (2010). Trek separation for Gaussian graphical models. *Ann. Statist.* **38** 1665–1685. MR2662356 https://doi.org/10.1214/09-AOS760

[71] TOFIGHI, D. and MACKINNON, D. P. (2016). Monte Carlo confidence intervals for complex functions if indirect effects. *Struct. Equ. Model.* **23** 194–205. MR3465641 https://doi.org/10.1080/10705511.2015.1057284

[72] VANDERWEELE, T. (2015). *Explanation in Causal Inference Methods for Mediation and Interaction*. Oxford Univ. Press, Oxford, New York.

[73] WHITTEMORE, A. S. (1978). Collapsibility of multidimensional contingency tables. *J. Roy. Statist. Soc. Ser. B* **40** 328–340. MR0522216

[74] ZWIERNIK, P. and SMITH, J. Q. (2012). Tree cumulants and the geometry of binary tree models. *Bernoulli* **18** 290–321. MR2888708 https://doi.org/10.3150/10-BEJ338

The Annals of Statistics
2025, Vol. 53, No. 2, 477–505
https://doi.org/10.1214/24-AOS2419
© Institute of Mathematical Statistics, 2025

LARGE-DIMENSIONAL INDEPENDENT COMPONENT ANALYSIS: STATISTICAL OPTIMALITY AND COMPUTATIONAL TRACTABILITY

BY ARNAB AUDDY[1,a] AND MING YUAN[2,b]

[1]*Department of Statistics, The Ohio State University,* [a]*auddy.1@osu.edu*
[2]*Department of Statistics, Columbia University,* [b]*ming.yuan@columbia.edu*

In this paper, we investigate the optimal statistical performance and the impact of computational constraints for independent component analysis (ICA). Our goal is twofold. On the one hand, we characterize the precise role of dimensionality on sample complexity and statistical accuracy, and how computational consideration may affect them. In particular, we show that the optimal sample complexity is linear in dimensionality, and interestingly, the commonly used sample kurtosis-based approaches are necessarily suboptimal. However, the optimal sample complexity becomes quadratic, up to a logarithmic factor, in the dimension if we restrict ourselves to estimates that can be computed with low-degree polynomial algorithms. On the other hand, we develop computationally tractable estimates that attain both the optimal sample complexity and minimax optimal rates of convergence. We study the asymptotic properties of the proposed estimates and establish their asymptotic normality that can be readily used for statistical inferences. Our method is fairly easy to implement and numerical experiments are presented to further demonstrate its practical merits.

REFERENCES

AMARI, S.-I. and CARDOSO, J.-F. (1997). Blind source separation-semiparametric statistical approach. *IEEE Trans*. *Signal Process*. **45** 2692–2700.
ANANDKUMAR, A., GE, R., HSU, D., KAKADE, S. M. and TELGARSKY, M. (2014a). Tensor decompositions for learning latent variable models. *J*. *Mach*. *Learn*. *Res*. **15** 2773–2832. MR3270750
ANANDKUMAR, A., GE, R., HSU, D., KAKADE, S. M. and TELGARSKY, M. (2014b). Tensor decompositions for learning latent variable models. *J*. *Mach*. *Learn*. *Res*. **15** 2773–2832. MR3270750
ANANDKUMAR, A., GE, R. and JANZAMIN, M. (2014c). Sample complexity analysis for learning overcomplete latent variable models through tensor methods. ArXiv preprint. Available at arXiv:1408.0553.
ANANDKUMAR, A., GE, R. and JANZAMIN, M. (2014d). Guaranteed non-orthogonal tensor decomposition via alternating rank-1 updates. arXiv preprint. Available at arXiv:1402.5180.
AUDDY, A. and YUAN, M. (2022). On estimating rank-one spiked tensors in the presence of heavy tailed errors. *IEEE Trans*. *Inf*. *Theory* **68** 8053–8075. MR4544931 https://doi.org/10.1109/tit.2022.3191883
AUDDY, A. and YUAN, M. (2023). Perturbation bounds for (nearly) orthogonally decomposable tensors with statistical applications. *Inf*. *Inference* **12** 1044–1072. MR4565757 https://doi.org/10.1093/imaiai/iaac033
AUDDY, A. and YUAN, M. (2025). Supplement to "Large-Dimensional Independent Component Analysis: Statistical Optimality and Computational Tractability." https://doi.org/10.1214/24-AOS2419SUPP
BELKIN, M., RADEMACHER, L. and VOSS, J. (2013). Blind signal separation in the presence of Gaussian noise. In *Conference on Learning Theory* 270–287. PMLR.
BELKIN, M., RADEMACHER, L. and VOSS, J. (2018). Eigenvectors of orthogonally decomposable functions. *SIAM J*. *Comput*. **47** 547–615. MR3794351 https://doi.org/10.1137/17M1122013
BHASKARA, A., CHARIKAR, M. and VIJAYARAGHAVAN, A. (2014). Uniqueness of tensor decompositions with applications to polynomial identifiability. In *Conference on Learning Theory* 742–778.
CARDOSO, J.-F. (1999). High-order contrasts for independent component analysis. *Neural Comput*. **11** 157–192.
CARDOSO, J.-F. and COMON, P. (1996). Independent component analysis, a survey of some algebraic methods. In 1996 *IEEE International Symposium on Circuits and Systems*. *Circuits and Systems Connecting the World*. *ISCAS* 96 **2** 93–96. IEEE, Los Alamitos.

MSC2020 subject classifications. 62H12, 62H25.
Key words and phrases. Independent component analysis, cumulants, tensors, minimax optimality, statistical-computational gaps.

CARDOSO, J.-F. and SOULOUMIAC, A. (1993). Blind beamforming for non-Gaussian signals. In *IEE Proceedings F (Radar and Signal Processing)* **140** 362–370. IET.

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

CHEN, A. and BICKEL, P. J. (2005). Consistent independent component analysis and prewhitening. *IEEE Trans. Signal Process.* **53** 3625–3632. MR2239886 https://doi.org/10.1109/TSP.2005.855098

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

CICHOCKI, A., MANDIC, D., DE LATHAUWER, L., ZHOU, G., ZHAO, Q., CAIAFA, C. and PHAN, H. A. (2015). Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE Signal Process. Mag.* **32** 145–163.

COMON, P. (1994). Independent component analysis, a new concept? *Signal Process.* **36** 287–314.

COMON, P. and JUTTEN, C. (2010). *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press, San Diego.

CONNOR, G., GOLDBERG, L. R. and KORAJCZYK, R. A. (2010). *Portfolio Risk Analysis*. Princeton Univ. Press, Princeton.

DELFOSSE, N. and LOUBATON, P. (1995). Adaptive blind separation of independent sources: A deflation approach. *Signal Process.* **45** 59–83.

ERIKSSON, J. and KOIVUNEN, V. (2003). Characteristic-function-based independent component analysis. *Signal Process.* **83** 2195–2208.

ERIKSSON, J. and KOIVUNEN, V. (2004). Identifiability, separability, and uniqueness of linear ica models. *IEEE Signal Process. Lett.* **11** 601–604.

GOYAL, N., VEMPALA, S. and XIAO, Y. (2014). Fourier pca and robust tensor decomposition. In *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing* 584–593.

HASTIE, T. and TIBSHIRANI, R. (2002). Independent components analysis through product density estimation. *Adv. Neural Inf. Process. Syst.* **15**.

HOPKINS, S. (2018). *Statistical Inference and the Sum of Squares Method*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—Cornell University. MR3864930

HSU, D. and SABATO, S. (2016). Loss minimization and parameter estimation with heavy tails. *J. Mach. Learn. Res.* **17** 18. MR3491112

HYVARINEN, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.* **10** 626–634.

HYVÄRINEN, A. and OJA, E. (2000). Independent component analysis: Algorithms and applications. *Neural Netw.* **13** 411–430.

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

JUTTEN, C. and HERAULT, J. (1991). Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Process.* **24** 1–10.

KE, Y., MINSKER, S., REN, Z., SUN, Q. and ZHOU, W.-X. (2019). User-friendly covariance estimation for heavy-tailed distributions. *Statist. Sci.* **34** 454–471. MR4017523 https://doi.org/10.1214/19-STS711

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

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

KOLTCHINSKII, V. and LOUNICI, K. (2016). Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 1976–2013. MR3573302 https://doi.org/10.1214/15-AIHP705

KOLTCHINSKII, V. and LOUNICI, K. (2017). Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli* **23** 110–133. MR3556768 https://doi.org/10.3150/15-BEJ730

KUNISKY, D., WEIN, A. S. and BANDEIRA, A. S. (2022). Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *Mathematical Analysis, Its Applications and Computation. Springer Proc. Math. Stat.* **385** 1–50. Springer, Cham. MR4461037 https://doi.org/10.1007/978-3-030-97127-4_1

LASSANCE, N., DEMIGUEL, V. and VRINS, F. (2022). Optimal portfolio diversification via independent component analysis. *Oper. Res.* **70** 55–72. MR4379572 https://doi.org/10.1287/opre.2021.2140

LASSANCE, N. and VRINS, F. (2021). Portfolio selection with parsimonious higher comoments estimation. *J. Bank. Financ.* **126** 106115.

LEE, T.-W., GIROLAMI, M. and SEJNOWSKI, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Comput.* **11** 417–441.

LUGOSI, G. and MENDELSON, S. (2020). Risk minimization by median-of-means tournaments. *J. Eur. Math. Soc.* (*JEMS*) **22** 925–965. MR4055993 https://doi.org/10.4171/jems/937

MINSKER, S. (2015). Geometric median and robust estimation in Banach spaces. *Bernoulli* **21** 2308–2335. MR3378468 https://doi.org/10.3150/14-BEJ645

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

MU, C., HSU, D. and GOLDFARB, D. (2015). Successive rank-one approximations for nearly orthogonally decomposable symmetric tensors. *SIAM J. Matrix Anal. Appl.* **36** 1638–1659. MR3432147 https://doi.org/10.1137/15M1010890

MU, C., HSU, D. and GOLDFARB, D. (2017). Greedy approaches to symmetric orthogonal tensor decomposition. *SIAM J. Matrix Anal. Appl.* **38** 1210–1226. MR3717819 https://doi.org/10.1137/16M1087734

NORDHAUSEN, K. and OJA, H. (2018). Independent component analysis: A statistical perspective. *Wiley Interdiscip. Rev.: Comput. Stat.* **10** e1440. MR3850449 https://doi.org/10.1002/wics.1440

PHAM, D. T. (1996). Blind separation of instantaneous mixture of sources via an independent component analysis. *IEEE Trans. Signal Process.* **44** 2768–2779.

PHAM, D. T. and GARAT, P. (1997). Blind separation of mixture of independent sources through a quasi-maximum likelihood approach. *IEEE Trans. Signal Process.* **45** 1712–1725.

ROBERTS, S. and EVERSON, R. (2001). *Independent Component Analysis: Principles and Practice*. Cambridge Univ. Press, Cambridge.

SAMAROV, A. and TSYBAKOV, A. (2004). Nonparametric independent component analysis. *Bernoulli* **10** 565–582. MR2076063 https://doi.org/10.3150/bj/1093265630

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

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

SIDIROPOULOS, N. D., DE LATHAUWER, L., FU, X., HUANG, K., PAPALEXAKIS, E. E. and FALOUTSOS, C. (2017). Tensor decomposition for signal processing and machine learning. *IEEE Trans. Signal Process.* **65** 3551–3582. MR3666587 https://doi.org/10.1109/TSP.2017.2690524

STONE, J. V. (2004). *Independent Component Analysis: A Tutorial Introduction*.

TROPP, J. A. et al. (2015). An introduction to matrix concentration inequalities. *Found. Trends Mach. Learn.* **8** 1–230.

VEMPALA, S. S. and XIAO, Y. (2015). Max vs min: Tensor decomposition and ica with nearly linear sample complexity. In *Conference on Learning Theory* 1710–1723. PMLR.

VOSS, J. R., BELKIN, M. and RADEMACHER, L. (2015). A pseudo-Euclidean iteration for optimal recovery in noisy ica. *Adv. Neural Inf. Process. Syst.* **28**.

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

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

ZARZOSO, V., COMON, P. and KALLEL, M. (2006). How fast is fastica? In 2006 14*th European Signal Processing Conference* 1–5. IEEE, Los Alamitos.

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

The Annals of Statistics
2025, Vol. 53, No. 2, 506–530
https://doi.org/10.1214/24-AOS2454
© Institute of Mathematical Statistics, 2025

FORWARD SELECTION AND POST-SELECTION INFERENCE IN FACTORIAL DESIGNS

BY LEI SHI[1,a], JINGSHEN WANG[1,b] AND PENG DING[2,c]

[1]*Division of Biostatistics, University of California, Berkeley,* [a]*leishi@berkeley.edu*
[2]*Department of Statistics, University of California, Berkeley,* [b]*jingshenwang@berkeley.edu,* [c]*pengdingpku@berkeley.edu*

Ever since the seminal work of R. A. Fisher and F. Yates, factorial designs have been an important experimental tool to simultaneously estimate the effects of multiple treatment factors. In factorial designs, the number of treatment combinations grows exponentially with the number of treatment factors, which motivates the forward selection strategy based on the sparsity, hierarchy and heredity principles for factorial effects. Although this strategy is intuitive and has been widely used in practice, its rigorous statistical theory has not been formally established. To fill this gap, we establish design-based theory for forward factor selection in factorial designs based on the potential outcome framework. We not only prove a consistency property for the factor selection procedure but also discuss statistical inference after factor selection. In particular, with selection consistency, we quantify the advantages of forward selection based on asymptotic efficiency gain in estimating factorial effects. With inconsistent selection in higher-order interactions, we propose two strategies and investigate their impact on subsequent inference. Our formulation differs from the existing literature on variable selection and post-selection inference because our theory is based solely on the physical randomization of the factorial design and does not rely on a correctly specified outcome model.

REFERENCES

[1] ANDREWS, I., KITAGAWA, T. and MCCLOSKEY, A. (2019). Inference on winners Technical Report, National Bureau of Economic Research.

[2] BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2010). Hierarchical selection of variables in sparse high-dimensional regression. In *Borrowing Strength*: *Theory Powering Applications—a Festschrift for Lawrence D*. *Brown*. *Inst*. *Math*. *Stat*. (*IMS*) *Collect*. **6** 56–69. IMS, Beachwood, OH. MR2798511

[3] BIEN, J., TAYLOR, J. and TIBSHIRANI, R. (2013). A lasso for hierarchical interactions. *Ann*. *Statist*. **41** 1111–1141. MR3113805 https://doi.org/10.1214/13-AOS1096

[4] BLACKWELL, M. and PASHLEY, N. E. (2023). Noncompliance and instrumental variables for 2^K factorial experiments. *J*. *Amer*. *Statist*. *Assoc*. **118** 1102–1114. MR4595480 https://doi.org/10.1080/01621459.2021.1978468

[5] BLONIARZ, A., LIU, H., ZHANG, C.-H., SEKHON, J. S. and YU, B. (2016). Lasso adjustments of treatment effect estimates in randomized experiments. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **113** 7383–7390. MR3531136 https://doi.org/10.1073/pnas.1510506113

[6] BOX, G. E. P., HUNTER, J. S. and HUNTER, W. G. (2005). *Statistics for Experimenters*: *Design*, *Innovation*, *and Discovery*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR2140250

[7] BRANSON, Z., DASGUPTA, T. and RUBIN, D. B. (2016). Improving covariate balance in 2^K factorial designs via rerandomization with an application to a New York City Department of Education high school study. *Ann*. *Appl*. *Stat*. **10** 1958–1976. MR3592044 https://doi.org/10.1214/16-AOAS959

[8] CAUGHEY, D., KATSUMATA, H. and YAMAMOTO, T. (2019). Item response theory for conjoint survey experiments Technical Report, Working Paper.

[9] DASGUPTA, T., PILLAI, N. S. and RUBIN, D. B. (2015). Causal inference from 2^K factorial designs by using potential outcomes. *J*. *R*. *Stat*. *Soc*. *Ser*. *B*. *Stat*. *Methodol*. **77** 727–753. MR3382595 https://doi.org/10.1111/rssb.12085

MSC2020 subject classifications. Primary 62K15, 62J05; secondary 62G05.
Key words and phrases. Causal inference, design-based inference, forward selection, post-selection inference.

[10] EGAMI, N. and IMAI, K. (2019). Causal interaction in factorial experiments: Application to conjoint analysis. *J. Amer. Statist. Assoc.* **114** 529–540. MR3963160 https://doi.org/10.1080/01621459.2018.1476246

[11] ESPINOSA, V., DASGUPTA, T. and RUBIN, D. B. (2016). A Bayesian perspective on the analysis of unreplicated factorial experiments using potential outcomes. *Technometrics* **58** 62–73. MR3463157 https://doi.org/10.1080/00401706.2015.1006337

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

[13] FISHER, R. A. (1935). *The Design of Experiments*: *Oliver and Boyd*, 1st ed. Edinburgh, London.

[14] FITHIAN, W., SUN, D. and TAYLOR, J. (2014). Optimal inference after model selection. Preprint. Available at arXiv:1410.2597.

[15] FREEDMAN, D. A. (2008). On regression adjustments to experimental data. *Adv. in Appl. Math.* **40** 180–193. MR2388610 https://doi.org/10.1016/j.aam.2006.12.003

[16] GERBER, A. S. and GREEN, D. P. (2012). *Field Experiments*: *Design*, *Analysis*, *and Interpretation*, Norton, New York, NY.

[17] GUO, X., WEI, L., WU, C. and WANG, J. (2021). Sharp inference on selected subgroups in observational studies. Preprint. Available at arXiv:2102.11338.

[18] HAINMUELLER, J., HOPKINS, D. J. and YAMAMOTO, T. (2014). Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Polit. Anal.* **22** 1–30.

[19] HAO, N., FENG, Y. and ZHANG, H. H. (2018). Model selection for high-dimensional quadratic regression via regularization. *J. Amer. Statist. Assoc.* **113** 615–625. MR3832213 https://doi.org/10.1080/01621459.2016.1264956

[20] HAO, N. and ZHANG, H. H. (2014). Interaction screening for ultrahigh-dimensional data. *J. Amer. Statist. Assoc.* **109** 1285–1301. MR3265697 https://doi.org/10.1080/01621459.2014.881741

[21] HARIS, A., WITTEN, D. and SIMON, N. (2016). Convex modeling of interactions with strong heredity. *J. Comput. Graph. Statist.* **25** 981–1004. MR3572025 https://doi.org/10.1080/10618600.2015.1067217

[22] KEMPTHORNE, O. (1952). *The Design and Analysis of Experiments*. Wiley, New York. MR0045368

[23] KUCHIBHOTLA, A. K., KOLASSA, J. E. and KUFFNER, T. A. (2022). Post-selection inference. *Annu. Rev. Stat. Appl.* **9** 505–527. MR4394918 https://doi.org/10.1146/annurev-statistics-100421-044639

[24] LI, W., NACHTSHEIM, C. J., WANG, K., REUL, R. and ALBRECHT, M. (2013). Conjoint analysis and discrete choice experiments for quality improvement. *J. Qual. Technol.* **45** 74–99.

[25] LI, X. and DING, P. (2017). General forms of finite population central limit theorems with applications to causal inference. *J. Amer. Statist. Assoc.* **112** 1759–1769. MR3750897 https://doi.org/10.1080/01621459.2017.1295865

[26] LIM, M. and HASTIE, T. (2015). Learning interactions via hierarchical group-lasso regularization. *J. Comput. Graph. Statist.* **24** 627–654. MR3397226 https://doi.org/10.1080/10618600.2014.938812

[27] LIN, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *Ann. Appl. Stat.* **7** 295–318. MR3086420 https://doi.org/10.1214/12-AOAS583

[28] LIU, H., REN, J. and YANG, Y. (2024). Randomization-based joint central limit theorem and efficient covariate adjustment in randomized block 2^K factorial experiments. *J. Amer. Statist. Assoc.* **119** 136–150. MR4713881 https://doi.org/10.1080/01621459.2022.2102985

[29] LU, J. (2016). On randomization-based and regression-based inferences for 2^K factorial designs. *Statist. Probab. Lett.* **112** 72–78. MR3475490 https://doi.org/10.1016/j.spl.2016.01.010

[30] LU, J. (2016). Covariate adjustment in randomization-based causal inference for 2^K factorial designs. *Statist. Probab. Lett.* **119** 11–20. MR3555264 https://doi.org/10.1016/j.spl.2016.07.010

[31] LUCE, R. D. and TUKEY, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *J. Math. Psych.* **1** 1–27.

[32] MENG, X.-L. and XIE, X. (2014). I got more data, my model is more refined, but my estimator is getting worse! Am I just dumb? *Econometric Rev.* **33** 218–250. MR3170847 https://doi.org/10.1080/07474938.2013.808567

[33] PASHLEY, N. E. and BIND, M.-A. C. (2023). Causal inference for multiple treatments using fractional factorial designs. *Canad. J. Statist.* **51** 444–468. MR4595237 https://doi.org/10.1002/cjs.11734

[34] SHAO, J. (1997). An asymptotic theory for linear model selection. *Statist. Sinica* **7** 221–264. MR1466682

[35] SHI, L. and DING, P. (2022). Berry–Esseen bounds for design-based causal inference with possibly diverging treatment levels and varying group sizes. Preprint. Available at arXiv:2209.12345.

[36] SHI, L., WANG, J. and DING, P. (2025). Supplement to "Forward selection and post-selection inference in factorial designs." https://doi.org/10.1214/24-AOS2454SUPP

[37] SPLAWA-NEYMAN, J. (1923/1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statist. Sci.* **5** 465–472. MR1092986

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

[39] WANG, H. (2009). Forward regression for ultra-high dimensional variable screening. *J. Amer. Statist. Assoc.* **104** 1512–1524. MR2750576 https://doi.org/10.1198/jasa.2008.tm08516

[40] WASSERMAN, L. and ROEDER, K. (2009). High-dimensional variable selection. *Ann. Statist.* **37** 2178–2201. MR2543689 https://doi.org/10.1214/08-AOS646

[41] WEI, W., ZHOU, Y., ZHENG, Z. and WANG, J. (2024). Inference on the best policies with many covariates. *J. Econometrics* **239** Paper No. 105460, 14. MR4708623 https://doi.org/10.1016/j.jeconom.2022.06.013

[42] WIECZOREK, J. and LEI, J. (2022). Model selection properties of forward selection and sequential cross-validation for high-dimensional regression. *Canad. J. Statist.* **50** 454–470. MR4429847 https://doi.org/10.1002/cjs.11635

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

[44] WU, Y., ZHENG, Z., ZHANG, G., ZHANG, Z. and WANG, C. (2022). Non-stationary a/b tests: Optimal variance reduction, bias correction, and valid inference. Bias Correction, and Valid Inference (May 20, 2022).

[45] YATES, F. (1937). The design and analysis of factorial experiments Technical Report No. Technical Communication 35, Imperial Bureau of Soil Science, London, U. K.

[46] YU, R. and DING, P. (2023). Balancing weights for causal inference in observational factorial studies. Preprint. Available at arXiv:2310.04660.

[47] YUAN, M., JOSEPH, V. R. and LIN, Y. (2007). An efficient variable selection approach for analyzing designed experiments. *Technometrics* **49** 430–439. MR2414515 https://doi.org/10.1198/004017007000000173

[48] ZHAO, A. and DING, P. (2022). Regression-based causal inference with factorial experiments: Estimands, model specifications and design-based properties. *Biometrika* **109** 799–815. MR4472849 https://doi.org/10.1093/biomet/asab051

[49] ZHAO, A. and DING, P. (2023). Covariate adjustment in multiarmed, possibly factorial experiments. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **85** 1–23. MR4726956 https://doi.org/10.1093/jrsssb/qkac003

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

[51] ZHAO, P. and YU, B. (2006). On model selection consistency of Lasso. *J. Mach. Learn. Res.* **7** 2541–2563. MR2274449

[52] ZHAO, S., WITTEN, D. and SHOJAIE, A. (2021). In defense of the indefensible: A very naïve approach to high-dimensional inference. *Statist. Sci.* **36** 562–577. MR4323053 https://doi.org/10.1214/20-sts815

[53] ZHIRKOV, K. (2022). Estimating and using individual marginal component effects from conjoint experiments. *Polit. Anal.* **30** 236–249.

The Annals of Statistics
2025, Vol. 53, No. 2, 531–560
https://doi.org/10.1214/24-AOS2464
© Institute of Mathematical Statistics, 2025

OBSERVABLE ADJUSTMENTS IN SINGLE-INDEX MODELS FOR REGULARIZED M-ESTIMATORS WITH BOUNDED P/N

BY PIERRE C. BELLEC[a]

Department of Statistics, Rutgers University, [a]*pierre.bellec@rutgers.edu*

We consider observations $(\boldsymbol{X}, \boldsymbol{y})$ from single index models with unknown link function, Gaussian covariates and a regularized M-estimator $\hat{\boldsymbol{\beta}}$ constructed from convex loss function and regularizer. In the regime where sample size n and dimension p are both increasing such that p/n has a finite limit, the behavior of the empirical distribution of $\hat{\boldsymbol{\beta}}$ and the predicted values $\boldsymbol{X}\hat{\boldsymbol{\beta}}$ has been previously characterized in a number of models: The empirical distributions are known to converge to proximal operators of the loss and penalty in a related Gaussian sequence model, which captures the interplay between ratio $\frac{p}{n}$, loss, regularization and the data generating process. This connection between $(\hat{\boldsymbol{\beta}}, \boldsymbol{X}\hat{\boldsymbol{\beta}})$ and the corresponding proximal operators involves mean-field parameters defined as solutions to a nonlinear system of equations. This system typically involve unobservable quantities such as the prior distribution on the index or the link function, so the mean-field parameters need to be estimated. Although estimators for the mean-field parameters have been proposed in specific cases, a general framework that applies simultaneously to a broad class of loss and penalty has so far been missing.

This paper develops a different theory to describe the empirical distribution of $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{X}\hat{\boldsymbol{\beta}}$: Approximations of $(\hat{\boldsymbol{\beta}}, \boldsymbol{X}\hat{\boldsymbol{\beta}})$ in terms of proximal operators are provided that only involve observable adjustments in place of the mean-field parameters. These proposed observable adjustments are data-driven, for example, do not require prior knowledge of the index or the link function. These new adjustments yield confidence intervals for individual components of the index, as well as estimators of the correlation of $\hat{\boldsymbol{\beta}}$ with the index, enabling parameter tuning to maximize the correlation. The interplay between loss, regularization and the model is captured in a data-driven manner, without relying on the nonlinear systems studied in previous works. The results are proved to hold both strongly convex regularizers and unregularized M-estimation.

REFERENCES

[1] BAYATI, M., ERDOGDU, M. A. and MONTANARI, A. (2013). Estimating lasso risk and noise level. In *Advances in Neural Information Processing Systems* 944–952.

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

[3] BEAN, D., BICKEL, P. J., EL KAROUI, N. and YU, B. (2013). Optimal m-estimation in high-dimensional regression. *Proc. Natl. Acad. Sci. USA* **110** 14563–14568.

[4] BELLEC, P. and KUCHIBHOTLA, A. (2019). First order expansion of convex regularized estimators. In *Advances in Neural Information Processing Systems* (*NeurIPS*) 3457–3468.

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

[6] BELLEC, P. C. (2025). Simultaneous analysis of approximate leave-one-out cross-validation and mean-field inference. Preprint. Available at arXiv:2501.02624.

[7] BELLEC, P. C. (2025). Supplement to "Observable adjustments in single-index models for regularized M-estimators with bounded p/n." https://doi.org/10.1214/24-AOS2464SUPP

MSC2020 subject classifications. 62J12, 62H15, 62E20.

Key words and phrases. Proportional regime, mean-field inference, confidence intervals in high-dimensions, approximate normality, parameter tuning.

[8] BELLEC, P. C. and KORIYAMA, T. (2023). Error estimation and adaptive tuning for unregularized robust M-estimator. arXiv preprint.

[9] BELLEC, P. C. and KORIYAMA, T. (2025). Phase transition for unregularized m-estimation in single index model. Preprint. Available at arXiv:2501.03163.

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

[11] BELLEC, P. C., SHEN, Y. and ZHANG, C.-H. (2022). Asymptotic normality of robust M-estimators with convex penalty. *Electron. J. Stat.* **16** 5591–5622. MR4497866 https://doi.org/10.1214/22-ejs2065

[12] BELLEC, P. C. and ZHANG, C.-H. (2021). Second-order Stein: SURE for SURE and other applications in high-dimensional inference. *Ann. Statist.* **49** 1864–1903. MR4319234 https://doi.org/10.1214/20-aos2005

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

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

[15] BRILLINGER, D. R. (1983). A generalized linear model with "Gaussian" regressor variables. In *A Festschrift for Erich L. Lehmann. Wadsworth Statist./Probab. Ser.* 97–114. Wadsworth, Belmont, CA. MR0689741

[16] CANDÈS, E. J. and SUR, P. (2020). The phase transition for the existence of the maximum likelihood estimate in high-dimensional logistic regression. *Ann. Statist.* **48** 27–42. MR4065151 https://doi.org/10.1214/18-AOS1789

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

[18] 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

[19] DICKER, L. H. (2014). Variance estimation in high-dimensional linear models. *Biometrika* **101** 269–284. MR3215347 https://doi.org/10.1093/biomet/ast065

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

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

[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] 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.

[24] GERBELOT, C., ABBARA, A. and KRZAKALA, F. (2020). Asymptotic errors for convex penalized linear regression beyond Gaussian matrices. Preprint. Available at arXiv:2002.04372.

[25] HAN, Q. and REN, H. (2022). Gaussian random projections of convex cones: Approximate kinematic formulae and applications. Preprint. Available at arXiv:2212.05545.

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

[27] LOUREIRO, B., GERBELOT, C., CUI, H., GOLDT, S., KRZAKALA, F., MEZARD, M. and ZDEBOROVÁ, L. (2021). Learning curves of generic features maps for realistic datasets with a teacher-student model. *Adv. Neural Inf. Process. Syst.* **34** 18137–18151.

[28] LOUREIRO, B., SICURO, G., GERBELOT, C., PACCO, A., KRZAKALA, F. and ZDEBOROVÁ, L. (2021). Learning Gaussian mixtures with generalized linear models: Precise asymptotics in high-dimensions. *Adv. Neural Inf. Process. Syst.* **34** 10144–10157.

[29] MAI, X., LIAO, Z. and COUILLET, R. (2019). A large scale analysis of logistic regression: Asymptotic performance and new insights. In *ICASSP* 2019-2019 *IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*) 3357–3361. IEEE.

[30] 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

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

[32] Plan, Y. and Vershynin, R. (2016). The generalized Lasso with non-linear observations. *IEEE Trans. Inf. Theory* **62** 1528–1537. MR3472264 https://doi.org/10.1109/TIT.2016.2517008

[33] Plan, Y., Vershynin, R. and Yudovina, E. (2017). High-dimensional estimation with geometric constraints. *Inf. Inference* **6** 1–40. MR3636866 https://doi.org/10.1093/imaiai/iaw015

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

Mai, X., Liao, Z. and Couillet, R. (2019). A large scale analysis of logistic regression: Asymptotic performance and new insights. In *ICASSP* 2019-2019 *IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*) 3357–3361. IEEE.

[35] Rush, C. (2020). An asymptotic rate for the lasso loss. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, 26–28 *Aug* 2020 (S. Chiappa and R. Calandra, eds.). *Proceedings of Machine Learning Research* **108** 3664–3673. PMLR https://proceedings.mlr.press/v108/rush20a.html.

[36] Salehi, F., Abbasi, E. and Hassibi, B. (2019). The impact of regularization on high-dimensional logistic regression. In *Advances in Neural Information Processing Systems* 12005–12015.

[37] Sawaya, K., Uematsu, Y. and Imaizumi, M. (2023). Statistical inference in high-dimensional generalized linear models with asymmetric link functions.

[38] Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. MR0630098

[39] Stojnic, M. (2013). A framework to characterize performance of lasso algorithms. Preprint. Available at arXiv:1303.7291.

[40] Sur, P. (2019). *A Modern Maximum Likelihood Theory for High-Dimensional Logistic Regression*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—Stanford University. MR4197622

[41] Sur, P. and Candès, E. J. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. *Proc. Natl. Acad. Sci. USA* **116** 14516–14525. MR3984492 https://doi.org/10.1073/pnas.1810420116

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

[43] Thrampoulidis, C., Abbasi, E. and Hassibi, B. (2015). Lasso with non-linear measurements is equivalent to one with linear measurements. In *Advances in Neural Information Processing Systems* 3420–3428.

[44] 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

[45] Tibshirani, R. J. (2013). The lasso problem and uniqueness. *Electron. J. Stat.* **7** 1456–1490. MR3066375 https://doi.org/10.1214/13-EJS815

[46] Tibshirani, R. J. and Taylor, J. (2012). Degrees of freedom in lasso problems. *Ann. Statist.* **40** 1198–1232. MR2985948 https://doi.org/10.1214/12-AOS1003

[47] 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

[48] Yadlowsky, S., Yun, T., McLean, C. and D'Amour. Sloe, A. (2021). A faster method for statistical inference in high-dimensional logistic regression. Preprint. Available at arXiv:2103.12725.

[49] Yang, Z., Balasubramanian, K. and Liu, H. (2017). High-dimensional non-Gaussian single index models via thresholded score function estimation. In *International Conference on Machine Learning* 3851–3860. PMLR.

[50] 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

[51] Zhao, P. and Yu, B. (2006). On model selection consistency of Lasso. *J. Mach. Learn. Res.* **7** 2541–2563. MR2274449

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

[53] Zou, H., Hastie, T. and Tibshirani, R. (2007). On the "degrees of freedom" of the lasso. *Ann. Statist.* **35** 2173–2192. MR2363967 https://doi.org/10.1214/009053607000000127

The Annals of Statistics
2025, Vol. 53, No. 2, 561–588
https://doi.org/10.1214/24-AOS2465
© Institute of Mathematical Statistics, 2025

A STATISTICAL FRAMEWORK FOR ANALYZING SHAPE IN A TIME SERIES OF RANDOM GEOMETRIC OBJECTS

BY ANNE VAN DELFT[1,a] AND ANDREW J. BLUMBERG[2,b]

[1]*Department of Statistics, Columbia University,* [a]*anne.vandelft@columbia.edu*
[2]*Irving Institute for Cancer Dynamics, Columbia University,* [b]*andrew.blumberg@columbia.edu*

We introduce a new framework to analyze shape descriptors that capture the geometric features of an ensemble of point clouds. At the core of our approach is the point of view that the data arises as sampled recordings from a metric space-valued stochastic process, possibly of nonstationary nature, thereby integrating geometric data analysis into the realm of functional time series analysis. Our framework allows for natural incorporation of spatial-temporal dynamics, heterogeneous sampling, and the study of convergence rates. Further, we derive complete invariants for classes of metric space-valued stochastic processes in the spirit of Gromov, and relate these invariants to so-called ball volume processes. Under mild dependence conditions, a weak invariance principle in $D([0,1] \times [0,\mathscr{R}])$ is established for sequential empirical versions of the latter, assuming the probabilistic structure possibly changes over time. Finally, we use this result to introduce novel test statistics for topological change, which are distribution-free in the limit under the hypothesis of stationarity. We explore these test statistics on time series of single-cell mRNA expression data, using shape descriptors coming from topological data analysis.

REFERENCES

[1] ADAMS, H., EMERSON, T., KIRBY, M., NEVILLE, R., PETERSON, C., SHIPMAN, P., CHEPUSHTANOVA, S., HANSON, E., MOTTA, F. et al. (2017). Persistence images: A stable vector representation of persistent homology. *J. Mach. Learn. Res.* **18** Paper No. 8, 35. MR3625712

[2] ASTON, J. A. D. and KIRCH, C. (2012). Evaluating stationarity via change-point alternatives with applications to fMRI data. *Ann. Appl. Stat.* **6** 1906–1948. MR3058688 https://doi.org/10.1214/12-AOAS565

[3] AUE, A., RICE, G. and SÖNMEZ, O. (2018). Detecting and dating structural breaks in functional data without dimension reduction. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 509–529. MR3798876 https://doi.org/10.1111/rssb.12257

[4] AUE, A. and VAN DELFT, A. (2020). Testing for stationarity of functional time series in the frequency domain. *Ann. Statist.* **48** 2505–2547. MR4152111 https://doi.org/10.1214/19-AOS1895

[5] BAUER, U. Ripser. https://github.com/Ripser/ripser.

[6] BAUER, U. (2021). Ripser: Efficient computation of Vietoris-Rips persistence barcodes. *J. Appl. Comput. Topol.* **5** 391–423. MR4298669 https://doi.org/10.1007/s41468-021-00071-5

[7] BERKES, I., GABRYS, R., HORVÁTH, L. and KOKOSZKA, P. (2009). Detecting changes in the mean of functional observations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 927–946. MR2750251 https://doi.org/10.1111/j.1467-9868.2009.00713.x

[8] BEUTNER, E. and ZÄHLE, H. (2012). Deriving the asymptotic distribution of U- and V-statistics of dependent data using weighted empirical processes. *Bernoulli* **18** 803–822. MR2948902 https://doi.org/10.3150/11-BEJ358

[9] BLOCK, A., JIA, Z., POLYANSKIY, Y. and RAKHLIN, A. (2022). Intrinsic dimension estimation using Wasserstein distance. *J. Mach. Learn. Res.* **23** Paper No. [313], 37. MR4577752

[10] BLUMBERG, A. J., BHAUMIK, P. and WALKER, S. J. (2018). Testing to distinguish measures on metric spaces. Available at arXiv:1802.01152.

MSC2020 subject classifications. Primary 62M99, 62R20, 62R40; secondary 60B05, 60F17, 62M10.
Key words and phrases. Topological data analysis, functional data analysis, persistent homology, locally stationary processes, U-statistics.

[11] BLUMBERG, A. J., GAL, I., MANDELL, M. A. and PANCIA, M. (2014). Robust statistics, hypothesis testing, and confidence intervals for persistent homology on metric measure spaces. *Found. Comput. Math.* **14** 745–789. MR3230014 https://doi.org/10.1007/s10208-014-9201-4

[12] BOROVKOVA, S., BURTON, R. and DEHLING, H. (2001). Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation. *Trans. Amer. Math. Soc.* **353** 4261–4318. MR1851171 https://doi.org/10.1090/S0002-9947-01-02819-7

[13] BOSQ, D. (2000). *Linear Processes in Function Spaces: Theory and Applications. Lecture Notes in Statistics* **149**. Springer, New York. MR1783138 https://doi.org/10.1007/978-1-4612-1154-9

[14] BOUZEBDA, S. and NEMOUCHI, B. (2023). Weak-convergence of empirical conditional processes and conditional U-processes involving functional mixing data. *Stat. Inference Stoch. Process.* **26** 33–88. MR4562252 https://doi.org/10.1007/s11203-022-09276-6

[15] BRADLEY, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** 107–144. MR2178042 https://doi.org/10.1214/154957805100000104

[16] BUBENIK, P. (2015). Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.* **16** 77–102. MR3317230

[17] BÜCHER, A., DETTE, H. and HEINRICHS, F. (2023). A portmanteau-type test for detecting serial correlation in locally stationary functional time series. *Stat. Inference Stoch. Process.* **26** 255–278. MR4594585 https://doi.org/10.1007/s11203-022-09285-5

[18] CARLSSON, G. (2009). Topology and data. *Bull. Amer. Math. Soc.* (*N.S.*) **46** 255–308. MR2476414 https://doi.org/10.1090/S0273-0979-09-01249-X

[19] CARLSSON, G. (2014). Topological pattern recognition for point cloud data. *Acta Numer.* **23** 289–368. MR3202240 https://doi.org/10.1017/S0962492914000051

[20] CARLSSON, G. and MÉMOLI, F. (2010). Characterization, stability and convergence of hierarchical clustering methods. *J. Mach. Learn. Res.* **11** 1425–1470. MR2645457

[21] CASINI, A. and PERRON, P. (2024). Change-point analysis of time series with evolutionary spectra. *J. Econometrics* **242** Paper No. 105811, 18. MR4769085 https://doi.org/10.1016/j.jeconom.2024.105811

[22] CHAZAL, F., GLISSE, M., LABRUÈRE, C. and MICHEL, B. (2015). Convergence rates for persistence diagram estimation in topological data analysis. *J. Mach. Learn. Res.* **16** 3603–3635. MR3450548

[23] COHEN-STEINER, D., EDELSBRUNNER, H. and HARER, J. (2007). Stability of persistence diagrams. *Discrete Comput. Geom.* **37** 103–120. MR2279866 https://doi.org/10.1007/s00454-006-1276-5

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

[25] DAS, S. and POLITIS, D. N. (2021). Predictive inference for locally stationary time series with an application to climate data. *J. Amer. Statist. Assoc.* **116** 919–934. MR4270034 https://doi.org/10.1080/01621459.2019.1708368

[26] DEHLING, H., FRIED, R., SHARIPOV, O. SH., VOGEL, D. and WORNOWIZKI, M. (2013). Estimation of the variance of partial sums of dependent processes. *Statist. Probab. Lett.* **83** 141–147. MR2998735 https://doi.org/10.1016/j.spl.2012.08.012

[27] DEHLING, H., GIRAUDO, D. and SHARIPOV, O. (2021). Convergence of the empirical two-sample U-statistics with β-mixing data. *Acta Math. Hungar.* **164** 377–412. MR4279342 https://doi.org/10.1007/s10474-021-01156-4

[28] DENG, A. and PERRON, P. (2008). A non-local perspective on the power properties of the CUSUM and CUSUM of squares tests for structural change. *J. Econometrics* **142** 212–240. MR2408735 https://doi.org/10.1016/j.jeconom.2007.04.002

[29] 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

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

[31] EDELSBRUNNER, H. and HARER, J. (2008). Persistent homology—a survey. In *Surveys on Discrete and Computational Geometry. Contemp. Math.* **453** 257–282. Amer. Math. Soc., Providence, RI. MR2405684 https://doi.org/10.1090/conm/453/08802

[32] EDELSBRUNNER, H. and HARER, J. L. (2010). *Computational Topology: An Introduction*. Amer. Math. Soc., Providence, RI. MR2572029 https://doi.org/10.1090/mbk/069

[33] FASY, B. T., LECCI, F., RINALDO, A., WASSERMAN, L., BALAKRISHNAN, S. and SINGH, A. (2014). Confidence sets for persistence diagrams. *Ann. Statist.* **42** 2301–2339. MR3269981 https://doi.org/10.1214/14-AOS1252

[34] GRASSBERGER, P. and PROCACCIA, I. (1983). Characterization of strange attractors. *Phys. Rev. Lett.* **50** 346–349. MR0689681 https://doi.org/10.1103/PhysRevLett.50.346

[35] Gromov, M. (1999). *Metric Structures for Riemannian and Non-Riemannian Spaces*. *Progress in Mathematics* **152**. Birkhäuser, Boston, MA. MR1699320

[36] Hacquard, O., Balasubramanian, K., Blanchard, G., Levrard, C. and Polonik, W. (2022). Topologically penalized regression on manifolds. *J. Mach. Learn. Res.* **23** Paper No. [161], 39. MR4577113

[37] Hong, Y., Linton, O., McCabe, B., Sun, J. and Wang, S. (2024). Kolmogorov-Smirnov type testing for structural breaks: A new adjusted-range based self-normalization approach. *J. Econometrics* **238** Paper No. 105603, 19. MR4665485 https://doi.org/10.1016/j.jeconom.2023.105603

[38] Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *Ann. Statist.* **38** 1845–1884. MR2662361 https://doi.org/10.1214/09-AOS768

[39] Hsing, T. and Wu, W. B. (2004). On weighted U-statistics for stationary processes. *Ann. Probab.* **32** 1600–1631. MR2060311 https://doi.org/10.1214/009117904000000333

[40] Hyndman, R. J. and Shang, H. L. (2009). Forecasting functional time series. *J. Korean Statist. Soc.* **38** 199–211. MR2750314 https://doi.org/10.1016/j.jkss.2009.06.002

[41] 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

[42] Kahle, M. (2011). Random geometric complexes. *Discrete Comput. Geom.* **45** 553–573. MR2770552 https://doi.org/10.1007/s00454-010-9319-3

[43] Kallenberg, O. (2002). *Foundations of Modern Probability*, 2nd ed. *Probability and Its Applications* (*New York*). Springer, New York. MR1876169 https://doi.org/10.1007/978-1-4757-4015-8

[44] Kerber, M., Morozov, D. and Nigmetov, A. (2017). Geometry helps to compare persistence diagrams. *ACM J. Exp. Algorithmics* **22** Art. 1.4, 20. MR3707737 https://doi.org/10.1145/3064175

[45] 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

[46] Lange, M., Bergen, V., Klein, M., Setty, M., Reuter, B., Bakhti, M., Lickert, H., Ansari, M., Schniering, J. et al. (2022). CellRank for directed single-cell fate mapping. *Nat. Methods*.

[47] Meilă, M. and Zhang, H. (2024). Manifold learning: What, how, and why. *Annu. Rev. Stat. Appl.* **11** 393–417. MR4742290

[48] Mileyko, Y., Mukherjee, S. and Harer, J. (2011). Probability measures on the space of persistence diagrams. *Inverse Probl.* **27** 124007, 22. MR2854323 https://doi.org/10.1088/0266-5611/27/12/124007

[49] Myers, A., Munch, E. and Khasawneh, F. A. (2019). Persistent homology of complex networks for dynamic state detection. *Phys. Rev. E* **100** 022314, 14. MR4001917 https://doi.org/10.1103/physreve.100.022314

[50] Niyogi, P., Smale, S. and Weinberger, S. (2008). Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.* **39** 419–441. MR2383768 https://doi.org/10.1007/s00454-008-9053-2

[51] Oudot, S. Y. (2015). *Persistence Theory*: *From Quiver Representations to Data Analysis*. *Mathematical Surveys and Monographs* **209**. Amer. Math. Soc., Providence, RI. MR3408277 https://doi.org/10.1090/surv/209

[52] Owada, T. and Adler, R. J. (2017). Limit theorems for point processes under geometric constraints (and topological crackle). *Ann. Probab.* **45** 2004–2055. MR3650420 https://doi.org/10.1214/16-AOP1106

[53] Öztürk, Ö. and Hettmansperger, T. P. (1997). Generalised weighted Cramér–von Mises distance estimators. *Biometrika* **84** 283–294. MR1467047 https://doi.org/10.1093/biomet/84.2.283

[54] Panaretos, V. M. and Tavakoli, S. (2013). Fourier analysis of stationary time series in function space. *Ann. Statist.* **41** 568–603. MR3099114 https://doi.org/10.1214/13-AOS1086

[55] Perea, J. A. (2019). Topological times series analysis. *Notices Amer. Math. Soc.* **66** 686–694. MR3929469

[56] Perea, J. A. and Harer, J. (2015). Sliding windows and persistence: An application of topological methods to signal analysis. *Found. Comput. Math.* **15** 799–838. MR3348174 https://doi.org/10.1007/s10208-014-9206-z

[57] Peyré, G. and Cuturi, M. (2019). Computational optimal transport. *Found. Trends Mach. Learn.* **11** 355–607.

[58] Rabadan, R. and Blumberg, A. J. (2020). *Topological Data Analysis for Genomics and Evolution*. Cambridge Univ. Press, Cambridge.

[59] Rigot, S. (2022). Differentiation of measures in metric spaces. In *New Trends on Analysis and Geometry in Metric Spaces*. *Lecture Notes in Math.* **2296** 93–116. Springer, Cham. MR4432546 https://doi.org/10.1007/978-3-030-84141-6_3

[60] Saul, L. K. and Roweis, S. T. An introduction to local linear embedding. https://cs.nyu.edu/~roweis/lle/papers/lleintroa4.pdf.

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

[62] SHAO, X. and ZHANG, X. (2010). Testing for change points in time series. *J. Amer. Statist. Assoc.* **105** 1228–1240. MR2752617 https://doi.org/10.1198/jasa.2010.tm10103

[63] STURM, K.-T. (2002). Nonlinear martingale theory for processes with values in metric spaces of nonpositive curvature. *Ann. Probab.* **30** 1195–1222. MR1920105 https://doi.org/10.1214/aop/1029867125

[64] SUBBA RAO, S. (2006). On some nonstationary, nonlinear random processes and their stationary approximations. *Adv. in Appl. Probab.* **38** 1155–1172. MR2285698 https://doi.org/10.1239/aap/1165414596

[65] TENENBAUM, J. B., DE SILVA, V. and LANGFORD, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science* **290** 2319–2323.

[66] 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

[67] VAN DELFT, A. and BLUMBERG, A. J. Measures determined by their values on balls and Gromov–Wasserstein convergence. Available at arXiv:2401.11125.

[68] VAN DELFT, A. and BLUMBERG, A. J (2025). Supplement to "A statistical framework for analyzing shape in a time series of random geometric objects." https://doi.org/10.1214/24-AOS2465SUPP

[69] 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

[70] VAN DELFT, A. and DETTE, H. (2024). A general framework to quantify deviations from structural assumptions in the analysis of nonstationary function-valued processes. *Ann. Statist.* **52** 550–579. MR4744187 https://doi.org/10.1214/24-aos2358

[71] 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

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

[73] VOGEL, D. and WENDLER, M. (2017). Studentized U-quantile processes under dependence with applications to change-point analysis. *Bernoulli* **23** 3114–3144. MR3654800 https://doi.org/10.3150/16-BEJ838

[74] VOGELSANG, T. J. (1999). Sources of nonmonotonic power when testing for a shift in mean of a dynamic time series. *J. Econometrics* **88** 283–299.

[75] 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

[76] XIAN, L., ADAMS, H., TOPAZ, C. M. and ZIEGELMEIER, L. (2022). Capturing dynamics of time-varying data via topology. *Found. Data Sci.* **4** 1–36. MR4622886 https://doi.org/10.3934/fods.2021033

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

The Annals of Statistics
2025, Vol. 53, No. 2, 589–614
https://doi.org/10.1214/24-AOS2470
© Institute of Mathematical Statistics, 2025

APPROXIMATION ERROR FROM DISCRETIZATIONS AND ITS APPLICATIONS

BY JUNLONG ZHAO[1,a], XIUMIN LIU[2,c], BIN DU[1,b] AND YUFENG LIU[3,d]

[1]*School of Statistics, Beijing Normal University,* [a]*zhaojunlong928@126.com,* [b]*201931011012@mail.bnu.edu.cn*
[2]*School of Mathematics and Statistics, Beijing Technology and Business University,* [c]*liuxiumin880407@yeah.net*
[3]*Department of Statistics and Operations Research, Department of Genetics, Department of Biostatistics, University of North Carolina at Chapel Hill,* [d]*yfliu@email.unc.edu*

Converting a continuous variable into a discrete one is a commonly used technique for solving various problems in both statistics and machine learning. It is well known that discretizations result in biases. However, this issue has not been studied systematically. In this paper, a general framework is proposed to understand and compare the approximation errors of different slicing strategies. Poincaré-type inequalities are first established for univariate discretizations and then generalized to the multivariate and other settings. It is shown that the bias is controlled by two factors: the distance between two specific distributions that are generated with and without discretizations respectively, and the smoothness of the functions involved. Several important applications are considered to illustrate the usefulness of the results. Our results help to understand the approximation error of some matrix used in the literature of dimension reduction. Furthermore, as an illustration of the usefulness of discretizations, we propose an algorithm for regression problems, by combining random forest with partial discretizations of responses. Simulation results confirm the advantages of this algorithm over the classical random forest.

REFERENCES

[1] ARLOT, S. and GENUER, R. (2014). Analysis of purely random forests bias. arXiv preprint. Available at arXiv:1407.3939.
[2] ATHEY, S., TIBSHIRANI, J. and WAGER, S. (2019). Generalized random forests. *Ann. Statist.* **47** 1148–1178. MR3909963 https://doi.org/10.1214/18-AOS1709
[3] AZAÏS, R., GÉGOUT-PETIT, A. and SARACCO, J. (2012). Optimal quantization applied to sliced inverse regression. *J. Statist. Plann. Inference* **142** 481–492. MR2843051 https://doi.org/10.1016/j.jspi.2011.08.006
[4] BAY, S. D. (2001). Multivariate discretization for set mining. *Knowl. Inf. Syst.* **3** 491–512.
[5] BREIMAN, L. (1996). Bagging predictors. *Mach. Learn.* **24** 123–140.
[6] BREIMAN, L. (2001). Random forests. *Mach. Learn.* **45** 5–32.
[7] BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. and STONE, C. J. (1984). *Classification and Regression Trees*. *Wadsworth Statistics/Probability Series*. Wadsworth Advanced Books and Software, Belmont, CA. MR0726392
[8] BROCKMANN, M., GASSER, T. and HERRMANN, E. (1993). Locally adaptive bandwidth choice for kernel regression estimators. *J. Amer. Statist. Assoc.* **88** 1302–1309. MR1245363
[9] CHING, J. Y., WONG, A. K. C. and CHAN, K. C. C. (1995). Class-dependent discretization for inductive learning from continuous and mixed-mode data. *IEEE Trans. Pattern Anal. Mach. Intell.* **17** 641–651.
[10] COOK, R. D. and ZHANG, X. (2014). Fused estimators of the central subspace in sufficient dimension reduction. *J. Amer. Statist. Assoc.* **109** 815–827. MR3223752 https://doi.org/10.1080/01621459.2013.866563
[11] DIMITROVA, E. S., VERA LICONA, M. P., MCGEE, J. and LAUBENBACHER, R. (2010). Discretization of time series data. *J. Comput. Biol.* **17** 853–868. MR2659477 https://doi.org/10.1089/cmb.2008.0023

MSC2020 subject classifications. 62G08, 62-08.
Key words and phrases. Bias, discretization, nonparametric regression, sliced inverse regression.

[12] DREZNER, Z. and ZEROM, D. (2016). A simple and effective discretization of a continuous random variable. *Comm. Statist. Simulation Comput.* **45** 3798–3810. MR3533618 https://doi.org/10.1080/03610918.2015.1071389

[13] 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

[14] FAN, J., GIJBELS, I., HU, T.-C. and HUANG, L.-S. (1996). A study of variable bandwidth selection for local polynomial regression. *Statist. Sinica* **6** 113–127. MR1379052

[15] FANG, K.-T. and PAN, J. (2023). A review of representative points of statistical distributions and their applications. *Mathematics* **11** 2930.

[16] FAYYAD, U. M. and IRANI, K. B. (1993). Multi-interval discretization of continuous-valued attributes for classification learning. In *XIII International Joint Conference on Artificial Intelligence* (*IJCAI*93) 1022–1029.

[17] GARCIA, S., LUENGO, J., SÁEZ, J. A., LOPEZ, V. and HERRERA, F. (2012). A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning. *IEEE Trans. Knowl. Data Eng.* **25** 734–750.

[18] GLUHOVSKY, I. and GLUHOVSKY, A. (2007). Smooth location-dependent bandwidth selection for local polynomial regression. *J. Amer. Statist. Assoc.* **102** 718–725. MR2370862 https://doi.org/10.1198/016214507000000086

[19] GUPTA, A., DATTA, S. and DAS, S. (2018). Fast automatic estimation of the number of clusters from the minimum inter-center distance for k-means clustering. *Pattern Recogn. Lett.* **116** 72–79.

[20] HO, K. M. and SCOTT, P. D. (1997). Zeta: A global method for discretization of continuous variables. In *Conference on Knowledge Discovery and Data Mining* 191–194.

[21] HSING, T. and CARROLL, R. J. (1992). An asymptotic theory for sliced inverse regression. *Ann. Statist.* **20** 1040–1061. MR1165605 https://doi.org/10.1214/aos/1176348669

[22] HSU, C. N., HUANG, H. J. and WONG, T. T. (2000). Why discretization works for naive Bayesian classifiers. In 17*th International Conference* (*ICML'*2000) 309–406.

[23] JIANG, B., YE, C. and LIU, J. S. (2015). Nonparametric K-sample tests via dynamic slicing. *J. Amer. Statist. Assoc.* **110** 642–653. MR3367254 https://doi.org/10.1080/01621459.2014.920257

[24] KOTSIANTIS, S. and KANELLOPOULOS, D. (2006). Discretization techniques: A recent survey. *GESTS Int. Trans. Comput. Sci. Eng.* **32** 47–58.

[25] LI, K.-C. (1991). Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.* **86** 316–342. MR1137117

[26] LI, R. P. and WANG, Z. O. (2002). An entropy-based discretization method for classification rules with inconsistency checking. In *Proceedings of the First International Conference on Machine Learning and Cybernetics* (*ICMLC*) 243–246.

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

[28] LIN, Y. and JEON, Y. (2006). Random forests and adaptive nearest neighbors. *J. Amer. Statist. Assoc.* **101** 578–590. MR2256176 https://doi.org/10.1198/016214505000001230

[29] LIQUET, B. and SARACCO, J. (2008). Application of the bootstrap approach to the choice of dimension and the α parameter in the SIR_α method. *Comm. Statist. Simulation Comput.* **37** 1198–1218. MR2528270 https://doi.org/10.1080/03610910801889011

[30] LIQUET, B. and SARACCO, J. (2012). A graphical tool for selecting the number of slices and the dimension of the model in SIR and SAVE approaches. *Comput. Statist.* **27** 103–125. MR2877813 https://doi.org/10.1007/s00180-011-0241-9

[31] LIU, H., HUSSAIN, F., TAN, C. L. and DASH, M. (2002). Discretization: An enabling technique. *Data Min. Knowl. Discov.* **6** 393–423. MR1917949 https://doi.org/10.1023/A:1016304305535

[32] LIU, H. and SETIONO, R. (1997). Feature selection and discretization. *IEEE Trans. Knowl. Data Eng.* **9** 642–645.

[33] MEHTA, S., PARTHASARATHY, S. and YANG, H. (2005). Toward unsupervised correlation preserving discretization. *IEEE Trans. Knowl. Data Eng.* **17** 1174–1185.

[34] POINCARE, H. (1890). Sur les Equations aux Derivees Partielles de la Physique Mathematique. *Amer. J. Math.* **12** 211–294. MR1505534 https://doi.org/10.2307/2369620

[35] QIAN, W., DING, S. and COOK, R. D. (2019). Sparse minimum discrepancy approach to sufficient dimension reduction with simultaneous variable selection in ultrahigh dimension. *J. Amer. Statist. Assoc.* **114** 1277–1290. MR4011779 https://doi.org/10.1080/01621459.2018.1497498

[36] RUPPERT, D. (1997). Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation. *J. Amer. Statist. Assoc.* **92** 1049–1062. MR1482136 https://doi.org/10.2307/2965570

[37] SANG, Y., QI, H., LI, K., JIN, Y., YAN, D. and GAO, S. (2014). An effective discretization method for disposing high-dimensional data. *Inform. Sci.* **270** 73–91. MR3190437 https://doi.org/10.1016/j.ins.2014.02.113

[38] SCORNET, E. (2016). Random forests and kernel methods. *IEEE Trans. Inf. Theory* **62** 1485–1500. MR3472261 https://doi.org/10.1109/TIT.2016.2514489

[39] SHEATHER, S. J. and JONES, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *J. Roy. Statist. Soc. Ser. B* **53** 683–690. MR1125725

[40] SHEENA, Y. (2019). Estimation of a continuous distribution on the real line by discretization methods. *Metrika* **82** 339–360. MR3946401 https://doi.org/10.1007/s00184-018-0683-y

[41] STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360. MR0594650

[42] TAN, K. M., WANG, Z., ZHANG, T., LIU, H. and COOK, R. D. (2018). A convex formulation for high-dimensional sparse sliced inverse regression. *Biometrika* **105** 769–782. MR3876161 https://doi.org/10.1093/biomet/asy049

[43] TORGO, L. and GAMA, J. (1996). Regression by classification. In *SBIA'96 Proceedings of the 13th Brazilian Symposium on Artificial Intelligence: Advances in Artificial Intelligence* 51–60.

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

[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. and WEBB, G. I. (2002). Non-disjoint discretization for naive-Bayes classifiers. In *ICML* **2** 666–673.

[47] YANG, Y. and WEBB, G. I. (2009). Discretization for naive-Bayes learning: Managing discretization bias and variance. *Mach. Learn.* **74** 39–74.

[48] ZADOR, P. L. (1964). *Development and Evaluation of Procedures for Quantizing Multivariate Distributions*. Stanford University.

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

[50] ZHANG, Y., DUCHI, J. and WAINWRIGHT, M. (2015). Divide and conquer kernel ridge regression: A distributed algorithm with minimax optimal rates. *J. Mach. Learn. Res.* **16** 3299–3340. MR3450540

[51] ZHAO, J., LIU, X., DU, B. and LIU, Y. (2025). Supplement to "Approximation error from discretizations and its applications." https://doi.org/10.1214/24-AOS2470SUPP

[52] ZHU, L., MIAO, B. and PENG, H. (2006). On sliced inverse regression with high-dimensional covariates. *J. Amer. Statist. Assoc.* **101** 630–643. MR2281245 https://doi.org/10.1198/016214505000001285

[53] ZHU, L. X. and NG, K. W. (1995). Asymptotics of sliced inverse regression. *Statist. Sinica* **5** 727–736. MR1347616

The Annals of Statistics
2025, Vol. 53, No. 2, 615–646
https://doi.org/10.1214/24-AOS2471
© Institute of Mathematical Statistics, 2025

EMBEDDING DISTRIBUTIONAL DATA

BY ERY ARIAS-CASTRO[1,a] AND WANLI QIAO[2,b]

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

[2]*Department of Statistics, George Mason University,* [b]*wqiao@gmu.edu*

We adapt concepts, methodology, and theory originally developed in the areas of multidimensional scaling and dimensionality reduction for Euclidean data to be applicable to distributional data. We focus on classical scaling and Isomap—prototypical methods that have played important roles in these areas—and showcase their use in the context of distributional data analysis. In the process, we highlight the crucial role that the ambient metric plays.

REFERENCES

[1] ABANDA, A., MORI, U. and LOZANO, J. A. (2019). A review on distance based time series classification. *Data Min. Knowl. Discov.* **33** 378–412. MR3914608 https://doi.org/10.1007/s10618-018-0596-4

[2] AGULLÓ-ANTOLÍN, M., CUESTA-ALBERTOS, J. A., LESCORNEL, H. and LOUBES, J.-M. (2015). A parametric registration model for warped distributions with Wasserstein's distance. *J. Multivariate Anal.* **135** 117–130. MR3306430 https://doi.org/10.1016/j.jmva.2014.12.005

[3] AMARI, S. (2016). *Information Geometry and Its Applications*. *Applied Mathematical Sciences* **194**. Springer, Tokyo. MR3495836 https://doi.org/10.1007/978-4-431-55978-8

[4] AMBROSIO, L. and GIGLI, N. (2013). A user's guide to optimal transport. In *Modelling and Optimisation of Flows on Networks*. *Lecture Notes in Math*. **2062** 1–155. Springer, Heidelberg. MR3050280 https://doi.org/10.1007/978-3-642-32160-3_1

[5] AMBROSIO, L., GIGLI, N. and SAVARÉ, G. (2005). *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. *Lectures in Mathematics ETH Zürich*. Birkhäuser, Basel. MR2129498

[6] ARIAS-CASTRO, E. and CHAU, P. A. (2023). Minimax estimation of distances on a surface and minimax manifold learning in the isometric-to-convex setting. *Inf. Inference* **12** Paper No. iaad046, 40. MR4669792 https://doi.org/10.1093/imaiai/iaad046

[7] ARIAS-CASTRO, E., JAVANMARD, A. and PELLETIER, B. (2020). Perturbation bounds for procrustes, classical scaling, and trilateration, with applications to manifold learning. *J. Mach. Learn. Res.* **21** Paper No. 15, 37. MR4071198

[8] ARIAS-CASTRO, E. and LE GOUIC, T. (2019). Unconstrained and curvature-constrained shortest-path distances and their approximation. *Discrete Comput. Geom.* **62** 1–28. MR3959920 https://doi.org/10.1007/s00454-019-00060-7

[9] ARIAS-CASTRO, E. and QIAO, W. (2025). Supplement to "Embedding distributional data." https://doi.org/10.1214/24-AOS2471SUPP

[10] AY, N., JOST, J., LÊ, H. V. and SCHWACHHÖFER, L. (2017). *Information Geometry*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete*. 3. *Folge*. *A Series of Modern Surveys in Mathematics* [*Results in Mathematics and Related Areas*. 3*rd Series*. *A Series of Modern Surveys in Mathematics*] **64**. Springer, Cham. MR3701408 https://doi.org/10.1007/978-3-319-56478-4

[11] BAKER, L. D. and MCCALLUM, A. K. (1998). Distributional clustering of words for text classification. In *ACM Conference on Research and Development in Information Retrieval* 96–103.

[12] BAKSHAEV, A. (2009). Goodness of fit and homogeneity tests on the basis of N-distances. *J. Statist. Plann. Inference* **139** 3750–3758. MR2553759 https://doi.org/10.1016/j.jspi.2009.05.014

[13] BERLINET, A. and THOMAS-AGNAN, C. (2004). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic, Boston, MA. With a preface by Persi Diaconis. MR2239907 https://doi.org/10.1007/978-1-4419-9096-9

MSC2020 subject classifications. Primary 62R10, 62R20, 62R30; secondary 62R10, 62H25, 62B11.

Key words and phrases. Functional data analysis (FDA), distributional data analysis (DDA), embedding problem, multidimensional scaling, dimensionality reduction, principal component analysis, classical scaling, Isomap, RKHS metric, Fisher metric, Wasserstein metric, optimal transport, information geometry.

[14] BERNSTEIN, M., DE SILVA, V., LANGFORD, J. C. and TENENBAUM, J. B. (2000). Graph approximations to geodesics on embedded manifolds. Technical Report, Dept. Psychology, Stanford Univ.

[15] BESSE, P. and RAMSAY, J. O. (1986). Principal components analysis of sampled functions. *Psychometrika* **51** 285–311. MR0848110 https://doi.org/10.1007/BF02293986

[16] BIGOT, J. (2020). Statistical data analysis in the Wasserstein space. In *Journées MAS* 2018—*Sampling and Processes*. *ESAIM Proc*. *Surveys* **68** 1–19. EDP Sci., Les Ulis. MR4142495 https://doi.org/10.1051/proc/202068001

[17] BIGOT, J., GAMBOA, F. and VIMOND, M. (2009). Estimation of translation, rotation, and scaling between noisy images using the Fourier–Mellin transform. *SIAM J. Imaging Sci*. **2** 614–645. MR2519925 https://doi.org/10.1137/070691231

[18] 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

[19] BRENIER, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Comm*. *Pure Appl*. *Math*. **44** 375–417. MR1100809 https://doi.org/10.1002/cpa.3160440402

[20] BRITO, P. and DIAS, S. (2022). *Analysis of Distributional Data*. CRC Press, Boca Raton.

[21] BU, Y., ZOU, S., LIANG, Y. and VEERAVALLI, V. V. (2018). Estimation of KL divergence: Optimal minimax rate. *IEEE Trans*. *Inf. Theory* **64** 2648–2674. MR3782280 https://doi.org/10.1109/TIT.2018.2805844

[22] 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

[23] CAIADO, J. (2010). *Classification and Clustering of Time Series*. Lambert Academic Publishing, Saarbrücken, Germany.

[24] CARTER, K. M., RAICH, R., FINN, W. G. and HERO, A. O. III (2009). FINE: Fisher information nonparametric embedding. *IEEE Trans*. *Pattern Anal*. *Mach*. *Intell*. **31** 2093–2098.

[25] CASTRO, P. E., LAWTON, W. H. and SYLVESTRE, E. (1986). Principal modes of variation for processes with continuous sample curves. *Technometrics* **28** 329–337.

[26] CHEN, D. and MÜLLER, H.-G. (2012). Nonlinear manifold representations for functional data. *Ann*. *Statist*. **40** 1–29. MR3013177 https://doi.org/10.1214/11-AOS936

[27] CHEN, Y. and LI, W. (2020). Optimal transport natural gradient for statistical manifolds with continuous sample space. *Inf. Geom*. **3** 1–32. MR4119003 https://doi.org/10.1007/s41884-020-00028-0

[28] 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

[29] CHIZAT, L., ROUSSILLON, P., LÉGER, F., VIALARD, F.-X. and PEYRÉ, G. (2020). Faster Wasserstein distance estimation with the Sinkhorn divergence. *Adv. Neural Inf. Process*. *Syst*. **33** 2257–2269.

[30] CLONINGER, A., HAMM, K., KHURANA, V. and MOOSMÜLLER, C. (2025). Linearized Wasserstein dimensionality reduction with approximation guarantees. *Appl*. *Comput*. *Harmon*. *Anal*. **74** Paper No. 101718, 31. MR4811552 https://doi.org/10.1016/j.acha.2024.101718

[31] CUTURI, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. *Adv. Neural Inf. Process*. *Syst*. **26**.

[32] DA FONSECA, M. and SAMENGO, I. (2016). Derivation of human chromatic discrimination ability from and information-theoretical notion of distance in color space. *Neural Comput*. **28** 2628–2655. MR3866416 https://doi.org/10.1162/neco_a_00903

[33] DALY, S. J. (1992). Visible differences predictor: An algorithm for the assessment of image fidelity. In *In Human Vision*, *Visual Processing*, *and Digital Display III* **1666** 2–15. SPIE, Bellingham.

[34] DE LEEUW, J. and MAIR, P. (2009). Multidimensional scaling using majorization: SMACOF in R. *J. Stat*. *Softw*. **31**.

[35] DEL BARRIO, E., GORDALIZA, P., LESCORNEL, H. and LOUBES, J.-M. (2019). Central limit theorem and bootstrap procedure for Wasserstein's variations with an application to structural relationships between distributions. *J. Multivariate Anal*. **169** 341–362. MR3875604 https://doi.org/10.1016/j.jmva.2018.09.014

[36] DEL MORAL, P. and NICLAS, A. (2018). A Taylor expansion of the square root matrix function. *J. Math*. *Anal*. *Appl*. **465** 259–266. MR3806701 https://doi.org/10.1016/j.jmaa.2018.05.005

[37] DONOHO, D. L. and GRIMES, C. (2005). Image manifolds which are isometric to Euclidean space. *J. Math*. *Imaging Vision* **23** 5–24. MR2208906 https://doi.org/10.1007/s10851-005-4965-4

[38] 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

[39] EILERS, P. H. C. (2004). Parametric time warping. *Anal*. *Chem*. **76** 404–411. https://doi.org/10.1021/ac034800e

[40] FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. *Springer Series in Statistics*. Springer, New York. MR2229687

[41] GAMBOA, F., LOUBES, J.-M. and MAZA, E. (2007). Semi-parametric estimation of shifts. *Electron. J. Stat.* **1** 616–640. MR2369028 https://doi.org/10.1214/07-EJS026

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

[43] GOSHTASBY, A. A. (2012). *Image Registration: Principles, Tools and Methods*. *Advances in Computer Vision and Pattern Recognition*. Springer, London. MR3014042 https://doi.org/10.1007/978-1-4471-2458-0

[44] GOWER, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53** 325–338. MR0214224 https://doi.org/10.1093/biomet/53.3-4.325

[45] GRETTON, A., BORGWARDT, K. M., RASCH, M., SCHÖLKOPF, B. and SMOLA, A. J. (2007). A kernel method for the two-sample problem. In *Advances in Neural Information Processing Systems* 513–520.

[46] HAJNAL, J. V. and HILL, D. L. (2001). *Medical Image Registration*. CRC Press, Boca Raton.

[47] HALL, P., MÜLLER, H.-G. and WANG, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34** 1493–1517. MR2278365 https://doi.org/10.1214/009053606000000272

[48] HAMM, K., HENSCHEID, N. and KANG, S. (2023). Wassmap: Wasserstein isometric mapping for image manifold learning. *SIAM J. Math. Data Sci.* **5** 475–501. MR4597927 https://doi.org/10.1137/22M1490053

[49] HAMM, K., MOOSMÜLLER, C., SCHMITZER, B. and THORPE, M. (2023). Manifold learning in Wasserstein space. arXiv preprint. Available at arXiv:2311.08549.

[50] HÄRDLE, W. and MARRON, J. S. (1990). Semiparametric comparison of regression curves. *Ann. Statist.* **18** 63–89. MR1041386 https://doi.org/10.1214/aos/1176347493

[51] HOTELLING, H. (1933). Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **24** 417–441.

[52] HOTELLING, H. (1933). Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **24** 498–520.

[53] JAMES, G. M., HASTIE, T. J. and SUGAR, C. A. (2000). Principal component models for sparse functional data. *Biometrika* **87** 587–602. MR1789811 https://doi.org/10.1093/biomet/87.3.587

[54] KATTA, S., PARIKH, H., RUDIN, C. and VOLFOVSKY, A. (2024). Interpretable causal inference for analyzing wearable, sensor, and distributional data. *Artif. Intell. Stat.* **238** 3340–3348.

[55] KLEBANOV, L. B., BENEŠ, V. and SAXL, I. (2005). *N-Distances and Their Applications*. Charles Univ. in Prague, the Karolinum Press, Prague, Czech Republic.

[56] KNEIP, A. and ENGEL, J. (1995). Model estimation in nonlinear regression under shape invariance. *Ann. Statist.* **23** 551–570. MR1332581 https://doi.org/10.1214/aos/1176324535

[57] KNEIP, A. and GASSER, T. (1988). Convergence and consistency results for self-modeling nonlinear regression. *Ann. Statist.* **16** 82–112. MR0924858 https://doi.org/10.1214/aos/1176350692

[58] KNEIP, A. and GASSER, T. (1992). Statistical tools to analyze data representing a sample of curves. *Ann. Statist.* **20** 1266–1305. MR1186250 https://doi.org/10.1214/aos/1176348769

[59] KROSHNIN, A., STEPANOV, E. and TREVISAN, D. (2022). Infinite multidimensional scaling for metric measure spaces. *ESAIM Control Optim. Calc. Var.* **28** Paper No. 58, 27. MR4467100 https://doi.org/10.1051/cocv/2022053

[60] KRUSKAL, J. B. and SEERY, J. B. (1980). Designing network diagrams. In *Conference on Social Graphics* 22–50.

[61] LAWTON, W., SYLVESTRE, E. and MAGGIO, M. (1972). Self modeling nonlinear regression. *Technometrics* **14** 513–532.

[62] LEE, J. M. (1997). *Riemannian Manifolds: An Introduction to Curvature*. *Graduate Texts in Mathematics* **176**. Springer, New York. MR1468735 https://doi.org/10.1007/b98852

[63] LEE, Y. K. and PARK, B. U. (2006). Estimation of Kullback–Leibler divergence by local likelihood. *Ann. Inst. Statist. Math.* **58** 327–340. MR2246160 https://doi.org/10.1007/s10463-005-0014-8

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

[65] LI, W. and ZHAO, J. (2023). Wasserstein information matrix. *Inf. Geom.* **6** 203–255. MR4611775 https://doi.org/10.1007/s41884-023-00099-9

[66] LIAO, T. W. (2005). Clustering of time series data: A survey. *Pattern Recognit.* **38** 1857–1874.

[67] LIM, S. and MÉMOLI, F. (2024). Classical multidimensional scaling on metric measure spaces. *Inf. Inference* **13** Paper No. iaae007, 76. MR4743459 https://doi.org/10.1093/imaiai/iaae007

[68] LIN, Z., KONG, D. and WANG, L. (2023). Causal inference on distribution functions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **85** 378–398. MR4725988 https://doi.org/10.1093/jrsssb/qkad008

[69] LU, D. and WENG, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.* **28** 823–870.
[70] MATABUENA, M. and PETERSEN, A. (2023). Distributional data analysis of accelerometer data from the NHANES database using nonparametric survey regression models. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **72** 294–313. MR4719277 https://doi.org/10.1093/jrsssc/qlad007
[71] MAYBANK, S. J. (2004). Detection of image structures using the Fisher information and the Rao metric. *IEEE Trans. Pattern Anal. Mach. Intell.* **26** 1579–1589.
[72] MAYBANK, S. J. (2006). Application of the Fisher–Rao metric to structure detection. *J. Math. Imaging Vision* **25** 49–62. MR2254438 https://doi.org/10.1007/s10851-006-4533-6
[73] MAYBANK, S. J. (2019). The Fisher–Rao metric in computer vision. In *Proceedings of the* 28*th ACM International Conference on Information and Knowledge Management* 3–3.
[74] MAYBANK, S. J. (2020). Fisher–Rao metric. In *Computer Vision: A Reference Guide* Springer, Berlin.
[75] MEI, Q., SHEN, X. and ZHAI, C. (2007). Automatic labeling of multinomial topic models. In *ACM Conference on Knowledge Discovery and Data Mining* 490–499.
[76] NEGRINI, E. and NURBEKYAN, L. (2024). Applications of no-collision transportation maps in manifold learning. *SIAM J. Math. Data Sci.* **6** 97–126. MR4708906 https://doi.org/10.1137/23M1567771
[77] OLIVA, J., PÓCZOS, B. and SCHNEIDER, J. (2013). Distribution to distribution regression. In *International Conference on Machine Learning* 1049–1057.
[78] OLIVEIRA, F. P. M. and TAVARES, J. M. R. S. (2014). Medical image registration: A review. *Comput. Methods Biomech. Biomed. Eng.* **17** 73–93. https://doi.org/10.1080/10255842.2012.670855
[79] PANARETOS, V. M. and ZEMEL, Y. (2019). Statistical aspects of Wasserstein distances. *Annu. Rev. Stat. Appl.* **6** 405–431. MR3939527 https://doi.org/10.1146/annurev-statistics-030718-104938
[80] PEARSON, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **2** 559–572.
[81] PÉREZ-CRUZ, F. (2008). Kullback–Leibler divergence estimation of continuous distributions. In *International Symposium on Information Theory* 1666–1670. IEEE Press, New York.
[82] PERRY, A., WEED, J., BANDEIRA, A. S., RIGOLLET, P. and SINGER, A. (2019). The sample complexity of multireference alignment. *SIAM J. Math. Data Sci.* **1** 497–517. MR4002723 https://doi.org/10.1137/18M1214317
[83] PERRY, A., WEIN, A. S., BANDEIRA, A. S. and MOITRA, A. (2018). Message-passing algorithms for synchronization problems over compact groups. *Comm. Pure Appl. Math.* **71** 2275–2322. MR3862091 https://doi.org/10.1002/cpa.21750
[84] PETER, A. and RANGARAJAN, A. (2006). Shape analysis using the Fisher–Rao Riemannian metric: Unifying shape representation and deformation. In *International Symposium on Biomedical Imaging: Nano to Macro* 1164–1167. IEEE Press, New York.
[85] 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
[86] PEYRÉ, G., CUTURI, M. et al. (2019). Computational optimal transport: With applications to data science. *Found. Trends Mach. Learn.* **11** 355–607.
[87] POLYANSKIY, Y. and WU, Y. (2019). Lecture Notes on Information Theory (Unpublished. Available from the first author's website).
[88] R CORE TEAM (2021). R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
[89] 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
[90] RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2168993
[91] RAO, C. R. (1945). Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **37** 81–91. MR0015748
[92] RAO, C. R. (1987). Differential metrics in probability spaces. *Differ. Geom. Stat. Inference* **10** 217–240.
[93] SANTAMBROGIO, F. (2015). *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*. *Progress in Nonlinear Differential Equations and Their Applications* **87**. Birkhäuser, Cham. MR3409718 https://doi.org/10.1007/978-3-319-20828-2
[94] SCHOENBERG, I. J. (1935). Remarks to Maurice Fréchet's article "Sur la définition axiomatique d'une classe d'espace distanciés vectoriellement applicable sur l'espace de Hilbert". *Ann. of Math.* (2) **36** 724–732. MR1503248 https://doi.org/10.2307/1968654
[95] SHANG, Y., RUML, W., ZHANG, Y. and FROMHERZ, M. P. (2003). Localization from mere connectivity. In *ACM International Symposium on Mobile Ad Hoc Networking and Computing* 201–212.
[96] SILVA, V. and TENENBAUM, J. B. (2002). Global versus local methods in nonlinear dimensionality reduction. *Adv. Neural Inf. Process. Syst.* **15** 705–712.

[97] SMOLA, A., GRETTON, A., SONG, L. and SCHÖLKOPF, B. (2007). A Hilbert space embedding for distributions. In *Algorithmic Learning Theory* 13–31. Springer, Berlin.
[98] SOBERÓN, J. and NAKAMURA, M. (2009). Niches and distributional areas: Concepts, methods, and assumptions. *Proc. Natl. Acad. Sci. USA* **106** 19644–19650.
[99] SRIPERUMBUDUR, B. K., GRETTON, A., FUKUMIZU, K., SCHÖLKOPF, B. and LANCKRIET, G. R. G. (2010). Hilbert space embeddings and metrics on probability measures. *J. Mach. Learn. Res.* **11** 1517–1561. MR2645460
[100] SRIVASTAVA, A. and KLASSEN, E. P. (2016). *Functional and Shape Data Analysis. Springer Series in Statistics*. Springer, New York. MR3821566
[101] SZABÓ, Z., SRIPERUMBUDUR, B. K., PÓCZOS, B. and GRETTON, A. (2016). Learning theory of distribution regression. *J. Mach. Learn. Res.* **17** Paper No. 152, 40. MR3555043
[102] SZÉKELY, G. J. and RIZZO, M. L. (2004). Testing for equal distributions in high dimension. *InterStat* **5** 1–6.
[103] TENENBAUM, J. (1997). Mapping a manifold of perceptual observations. *Adv. Neural Inf. Process. Syst.* **10**.
[104] TENENBAUM, J. B., DE SILVA, V. and LANGFORD, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science* **290** 2319–2323. https://doi.org/10.1126/science.290.5500.2319
[105] TORGERSON, W. S. (1985). *Theory and Methods of Scaling*. Robert E. Krieger Publishing Co., Inc., Melbourne, FL. With a foreword by Harold Gulliksen. Reprint of the 1958 original. MR0893591
[106] TRIGANO, T., ISSERLES, U. and RITOV, Y. (2011). Semiparametric curve alignment and shift density estimation for biological data. *IEEE Trans. Signal Process.* **59** 1970–1984. MR2816476 https://doi.org/10.1109/TSP.2011.2113179
[107] TUCKER, J. D., WU, W. and SRIVASTAVA, A. (2014). Analysis of proteomics data: Phase amplitude separation using an extended Fisher–Rao metric. *Electron. J. Stat.* **8** 1724–1733. MR3273587 https://doi.org/10.1214/14-EJS900B
[108] VAN DEN BRANDEN LAMBRECHT, C. J. and VERSCHEURE, O. (1996). Perceptual quality measure using a spatiotemporal model of the human visual system. In *Digital Video Compression: Algorithms and Technologies* 1996 **2668** 450–461. SPIE, Bellingham.
[109] VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256
[110] VIMOND, M. (2010). Efficient estimation for a subclass of shape invariant models. *Ann. Statist.* **38** 1885–1912. MR2662362 https://doi.org/10.1214/07-AOS566
[111] WANG, J.-L., CHIOU, J.-M. and MÜLLER, H.-G. (2016). Functional data analysis. *Annu. Rev. Stat. Appl.* **3** 257–295.
[112] WANG, K. and GASSER, T. (1999). Synchronizing sample curves nonparametrically. *Ann. Statist.* **27** 439–460. MR1714722 https://doi.org/10.1214/aos/1018031210
[113] WANG, L. and SINGER, A. (2013). Exact and stable recovery of rotations for robust synchronization. *Inf. Inference* **2** 145–193. MR3311446 https://doi.org/10.1093/imaiai/iat005
[114] WANG, Q., KULKARNI, S. R. and VERDÚ, S. (2005). Divergence estimation of continuous distributions based on data-dependent partitions. *IEEE Trans. Inf. Theory* **51** 3064–3074. MR2239136 https://doi.org/10.1109/TIT.2005.853314
[115] WANG, Q., KULKARNI, S. R. and VERDÚ, S. (2009). Divergence estimation for multidimensional densities via k-nearest-neighbor distances. *IEEE Trans. Inf. Theory* **55** 2392–2405. MR2729888 https://doi.org/10.1109/TIT.2009.2016060
[116] WANG, Z., SIMONCELLI, E. P. and BOVIK, A. C. (2003). Multiscale structural similarity for image quality assessment. In *Asilomar Conference on Signals, Systems & Computers, Vol.* 2 1398–1402. IEEE Press, New York.
[117] WINKLER, S. (1998). A perceptual distortion metric for digital color images. In *International Conference on Image Processing* 399–403. IEEE Press, New York.
[118] WU, W. and SRIVASTAVA, A. (2014). Analysis of spike train data: Alignment and comparisons using the extended Fisher–Rao metric. *Electron. J. Stat.* **8** 1776–1785. MR3273594 https://doi.org/10.1214/14-EJS865B
[119] YANG, H., BALADANDAYUTHAPANI, V., RAO, A. U. K. and MORRIS, J. S. (2020). Quantile function on scalar regression analysis for distributional data. *J. Amer. Statist. Assoc.* **115** 90–106. MR4078447 https://doi.org/10.1080/01621459.2019.1609969
[120] YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. MR2160561 https://doi.org/10.1198/016214504000001745
[121] ZHANG, C., KOKOSZKA, P. and PETERSEN, A. (2022). Wasserstein autoregressive models for density time series. *J. Time Series Anal.* **43** 30–52. MR4400283 https://doi.org/10.1111/jtsa.12590

[122] ZHAO, P. and LAI, L. (2020). Minimax optimal estimation of KL divergence for continuous distributions. *IEEE Trans. Inf. Theory* **66** 7787–7811. MR4191027 https://doi.org/10.1109/TIT.2020.3009923

[123] ZHUANG, Y., CHEN, X. and YANG, Y. (2022). Wasserstein K-means for clustering probability distributions. *Adv. Neural Inf. Process. Syst.* **35** 11382–11395.

[124] ZINGER, A. A., KAKOSYAN, A. V. and KLEBANOV, L. B. (1992). A characterization of distributions by mean values of statistics and certain probabilistic metrics. *J. Math. Sci.* **59** 914–920. MR1163396 https://doi.org/10.1007/BF01099119

The Annals of Statistics
2025, Vol. 53, No. 2, 647–675
https://doi.org/10.1214/24-AOS2476
© Institute of Mathematical Statistics, 2025

A NEW CENTRAL LIMIT THEOREM FOR THE AUGMENTED IPW ESTIMATOR: VARIANCE INFLATION, CROSS-FIT COVARIANCE AND BEYOND

BY KUANHAO JIANG[1,a], RAJARSHI MUKHERJEE[2,d], SUBHABRATA SEN[1,b] AND PRAGYA SUR[1,c]

[1]*Department of Statistics, Harvard University,* [a]*kuanhaojiang@g.harvard.edu,* [b]*subhabratasen@fas.harvard.edu,* [c]*pragya@fas.harvard.edu*

[2]*Department of Biostatistics, Harvard T.H. Chan School of Public Health,* [d]*ram521@mail.harvard.edu*

Estimation of the average treatment effect (ATE) is a central problem in causal inference. In recent times, inference for the ATE in the presence of high-dimensional covariates has been extensively studied. Among diverse approaches that have been proposed, augmented inverse propensity weighting (AIPW) with cross-fitting has emerged a popular choice in practice. In this work, we study this cross-fit AIPW estimator under well-specified outcome regression and propensity score models in a high-dimensional regime where the number of features and samples are both large and comparable. Under assumptions on the covariate distribution, we establish a new central limit theorem for the suitably scaled cross-fit AIPW that applies without *any* sparsity assumptions on the underlying high-dimensional parameters. Our CLT uncovers two crucial phenomena among others: (i) the AIPW exhibits a substantial variance inflation that can be precisely quantified in terms of the signal-to-noise ratio and other problem parameters, (ii) the asymptotic covariance between the precross-fit estimators is nonnegligible even on the $\sqrt{n}$ scale. These findings are strikingly different from their classical counterparts. On the technical front, our work utilizes a novel interplay between three distinct tools—approximate message passing theory, the theory of deterministic equivalents and the leave-one-out approach. We believe our proof techniques should be useful for analyzing other two-stage estimators in this high-dimensional regime. We complement our theoretical results with simulations that demonstrate both the finite sample efficacy of our CLT and its robustness to our assumptions. Finally, we provide some theoretical evidence for the universality of our CLT to the law of the covariates, and explore the effects of certain forms of model misspecification.

REFERENCES

[1] ABADIE, A. and IMBENS, G. W. (2011). Bias-corrected matching estimators for average treatment effects. *J. Bus. Econom. Statist.* **29** 1–11. MR2789386 https://doi.org/10.1198/jbes.2009.07333

[2] ABADIE, A. and IMBENS, G. W. (2016). Matching on the estimated propensity score. *Econometrica* **84** 781–807. MR3481379 https://doi.org/10.3982/ECTA11293

[3] ANATOLYEV, S. (2019). Many instruments and/or regressors: A friendly guide. *J. Econ. Surv.* **33** 689–726.

[4] ANDREWS, D. W. and STOCK, J. H. (2005). *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge University Press.

[5] ATHEY, S. and IMBENS, G. W. (2019). Machine learning methods that economists should know about. *Ann. Rev. Econ.* **11** 685–725.

[6] ATHEY, S., IMBENS, G. W. and WAGER, S. (2018). Approximate residual balancing: Debiased inference of average treatment effects in high dimensions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 597–623. MR3849336 https://doi.org/10.1111/rssb.12268

MSC2020 subject classifications. Primary 62E20; secondary 62F12.

Key words and phrases. High-dimensional inference, causal inference, augmented inverse probability weighting, average treatment effect estimation, variance inflation, approximate message passing, leave-one-out, cross-fitting.

[7] BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2567175 https://doi.org/10.1007/978-1-4419-0661-8

[8] BANG, H. and ROBINS, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61** 962–972. MR2216189 https://doi.org/10.1111/j.1541-0420.2005.00377.x

[9] BAYATI, M., ERDOGDU, M. A. and MONTANARI, A. (2013). Estimating lasso risk and noise level. *Adv. Neural Inf. Process. Syst.* **26**.

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

[11] BEAN, D., BICKEL, P. J., EL KAROUI, N. and YU, B. (2013). Optimal M-estimation in high-dimensional regression. *Proc. Natl. Acad. Sci. USA* **110** 14563–14568.

[12] BEKKER, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica* **62** 657–681. MR1281697 https://doi.org/10.2307/2951662

[13] BELLEC, P. C. (2022). Observable adjustments in single-index models for regularized M-estimators. ArXiv preprint. Available at arXiv:2204.06990.

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

[15] BELLEC, P. C., SHEN, Y. and ZHANG, C.-H. (2022). Asymptotic normality of robust M-estimators with convex penalty. *Electron. J. Stat.* **16** 5591–5622. MR4497866 https://doi.org/10.1214/22-ejs2065

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

[17] BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Rev. Econ. Stud.* **81** 608–650. MR3207983 https://doi.org/10.1093/restud/rdt044

[18] BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. *Johns Hopkins Series in the Mathematical Sciences*. Johns Hopkins Univ. Press, Baltimore, MD. MR1245941

[19] BOLTHAUSEN, E. (2014). An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model. *Comm. Math. Phys.* **325** 333–366. MR3147441 https://doi.org/10.1007/s00220-013-1862-3

[20] BRADIC, J., WAGER, S. and ZHU, Y. (2019). Sparsity double robust inference of average treatment effects. ArXiv preprint. Available at arXiv:1905.00744.

[21] BU, Z., KLUSOWSKI, J., RUSH, C. and SU, W. (2019). Algorithmic analysis and statistical estimation of slope via approximate message passing. *Adv. Neural Inf. Process. Syst.* **32**.

[22] 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

[23] CANDÈS, E. J. and SUR, P. (2020). The phase transition for the existence of the maximum likelihood estimate in high-dimensional logistic regression. *Ann. Statist.* **48** 27–42. MR4065151 https://doi.org/10.1214/18-AOS1789

[24] CARMONA, P. and HU, Y. (2006). Universality in Sherrington–Kirkpatrick's spin glass model. *Ann. Inst. Henri Poincaré Probab. Stat.* **42** 215–222. MR2199799 https://doi.org/10.1016/j.anihpb.2005.04.001

[25] CATTANEO, M. D., JANSSON, M. and MA, X. (2019). Two-step estimation and inference with possibly many included covariates. *Rev. Econ. Stud.* **86** 1095–1122. MR3945564 https://doi.org/10.1093/restud/rdy053

[26] CATTANEO, M. D., JANSSON, M. and NEWEY, W. K. (2018). Inference in linear regression models with many covariates and heteroscedasticity. *J. Amer. Statist. Assoc.* **113** 1350–1361. MR3862362 https://doi.org/10.1080/01621459.2017.1328360

[27] CATTANEO, M. D., JANSSON, M. and NEWEY, W. K. (2018). Alternative asymptotics and the partially linear model with many regressors. *Econometric Theory* **34** 277–301. MR3783998 https://doi.org/10.1017/S026646661600013X

[28] 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

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

[30] CELENTANO, M. and WAINWRIGHT, M. J. (2023). Challenges of the inconsistency regime: Novel debiasing methods for missing data models. ArXiv preprint. Available at arXiv:2309.01362.

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

[32] CHERNOZHUKOV, V., CHETVERIKOV, D., DEMIRER, M., DUFLO, E., HANSEN, C. and NEWEY, W. (2017). Double/debiased/Neyman machine learning of treatment effects. *Amer. Econ. Rev.* **107** 261–65.

[33] CHERNOZHUKOV, V., NEWEY, W. K., QUINTAS-MARTINEZ, V. and SYRGKANIS, V. (2021). Automatic debiased machine learning via neural nets for generalized linear regression. ArXiv preprint. Available at arXiv:2104.14737.

[34] CHERNOZHUKOV, V., NEWEY, W. K. and SINGH, R. (2022). Debiased machine learning of global and local parameters using regularized Riesz representers. *Econom. J.* **25** 576–601. MR4565439 https://doi.org/10.1093/ectj/utac002

[35] COUILLET, R., DEBBAH, M. and SILVERSTEIN, J. W. (2011). A deterministic equivalent for the analysis of correlated MIMO multiple access channels. *IEEE Trans. Inf. Theory* **57** 3493–3514. MR2817033 https://doi.org/10.1109/TIT.2011.2133151

[36] COUILLET, R. and LIAO, Z. (2021). *Random Matrix Theory for Machine Learning*.

[37] 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

[38] DICKER, L. H. and ERDOGDU, M. A. (2016). Maximum likelihood for variance estimation in high-dimensional linear models. In *Artificial Intelligence and Statistics* 159–167. PMLR.

[39] DOBRIBAN, E. and WAGER, S. (2018). High-dimensional asymptotics of prediction: Ridge regression and classification. *Ann. Statist.* **46** 247–279. MR3766952 https://doi.org/10.1214/17-AOS1549

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

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

[42] 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

[43] EL KAROUI, N. and PURDOM, E. (2018). Can we trust the bootstrap in high-dimensions? The case of linear models. *J. Mach. Learn. Res.* **19** Paper No. 5. MR3862412

[44] 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.

[45] 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

[46] FENG, O. Y., VENKATARAMANAN, R., RUSH, C. and SAMWORTH, R. J. (2022). A unifying tutorial on approximate message passing. *Found. Trends Mach. Learn.* **15** 335–536.

[47] FONG, C., HAZLETT, C. and IMAI, K. (2018). Covariate balancing propensity score for a continuous treatment: Application to the efficacy of political advertisements. *Ann. Appl. Stat.* **12** 156–177. MR3773389 https://doi.org/10.1214/17-AOAS1101

[48] GIRKO, V. L. (2012). *Theory of Stochastic Canonical Equations*: *Volumes I and II*, **535**. Springer Science & Business Media.

[49] HACHEM, W., LOUBATON, P. and NAJIM, J. (2007). Deterministic equivalents for certain functionals of large random matrices. *Ann. Appl. Probab.* **17** 875–930. MR2326235 https://doi.org/10.1214/105051606000000925

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

[51] HAHN, J. (2002). Optimal inference with many instruments. *Econometric Theory* **18** 140–168. MR1885354 https://doi.org/10.1017/S0266466602181084

[52] HERNÁN, M. A. and ROBINS, J. M. (2010). Causal inference.

[53] HIRANO, K., IMBENS, G. W. and RIDDER, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71** 1161–1189. MR1995826 https://doi.org/10.1111/1468-0262.00442

[54] HORVITZ, D. G. and THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47** 663–685. MR0053460

[55] 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*) 375–379. IEEE Press, New York.

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

[57] IMAI, K. and RATKOVIC, M. (2014). Covariate balancing propensity score. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 243–263. MR3153941 https://doi.org/10.1111/rssb.12027

[58] IMBENS, G. W. and RUBIN, D. B. (2015). *Causal Inference—for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge Univ. Press, New York. MR3309951 https://doi.org/10.1017/CBO9781139025751

[59] JANSON, L., FOYGEL BARBER, R. and CANDÈS, E. (2017). EigenPrism: Inference for high dimensional signal-to-noise ratios. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1037–1065. MR3689308 https://doi.org/10.1111/rssb.12203

[60] JAVANMARD, A. and MONTANARI, A. (2013). State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Inf. Inference* **2** 115–144. MR3311445 https://doi.org/10.1093/imaiai/iat004

[61] 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

[62] JIANG, K., MUKHERJEE, R., SEN, S. and SUR, P. (2025). Supplement to "A new central limit theorem for the augmented IPW estimator: Variance inflation, cross-fit covariance and beyond." https://doi.org/10.1214/24-AOS2476SUPPA

[63] JIANG, K., MUKHERJEE, R., SEN, S. and SUR, P. (2025). Codes for "A new central limit theorem for the augmented IPW estimator: Variance inflation, cross-fit covariance and beyond." https://doi.org/10.1214/24-AOS2476SUPPB

[64] 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

[65] KALLUS, N. and MAO, X. (2020). On the role of surrogates in the efficient estimation of treatment effects with limited outcome data. ArXiv preprint. Available at arXiv:2003.12408.

[66] KENNEDY, E. H. (2022). Semiparametric doubly robust targeted double machine learning: a review. ArXiv preprint. Available at arXiv:2203.06469.

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

[68] LEI, L., BICKEL, P. J. and EL KAROUI, N. (2018). Asymptotics for high dimensional regression M-estimates: Fixed design results. *Probab. Theory Related Fields* **172** 983–1079. MR3877551 https://doi.org/10.1007/s00440-017-0824-7

[69] LI, F., MORGAN, K. L. and ZASLAVSKY, A. M. (2018). Balancing covariates via propensity score weighting. *J. Amer. Statist. Assoc.* **113** 390–400. MR3803473 https://doi.org/10.1080/01621459.2016.1260466

[70] LI, Y. and SUR, P. (2023). Spectrum-Aware Adjustment: a New Debiasing Framework with Applications to Principal Components Regression. ArXiv preprint. Available at arXiv:2309.07810.

[71] LI, Y. and WEI, Y. (2021). Minimum L1-norm interpolators: Precise asymptotics and multiple descent. ArXiv preprint. Available at arXiv:2110.09502.

[72] LIANG, T., SEN, S. and SUR, P. (2023). High-dimensional asymptotics of Langevin dynamics in spiked matrix models. *Inf. Inference* **12** Paper No. iaad042. MR4655761 https://doi.org/10.1093/imaiai/iaad042

[73] 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

[74] LIAO, Z. (2019). A random matrix framework for large dimensional machine learning and neural networks. Ph.D. thesis.

[75] MA, C., WANG, K., CHI, Y. and CHEN, Y. (2018). Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval and matrix completion. In *International Conference on Machine Learning* 3345–3354. PMLR.

[76] 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

[77] 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

[78] MÉZARD, M. and MONTANARI, A. (2009). *Information, Physics, and Computation*. *Oxford Graduate Texts*. Oxford Univ. Press, Oxford. MR2518205 https://doi.org/10.1093/acprof:oso/9780198570837.001.0001

[79] MIGNACCO, F., KRZAKALA, F., LU, Y., URBANI, P. and ZDEBOROVA, L. (2020). The role of regularization in classification of high-dimensional noisy Gaussian mixture. In *International Conference on Machine Learning* 6874–6883. PMLR.

[80] MONTANARI, A. (2012). Graphical models concepts in compressed sensing. In *Compressed Sensing* (Y. Eldar and G. Kutyniok, eds.). Cambridge Univ. Press, Cambridge. MR2963574 https://doi.org/10.1017/CBO9780511794308.010

[81] 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.

[82] MONTANARI, A. and SEN, S. (2022). A Short Tutorial on Mean-Field Spin Glass Techniques for Non-Physicists. ArXiv preprint. Available at arXiv:2204.02909.

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

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

[85] NING, Y., PENG, S. 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

[86] PATIL, P., WEI, Y., RINALDO, A. and TIBSHIRANI, R. (2021). Uniform consistency of cross-validation estimators for high-dimensional ridge regression. In *International Conference on Artificial Intelligence and Statistics* 3178–3186. PMLR.

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

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

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

[90] 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.

[91] ROBINS, J. M. and ROTNITZKY, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *J. Amer. Statist. Assoc.* **90** 122–129. MR1325119

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

[93] ROBINS, J. M., ROTNITZKY, A. and ZHAO, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J. Amer. Statist. Assoc.* **90** 106–121. MR1325118

[94] 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

[95] ROSENBAUM, P. R. and RUBIN, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *J. Amer. Statist. Assoc.* **79** 516–524.

[96] RUBIN, D. B. (1973). The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics* 185–203.

[97] RUBIN, D. B. and THOMAS, N. (1996). Matching using estimated propensity scores: Relating theory to practice. *Biometrics* 249–264.

[98] SALEHI, F., ABBASI, E. and HASSIBI, B. (2019). The impact of regularization on high-dimensional logistic regression. *Adv. Neural Inf. Process. Syst.* **32**.

[99] SCHARFSTEIN, D. O., ROTNITZKY, A. and ROBINS, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J. Amer. Statist. Assoc.* **94** 1096–1146. MR1731478 https://doi.org/10.2307/2669923

[100] 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.

[101] SNOWDEN, J. M., ROSE, S. and MORTIMER, K. M. (2011). Implementation of G-computation on a simulated data set: Demonstration of a causal inference technique. *Amer. J. Epidemiol.* **173** 731–738. https://doi.org/10.1093/aje/kwq472

[102] STUART, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statist. Sci.* **25** 1–21. MR2741812 https://doi.org/10.1214/09-STS313

[103] SUN, B. and TAN, Z. (2022). High-dimensional model-assisted inference for local average treatment effects with instrumental variables. *J. Bus. Econom. Statist.* **40** 1732–1744. MR4492065 https://doi.org/10.1080/07350015.2021.1970575

[104] SUR, P. and CANDÈS, E. J. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. *Proc. Natl. Acad. Sci. USA* **116** 14516–14525. MR3984492 https://doi.org/10.1073/pnas.1810420116

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

[106] TAN, Z. (2020). Regularized calibrated estimation of propensity scores with model misspecification and high-dimensional data. *Biometrika* **107** 137–158. MR4064145 https://doi.org/10.1093/biomet/asz059

[107] 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

[108] 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

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

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

[111] VANSTEELANDT, S. and KEIDING, N. (2011). Invited commentary: G-computation–lost in translation? *Amer. J. Epidemiol.* **173** 739–742. https://doi.org/10.1093/aje/kwq474

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

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

[114] WAGER, S., DU, W., TAYLOR, J. and TIBSHIRANI, R. J. (2016). High-dimensional regression adjustments in randomized experiments. *Proc. Natl. Acad. Sci. USA* **113** 12673–12678. MR3576188 https://doi.org/10.1073/pnas.1614732113

[115] WANG, S., GITTENS, A. and MAHONEY, M. W. (2017). Sketched ridge regression: Optimization perspective, statistical perspective, and model averaging. *J. Mach. Learn. Res.* **18** Paper No. 218. MR3827106

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

[117] WANG, S., ZHOU, W., LU, H., MALEKI, A. and MIRROKNI, V. (2018). Approximate leave-one-out for fast parameter tuning in high dimensions. In *International Conference on Machine Learning* 5228–5237. PMLR.

[118] WANG, Y. and SHAH, R. D. (2024). Debiased inverse propensity score weighting for estimation of average treatment effects with high-dimensional confounders. *Ann. Statist.* **52** 1978–2003. MR4829477 https://doi.org/10.1214/24-aos2409

[119] XU, J., MALEKI, A., RAD, K. R. and HSU, D. (2021). Consistent risk estimation in moderately high-dimensional linear regression. *IEEE Trans. Inf. Theory* **67** 5997–6030. MR4345048 https://doi.org/10.1109/TIT.2021.3095375

[120] YADLOWSKY, S. (2022). Explaining Practical Differences Between Treatment Effect Estimators with High Dimensional Asymptotics. ArXiv preprint. Available at arXiv:2203.12538.

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

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

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

[124] ZHOU, L., KOEHLER, F., SUR, P., SUTHERLAND, D. J. and SREBRO, N. (2022). A non-asymptotic Moreau envelope theory for high-dimensional generalized linear models. *Adv. Neural Inf. Process. Syst.* **35** 21286–21299.

[125] ZUBIZARRETA, J. R. (2015). Stable weights that balance covariates for estimation with incomplete outcome data. *J. Amer. Statist. Assoc.* **110** 910–922. MR3420672 https://doi.org/10.1080/01621459.2015.1023805

The Annals of Statistics
2025, Vol. 53, No. 2, 676–702
https://doi.org/10.1214/24-AOS2477
© Institute of Mathematical Statistics, 2025

SPARSE ANOMALY DETECTION ACROSS REFERENTIALS: A RANK-BASED HIGHER CRITICISM APPROACH

BY IVO V. STOEPKER[1,a], RUI M. CASTRO[1,b] AND ERY ARIAS-CASTRO[2,c]

[1] *Department of Mathematics and Computer Science, Technische Universiteit Eindhoven,* [a]*i.v.stoepker@tue.nl,* [b]*rmcastro@tue.nl*

[2] *Department of Mathematics and Halıcıoğlu Data Science Institute, University of California,* [c]*eariascastro@ucsd.edu*

Detecting anomalies in large sets of observations is crucial in various applications, such as epidemiological studies, gene expression studies, and systems monitoring. We consider settings where the units of interest result in multiple independent observations from potentially distinct referentials. Scan statistics and related methods are commonly used in such settings, but rely on stringent modeling assumptions for proper calibration. We instead propose a rank-based variant of the higher criticism statistic that only requires independent observations originating from ordered spaces. We show under what conditions the resulting methodology is able to detect the presence of anomalies. These conditions are stated in a general, nonparametric manner, and depend solely on the probabilities of anomalous observations exceeding nominal observations. The analysis requires a refined understanding of the distribution of the ranks under the presence of anomalies, and in particular of the rank-induced dependencies. The methodology is robust against heavy-tailed distributions through the use of ranks. Within the exponential family and a family of convolutional models, we analytically quantify the asymptotic performance of our methodology and the performance of the oracle, and show the difference is small for many common models. Simulations confirm these results. We show the applicability of the methodology through an analysis of quality control data of a pharmaceutical manufacturing process.

REFERENCES

AGGARWAL, C. C. (2017). *Outlier Analysis*, 2nd ed. Springer, Cham. MR3586374 https://doi.org/10.1007/978-3-319-47578-3

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

ARIAS-CASTRO, E., CANDÈS, E. J. and PLAN, Y. (2011). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *Ann. Statist.* **39** 2533–2556. MR2906877 https://doi.org/10.1214/11-AOS910

ARIAS-CASTRO, E., CASTRO, R. M., TÁNCZOS, E. and WANG, M. (2018). Distribution-free detection of structured anomalies: Permutation and rank-based scans. *J. Amer. Statist. Assoc.* **113** 789–801. MR3832227 https://doi.org/10.1080/01621459.2017.1286240

ARIAS-CASTRO, E. and HUANG, R. (2020). The sparse variance contamination model. *Statistics* **54** 1081–1093. MR4178899 https://doi.org/10.1080/02331888.2020.1823394

ARIAS-CASTRO, E., HUANG, R. and VERZELEN, N. (2020). Detection of sparse positive dependence. *Electron. J. Stat.* **14** 702–730. MR4057605 https://doi.org/10.1214/19-EJS1675

ARIAS-CASTRO, E. and LIU, Y. (2017). Distribution-free detection of a submatrix. *J. Multivariate Anal.* **156** 29–38. MR3624683 https://doi.org/10.1016/j.jmva.2017.01.013

ARIAS-CASTRO, E. and WANG, M. (2017). Distribution-free tests for sparse heterogeneous mixtures. *TEST* **26** 71–94. MR3613606 https://doi.org/10.1007/s11749-016-0499-x

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

MSC2020 subject classifications. Primary 62G10, 62G20, 62G32; secondary 62J15.

Key words and phrases. Rank-based testing, sparse anomaly detection, distribution-free testing, high-dimensional inference, minimax hypothesis testing.

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

BERK, R. H. and JONES, D. H. (1979). Goodness-of-fit test statistics that dominate the Kolmogorov statistics. *Z. Wahrsch. Verw. Gebiete* **47** 47–59. MR0521531 https://doi.org/10.1007/BF00533250

BUTUCEA, C. and INGSTER, Y. I. (2013). Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli* **19** 2652–2688. MR3160567 https://doi.org/10.3150/12-BEJ470

CAI, T. T., JENG, X. J. and JIN, J. (2011). Optimal detection of heterogeneous and heteroscedastic mixtures. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 629–662. MR2867452 https://doi.org/10.1111/j.1467-9868.2011.00778.x

CAI, T. T., JIN, J. and LOW, M. G. (2007). Estimation and confidence sets for sparse normal mixtures. *Ann. Statist.* **35** 2421–2449. MR2382653 https://doi.org/10.1214/009053607000000334

CAI, T. T. and WU, Y. (2014). Optimal detection of sparse mixtures against a given null distribution. *IEEE Trans. Inf. Theory* **60** 2217–2232. MR3181520 https://doi.org/10.1109/TIT.2014.2304295

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

DELAIGLE, A. and HALL, P. (2009). Higher criticism in the context of unknown distribution, non-independence and classification. In *Perspectives in Mathematical Sciences. I. Stat. Sci. Interdiscip. Res.* **7** 109–138. World Sci. Publ., Hackensack, NJ. MR2581742 https://doi.org/10.1142/9789814273633_0006

DELAIGLE, A., HALL, P. and JIN, J. (2011). Robustness and accuracy of methods for high dimensional data analysis based on Student's t-statistic. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 283–301. MR2815777 https://doi.org/10.1111/j.1467-9868.2010.00761.x

DOBRIBAN, E. (2022). Consistency of invariance-based randomization tests. *Ann. Statist.* **50** 2443–2466. MR4474497 https://doi.org/10.1214/22-aos2200

DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. MR2065195 https://doi.org/10.1214/009053604000000265

DONOHO, D. and JIN, J. (2015). Higher criticism for large-scale inference, especially for rare and weak effects. *Statist. Sci.* **30** 1–25. MR3317751 https://doi.org/10.1214/14-STS506

FRIEDMAN, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. Statist. Assoc.* **32** 675–701.

GIBBONS, J. D. and CHAKRABORTI, S. (2003). *Nonparametric Statistical Inference*, 4th ed. *Statistics: Textbooks and Monographs* **168**. Dekker, New York. MR2064386

GONTSCHARUK, V., LANDWEHR, S. and FINNER, H. (2015). The intermediates take it all: Asymptotics of higher criticism statistics and a powerful alternative based on equal local levels. *Biom. J.* **57** 159–180. MR3298224 https://doi.org/10.1002/bimj.201300255

GORDON, G. J., JENSEN, R. V., HSIAO, L.-L., GULLANS, S. R., BLUMENSTOCK, J. E., RAMASWAMY, S., RICHARDS, W. G., SUGARBAKER, D. J. and BUENO, R. (2002). Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma. *Cancer Res.* **62** 4963–4967.

HÁJEK, J., ŠIDÁK, Z. and SEN, P. K. (1999). *Theory of Rank Tests*, 2nd ed. *Probability and Mathematical Statistics*. Academic Press, San Diego, CA. MR1680991

HALL, P. and JIN, J. (2008). Properties of higher criticism under strong dependence. *Ann. Statist.* **36** 381–402. MR2387976 https://doi.org/10.1214/009053607000000767

HALL, P. and JIN, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.* **38** 1686–1732. MR2662357 https://doi.org/10.1214/09-AOS764

HETTMANSPERGER, T. P. (1984). *Statistical Inference Based on Ranks. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. Wiley, New York. MR0758442

INGSTER, Y. I. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods Statist.* **6** 47–69. MR1456646

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

INGSTER, Y. I., TSYBAKOV, A. B. and VERZELEN, N. (2010). Detection boundary in sparse regression. *Electron. J. Stat.* **4** 1476–1526. MR2747131 https://doi.org/10.1214/10-EJS589

JENG, X. J., CAI, T. T. and LI, H. (2013). Simultaneous discovery of rare and common segment variants. *Biometrika* **100** 157–172. MR3034330 https://doi.org/10.1093/biomet/ass059

JIN, J., STARCK, J.-L., DONOHO, D. L., AGHANIM, N. and FORNI, O. (2005). Cosmological non-Gaussian signature detection: Comparing performance of different statistical tests. *EURASIP J. Appl. Signal Process.* **15** 2470–2485. MR2210857

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

KULLDORFF, M., HEFFERNAN, R., HARTMAN, J., ASSUNCAO, R. and MOSTASHARI, F. (2005). A space-time permutation scan statistic for disease outbreak detection. *PLoS Med*. **2** e59. https://doi.org/10.1371/journal.pmed.0020059

KURT, M. N., YILMAZ, Y. and WANG, X. (2020). Real-time nonparametric anomaly detection in high-dimensional settings. *IEEE Trans*. *Pattern Anal*. *Mach*. *Intell*. **43** 2463-2479. https://doi.org/10.1109/tpami.2020.2970410

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

LEHMANN, E. L. (1975). *Nonparametrics*: *Statistical Methods Based on Ranks*. *Holden-Day Series in Probability and Statistics*. Holden-Day, San Francisco, CA. MR0395032

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

LIU, Y. and XIE, J. (2018). Powerful test based on conditional effects for genome-wide screening. *Ann*. *Appl*. *Stat*. **12** 567–585. MR3773405 https://doi.org/10.1214/17-AOAS1103

LIU, Y. and XIE, J. (2019). Accurate and efficient P-value calculation via Gaussian approximation: A novel Monte-Carlo method. *J*. *Amer*. *Statist*. *Assoc*. **114** 384–392. MR3941262 https://doi.org/10.1080/01621459.2017.1407776

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

MOSCOVICH, A., NADLER, B. and SPIEGELMAN, C. (2016). On the exact Berk-Jones statistics and their p-value calculation. *Electron*. *J*. *Stat*. **10** 2329–2354. MR3544289 https://doi.org/10.1214/16-EJS1172

MUKHERJEE, R., PILLAI, N. S. and LIN, X. (2015). Hypothesis testing for high-dimensional sparse binary regression. *Ann*. *Statist*. **43** 352–381. MR3311863 https://doi.org/10.1214/14-AOS1279

R CORE TEAM (2024). R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

SABATTI, C., SERVICE, S. K., HARTIKAINEN, A. L., POUTA, A., RIPATTI, S., BRODSKY, J., JONES, C. G., ZAITLEN, N. A., VARILO, T. et al. (2009). Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat*. *Genet*. **41** 35–46.

STEPANOVA, N. and PAVLENKO, T. (2018). Goodness-of-fit tests based on sup-functionals of weighted empirical processes. *Theory Probab*. *Appl*. **63** 292–317. MR3796493 https://doi.org/10.4213/tvp5160

STOEPKER, I. V., CASTRO, R. M. and ARIAS-CASTRO, E. (2025a). Supplement to “Sparse Anomaly Detection Across Referentials: a Rank-Based Higher Criticism Approach.” https://doi.org/10.1214/24-AOS2477SUPPA

STOEPKER, I. V., CASTRO, R. M. and ARIAS-CASTRO, E. (2025b). Companion code to “Sparse Anomaly Detection Across Referentials: a Rank-Based Higher Criticism Approach.” https://doi.org/10.1214/24-AOS2477SUPPB

STOEPKER, I. V., CASTRO, R. M., ARIAS-CASTRO, E. and VAN DEN HEUVEL, E. (2024c). Anomaly detection for a large number of streams: A permutation-based higher criticism approach. *J*. *Amer*. *Statist*. *Assoc*. **119** 461–474. MR4713906 https://doi.org/10.1080/01621459.2022.2126361

TAHA, A. and HADI, A. S. (2020). Anomaly detection methods for categorical data: A review. *ACM Comput*. *Surv*. **52** 1–35.

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*, 8th ed. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

WALTHER, G. (2010). Optimal and fast detection of spatial clusters with scan statistics. *Ann*. *Statist*. **38** 1010–1033. MR2604703 https://doi.org/10.1214/09-AOS732

WANG, Y., KLIJN, J. G., ZHANG, Y., SIEUWERTS, A. M., LOOK, M. P., YANG, F., TALANTOV, D., TIMMERMANS, M., MEIJER-VAN GELDER, M. E. et al. (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* **365** 671–679.

WU, Z., SUN, Y., HE, S., CHO, J., ZHAO, H. and JIN, J. (2014). Detection boundary and higher criticism approach for rare and weak genetic effects. *Ann*. *Appl*. *Stat*. **8** 824–851. MR3262536 https://doi.org/10.1214/14-AOAS724

ŽAGAR, J. and MIHELIČ, J. (2022). Big data collection in pharmaceutical manufacturing and its use for product quality predictions. *Sci*. *Data* **9** 1–11.

ZOU, S., LIANG, Y., POOR, H. V. and SHI, X. (2017). Nonparametric detection of anomalous data streams. *IEEE Trans*. *Signal Process*. **65** 5785–5797. MR3708948 https://doi.org/10.1109/TSP.2017.2733472

The Annals of Statistics
2025, Vol. 53, No. 2, 703–723
https://doi.org/10.1214/24-AOS2478
© Institute of Mathematical Statistics, 2025

STUDENTIZED TESTS OF INDEPENDENCE: RANDOM-LIFTER APPROACH

BY ZHE GAO[1,a], ROULIN WANG[2,c], XUEQIN WANG[1,b] AND HEPING ZHANG[3,d]

[1]*School of Management, International Institute of Finance, University of Science and Technology of China,* [a]*gaozh8@mail.ustc.edu.cn,* [b]*wangxq20@ustc.edu.cn*

[2]*School of Statistics, KLATASDS-MOE, East China Normal University,* [c]*rlwang@sfs.ecnu.edu.cn*

[3]*School of Public Health, Yale University,* [d]*heping.zhang@yale.edu*

The exploration of associations between random objects with complex geometric structures has catalyzed the development of various novel statistical tests encompassing distance-based and kernel-based statistics. These methods have various strengths and limitations. One problem is that their test statistics tend to converge to asymptotic null distributions involving second-order Wiener chaos, which are hard to compute and need approximation or permutation techniques that use much computing power to build rejection regions. In this work, we take an entirely different and novel strategy by using the so-called "random-lifter." This method is engineered to yield test statistics with the standard normal limit under null distributions without the need for sample splitting. In other words, we set our sights on having simple limiting distributions and finding the proper statistics through reverse engineering. We use the Central Limit Theorems (CLTs) for degenerate U-statistics derived from our novel association measures to do this. As a result, the asymptotic distributions of our proposed tests are straightforward to compute. Our test statistics also have the minimax property. We further substantiate that our method maintains competitive power against existing methods with minimal adjustments to constant factors. Both numerical simulations and real-data analysis corroborate the efficacy of the random-lifter method.

REFERENCES

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

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

[3] FUKUMIZU, K., GRETTON, A., SUN, X. and SCHÖLKOPF, B. (2007). Kernel measures of conditional dependence. In *Advances in Neural Information Processing Systems* (J. Platt, D. Koller, Y. Singer and S. Roweis, eds.) **20**. Curran Associates, Red Hook.

[4] GAO, H. and SHAO, X. (2023). Two sample testing in high dimension via maximum mean discrepancy. *J. Mach. Learn. Res.* **24** Paper No. [304], 33 pp. MR4664741

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

[6] GAO, Z., WANG, R., WANG, X. and ZHANG, H. (2025). Supplement to "Studentized Tests of Independence: Random-Lifter approach." https://doi.org/10.1214/24-AOS2478SUPPA, https://doi.org/10.1214/24-AOS2478SUPPB

[7] GRETTON, A., BOUSQUET, O., SMOLA, A. and SCHÖLKOPF, B. (2005). Measuring statistical dependence with Hilbert-Schmidt norms. In *Algorithmic Learning Theory*. *Lecture Notes in Computer Science* **3734** 63–77. Springer, Berlin. MR2255909 https://doi.org/10.1007/11564089_7

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

[9] GRETTON, A., HERBRICH, R., SMOLA, A., BOUSQUET, O. and SCHÖLKOPF, B. (2005). Kernel methods for measuring independence. *J. Mach. Learn. Res.* **6** 2075–2129. MR2249882

MSC2020 subject classifications. Primary 62H20; secondary 62G10.
Key words and phrases. Random-lifter, test of independence, second-order Wiener chaos, Studentized test.

[10] GRETTON, A., SMOLA, A., BOUSQUET, O., HERBRICH, R., BELITSKI, A., AUGATH, M., MURAYAMA, Y., PAULS, J., SCHÖLKOPF, B. et al. (2005). Kernel constrained covariance for dependence measurement. In *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics* (R. G. Cowell and Z. Ghahramani, eds.). *Proceedings of Machine Learning Research* **R5** 112–119. PMLR. Reissued by PMLR on 30 March 2021.

[11] KIM, I., BALAKRISHNAN, S. and WASSERMAN, L. (2020). Robust multivariate nonparametric tests via projection averaging. *Ann. Statist.* **48** 3417–3441. MR4185814 https://doi.org/10.1214/19-AOS1936

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

[13] LI, T. and YUAN, M. (2024). On the optimality of Gaussian kernel based nonparametric tests against smooth alternatives. *J. Mach. Learn. Res.* **25** 1–62.

[14] LOPEZ-PAZ, D., HENNIG, P. and SCHÖLKOPF, B. (2013). The randomized dependence coefficient. *Adv. Neural Inf. Process. Syst.* **26**.

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

[16] MOIGNARD, V., WOODHOUSE, S., HAGHVERDI, L., LILLY, A. J., TANAKA, Y., WILKINSON, A. C., BUETTNER, F., MACAULAY, I. C., JAWAID, W. et al. (2015). Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nat. Biotechnol.* **33** 269–276.

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

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

[19] RIGOLLET, P. and TSYBAKOV, A. B. (2007). Linear and convex aggregation of density estimators. *Math. Methods Statist.* **16** 260–280. MR2356821 https://doi.org/10.3103/S1066530707030052

[20] SCHWEIZER, B. and WOLFF, E. F. (1981). On nonparametric measures of dependence for random variables. *Ann. Statist.* **9** 879–885. MR0619291

[21] SCOTT, D. W. (1979). On optimal and data-based histograms. *Biometrika* **66** 605–610. MR0556742 https://doi.org/10.1093/biomet/66.3.605

[22] SHEKHAR, S., KIM, I. and RAMDAS, A. (2023). A permutation-free kernel independence test. *J. Mach. Learn. Res.* **24** Paper No. [369], 68 pp. MR4690318

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

[24] SILVERMAN, B. W. (2018). *Density Estimation for Statistics and Data Analysis*. *Monographs on Statistics and Applied Probability*. CRC Press, London. MR0848134 https://doi.org/10.1007/978-1-4899-3324-9

[25] SRIPERUMBUDUR, B. K., GRETTON, A., FUKUMIZU, K., SCHÖLKOPF, B. and LANCKRIET, G. R. G. (2010). Hilbert space embeddings and metrics on probability measures. *J. Mach. Learn. Res.* **11** 1517–1561. MR2645460

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

[27] 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

[28] 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.* **119** 2772–2784.

[29] WEN, C., WANG, X., GAO, Z. and JIANG, Y. (2023). High-dimensional robust test of independence via Grothendieck's covariance. Technical report.

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

[31] ZHANG, Q., FILIPPI, S., GRETTON, A. and SEJDINOVIC, D. (2018). Large-scale kernel methods for independence testing. *Stat. Comput.* **28** 113–130. MR3741641 https://doi.org/10.1007/s11222-016-9721-7

The Annals of Statistics
2025, Vol. 53, No. 2, 724–748
https://doi.org/10.1214/24-AOS2479
This research was funded, in whole or in part, by UKRI Engineering and Physical Sciences Research Council, EP/T02772X/2. A CC BY 4.0 license is applied to this article arising from this submission, in accordance with the grant's open access conditions.

RESIDUAL PERMUTATION TEST FOR REGRESSION COEFFICIENT TESTING

BY KAIYUE WEN[1,a], TENGYAO WANG[2,b] AND YUHAO WANG[3,c]

[1]*Department of Computer Science, Stanford University,* [a]*kaiyuew@stanford.edu*
[2]*Department of Statistics, London School of Economics and Political Science,* [b]*t.wang59@lse.ac.uk*
[3]*Institute for Interdisciplinary Information Sciences, Tsinghua University,* [c]*yuhaow@tsinghua.edu.cn*

We consider the problem of testing whether a single coefficient is equal to zero in linear models when the dimension of covariates p can be up to a constant fraction of sample size n. In this regime, an important topic is to propose tests with *finite-sample valid* size control without requiring the noise to follow strong distributional assumptions. In this paper, we propose a new method, called the *residual permutation test* (RPT), which is constructed by projecting the regression residuals onto the space orthogonal to the union of the column spaces of the original and permuted design matrices. RPT can be proved to achieve finite-sample size validity under fixed design with just exchangeable noises, whenever $p < n/2$. Moreover, RPT is shown to be asymptotically powerful for heavy-tailed noises with bounded $(1+t)$th order moment when the true coefficient is at least of order $n^{-t/(1+t)}$ for $t \in [0, 1]$. We further proved that this signal size requirement is essentially rate-optimal in the minimax sense. Numerical studies confirm that RPT performs well in a wide range of simulation settings with normal and heavy-tailed noise distributions.

REFERENCES

ARIAS-CASTRO, E., CANDÈS, E. J. and PLAN, Y. (2011). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *Ann. Statist.* **39** 2533–2556. MR2906877 https://doi.org/10.1214/11-AOS910

BALASUBRAMANIAN, V., HO, S.-S. and VOVK, V. (2014). Conformal prediction for reliable machine learning: Theory, adaptations and applications. Newnes.

BARBER, R. F. and CANDÈS, E. J. (2015). Controlling the false discovery rate via knockoffs. *Ann. Statist.* **43** 2055–2085. MR3375876 https://doi.org/10.1214/15-AOS1337

BASU, D. (1980). Randomization analysis of experimental data: The Fisher randomization test. *J. Amer. Statist. Assoc.* **75** 575–595. MR0590687

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

BICKEL, P. J. and SAKOV, A. (2008). On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statist. Sinica* **18** 967–985. MR2440400

BRADIC, J., CHERNOZHUKOV, V., NEWEY, W. K. and ZHU, Y. (2019). Minimax semiparametric learning with approximate sparsity. Preprint. Available at arXiv:1912.12213.

BUBECK, S., CESA-BIANCHI, N. and LUGOSI, G. (2013). Bandits with heavy tail. *IEEE Trans. Inf. Theory* **59** 7711–7717. MR3124669 https://doi.org/10.1109/TIT.2013.2277869

CANAY, I. A., ROMANO, J. P. and SHAIKH, A. M. (2017). Randomization tests under an approximate symmetry assumption. *Econometrica* **85** 1013–1030. MR3664187 https://doi.org/10.3982/ECTA13081

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

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

CAUGHEY, D., DAFOE, A. and MIRATRIX, L. (2017). Beyond the sharp null: Randomization inference, bounded null hypotheses, and confidence intervals for maximum effects. Preprint. Available at arXiv:1709.07339.

MSC2020 subject classifications. Primary 62J99; secondary 62F03.
Key words and phrases. Permutation test, finite-sample validity, heavy-tail distribution, high-dimensional data.

CAUGHEY, D., DAFOE, A., LI, X. and MIRATRIX, L. (2023). Randomisation inference beyond the sharp null: Bounded null hypotheses and quantiles of individual treatment effects. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **85** 1471–1491. MR4718545 https://doi.org/10.1093/jrsssb/qkad080

CHATTERJEE, S. B. (1999). *Generalised Bootstrap Techniques*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—Indian Statistical Institute—Kolkata. MR4380488

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

CONT, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quant. Finance* **1** 223.

FISHER, R. A. (1935). *The Design of Experiments*. Oliver & Boyd, Edinburgh.

D'HAULTFŒUILLE, X. and TUVAANDORJ, P. (2024). A robust permutation test for subvector inference in linear regressions. *Quant. Econ.* **15** 27–87. MR4703622

DEZEURE, R., BÜHLMANN, P., MEIER, L. and MEINSHAUSEN, N. (2015). High-dimensional inference: Confidence intervals, p-values and R-software hdi. *Statist. Sci.* **30** 533–558. MR3432840 https://doi.org/10.1214/15-STS527

DICICCIO, C. J. and ROMANO, J. P. (2017). Robust permutation tests for correlation and regression coefficients. *J. Amer. Statist. Assoc.* **112** 1211–1220. MR3735371 https://doi.org/10.1080/01621459.2016.1202117

DOANE, D. P. and SEWARD, L. E. (2016). Applied statistics in business and economics, 5th. Mcgraw-Hill.

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

EKLUND, A., NICHOLS, T. E. and KNUTSSON, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proc. Natl. Acad. Sci. USA* **113** 7900–7905.

FAN, J., LI, Q. and WANG, Y. (2017). Estimation of high dimensional mean regression in the absence of symmetry and light tail assumptions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 247–265. MR3597972 https://doi.org/10.1111/rssb.12166

FAN, J., WANG, W. and ZHU, Z. (2021). A shrinkage principle for heavy-tailed data: High-dimensional robust low-rank matrix recovery. *Ann. Statist.* **49** 1239–1266. MR4298863 https://doi.org/10.1214/20-aos1980

FREEDMAN, D. and LANE, D. (1983). A nonstochastic interpretation of reported significance levels. *J. Bus. Econom. Statist.* **1** 292–298.

FREEDMAN, D. A. (1981). Bootstrapping regression models. *Ann. Statist.* **9** 1218–1228. MR0630104

HARTIGAN, J. A. (1970). Exact confidence intervals in regression problems with independent symmetric errors. *Ann. Math. Stat.* **41** 1992–1998. MR0267701 https://doi.org/10.1214/aoms/1177696700

IMBENS, G. W. and ROSENBAUM, P. R. (2005). Robust, accurate confidence intervals with a weak instrument: Quarter of birth and education. *J. Roy. Statist. Soc. Ser. A* **168** 109–126. MR2113230 https://doi.org/10.1111/j.1467-985X.2004.00339.x

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

KENNEDY, P. E. (1995). Randomization tests in econometrics. *J. Bus. Econom. Statist.* **13** 85–94. MR1323048 https://doi.org/10.2307/1392523

KIM, I., NEYKOV, M., BALAKRISHNAN, S. and WASSERMAN, L. (2022). Local permutation tests for conditional independence. *Ann. Statist.* **50** 3388–3414. MR4524501 https://doi.org/10.1214/22-aos2233

LAZIC, S. E. (2008). Why we should use simpler models if the data allow this: Relevance for ANOVA designs in experimental biology. *BMC Physiol.* **8** 1–7.

LEI, J., ROBINS, J. and WASSERMAN, L. (2013). Distribution-free prediction sets. *J. Amer. Statist. Assoc.* **108** 278–287. MR3174619 https://doi.org/10.1080/01621459.2012.751873

LEI, L. and BICKEL, P. J. (2021). An assumption-free exact test for fixed-design linear models with exchangeable errors. *Biometrika* **108** 397–412. MR4259139 https://doi.org/10.1093/biomet/asaa079

LOH, P.-L. and TAN, X. L. (2018). High-dimensional robust precision matrix estimation: Cellwise corruption under ϵ-contamination. *Electron. J. Stat.* **12** 1429–1467. MR3804842 https://doi.org/10.1214/18-EJS1427

LOPES, M. (2014). A residual bootstrap for high-dimensional regression with near low-rank designs. *Adv. Neural Inf. Process. Syst.* **27**.

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

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

LYKOURIS, T., MIRROKNI, V. and PAES LEME, R. (2018). Stochastic bandits robust to adversarial corruptions. In *STOC'18—Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing* 114–122. ACM, New York. MR3826238 https://doi.org/10.1145/3188745.3188918

MAMMEN, E. (1993). Bootstrap and wild bootstrap for high-dimensional linear models. *Ann. Statist.* **21** 255–285. MR1212176 https://doi.org/10.1214/aos/1176349025

MEINSHAUSEN, N. (2015). Group bound: Confidence intervals for groups of variables in sparse high dimensional regression without assumptions on the design. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 923–945. MR3414134 https://doi.org/10.1111/rssb.12094

MILLER, R. G. JR. (1974). An unbalanced jackknife. *Ann. Statist.* **2** 880–891. MR0356353

PAOLELLA, M. S. (2019). *Linear Models and Time-Series Analysis: Regression, ANOVA, ARMA and GARCH. Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR3839332

PENSIA, A., JOG, V. and LOH, P.-L. (2024). Robust regression with covariate filtering: Heavy tails and adversarial contamination. *J. Amer. Statist. Assoc.* 1–12.

PITMAN, E. J. (1937a). Significance tests which may be applied to samples from any populations. *Suppl. J. R. Stat. Soc.* **4** 119–130.

PITMAN, E. J. (1937b). Significance tests which may be applied to samples from any populations. II. The correlation coefficient test. *Suppl. J. R. Stat. Soc.* **4** 225–232.

PITMAN, E. J. (1938). Significance tests which may be applied to samples from any populations: III. The analysis of variance test. *Biometrika* **29** 322–335.

ROMANO, J. P. (1990). On the behavior of randomization tests without a group invariance assumption. *J. Amer. Statist. Assoc.* **85** 686–692. MR1138350

ROMANO, Y., PATTERSON, E. and CANDES, E. (2019). Conformalized quantile regression. *Adv. Neural Inf. Process. Syst.* **32**.

ROSENBAUM, P. R. (1984). Conditional permutation tests and the propensity score in observational studies. *J. Amer. Statist. Assoc.* **79** 565–574. MR0763575

ROSENBAUM, P. R. (2002). Covariance adjustment in randomized experiments and observational studies. *Statist. Sci.* **17** 286–327. MR1962487 https://doi.org/10.1214/ss/1042727942

SHAH, R. D. and BÜHLMANN, P. (2023). Double-estimation-friendly inference for high-dimensional misspecified models. *Statist. Sci.* **38** 68–91. MR4535395 https://doi.org/10.1214/22-sts850

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

STIGLER, S. M. (2016). *The Seven Pillars of Statistical Wisdom*. Harvard Univ. Press, Cambridge, MA. MR3585675 https://doi.org/10.4159/9780674970199

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

TOULIS, P. (2019). Invariant inference via residual randomization. Preprint. Available at arXiv:1908.04218.

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

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

WANG, L. (2013). The L_1 penalized LAD estimator for high dimensional linear regression. *J. Multivariate Anal.* **120** 135–151. MR3072722 https://doi.org/10.1016/j.jmva.2013.04.001

WANG, L., PENG, B. and LI, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *J. Amer. Statist. Assoc.* **110** 1658–1669. MR3449062 https://doi.org/10.1080/01621459.2014.988215

WEN, K., WANG, T. and WANG, Y. (2025). Supplement to "Residual permutation test for regression coefficient testing." https://doi.org/10.1214/24-AOS2479SUPP

YOUNG, A. (2019). Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *Q. J. Econ.* **134** 557–598.

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

ZHU, Y. and BRADIC, J. (2018). Linear hypothesis testing in dense high-dimensional linear models. *J. Amer. Statist. Assoc.* **113** 1583–1600. MR3902231 https://doi.org/10.1080/01621459.2017.1356319

The Annals of Statistics
2025, Vol. 53, No. 2, 749–773
https://doi.org/10.1214/24-AOS2480
© Institute of Mathematical Statistics, 2025

ARK: ROBUST KNOCKOFFS INFERENCE WITH COUPLING

BY YINGYING FAN[1,a], LAN GAO[2,c] AND JINCHI LV[1,b]

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

[2]*Department of Business Analytics and Statistics, Haslam College of Business, University of Tennessee,* [c]*lgao13@utk.edu*

We investigate the robustness of the model-X knockoffs framework with respect to the misspecified or estimated feature distribution. We achieve such a goal by theoretically studying the feature selection performance of a practically implemented knockoffs algorithm, which we name as the approximate knockoffs (ARK) procedure, under the measures of the false discovery rate (FDR) and k-familywise error rate (k-FWER). The approximate knockoffs procedure differs from the model-X knockoffs procedure only in that the former uses the misspecified or estimated feature distribution. A key technique in our theoretical analyses is to couple the approximate knockoffs procedure with the model-X knockoffs procedure so that random variables in these two procedures can be close in realizations. We prove that if such coupled model-X knockoffs procedure exists, the approximate knockoffs procedure can achieve the asymptotic FDR or k-FWER control at the target level. We showcase three specific constructions of such coupled model-X knockoff variables, verifying their existence and justifying the robustness of the model-X knockoffs framework. Additionally, we formally connect our concept of knockoff variable coupling to a type of Wasserstein distance.

REFERENCES

BAI, X., REN, J., FAN, Y. and SUN, F. (2021). KIMI: Knockoff inference for motif identification from molecular sequences with controlled false discovery rate. *Bioinformatics* **37** 759–766.

BARBER, R. F. and CANDÈS, E. J. (2015). Controlling the false discovery rate via knockoffs. *Ann. Statist.* **43** 2055–2085. MR3375876 https://doi.org/10.1214/15-AOS1337

BARBER, R. F. and CANDÈS, E. J. (2019). A knockoff filter for high-dimensional selective inference. *Ann. Statist.* **47** 2504–2537. MR3988764 https://doi.org/10.1214/18-AOS1755

BARBER, R. F., CANDÈS, E. J. and SAMWORTH, R. J. (2020). Robust inference with knockoffs. *Ann. Statist.* **48** 1409–1431. MR4124328 https://doi.org/10.1214/19-AOS1852

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

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

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

CAO, Y., SUN, X. and YAO, Y. (2024). Controlling the false discovery rate in transformational sparsity: Split knockoffs. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **86** 386–410. MR4754089 https://doi.org/10.1093/jrsssb/qkad126

CHI, C. M., FAN, Y., ING, C. K. and LV, J. (2023). High-dimensional knockoffs inference for time series data. arXiv preprint. Available at arXiv:2112.09851.

DAI, C., LIN, B., XING, X. and LIU, J. S. (2023). False discovery rate control via data splitting. *J. Amer. Statist. Assoc.* **118** 2503–2520. MR4681600 https://doi.org/10.1080/01621459.2022.2060113

DAI, X., LYU, X. and LI, L. (2023). Kernel knockoffs selection for nonparametric additive models. *J. Amer. Statist. Assoc.* **118** 2158–2170. MR4646633 https://doi.org/10.1080/01621459.2022.2039671

MSC2020 subject classifications. Primary 62G35, 62F07, 62E17; secondary 62G10, 62F35.

Key words and phrases. Knockoffs inference, Wasserstein distance, false discovery rate control, familywise error rate control, coupling, robustness, high dimensionality.

FAN, J., LIAO, Y. and LIU, H. (2016). An overview of the estimation of large covariance and precision matrices. *Econom. J.* **19** C1–C32. MR3501529 https://doi.org/10.1111/ectj.12061

FAN, Y., DEMIRKAYA, E., LI, G. and LV, J. (2020a). RANK: Large-scale inference with graphical nonlinear knockoffs. *J. Amer. Statist. Assoc.* **115** 362–379. MR4078469 https://doi.org/10.1080/01621459.2018.1546589

FAN, Y., GAO, L. and LV, J. (2025). Supplement to "ARK: Robust Knockoffs Inference with Coupling." https://doi.org/10.1214/24-AOS2480SUPP

FAN, Y. and LV, J. (2016). Innovated scalable efficient estimation in ultra-large Gaussian graphical models. *Ann. Statist.* **44** 2098–2126. MR3546445 https://doi.org/10.1214/15-AOS1416

FAN, Y., LV, J., SHARIFVAGHEFI, M. and UEMATSU, Y. (2020b). IPAD: Stable interpretable forecasting with knockoffs inference. *J. Amer. Statist. Assoc.* **115** 1822–1834. MR4189760 https://doi.org/10.1080/01621459.2019.1654878

GUO, X., REN, H., ZOU, C. and LI, R. (2023). Threshold selection in feature screening for error rate control. *J. Amer. Statist. Assoc.* **118** 1773–1785. MR4646605 https://doi.org/10.1080/01621459.2021.2011735

HUANG, D. and JANSON, L. (2020). Relaxing the assumptions of knockoffs by conditioning. *Ann. Statist.* **48** 3021–3042. MR4152633 https://doi.org/10.1214/19-AOS1920

JANSON, L. and SU, W. (2016). Familywise error rate control via knockoffs. *Electron. J. Stat.* **10** 960–975. MR3486422 https://doi.org/10.1214/16-EJS1129

JORDON, J., YOON, J. and VAN DER SCHAAR, M. (2018). KnockoffGAN: Generating knockoffs for feature selection using generative adversarial networks. In *International Conference on Learning Representations*.

LEHMANN, E. L. and ROMANO, J. P. (2005). Generalizations of the familywise error rate. *Ann. Statist.* **33** 1138–1154. MR2195631 https://doi.org/10.1214/009053605000000084

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

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

LIU, H., LAFFERTY, J. and WASSERMAN, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *J. Mach. Learn. Res.* **10** 2295–2328. MR2563983

LU, Y., FAN, Y., LV, J. and NOBLE, W. S. (2018). DeepPINK: Reproducible feature selection in deep neural networks. In *Advances in Neural Information Processing Systems* (*NeurIPS* 2018) 31.

NIU, Z., CHAKRABORTY, A., DUKES, O. and KATSEVICH, E. (2024). Reconciling model-X and doubly robust approaches to conditional independence testing. *Ann. Statist.* **52** 895–921. MR4784063 https://doi.org/10.1214/24-aos2372

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

REN, Z. and BARBER, R. F. (2024). Derandomised knockoffs: Leveraging e-values for false discovery rate control. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **86** 122–154. MR4754084 https://doi.org/10.1093/jrsssb/qkad085

REN, Z., WEI, Y. and CANDÈS, E. (2023). Derandomizing knockoffs. *J. Amer. Statist. Assoc.* **118** 948–958. MR4595468 https://doi.org/10.1080/01621459.2021.1962720

ROMANO, Y., SESIA, M. and CANDÈS, E. (2020). Deep knockoffs. *J. Amer. Statist. Assoc.* **115** 1861–1872. MR4189763 https://doi.org/10.1080/01621459.2019.1660174

SESIA, M., SABATTI, C. and CANDÈS, E. J. (2019). Gene hunting with hidden Markov model knockoffs. *Biometrika* **106** 1–18. MR3912377 https://doi.org/10.1093/biomet/asy033

SPECTOR, A. and JANSON, L. (2022). Powerful knockoffs via minimizing reconstructability. *Ann. Statist.* **50** 252–276. MR4382016 https://doi.org/10.1214/21-aos2104

WANG, W. and JANSON, L. (2022). A high-dimensional power analysis of the conditional randomization test and knockoffs. *Biometrika* **109** 631–645. MR4472839 https://doi.org/10.1093/biomet/asab052

WEINSTEIN, A., SU, W. J., BOGDAN, M., FOYGEL BARBER, R. and CANDÈS, E. J. (2023). A power analysis for model-X knockoffs with ℓ_p-regularized statistics. *Ann. Statist.* **51** 1005–1029. MR4630938 https://doi.org/10.1214/23-aos2274

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

ZHU, Z., FAN, Y., KONG, Y., LV, J. and SUN, F. (2021). DeepLINK: Deep learning inference using knockoffs with applications to genomics. *Proc. Natl. Acad. Sci. USA* **118** Paper No. e2104683118, 12. MR4324622 https://doi.org/10.1073/pnas.2104683118

The Annals of Statistics
2025, Vol. 53, No. 2, 774–801
https://doi.org/10.1214/24-AOS2484
© Institute of Mathematical Statistics, 2025

LOW COORDINATE DEGREE ALGORITHMS I: UNIVERSALITY OF COMPUTATIONAL THRESHOLDS FOR HYPOTHESIS TESTING

BY DMITRIY KUNISKY[a]

Department of Applied Mathematics and Statistics, Johns Hopkins University, [a]*kunisky@jhu.edu*

We study when *low coordinate degree functions (LCDF)*—linear combinations of functions depending on small subsets of entries of a vector—can hypothesis test between high-dimensional probability measures. These functions are a generalization, proposed in Hopkins' 2018 thesis but seldom studied since, of *low degree polynomials (LDP)*, a class widely used in recent literature as a proxy for all efficient algorithms for tasks in statistics and optimization. Instead of the orthogonal polynomial decompositions used in LDP calculations, our analysis of LCDF is based on the *Efron–Stein* or *ANOVA decomposition*, making it much more broadly applicable. By way of illustration, we prove *channel universality* for the success of LCDF in testing for the presence of sufficiently "dilute" random signals through noisy channels: the efficacy of LCDF depends on the channel only through the scalar *Fisher information* for a class of channels including nearly arbitrary additive i.i.d. noise and nearly arbitrary exponential families. As applications, we extend lower bounds against LDP for spiked matrix and tensor models under additive Gaussian noise to lower bounds against LCDF under general noisy channels. We also give a simple and unified treatment of the effect of *censoring* models by erasing observations at random and of *quantizing* models by taking the sign of the observations. These results are the first computational lower bounds against any large class of algorithms for all of these models when the channel is not one of a few special cases, and thereby give the first substantial evidence for the universality of several statistical-to-computational gaps.

REFERENCES

[1] ABBE, E. (2017). Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** Paper No. 177, 86 pp. MR3827065

[2] ACHLIOPTAS, D. and COJA-OGHLAN, A. (2008). Algorithmic barriers from phase transitions. In 49*th Annual IEEE Symposium on Foundations of Computer Science* (*FOCS* 2008) 793–802. IEEE.

[3] ANDERSON, G. W., GUIONNET, A. and ZEITOUNI, O. (2010). *An Introduction to Random Matrices*. *Cambridge Studies in Advanced Mathematics* **118**. Cambridge Univ. Press, Cambridge. MR2760897

[4] AUBIN, B., MAILLARD, A., KRZAKALA, F., MACRIS, N., ZDEBOROVÁ, L. et al. (2018). The committee machine: Computational to statistical gaps in learning a two-layers neural network. *Adv. Neural Inf. Process. Syst.* **31**.

[5] BANDEIRA, A. S., ALAOUI, A. E., HOPKINS, S. B., SCHRAMM, T., WEIN, A. S. and ZADIK, I. (2022). The Franz-Parisi criterion and computational trade-offs in high dimensional statistics. Preprint. Available at arXiv:2205.09727.

[6] BANDEIRA, A. S., BANKS, J., KUNISKY, D., MOORE, C. and WEIN, A. S. (2021). Spectral planting and the hardness of refuting cuts, colorability, and communities in random graphs. In 34*th Annual Conference on Learning Theory* (*COLT* 2021) 410–473. PMLR.

[7] BANDEIRA, A. S., KUNISKY, D. and WEIN, A. S. (2020). Computational hardness of certifying bounds on constrained PCA problems. In 11*th Innovations in Theoretical Computer Science Conference*. *LIPIcs. Leibniz Int. Proc. Inform.* **151** Art. No. 78, 29 pp. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR4048181 https://doi.org/10.4230/LIPIcs.ITCS.2020.78

MSC2020 subject classifications. Primary 62F15, 62C05; secondary 68Q17.

Key words and phrases. Computational complexity, low-degree polynomials, spiked matrix models, spiked tensor models, statistical-to-computational gaps.

[8] BANKS, J., MOORE, C., VERSHYNIN, R., VERZELEN, N. and XU, J. (2018). Information-theoretic bounds and phase transitions in clustering, sparse PCA, and submatrix localization. *IEEE Trans. Inf. Theory* **64** 4872–4994. MR3819345 https://doi.org/10.1109/tit.2018.2810020

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

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

[11] BERTHET, Q. and RIGOLLET, P. (2013). Complexity theoretic lower bounds for sparse principal component detection. In 26*th Annual Conference on Learning Theory* (*COLT* 2013) 1046–1066.

[12] BRENNAN, M. and BRESLER, G. (2019). Optimal average-case reductions to sparse PCA: From weak assumptions to strong hardness. In 32*nd Annual Conference on Learning Theory* (*COLT* 2019) 469–470. PMLR.

[13] BRENNAN, M. and BRESLER, G. (2020). Reducibility and statistical-computational gaps from secret leakage. In 33*rd Annual Conference on Learning Theory* (*COLT* 2020) 648–847. PMLR.

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

[15] BRENNAN, M., BRESLER, G. and HULEIHEL, W. (2019). Universality of computational lower bounds for submatrix detection. In *Conference on Learning Theory* 417–468. PMLR.

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

[17] CHEN, Z., MOSSEL, E. and ZADIK, I. (2023). Almost-linear planted cliques elude the Metropolis process. In *Proceedings of the* 2023 *Annual ACM-SIAM Symposium on Discrete Algorithms* (*SODA*) 4504–4539. SIAM, Philadelphia, PA. MR4538129 https://doi.org/10.1137/1.9781611977554.ch171

[18] CHERTKOV, M., KROC, L., KRZAKALA, F., VERGASSOLA, M. and ZDEBOROVÁ, L. (2010). Inference in particle tracking experiments by passing messages between images. *Proc. Natl. Acad. Sci. USA* **107** 7663–7668.

[19] COJA-OGHLAN, A., GEBHARD, O., HAHN-KLIMROTH, M., WEIN, A. S. and ZADIK, I. (2022). Statistical and computational phase transitions in group testing. In *Conference on Learning Theory* 4764–4781. PMLR.

[20] CORWIN, I. (2012). The Kardar–Parisi–Zhang equation and universality class. *Random Matrices Theory Appl.* **1** 1130001, 76 pp. MR2930377 https://doi.org/10.1142/S2010326311300014

[21] DECELLE, A., KRZAKALA, F., MOORE, C. and ZDEBOROVÁ, L. (2011). Inference and phase transitions in the detection of modules in sparse networks. *Phys. Rev. Lett.* **107** 065701. https://doi.org/10.1103/PhysRevLett.107.065701

[22] DHAWAN, A., MAO, C. and WEIN, A. S. (2025). Detection of dense subhypergraphs by low-degree polynomials. *Random Structures Algorithms* **66** Paper No. e21279, 16 pp. MR4853082 https://doi.org/10.1002/rsa.21279

[23] DIAKONIKOLAS, I. and KANE, D. (2022). Non-Gaussian component analysis via lattice basis reduction. In *Conference on Learning Theory* 4535–4547. PMLR.

[24] DING, J., DU, H. and LI, Z. (2023). Low-degree hardness of detection for correlated Erdős–Rényi graphs. Preprint. Available at arXiv:2311.15931.

[25] DING, J., WU, Y., XU, J. and YANG, D. (2023). The planted matching problem: Sharp threshold and infinite-order phase transition. *Probab. Theory Related Fields* **187** 1–71. MR4634336 https://doi.org/10.1007/s00440-023-01208-6

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

[27] DONSKER, M. D. (1951). An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.* **6** 12. MR0040613

[28] DYER, M., FRIEZE, A. and JERRUM, M. (2002). On counting independent sets in sparse graphs. *SIAM J. Comput.* **31** 1527–1541. MR1936657 https://doi.org/10.1137/S0097539701383844

[29] ERDŐS, L., SCHLEIN, B. and YAU, H.-T. (2011). Universality of random matrices and local relaxation flow. *Invent. Math.* **185** 75–119. MR2810797 https://doi.org/10.1007/s00222-010-0302-7

[30] FÜREDI, Z. and KOMLÓS, J. (1981). The eigenvalues of random symmetric matrices. *Combinatorica* **1** 233–241. MR0637828 https://doi.org/10.1007/BF02579329

[31] GAMARNIK, D., JAGANNATH, A. and WEIN, A. S. (2020). Low-degree hardness of random optimization problems. In 2020 *IEEE* 61*st Annual Symposium on Foundations of Computer Science* 131–140. IEEE Computer Soc., Los Alamitos, CA. MR4232029 https://doi.org/10.1109/FOCS46700.2020.00021

[32] GAMARNIK, D. and SUDAN, M. (2014). Limits of local algorithms over sparse random graphs [extended abstract]. In *ITCS'*14*—Proceedings of the* 2014 *Conference on Innovations in Theoretical Computer Science* 369–375. ACM, New York. MR3359490

[33] GAMARNIK, D. and ZADIK, I. (2024). The landscape of the planted clique problem: Dense subgraphs and the overlap gap property. *Ann. Appl. Probab.* **34** 3375–3434. MR4782508 https://doi.org/10.1214/23-AAP2003

[34] GHEISSARI, R., JAGANNATH, A. and XU, Y. (2023). Finding planted cliques using Markov chain Monte Carlo. Preprint. Available at arXiv:2311.07540.

[35] HOLMGREN, J. and WEIN, A. S. (2021). Counterexamples to the low-degree conjecture. In 12*th Innovations in Theoretical Computer Science Conference*. *LIPIcs*. *Leibniz Int*. *Proc*. *Inform*. **185** Art. No. 75, 9 pp. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR4214319

[36] HOPKINS, S. B. (2018). Statistical inference and the sum of squares method. Ph.D. thesis, Cornell Univ. MR3864930

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

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

[39] HUANG, H. and MOSSEL, E. (2024). Low degree hardness for broadcasting on trees. Preprint. Available at arXiv:2402.13359.

[40] JERRUM, M. (1992). Large cliques elude the Metropolis process. *Random Structures Algorithms* **3** 347–359. MR1179827 https://doi.org/10.1002/rsa.3240030402

[41] 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

[42] KANE, D., LIU, S., LOVETT, S. and MAHAJAN, G. (2022). Computational-statistical gap in reinforcement learning. In *Conference on Learning Theory* 1282–1302. PMLR.

[43] KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* (3) **83** 016107, 10 pp. MR2788206 https://doi.org/10.1103/PhysRevE.83.016107

[44] KOEHLER, F. and MOSSEL, E. (2021). Reconstruction on trees and low-degree polynomials. Preprint. Available at arXiv:2109.06915.

[45] KRZAKALA, F., XU, J. and ZDEBOROVÁ, L. (2016). Mutual information in rank-one matrix estimation. In 2016 *IEEE Information Theory Workshop* (*ITW*) 71–75. IEEE.

[46] KUNISKY, D. (2021). Hypothesis testing with low-degree polynomials in the Morris class of exponential families. In 34*th Annual Conference on Learning Theory* (*COLT* 2021) 2822–2848. PMLR.

[47] KUNISKY, D. (2021). Spectral barriers in certification problems. Ph.D. thesis, New York Univ. MR4314625

[48] KUNISKY, D. (2025). Supplement to "Low coordinate degree algorithms I: Universality of computational thresholds for hypothesis testing." https://doi.org/10.1214/24-AOS2484SUPP

[49] KUNISKY, D., WEIN, A. S. and BANDEIRA, A. S. (2022). Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *Mathematical Analysis*, *Its Applications and Computation*. *Springer Proc*. *Math*. *Stat*. **385** 1–50. Springer, Cham. MR4461037 https://doi.org/10.1007/978-3-030-97127-4_1

[50] LELARGE, M., MASSOULIÉ, L. and XU, J. (2015). Reconstruction in the labelled stochastic block model. *IEEE Trans*. *Netw*. *Sci*. *Eng*. **2** 152–163. MR3453283 https://doi.org/10.1109/TNSE.2015.2490580

[51] LESIEUR, T., KRZAKALA, F. and ZDEBOROVÁ, L. (2015). MMSE of probabilistic low-rank matrix estimation: Universality with respect to the output channel. In 53*rd Annual Allerton Conference on Communication*, *Control*, *and Computing* (*Allerton* 2015) 680–687. IEEE.

[52] LESIEUR, T., KRZAKALA, F. and ZDEBOROVÁ, L. (2015). Phase transitions in sparse PCA. In *IEEE International Symposium on Information Theory* (*ISIT* 2015) 1635–1639. IEEE.

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

[54] MOHARRAMI, M., MOORE, C. and XU, J. (2021). The planted matching problem: Phase transitions and exact results. *Ann. Appl. Probab.* **31** 2663–2720. MR4350971 https://doi.org/10.1214/20-aap1660

[55] MONTANARI, A., REICHMAN, D. and ZEITOUNI, O. (2017). On the limitation of spectral methods: From the Gaussian hidden clique problem to rank one perturbations of Gaussian tensors. *IEEE Trans. Inf. Theory* **63** 1572–1579. MR3625981 https://doi.org/10.1109/TIT.2016.2637959

[56] MONTANARI, A. and WEIN, A. S. (2025). Equivalence of approximate message passing and low-degree polynomials in rank-one matrix estimation. *Probab. Theory Related Fields* **191** 181–233. MR4869255 https://doi.org/10.1007/s00440-024-01322-z

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

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

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

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

[61] RUSH, C., SKERMAN, F., WEIN, A. S. and YANG, D. (2023). Is it easier to count communities than find them? In 14*th Innovations in Theoretical Computer Science Conference*. *LIPIcs*. *Leibniz Int. Proc. Inform.* **251** Art. No. 94, 23 pp. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR4587108 https://doi.org/10.4230/lipics.itcs.2023.94

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

[63] WEIN, A. S. (2021). Optimal low-degree hardness of maximum independent set. *Math. Stat. Learn.* **4** 221–251. MR4383734

[64] ZADIK, I., SONG, M. J., WEIN, A. S. and BRUNA, J. (2022). Lattice-based methods surpass sum-of-squares in clustering. In *Conference on Learning Theory* 1247–1248. PMLR.

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

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

The Annals of Statistics
2025, Vol. 53, No. 2, 802–821
https://doi.org/10.1214/24-AOS2486
© Institute of Mathematical Statistics, 2025

DEEP APPROXIMATE POLICY ITERATION

BY YULING JIAO[1,2,6,a], LICAN KANG[5,d], JIN LIU[3,b], XILIANG LU[1,4,6,c] AND JERRY ZHIJIAN YANG[1,5,6,e]

[1]*School of Mathematics and Statistics, Wuhan University,* [a]*yulingjiaomath@whu.edu.cn*
[2]*School of Artificial Intelligence, Wuhan University*
[3]*School of Data Science, The Chinese University of Hong Kong,* [b]*liujinlab@cuhk.edu.cn*
[4]*National Center for Applied Mathematics in Hubei, Wuhan University,* [c]*xllv.math@whu.edu.cn*
[5]*Institute for Math and AI, Wuhan University,* [d]*kanglican@whu.edu.cn*
[6]*Hubei Key Laboratory of Computational Science, Wuhan University,* [e]*zjyang.math@whu.edu.cn*

In this paper, we consider deep approximate policy iteration (DAPI) with the Bellman residual minimization in reinforcement learning. In each iteration of DAPI, we apply convolutional neural networks (CNNs) with ReLU activation, called ReLU CNNs, to estimate the fixed point of the Bellman equation by minimizing an unbiased minimax loss. To bound the estimation error in each iteration, we control the statistical and approximation errors using the tools of the empirical process theory with dependent data and deep approximation theory, respectively. We establish a novel statistical error bound for ReLU CNNs on dependent data that is $\mathcal{C}$-mixing, and an approximation error bound for ReLU CNNs on Hölder class. Combining with error propagation, we obtain a nonasymptotic error bound between the optimal action-value function Q^* and the estimated Q function induced by the greedy policy in DAPI. This bound depends on the sample size and ambient dimension of the data, as well as the size, weight bound, and depth of the CNNs, providing prior guidance on how to set these hyperparameters to achieve the desired convergence rate when training DAPI in practice. Moreover, this bound circumvents the curse of dimensionality if the distribution of state-action pairs is supported on a set with a low intrinsic dimension.

REFERENCES

ANDRYCHOWICZ, O. M., BAKER, B., CHOCIEJ, M., JOZEFOWICZ, R., MCGREW, B., PACHOCKI, J., PETRON, A., PLAPPERT, M., POWELL, G. et al. (2020). Learning dexterous in-hand manipulation. *Int. J. Robot. Res.* **39** 3–20.

ANTHONY, M. and BARTLETT, P. L. (1999). *Neural Network Learning: Theoretical Foundations*. Cambridge Univ. Press, Cambridge. MR1741038 https://doi.org/10.1017/CBO9780511624216

ANTOS, A., SZEPESVÁRI, C. and MUNOS, R. (2007). Fitted Q-iteration in continuous action-space MDPs. *Adv. Neural Inf. Process. Syst.* **20**.

ANTOS, A., SZEPESVÁRI, C. and MUNOS, R. (2008). Learning near-optimal policies with Bellman-residual minimization based fitted policy iteration and a single sample path. *Mach. Learn.* **71** 89–129.

BAGNELL, J. A. and SCHNEIDER, J. (2003). Covariant policy search. In *Proceedings of the* 18*th International Joint Conference on Artificial Intelligence* 1019–1024.

BAUER, B. and KOHLER, M. (2019). On deep learning as a remedy for the curse of dimensionality in nonparametric regression. *Ann. Statist.* **47** 2261–2285. MR3953451 https://doi.org/10.1214/18-AOS1747

BAXTER, J. and BARTLETT, P. L. (2001). Infinite-horizon policy-gradient estimation. *J. Artificial Intelligence Res.* **15** 319–350. MR1884081 https://doi.org/10.1613/jair.806

BERTSEKAS, D. and SHREVE, S. E. (1996). *Stochastic Optimal Control: The Discrete-Time Case* **5**. Athena Scientific, Belmont, MA.

BRADTKE, S. J. and BARTO, A. G. (1996). Linear least-squares algorithms for temporal difference learning. *Mach. Learn.* **22** 33–57.

MSC2020 subject classifications. Primary 62G05; secondary 68T07.
Key words and phrases. Approximate policy iteration, ReLU CNNs, $\mathcal{C}$-mixing, approximation error, low intrinsic dimension.

BRAKEL, D. B. P., GOYAL, K. X. A., PINEAU, R. L. J., COURVILLE, A. and BENGIO, Y. (2017). An actor-critic algorithm for sequence prediction. In *Open Review. Net. International Conference on Learning Representations* 1–17.

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

DEHLING, H. and PHILIPP, W. (2002). Empirical process techniques for dependent data. In *Empirical Process Techniques for Dependent Data* 3–113. Birkhäuser, Boston, MA. MR1958777 https://doi.org/10.1007/978-1-4612-0099-4_1

FALCONER, K. (2013). *Fractal Geometry*: *Mathematical Foundations and Applications*. Wiley, New York.

FAN, J., WANG, Z., XIE, Y. and YANG, Z. (2020). A theoretical analysis of deep Q-learning. In *Learning for Dynamics and Control* 486–489. PMLR.

FARAHMAND, A., GHAVAMZADEH, M., MANNOR, S. and SZEPESVÁRI, C. (2008). Regularized policy iteration. *Adv. Neural Inf. Process. Syst.* **21**.

FARAHMAND, A., GHAVAMZADEH, M., SZEPESVÁRI, C. and MANNOR, S. (2016). Regularized policy iteration with nonparametric function spaces. *J. Mach. Learn. Res.* **17** 139. MR3555030

FARAHMAND, A.-M., SZEPESVÁRI, C. and MUNOS, R. (2010). Error propagation for approximate policy and value iteration. In *Advances in Neural Information Processing Systems* **23**.

FARRELL, M. H., LIANG, T. and MISRA, S. (2021). Deep neural networks for estimation and inference. *Econometrica* **89** 181–213. MR4220387 https://doi.org/10.3982/ecta16901

FEFFERMAN, C. (2006). Whitney's extension problem for C^m. *Ann. of Math.* (2) **164** 313–359. MR2233850 https://doi.org/10.4007/annals.2006.164.313

FRANÇOIS-LAVET, V., HENDERSON, P., ISLAM, R., BELLEMARE, M. G., PINEAU, J. et al. (2018). An introduction to deep reinforcement learning. *Found. Trends Mach. Learn.* **11** 219–354.

GEIST, M., SCHERRER, B. and PIETQUIN, O. (2019). A theory of regularized Markov decision processes. In *International Conference on Machine Learning* 2160–2169. PMLR.

HANG, H. and STEINWART, I. (2017). A Bernstein-type inequality for some mixing processes and dynamical systems with an application to learning. *Ann. Statist.* **45** 708–743. MR3650398 https://doi.org/10.1214/16-AOS1465

HENDERSON, P., ISLAM, R., BACHMAN, P., PINEAU, J., PRECUP, D. and MEGER, D. (2018). Deep reinforcement learning that matters. In *Proceedings of the AAAI Conference on Artificial Intelligence* **32**.

HOWARD, R. A. (1960). Dynamic programming and Markov processes.

JIAO, Y., KANG, L., LIU, J., LU, X. and YANG, J. Z. (2025). Supplement to "Deep Approximate Policy Iteration." https://doi.org/10.1214/24-AOS2486SUPP

JIAO, Y., SHEN, G., LIN, Y. and HUANG, J. (2023). Deep nonparametric regression on approximate manifolds: Nonasymptotic error bounds with polynomial prefactors. *Ann. Statist.* **51** 691–716. MR4600998 https://doi.org/10.1214/23-aos2266

JUNG, T. and POLANI, D. (2006). Least squares SVM for least squares TD learning. In *ECAI* 499–503. Citeseer.

KAELBLING, L. P., LITTMAN, M. L. and MOORE, A. W. (1996). Reinforcement learning: A survey. *J. Artificial Intelligence Res.* **4** 237–285.

KAKADE, S. and LANGFORD, J. (2002). Approximately optimal approximate reinforcement learning. In *Proceedings of the Nineteenth International Conference on Machine Learning* 267–274.

KAKADE, S. M. (2001). A natural policy gradient. *Adv. Neural Inf. Process. Syst.* **14**.

KIRAN, B. R., SOBH, I., TALPAERT, V., MANNION, P., AL SALLAB, A. A., YOGAMANI, S. and PÉREZ, P. (2021). Deep reinforcement learning for autonomous driving: A survey. *IEEE Trans. Intell. Transp. Syst.*.

KLARTAG, B. and MENDELSON, S. (2005). Empirical processes and random projections. *J. Funct. Anal.* **225** 229–245. MR2149924 https://doi.org/10.1016/j.jfa.2004.10.009

KOHLER, M., KRZYŻAK, A. and WALTER, B. (2022). On the rate of convergence of image classifiers based on convolutional neural networks. *Ann. Inst. Statist. Math.* **74** 1085–1108. MR4498394 https://doi.org/10.1007/s10463-022-00828-4

KOHLER, M. and LANGER, S. (2020). Statistical theory for image classification using deep convolutional neural networks with cross-entropy loss. arXiv preprint. Available at arXiv:2011.13602.

KOLTER, J. Z. and NG, A. Y. (2009). Regularization and feature selection in least-squares temporal difference learning. In *Proceedings of the* 26*th Annual International Conference on Machine Learning* 521–528.

LAGOUDAKIS, M. G. and PARR, R. (2004). Least-squares policy iteration. *J. Mach. Learn. Res.* **4** 1107–1149. MR2125347 https://doi.org/10.1162/1532443041827907

LAZARIC, A., GHAVAMZADEH, M. and MUNOS, R. (2016). Analysis of classification-based policy iteration algorithms. *J. Mach. Learn. Res.* **17** 19. MR3491113

LEVINE, S., FINN, C., DARRELL, T. and ABBEEL, P. (2016). End-to-end training of deep visuomotor policies. *J. Mach. Learn. Res.* **17** 39. MR3491133

LEVINE, S., PASTOR, P., KRIZHEVSKY, A., IBARZ, J. and QUILLEN, D. (2018). Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. *Int. J. Robot. Res.* **37** 421–436.
LI, Y. (2017). Deep reinforcement learning: An overview. arXiv preprint. Available at arXiv:1701.07274.
LIU, H., CHEN, M., ZHAO, T. and LIAO, W. (2021). Besov function approximation and binary classification on low-dimensional manifolds using convolutional residual networks. In *International Conference on Machine Learning* 6770–6780. PMLR.
LU, J., SHEN, Z., YANG, H. and ZHANG, S. (2021). Deep network approximation for smooth functions. *SIAM J. Math. Anal.* **53** 5465–5506. MR4319100 https://doi.org/10.1137/20M134695X
MASSOUD FARAHMAND, A., GHAVAMZADEH, M., SZEPESVÁRI, C. and MANNOR, S. (2009). Regularized fitted Q-iteration for planning in continuous-space Markovian decision problems. In 2009 *American Control Conference* 725–730. IEEE, Piscatawa, NJ.
MAUME-DESCHAMPS, V. (2006). Exponential inequalities and functional estimations for weak dependent data; applications to dynamical systems. *Stoch. Dyn.* **6** 535–560. MR2285515 https://doi.org/10.1142/S0219493706001876
MNIH, V., KAVUKCUOGLU, K., SILVER, D., RUSU, A. A., VENESS, J., BELLEMARE, M. G., GRAVES, A., RIEDMILLER, M., FIDJELAND, A. K. et al. (2015). Human-level control through deep reinforcement learning. *Nature* **518** 529–533.
MOHRI, M. and ROSTAMIZADEH, A. (2008). Rademacher complexity bounds for non-IID processes. *Adv. Neural Inf. Process. Syst.* **21** 1097–1104.
MOHRI, M. and ROSTAMIZADEH, A. (2010). Stability bounds for stationary ϕ-mixing and β-mixing processes. *J. Mach. Learn. Res.* **11** 789–814. MR2600630
MUNOS, R. (2003). Error bounds for approximate policy iteration. In *ICML* **3** 560–567.
MUNOS, R. and SZEPESVÁRI, C. (2008). Finite-time bounds for fitted value iteration. *J. Mach. Learn. Res.* **9** 815–857. MR2417255
MURPHY, K. P. (2022). *Probabilistic Machine Learning: An Introduction*. MIT Press, Cambridge.
MURPHY, S. A. (2005). A generalization error for Q-learning. *J. Mach. Learn. Res.* **6** 1073–1097. MR2249849
NAKADA, R. and IMAIZUMI, M. (2020). Adaptive approximation and generalization of deep neural network with intrinsic dimensionality. *J. Mach. Learn. Res.* **21** 174. MR4209460
OONO, K. and SUZUKI, T. (2019). Approximation and non-parametric estimation of ResNet-type convolutional neural networks. In *International Conference on Machine Learning* 4922–4931. PMLR.
PETERSEN, P. and VOIGTLAENDER, F. (2018). Optimal approximation of piecewise smooth functions using deep ReLU neural networks. *Neural Netw.* **108** 296–330. https://doi.org/10.1016/j.neunet.2018.08.019
PETERSEN, P. and VOIGTLAENDER, F. (2020). Equivalence of approximation by convolutional neural networks and fully-connected networks. *Proc. Amer. Math. Soc.* **148** 1567–1581. MR4069195 https://doi.org/10.1090/proc/14789
RALAIVOLA, L. and AMINI, M.-R. (2015). Entropy-based concentration inequalities for dependent variables. In *International Conference on Machine Learning* 2436–2444. PMLR.
RANZATO, M., CHOPRA, S., AULI, M. and ZAREMBA, W. (2015). Sequence level training with recurrent neural networks. arXiv preprint. Available at arXiv:1511.06732.
ROY, A., BALASUBRAMANIAN, K. and ERDOGDU, M. A. (2021). On empirical risk minimization with dependent and heavy-tailed data. *Adv. Neural Inf. Process. Syst.* **34**.
SALLAB, A. E., ABDOU, M., PEROT, E. and YOGAMANI, S. (2017). Deep reinforcement learning framework for autonomous driving. *Electron. Imaging* **2017** 70–76.
SCHERRER, B., GHAVAMZADEH, M., GABILLON, V., LESNER, B. and GEIST, M. (2015). Approximate modified policy iteration and its application to the game of Tetris. *J. Mach. Learn. Res.* **16** 1629–1676. MR3417794
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
SCHULMAN, J., LEVINE, S., ABBEEL, P., JORDAN, M. and MORITZ, P. (2015). Trust region policy optimization. In *International Conference on Machine Learning* 1889–1897. PMLR.
SHALEV-SHWARTZ, S., SHAMMAH, S. and SHASHUA, A. (2016). Safe, multi-agent, reinforcement learning for autonomous driving. arXiv preprint. Available at arXiv:1610.03295.
SHEN, Z., YANG, H. and ZHANG, S. (2019). Nonlinear approximation via compositions. *Neural Netw.* **119** 74–84.
SHEN, Z., YANG, H. and ZHANG, S. (2020). Deep network approximation characterized by number of neurons. *Commun. Comput. Phys.* **28** 1768–1811. MR4188521 https://doi.org/10.4208/cicp.oa-2020-0149
SILVER, D., HUANG, A., MADDISON, C. J., GUEZ, A., SIFRE, L., VAN DEN DRIESSCHE, G., SCHRITTWIESER, J., ANTONOGLOU, I., PANNEERSHELVAM, V. et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature* **529** 484–489.
STEINWART, I., HUSH, D. and SCOVEL, C. (2009). Learning from dependent observations. *J. Multivariate Anal.* **100** 175–194. MR2460486 https://doi.org/10.1016/j.jmva.2008.04.001

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

TAYLOR, G. and PARR, R. (2009). Kernelized value function approximation for reinforcement learning. In *Proceedings of the* 26*th Annual International Conference on Machine Learning* 1017–1024.

TOSATTO, S., PIROTTA, M., D'ERAMO, C. and RESTELLI, M. (2017). Boosted fitted q-iteration. In *International Conference on Machine Learning* 3434–3443. PMLR.

XIE, T. and JIANG, N. (2020). Q* approximation schemes for batch reinforcement learning: A theoretical comparison. In *Conference on Uncertainty in Artificial Intelligence* 550–559. PMLR.

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

XU, X., HU, D. and LU, X. (2007). Kernel-based least squares policy iteration for reinforcement learning. *IEEE Trans*. *Neural Netw.* **18** 973–992.

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

YAROTSKY, D. (2018). Optimal approximation of continuous functions by very deep ReLU networks. In *Conference on Learning Theory* 639–649. PMLR.

YU, B. (1994). Rates of convergence for empirical processes of stationary mixing sequences. *Ann*. *Probab*. **22** 94–116. MR1258867

ZHOU, D.-X. (2020a). Theory of deep convolutional neural networks: Downsampling. *Neural Netw.* **124** 319–327.

ZHOU, D.-X. (2020b). Universality of deep convolutional neural networks. *Appl*. *Comput*. *Harmon*. *Anal*. **48** 787–794. MR4047545 https://doi.org/10.1016/j.acha.2019.06.004

The Annals of Statistics
2025, Vol. 53, No. 2, 822–853
https://doi.org/10.1214/25-AOS2489
© Institute of Mathematical Statistics, 2025

THE GENERALIZATION ERROR OF MAX-MARGIN LINEAR CLASSIFIERS: BENIGN OVERFITTING AND HIGH DIMENSIONAL ASYMPTOTICS IN THE OVERPARAMETRIZED REGIME

BY ANDREA MONTANARI[1,a], FENG RUAN[2,b], YOUNGTAK SOHN[3,c] AND JUN YAN[4,d]

[1]*Department of Statistics and Department of Electrical Engineering, Stanford University,* [a]*montanar@stanford.edu*
[2]*Department of Statistics and Data Science, Northwestern University,* [b]*fengruan@northwestern.edu*
[3]*Department of Mathematics, Massachusetts Institute of Technology,* [c]*youngtak@mit.edu*
[4]*D. E. Shaw & Co., L.P.,* [d]*yanjun9306@gmail.com*

Modern machine learning classifiers often exhibit vanishing classification error on the training set. They achieve this by learning nonlinear representations of the inputs that map the data into linearly separable classes.

Motivated by these phenomena, we revisit high-dimensional maximum margin classification for linearly separable data. We consider a stylized setting in which data $(y_i, \boldsymbol{x}_i)$, $i \le n$ are i.i.d. with $\boldsymbol{x}_i \sim \mathsf{N}(\mathbf{0}, \boldsymbol{\Sigma})$ a p-dimensional Gaussian feature vector, and $y_i \in \{+1, -1\}$ a label whose distribution depends on a linear combination of the covariates $\langle \boldsymbol{\theta}_*, \boldsymbol{x}_i \rangle$. Recent universality results can be used to show that the results derived in the Gaussian setting also apply when $\boldsymbol{x}_i = \boldsymbol{\phi}(z_i)$ for standard Gaussian z_i and $\boldsymbol{\phi}$ a nonlinear featurization map.

We consider the proportional asymptotics $n, p \to \infty$ with $p/n \to \psi$, and derive exact expressions for the limiting generalization error. We use this theory to derive two results of independent interest: (i) Sufficient conditions on $(\boldsymbol{\Sigma}, \boldsymbol{\theta}_*)$ for "benign overfitting" that are parallel to previously derived conditions in the case of linear regression; (ii) An asymptotically exact expression for the generalization error when max-margin classification is used in conjunction with feature vectors produced by random one-layer neural networks.

REFERENCES

[1] AMELUNXEN, D., LOTZ, M., MCCOY, M. B. and TROPP, J. A. (2014). Living on the edge: Phase transitions in convex programs with random data. *Inf. Inference* **3** 224–294. MR3311453 https://doi.org/10.1093/imaiai/iau005

[2] ANTHONY, M. and BARTLETT, P. L. (2009). *Neural Network Learning*: *Theoretical Foundations*. Cambridge Univ. Press, Cambridge. MR1741038 https://doi.org/10.1017/CBO9780511624216

[3] ARORA, S., COHEN, N., HU, W. and LUO, Y. (2019). Implicit regularization in deep matrix factorization. Preprint. Available at arXiv:1905.13655.

[4] BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2567175 https://doi.org/10.1007/978-1-4419-0661-8

[5] BALCAN, M.-F., BLUM, A. and VEMPALA, S. (2004). Kernels as features: On kernels, margins, and low-dimensional mappings. *Mach*. *Learn*. **65** 79–94.

[6] BARBIER, J., DIA, M., MACRIS, N., KRZAKALA, F. and ZDEBOROVÁ, L. (2018). Rank-one matrix estimation: Analysis of algorithmic and information theoretic limits by the spatial coupling method. Preprint. Available at arXiv:1812.02537.

[7] BARBIER, J., KRZAKALA, F., MACRIS, N., MIOLANE, L. and ZDEBOROVÁ, L. (2019). Optimal errors and phase transitions in high-dimensional generalized linear models. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **116** 5451–5460. MR3939767 https://doi.org/10.1073/pnas.1802705116

[8] BARBIER, J. and MACRIS, N. (2019). The adaptive interpolation method: A simple scheme to prove replica formulas in Bayesian inference. *Probab. Theory Related Fields* **174** 1133–1185. MR3980313 https://doi.org/10.1007/s00440-018-0879-0

MSC2020 subject classifications. Primary 62H30, 62J12; secondary 62F12.
Key words and phrases. Max-margin classification, high-dimensional asymptotics, benign overfitting, overparametrization.

[9] BARTLETT, P. L. (1998). The sample complexity of pattern classification with neural networks: The size of the weights is more important than the size of the network. *IEEE Trans. Inf. Theory* **44** 525–536. MR1607706 https://doi.org/10.1109/18.661502

[10] 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

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

[12] 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

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

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

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

[16] 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

[17] BELKIN, M., HSU, D. and XU, J. (2020). Two models of double descent for weak features. *SIAM J. Math. Data Sci.* **2** 1167–1180. MR4186534 https://doi.org/10.1137/20M1336072

[18] BELKIN, M., HSU, D. J. and MITRA, P. (2018). Overfitting or perfect fitting? Risk bounds for classification and regression rules that interpolate. *Adv. Neural Inf. Process. Syst.* 2300–2311.

[19] BELKIN, M., MA, S. and MANDAL, S. (2018). To understand deep learning we need to understand kernel learning. Preprint. Available at arXiv:1802.01396.

[20] CANDÈS, E. J. and SUR, P. (2020). The phase transition for the existence of the maximum likelihood estimate in high-dimensional logistic regression. *Ann. Statist.* **48** 27–42. MR4065151 https://doi.org/10.1214/18-AOS1789

[21] CHANDRASEKARAN, V., RECHT, B., PARRILO, P. A. and WILLSKY, A. S. (2012). The convex geometry of linear inverse problems. *Found. Comput. Math.* **12** 805–849. MR2989474 https://doi.org/10.1007/s10208-012-9135-7

[22] CHATTERJI, N. S. and LONG, P. M. (2021). Finite-sample analysis of interpolating linear classifiers in the overparameterized regime. *J. Mach. Learn. Res.* **22** Paper No. 129, 30 pp. MR4279780

[23] CHENG, C. and MONTANARI, A. (2024). Dimension free ridge regression. *Ann. Statist.* **52** 2879–2912. MR4842830 https://doi.org/10.1214/24-aos2449

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

[25] CHIZAT, L. and BACH, F. (2018). A note on lazy training in supervised differentiable programming. Preprint. Available at arXiv:1812.07956.

[26] COVER, T. M. (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Trans. Electron. Comput.* **3** 326–334.

[27] 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

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

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

[30] DONOHO, D. L., JOHNSTONE, I. and MONTANARI, A. (2013). Accurate prediction of phase transitions in compressed sensing via a connection to minimax denoising. *IEEE Trans. Inf. Theory* **59** 3396–3433. MR3061255 https://doi.org/10.1109/TIT.2013.2239356

[31] DU, S. S., ZHAI, X., POCZOS, B. and SINGH, A. (2018). Gradient descent provably optimizes over-parameterized neural networks. Preprint. Available at arXiv:1810.02054.

[32] 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

[33] 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.

[34] ENGEL, A. and VAN DEN BROECK, C. (2001). *Statistical Mechanics of Learning*. Cambridge Univ. Press, Cambridge. MR1827275 https://doi.org/10.1017/CBO9781139164542

[35] FAN, Z. and MONTANARI, A. (2019). The spectral norm of random inner-product kernel matrices. *Probab. Theory Related Fields* **173** 27–85. MR3916104 https://doi.org/10.1007/s00440-018-0830-4

[36] FREI, S., CHATTERJI, N. S. and BARTLETT, P. (2022). Benign overfitting without linearity: Neural network classifiers trained by gradient descent for noisy linear data. In *Conference on Learning Theory* 2668–2703. PMLR.

[37] FREI, S., VARDI, G., BARTLETT, P. L. and SREBRO, N. (2023). Benign overfitting in linear classifiers and leaky relu networks from kkt conditions for margin maximization. Preprint. Available at arXiv:2303.01462.

[38] GARDNER, E. (1988). The space of interactions in neural network models. *J. Phys. A* **21** 257–270. MR0939730

[39] GEIGER, M., JACOT, A., SPIGLER, S., GABRIEL, F., SAGUN, L., D'ASCOLI, S., BIROLI, G., HONGLER, C. and WYART, M. (2020). Scaling description of generalization with number of parameters in deep learning. *J. Stat. Mech. Theory Exp.* **2020** 023401, 23 pp. MR4137317 https://doi.org/10.1088/1742-5468/ab633c

[40] GERACE, F., LOUREIRO, B., KRZAKALA, F., MÉZARD, M. and ZDEBOROVÁ, L. (2021). Generalisation error in learning with random features and the hidden manifold model. *J. Stat. Mech. Theory Exp.* **2021** Paper No. 124013, 40 pp. MR4412841 https://doi.org/10.1088/1742-5468/ac3ae6

[41] GOLDT, S., MEZARD, M., KRZAKALA, F. and ZDEBOROVÁ, L. Modelling the influence of data structure on learning in neural networks: the hidden manifold model.

[42] 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

[43] GUNASEKAR, S., LEE, J., SOUDRY, D. and SREBRO, N. (2018). Characterizing implicit bias in terms of optimization geometry. Preprint. Available at arXiv:1802.08246.

[44] GUNASEKAR, S., LEE, J. D., SOUDRY, D. and SREBRO, N. (2018). Implicit bias of gradient descent on linear convolutional networks. *Adv. Neural Inf. Process. Syst.* 9461–9471.

[45] 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

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

[47] HSU, D., MUTHUKUMAR, V. and XU, J. (2022). On the proliferation of support vectors in high dimensions. *J. Stat. Mech. Theory Exp.* **2022** Paper No. 114011, 27 pp. MR4535582 https://doi.org/10.1088/1742-5468/ac98a9

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

[49] JACOT, A., GABRIEL, F. and HONGLER, C. (2018). Neural tangent kernel: Convergence and generalization in neural networks. *Adv. Neural Inf. Process. Syst.* 8571–8580.

[50] 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

[51] KAKADE, S. M., SRIDHARAN, K. and TEWARI, A. (2009). On the complexity of linear prediction: Risk bounds, margin bounds, and regularization. *Adv. Neural Inf. Process. Syst.* 793–800.

[52] KINI, G. R., PARASKEVAS, O., OYMAK, S. and THRAMPOULIDIS, C. (2021). Label-imbalanced and group-sensitive classification under overparameterization. *Adv. Neural Inf. Process. Syst.* **34** 18970–18983.

[53] KINI, G. R. and THRAMPOULIDIS, C. (2020). Analytic study of double descent in binary classification: The impact of loss. In 2020 *IEEE International Symposium on Information Theory* (*ISIT*) 2527–2532. IEEE.

[54] KOEHLER, F., ZHOU, L., SUTHERLAND, D. J. and SREBRO, N. (2021). Uniform convergence of interpolators: Gaussian width, norm bounds and benign overfitting. *Adv. Neural Inf. Process. Syst.* **34** 20657–20668.

[55] KOLTCHINSKII, V. (2011). *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems*. *Lecture Notes in Math.* **2033**. Springer, Heidelberg. MR2829871 https://doi.org/10.1007/978-3-642-22147-7

[56] KOLTCHINSKII, V. and PANCHENKO, D. (2002). Empirical margin distributions and bounding the generalization error of combined classifiers. *Ann. Statist.* **30** 1–50. MR1892654 https://doi.org/10.1214/aos/1015362182

[57] LECUÉ, G., SHANG, Z. and SHANG, Z. (2022). A geometrical viewpoint on the benign overfitting property of the minimum l_2-norm interpolant estimator and its universality. Preprint. Available at arXiv:2203.05873.

[58] LELARGE, M. and MIOLANE, L. (2019). Fundamental limits of symmetric low-rank matrix estimation. *Probab. Theory Related Fields* **173** 859–929. MR3936148 https://doi.org/10.1007/s00440-018-0845-x

[59] LI, Y., MA, T. and ZHANG, H. (2017). Algorithmic regularization in over-parameterized matrix sensing and neural networks with quadratic activations. Preprint. Available at arXiv:1712.09203.

[60] LIANG, T. and RAKHLIN, A. (2020). Just interpolate: Kernel "ridgeless" regression can generalize. *Ann. Statist.* **48** 1329–1347. MR4124325 https://doi.org/10.1214/19-AOS1849

[61] 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

[62] 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

[63] 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

[64] MÉZARD, M. and MONTANARI, A. (2009). *Information, Physics, and Computation. Oxford Graduate Texts*. Oxford Univ. Press, Oxford. MR2518205 https://doi.org/10.1093/acprof:oso/9780198570837.001.0001

[65] MÉZARD, M., PARISI, G. and VIRASORO, M. A. (1987). *Spin Glass Theory and Beyond. World Scientific Lecture Notes in Physics* **9**. World Scientific Co., Inc., Teaneck, NJ. MR1026102

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

[67] MONTANARI, A. (2018). Mean field asymptotics in high-dimensional statistics: From exact results to efficient algorithms. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro* 2018. *Vol. IV. Invited Lectures* 2973–2994. World Sci. Publ., Hackensack, NJ. MR3966519

[68] MONTANARI, A., RUAN, F., SAEED, B. and SOHN, Y. (2023). Universality of max-margin classifiers. Preprint. Available at arXiv:2310.00176.

[69] MONTANARI, A., RUAN, F., SOHN, Y. and YAN, J. (2025). Supplement to "The generalization error of max-margin linear classifiers: Benign overfitting and high dimensional asymptotics in the overparametrized regime." https://doi.org/10.1214/25-AOS2489SUPP

[70] MONTANARI, A. and SAEED, B. N. (2022). Universality of empirical risk minimization. In *Conference on Learning Theory* 4310–4312. PMLR.

[71] MONTANARI, A. and ZHONG, Y. (2022). The interpolation phase transition in neural networks: Memorization and generalization under lazy training. *Ann. Statist.* **50** 2816–2847. MR4500626 https://doi.org/10.1214/22-aos2211

[72] MUTHUKUMAR, V., NARANG, A., SUBRAMANIAN, V., BELKIN, M., HSU, D. and SAHAI, A. (2021). Classification vs regression in overparameterized regimes: Does the loss function matter? *J. Mach. Learn. Res.* **22** Paper No. 222, 69 pp. MR4329801

[73] MUTHUKUMAR, V., VODRAHALLI, K. and SAHAI, A. (2019). Harmless interpolation of noisy data in regression. Preprint. Available at arXiv:1903.09139.

[74] NEYSHABUR, B., BHOJANAPALLI, S., MCALLESTER, D. and SREBRO, N. (2017). Exploring generalization in deep learning. *Adv. Neural Inf. Process. Syst.* **30**.

[75] OYMAK, S., THRAMPOULIDIS, C. and HASSIBI, B. The squared-error of generalized lasso: A precise analysis. In 2013 51*st Annual Allerton Conference on Communication, Control, and Computing* (*Allerton*) 1002–1009.

[76] RADFORD, M. N. (1996). Priors for infinite networks. In *Bayesian Learning for Neural Networks* 29–53. Springer, Berlin.

[77] RAHIMI, A. and RECHT, B. (2008). Random features for large-scale kernel machines. *Adv. Neural Inf. Process. Syst.* 1177–1184.

[78] RAKHLIN, A. and ZHAI, X. (2018). Consistency of interpolation with Laplace kernels is a high-dimensional phenomenon. Preprint. Available at arXiv:1812.11167.

[79] SALEHI, F., ABBASI, E. and HASSIBI, B. (2019). The impact of regularization on high-dimensional logistic regression. Preprint. Available at arXiv:1906.03761.

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

[81] SHCHERBINA, M. and TIROZZI, B. (2003). Rigorous solution of the Gardner problem. *Comm. Math. Phys.* **234** 383–422. MR1964377 https://doi.org/10.1007/s00220-002-0783-3

[82] 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** Paper No. 70, 57 pp. MR3899772

[83] STOJNIC, M. (2010). ℓ_1 optimization and its various thresholds in compressed sensing. In 2010 *IEEE International Conference on Acoustics, Speech and Signal Processing* 3910–3913.

[84] STOJNIC, M. (2013). A framework to characterize performance of lasso algorithms. Preprint. Available at arXiv:1303.7291.

[85] SUR, P. and CANDÈS, E. J. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. *Proc. Natl. Acad. Sci. USA* **116** 14516–14525. MR3984492 https://doi.org/10.1073/pnas.1810420116

[86] TAHERI, H., PEDARSANI, R. and THRAMPOULIDIS, C. (2020). Fundamental limits of ridge-regularized empirical risk minimization in high dimensions. Preprint. Available at arXiv:2006.08917.

[87] 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

[88] 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.

[89] TSIGLER, A. and BARTLETT, P. L. (2023). Benign overfitting in ridge regression. *J. Mach. Learn. Res.* **24** Paper No. [123], 76 pp. MR4583284

[90] VERSHYNIN, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. *Cambridge Series in Statistical and Probabilistic Mathematics* **47**. Cambridge Univ. Press, Cambridge. MR3837109 https://doi.org/10.1017/9781108231596

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

[92] WAHBA, G. (2002). Soft and hard classification by reproducing kernel Hilbert space methods. *Proc. Natl. Acad. Sci. USA* **99** 16524–16530. MR1947755 https://doi.org/10.1073/pnas.242574899

[93] ZHANG, C., BENGIO, S., HARDT, M., RECHT, B. and VINYALS, O. (2016). Understanding deep learning requires rethinking generalization. Preprint. Available at arXiv:1611.03530.

[94] ZHOU, L., KOEHLER, F., SUR, P., SUTHERLAND, D. J. and SREBRO, N. (2022). A non-asymptotic Moreau envelope theory for high-dimensional generalized linear models. Preprint. Available at arXiv:2210.12082.

The Annals of Statistics
2025, Vol. 53, No. 2, 854–878
https://doi.org/10.1214/25-AOS2490
© Institute of Mathematical Statistics, 2025

PRECISE ERROR RATES FOR COMPUTATIONALLY EFFICIENT TESTING

BY ANKUR MOITRA[1,a] AND ALEXANDER S. WEIN[2,b]

[1]*Department of Mathematics and CSAIL, Massachusetts Institute of Technology,* [a]*moitra@mit.edu*
[2]*Department of Mathematics, University of California, Davis,* [b]*aswein@ucdavis.edu*

We revisit the fundamental question of simple-versus-simple hypothesis testing with an eye toward computational complexity, as the statistically optimal likelihood ratio test is often computationally intractable in high-dimensional settings. In the classical spiked Wigner model with a general i.i.d. spike prior, we show (conditional on a conjecture) that an existing test based on linear spectral statistics achieves the best possible trade-off curve between type-I and type-II error rates among all computationally efficient tests, even though there are exponential-time tests that do better. This result is conditional on an appropriate complexity-theoretic conjecture, namely a natural strengthening of the well-established low-degree conjecture. Our result shows that the spectrum is a sufficient statistic for computationally bounded tests (but not for all tests).

To our knowledge, our approach gives the first tool for reasoning about the precise asymptotic testing error achievable with efficient computation. The main ingredients required for our hardness result are a sharp bound on the norm of the low-degree likelihood ratio along with (counterintuitively) a positive result on achievability of testing. This strategy appears to be new even in the setting of unbounded computation, in which case it gives an alternate way to analyze the fundamental statistical limits of testing.

REFERENCES

[1] AIZENMAN, M., LEBOWITZ, J. L. and RUELLE, D. (1987). Some rigorous results on the Sherrington–Kirkpatrick spin glass model. *Comm. Math. Phys.* **112** 3–20. MR0904135

[2] ANDERSON, G. W. and ZEITOUNI, O. (2006). A CLT for a band matrix model. *Probab. Theory Related Fields* **134** 283–338. MR2222385 https://doi.org/10.1007/s00440-004-0422-3

[3] BAI, Z., WANG, X. and ZHOU, W. (2009). CLT for linear spectral statistics of Wigner matrices. *Electron. J. Probab.* **14** 2391–2417. MR2556016 https://doi.org/10.1214/EJP.v14-705

[4] BAI, Z. and YAO, J. (2008). Central limit theorems for eigenvalues in a spiked population model. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 447–474. MR2451053 https://doi.org/10.1214/07-AIHP118

[5] BAI, Z. D. and SILVERSTEIN, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** 553–605. MR2040792 https://doi.org/10.1214/aop/1078415845

[6] 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

[7] BAIK, J. and LEE, J. O. (2016). Fluctuations of the free energy of the spherical Sherrington-Kirkpatrick model. *J. Stat. Phys.* **165** 185–224. MR3554380 https://doi.org/10.1007/s10955-016-1610-0

[8] BAIK, J. and LEE, J. O. (2017). Fluctuations of the free energy of the spherical Sherrington-Kirkpatrick model with ferromagnetic interaction. *Ann. Henri Poincaré* **18** 1867–1917. MR3649446 https://doi.org/10.1007/s00023-017-0562-5

[9] BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. MR2279680 https://doi.org/10.1016/j.jmva.2005.08.003

[10] BANDEIRA, A. S., BANKS, J., KUNISKY, D., MOORE, C. and WEIN, A. S. (2021). Spectral planting and the hardness of refuting cuts, colorability, and communities in random graphs. In *Conference on Learning Theory* 410–473. PMLR.

MSC2020 subject classifications. Primary 62F03; secondary 68Q17.
Key words and phrases. Hypothesis testing, spiked Wigner model, low-degree likelihood ratio.

[11] BANDEIRA, A. S., EL ALAOUI, A., HOPKINS, S., SCHRAMM, T., WEIN, A. S. and ZADIK, I. (2022). The Franz-Parisi criterion and computational trade-offs in high dimensional statistics. *Adv. Neural Inf. Process. Syst.* **35** 33831–33844.

[12] BANDEIRA, A. S., KUNISKY, D. and WEIN, A. S. (2020). Computational hardness of certifying bounds on constrained PCA problems. In 11*th Innovations in Theoretical Computer Science Conference* (*ITCS* 2020). Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

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

[14] BANKS, J., MOORE, C., VERSHYNIN, R., VERZELEN, N. and XU, J. (2018). Information-theoretic bounds and phase transitions in clustering, sparse PCA, and submatrix localization. *IEEE Trans. Inf. Theory* **64** 4872–4994. MR3819345 https://doi.org/10.1109/tit.2018.2810020

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

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

[17] BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** 494–521. MR2782201 https://doi.org/10.1016/j.aim.2011.02.007

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

[19] BLUM, L., SHUB, M. and SMALE, S. (1989). On a theory of computation and complexity over the real numbers: NP-completeness, recursive functions and universal machines. *Bull. Amer. Math. Soc.* (*N.S.*) **21** 1–46. MR0974426 https://doi.org/10.1090/S0273-0979-1989-15750-9

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

[21] CAPITAINE, M., DONATI-MARTIN, C. and FÉRAL, D. (2009). The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations. *Ann. Probab.* **37** 1–47. MR2489158 https://doi.org/10.1214/08-AOP394

[22] CHUNG, H. W. and LEE, J. O. (2022). Weak detection in the spiked Wigner model. *IEEE Trans. Inf. Theory* **68** 7427–7453. MR4524648 https://doi.org/10.1109/tit.2022.3185232

[23] DECELLE, A., KRZAKALA, F., MOORE, C. and ZDEBOROVÁ, L. (2011). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* **84** 066106.

[24] DIAKONIKOLAS, I. and KANE, D. (2022). Non-Gaussian component analysis via lattice basis reduction. In *Conference on Learning Theory* 4535–4547. PMLR.

[25] DING, X. and WANG, Z. (2023). Global and local CLTs for linear spectral statistics of general sample covariance matrices when the dimension is much larger than the sample size with applications. Preprint. Available at arXiv:2308.08646.

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

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

[28] EL ALAOUI, A., KRZAKALA, F. and JORDAN, M. (2020). Fundamental limits of detection in the spiked Wigner model. *Ann. Statist.* **48** 863–885. MR4102679 https://doi.org/10.1214/19-AOS1826

[29] FELDMAN, V., GRIGORESCU, E., REYZIN, L., VEMPALA, S. S. and XIAO, Y. (2017). Statistical algorithms and a lower bound for detecting planted cliques. *J. ACM* **64** Art. 8, 37 pp. MR3664576 https://doi.org/10.1145/3046674

[30] FÉRAL, D. and PÉCHÉ, S. (2007). The largest eigenvalue of rank one deformation of large Wigner matrices. *Comm. Math. Phys.* **272** 185–228. MR2291807 https://doi.org/10.1007/s00220-007-0209-3

[31] FLETCHER, A. K. and RANGAN, S. (2018). Iterative reconstruction of rank-one matrices in noise. *Inf. Inference* **7** 531–562. MR3858334 https://doi.org/10.1093/imaiai/iax014

[32] GAMARNIK, D., JAGANNATH, A. and WEIN, A. S. (2020). Low-degree hardness of random optimization problems. In 2020 *IEEE* 61*st Annual Symposium on Foundations of Computer Science* 131–140. IEEE Computer Soc., Los Alamitos, CA. MR4232029 https://doi.org/10.1109/FOCS46700.2020.00021

[33] GAMARNIK, D., MOORE, C. and ZDEBOROVÁ, L. (2022). Disordered systems insights on computational hardness. *J. Stat. Mech. Theory Exp.* **2022** Paper No. 114015, 41 pp. MR4535586 https://doi.org/10.1088/1742-5468/ac9cc8

[34] GAMARNIK, D. and SUDAN, M. (2017). Limits of local algorithms over sparse random graphs. *Ann. Probab.* **45** 2353–2376. MR3693964 https://doi.org/10.1214/16-AOP1114

[35] HOLMGREN, J. and WEIN, A. S. (2021). Counterexamples to the low-degree conjecture. In 12*th Innovations in Theoretical Computer Science Conference* (*ITCS* 2021) **185**.

[36] HOPKINS, S. (2018). Statistical inference and the sum of squares method. Ph.D. Thesis, Cornell Univ. MR3864930

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

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

[39] JERRUM, M. (1992). Large cliques elude the Metropolis process. *Random Structures Algorithms* **3** 347–359. MR1179827 https://doi.org/10.1002/rsa.3240030402

[40] 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

[41] JUNG, J. H., CHUNG, H. W. and LEE, J. O. (2020). Weak detection in the spiked Wigner model with general rank. Preprint. Available at arXiv:2001.05676.

[42] JUNG, J. H., CHUNG, H. W. and LEE, J. O. (2021). Detection of signal in the spiked rectangular models. In *International Conference on Machine Learning* 5158–5167. PMLR.

[43] KOEHLER, F. and MOSSEL, E. (2021). Reconstruction on trees and low-degree polynomials. Preprint. Available at arXiv:2109.06915.

[44] KOTHARI, P., VEMPALA, S. S., WEIN, A. S. and XU, J. (2023). Is planted coloring easier than planted clique? In *The Thirty Sixth Annual Conference on Learning Theory* 5343–5372. PMLR.

[45] KUNISKY, D. (2021). Hypothesis testing with low-degree polynomials in the Morris class of exponential families. In *Conference on Learning Theory* 2822–2848. PMLR.

[46] KUNISKY, D., WEIN, A. S. and BANDEIRA, A. S. (2022). Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *Mathematical Analysis, Its Applications and Computation. Springer Proc. Math. Stat.* **385** 1–50. Springer, Cham. MR4461037 https://doi.org/10.1007/978-3-030-97127-4_1

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

[48] LESIEUR, T., KRZAKALA, F. and ZDEBOROVÁ, L. (2015). Phase transitions in sparse PCA. In *International Symposium on Information Theory* (*ISIT*) 1635–1639. IEEE.

[49] LIU, Z., HU, J., BAI, Z. and SONG, H. (2023). A CLT for the LSS of large-dimensional sample covariance matrices with diverging spikes. *Ann. Statist.* **51** 2246–2271. MR4678803 https://doi.org/10.1214/23-aos2333

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

[51] MAÏDA, M. (2007). Large deviations for the largest eigenvalue of rank one deformations of Gaussian ensembles. *Electron. J. Probab.* **12** 1131–1150. MR2336602 https://doi.org/10.1214/EJP.v12-438

[52] MIOLANE, L. (2018). Phase transitions in spiked matrix estimation: Information-theoretic analysis. Preprint. Available at arXiv:1806.04343.

[53] MOITRA, A. and WEIN, A. S. (2025). Supplement to "Precise error rates for computationally efficient testing." https://doi.org/10.1214/25-AOS2490SUPP

[54] MONTANARI, A., REICHMAN, D. and ZEITOUNI, O. (2017). On the limitation of spectral methods: From the Gaussian hidden clique problem to rank one perturbations of Gaussian tensors. *IEEE Trans. Inf. Theory* **63** 1572–1579. MR3625981 https://doi.org/10.1109/TIT.2016.2637959

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

[56] MONTANARI, A. and WEIN, A. S. (2022). Equivalence of approximate message passing and low-degree polynomials in rank-one matrix estimation. Preprint. Available at arXiv:2212.06996.

[57] 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

[58] ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2013). Asymptotic power of sphericity tests for high-dimensional data. *Ann. Statist.* **41** 1204–1231. MR3113808 https://doi.org/10.1214/13-AOS1100

[59] ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2014). Signal detection in high dimension: The multi-spiked case. *Ann. Statist.* **42** 225–254. MR3189485 https://doi.org/10.1214/13-AOS1181

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

[61] PÉCHÉ, S. (2006). The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probab. Theory Related Fields* **134** 127–173. MR2221787 https://doi.org/10.1007/s00440-005-0466-z

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

[63] PIZZO, A., RENFREW, D. and SOSHNIKOV, A. (2013). On finite rank deformations of Wigner matrices. *Ann*. *Inst*. *Henri Poincaré Probab*. *Stat*. **49** 64–94. MR3060148 https://doi.org/10.1214/11-AIHP459

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

[65] RUSH, C., SKERMAN, F., WEIN, A. S. and YANG, D. (2023). Is it easier to count communities than find them? In 14*th Innovations in Theoretical Computer Science Conference*. *LIPIcs*. *Leibniz Int*. *Proc*. *Inform*. **251** Art. No. 94, 23 pp. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern. MR4587108 https://doi.org/10.4230/lipics.itcs.2023.94

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

[67] WANG, Z. and YAO, J. (2023). Central limit theorem for linear spectral statistics of block-Wigner-type matrices. *Random Matrices Theory Appl*. **12** Paper No. 2350006, 57 pp. MR4672878 https://doi.org/10.1142/S2010326323500065

[68] ZADIK, I., SONG, M. J., WEIN, A. S. and BRUNA, J. (2022). Lattice-based methods surpass sum-of-squares in clustering. In *Conference on Learning Theory* 1247–1248. PMLR.

The Annals of Statistics
2025, Vol. 53, No. 2, 879–906
https://doi.org/10.1214/25-AOS2491
© Institute of Mathematical Statistics, 2025

ASYMPTOTIC EQUIVALENCE OF LOCALLY STATIONARY PROCESSES AND BIVARIATE GAUSSIAN WHITE NOISE

BY CRISTINA BUTUCEA[1,a], ALEXANDER MEISTER[2,b] AND ANGELIKA ROHDE[3,c]

[1]*CREST, ENSAE, Institut Polytechnique de Paris,* [a]*cristina.butucea@ensae.fr*
[2]*Institut für Mathematik, Universität Rostock,* [b]*alexander.meister@uni-rostock.de*
[3]*Mathematisches Institut, Albert-Ludwigs-Universität Freiburg,* [c]*angelika.rohde@stochastik.uni-freiburg.de*

We consider a general class of statistical experiments, in which an n-dimensional centered Gaussian random variable is observed and its covariance matrix is the parameter of interest. The covariance matrix is assumed to be well-approximable in a linear space of lower dimension K_n with eigenvalues uniformly bounded away from zero and infinity. We prove asymptotic equivalence of this experiment and a class of K_n-dimensional Gaussian models with informative expectation in Le Cam's sense when n tends to infinity and K_n is allowed to increase moderately in n at a polynomial rate. For this purpose we derive a new localization technique for non-i.i.d. data and a novel high-dimensional Central Limit Law in total variation distance. These results are key ingredients to show asymptotic equivalence between the experiments of locally stationary Gaussian time series and a bivariate Wiener process with the log spectral density as its drift. On this way a novel class of matrices is introduced which generalizes circulant Toeplitz matrices traditionally used for strictly stationary time series.

REFERENCES

AUE, A. and VAN DELFT, A. (2020). Testing for stationarity of functional time series in the frequency domain. *Ann. Statist.* **48** 2505–2547. MR4152111 https://doi.org/10.1214/19-AOS1895

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

BALLY, V., CARAMELLINO, L. and POLY, G. (2020). Regularization lemmas and convergence in total variation. *Electron. J. Probab.* **25** Paper No. 74, 20. MR4119120 https://doi.org/10.1214/20-ejp481

BASU, S. and SUBBA RAO, S. (2023). Graphical models for nonstationary time series. *Ann. Statist.* **51** 1453–1483. MR4657306 https://doi.org/10.1214/22-AOS2205

BONIS, T. (2024). Improved rates of convergence for the multivariate central limit theorem in Wasserstein distance. *Electron. J. Probab.* **29** Paper No. 78, 18. MR4757537 https://doi.org/10.1214/24-ejp1134

BORNEMANN, F. (2010). On the numerical evaluation of distributions in random matrix theory: A review. *Markov Process. Related Fields* **16** 803–866. MR2895091

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

BUTUCEA, C., GUŢĂ, M. and NUSSBAUM, M. (2018). Local asymptotic equivalence of pure states ensembles and quantum Gaussian white noise. *Ann. Statist.* **46** 3676–3706. MR3852665 https://doi.org/10.1214/17-AOS1672

BUTUCEA, C., MEISTER, A. and ROHDE, A. (2025). Supplement to "Asymptotic equivalence of locally stationary processes and bivariate Gaussian white noise." https://doi.org/10.1214/25-AOS2491SUPP

CARTER, A. V. (2007). Asymptotic approximation of nonparametric regression experiments with unknown variances. *Ann. Statist.* **35** 1644–1673. MR2351100 https://doi.org/10.1214/009053606000001613

CHAE, M. (2024). Wasserstein upper bounds of L^p-norms for multivariate densities in Besov spaces. *Statist. Probab. Lett.* **210** Paper No. 110131, 7. MR4729803 https://doi.org/10.1016/j.spl.2024.110131

DAHLHAUS, R. (1996). On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Process. Appl.* **62** 139–168. MR1388767 https://doi.org/10.1016/0304-4149(95)00090-9

MSC2020 subject classifications. Primary 62B15, 62M10; secondary 60F05.

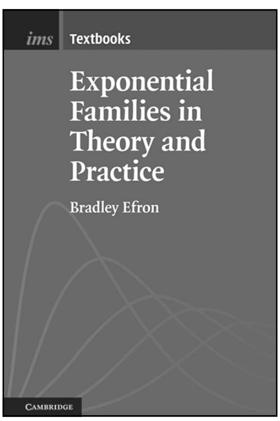
Key words and phrases. Covariance estimation, quantitative CLT in total variation, Le Cam distance, locally stationary time series.

DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37. MR1429916 https://doi.org/10.1214/aos/1034276620
DAHLHAUS, R. (2000). A likelihood approximation for locally stationary processes. *Ann. Statist.* **28** 1762–1794. MR1835040 https://doi.org/10.1214/aos/1015957480
DALALYAN, A. and REISS, M. (2006). Asymptotic statistical equivalence for scalar ergodic diffusions. *Probab. Theory Related Fields* **134** 248–282. MR2222384 https://doi.org/10.1007/s00440-004-0416-1
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
DING, X. and ZHOU, Z. (2023). AutoRegressive approximations to nonstationary time series with inference and applications. *Ann. Statist.* **51** 1207–1231. MR4630946 https://doi.org/10.1214/23-aos2288
FREDHOLM, I. (1903). Sur une classe d'équations fonctionnelles. *Acta Math.* **27** 365–390. MR1554993 https://doi.org/10.1007/BF02421317
GOLUBEV, G. K., NUSSBAUM, M. and ZHOU, H. H. (2010). Asymptotic equivalence of spectral density estimation and Gaussian white noise. *Ann. Statist.* **38** 181–214. MR2589320 https://doi.org/10.1214/09-AOS705
GRAMA, I. G. and NEUMANN, M. H. (2006). Asymptotic equivalence of nonparametric autoregression and nonparametric regression. *Ann. Statist.* **34** 1701–1732. MR2283714 https://doi.org/10.1214/009053606000000560
JÄHNISCH, M. and NUSSBAUM, M. (2003). Asymptotic equivalence for a model of independent non identically distributed observations. *Statist. Decisions* **21** 197–218. MR2045142 https://doi.org/10.1524/stnd.21.3.197.23430
LE CAM, L. (1964). Sufficiency and approximate sufficiency. *Ann. Math. Statist.* **35** 1419–1455. MR0207093 https://doi.org/10.1214/aoms/1177700372
MEISTER, A. (2011). Asymptotic equivalence of functional linear regression and a white noise inverse problem. *Ann. Statist.* **39** 1471–1495. MR2850209 https://doi.org/10.1214/10-AOS872
NOURDIN, I. and PECCATI, G. (2012). *Normal Approximations with Malliavin Calculus: From Stein's Method to Universality. Cambridge Tracts in Mathematics* **192**. Cambridge Univ. Press, Cambridge. MR2962301 https://doi.org/10.1017/CBO9781139084659
NUSSBAUM, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24** 2399–2430. MR1425959 https://doi.org/10.1214/aos/1032181160
PECCATI, G. and TAQQU, M. S. (2011). *Wiener Chaos: Moments, Cumulants and Diagrams. Bocconi & Springer Series* **1**. Springer, Milan. MR2791919 https://doi.org/10.1007/978-88-470-1679-8
RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning. Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR2514435
REISS, M. (2008). Asymptotic equivalence for nonparametric regression with multivariate and random design. *Ann. Statist.* **36** 1957–1982. MR2435461 https://doi.org/10.1214/07-AOS525
REISS, M. (2011). Asymptotic equivalence for inference on the volatility from noisy observations. *Ann. Statist.* **39** 772–802. MR2816338 https://doi.org/10.1214/10-AOS855
RICHTER, S. and DAHLHAUS, R. (2019). Cross validation for locally stationary processes. *Ann. Statist.* **47** 2145–2173. MR3953447 https://doi.org/10.1214/18-AOS1743
ROHDE, A. (2004). On the asymptotic equivalence and rate of convergence of nonparametric regression and Gaussian white noise. *Statist. Decisions* **22** 235–243. MR2125610 https://doi.org/10.1524/stnd.22.3.235.57063
SCHMIDT-HIEBER, J. (2014). Asymptotic equivalence for regression under fractional noise. *Ann. Statist.* **42** 2557–2585. MR3277671 https://doi.org/10.1214/14-AOS1262
TECUAPETLA-GÓMEZ, I. and NUSSBAUM, M. (2012). On large deviations in testing simple hypotheses for locally stationary Gaussian processes. *Stat. Inference Stoch. Process.* **15** 225–239. MR2979651 https://doi.org/10.1007/s11203-012-9071-9
TRUQUET, L. (2019). Local stationarity and time-inhomogeneous Markov chains. *Ann. Statist.* **47** 2023–2050. MR3953443 https://doi.org/10.1214/18-AOS1739
VAN DELFT, A. and DETTE, H. (2024). A general framework to quantify deviations from structural assumptions in the analysis of nonstationary function-valued processes. *Ann. Statist.* **52** 550–579. MR4744187 https://doi.org/10.1214/24-aos2358
VAN DER VAART, A. and VAN ZANTEN, H. (2011). Information rates of nonparametric Gaussian process methods. *J. Mach. Learn. Res.* **12** 2095–2119. MR2819028
VOGT, M. (2012). Nonparametric regression for locally stationary time series. *Ann. Statist.* **40** 2601–2633. MR3097614 https://doi.org/10.1214/12-AOS1043
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

NEW BOOK

The Institute of Mathematical Statistics presents

IMS TEXTBOOKS



Exponential Families in Theory and Practice

Bradley Efron, Stanford University

During the past half-century, exponential families have attained a position at the center of parametric statistical inference. Theoretical advances have been matched, and more than matched, in the world of applications, where logistic regression by itself has become the go-to methodology in medical statistics, computer-based prediction algorithms, and the social sciences. This book is based on a one-semester graduate course for first year Ph.D. and advanced master's students. After presenting the basic structure of univariate and multivariate exponential families, their application to generalized linear models including logistic and Poisson regression is described in detail, emphasizing geometrical ideas, computational practice, and the analogy with ordinary linear regression. Connections are made with a variety of current statistical methodologies: missing data, survival analysis and proportional hazards, false discovery rates, bootstrapping, and empirical Bayes analysis. The book connects exponential family theory with its applications in a way that doesn't require advanced mathematical preparation.

Hardback $105.00

Paperback $39.99

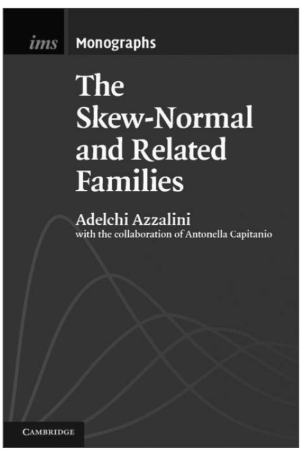
IMS members are entitled to a 40% discount: email ims@imstat.org to request your code

www.imstat.org/cup/

Cambridge University Press, with the Institute of Mathematical Statistics, established the *IMS Monographs* and *IMS Textbooks* series of high-quality books. The series editors are Mark Handcock, Ramon van Handel, Arnaud Doucet, and John Aston.

The Institute of Mathematical Statistics presents

IMS MONOGRAPHS



The Skew-Normal and Related Families

Adelchi Azzalini
in collaboration with Antonella Capitanio

Interest in the skew-normal and related families of distributions has grown enormously over recent years, as theory has advanced, challenges of data have grown, and computational tools have made substantial progress. This comprehensive treatment, blending theory and practice, will be the standard resource for statisticians and applied researchers. Assuming only basic knowledge of (non-measure-theoretic) probability and statistical inference, the book is accessible to the wide range of researchers who use statistical modelling techniques. Guiding readers through the main concepts and results, it covers both the probability and the statistics sides of the subject, in the univariate and multivariate settings. The theoretical development is complemented by numerous illustrations and applications to a range of fields including quantitative finance, medical statistics, environmental risk studies, and industrial and business efficiency.

The author's freely available R package sn, available from CRAN, equips readers to put the methods into action with their own data.

IMS member? Claim your 40% discount:
www.cambridge.org/ims

Hardback price US$48.00
(non-member price $80.00)

Cambridge University Press, in conjunction with the Institute of Mathematical Statistics, established the IMS Monographs and IMS Textbooks series of high-quality books. The Series Editors are Xiao-Li Meng, Susan Holmes, Ben Hambly, D. R. Cox and Alan Agresti.