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

# THE ANNALS *of* STATISTICS

*AN OFFICIAL JOURNAL OF THE*
INSTITUTE OF MATHEMATICAL STATISTICS

**Articles**



Vol. 48, No. 5—October 2020

THE ANNALS OF STATISTICS Vol. 48, No. 5, pp. 2505–3111 October 2020

# INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

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

---




*The Annals of Statistics* **[ISSN 0090-5364 (print); ISSN 2168-8966 (online)],** Volume 48, Number 5, October 2020. Published bimonthly by the Institute of Mathematical Statistics, 3163 Somerset Drive, Cleveland, Ohio 44122, 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, 9650 Rockville Pike, Suite L 2310, Bethesda, Maryland 20814-3998, USA.

*The Annals of Statistics*
2020, Vol. 48, No. 5, 2505–2547
https://doi.org/10.1214/19-AOS1895


# TESTING FOR STATIONARITY OF FUNCTIONAL TIME SERIES IN THE FREQUENCY DOMAIN

BY ALEXANDER AUE[1] AND ANNE VAN DELFT[2]

[1] *Department of Statistics, University of California, aaue@ucdavis.edu*
[2] *Fakultät für Mathematik, Ruhr-Universität Bochum, Anne.vanDelft@rub.de*

Interest in functional time series has spiked in the recent past with papers covering both methodology and applications being published at a much increased pace. This article contributes to the research in this area by proposing a new stationarity test for functional time series based on frequency domain methods. The proposed test statistics is based on joint dimension reduction via functional principal components analysis across the spectral density operators at all Fourier frequencies, explicitly allowing for frequency-dependent levels of truncation to adapt to the dynamics of the underlying functional time series. The properties of the test are derived both under the null hypothesis of stationary functional time series and under the smooth alternative of locally stationary functional time series. The methodology is theoretically justified through asymptotic results. Evidence from simulation studies and an application to annual temperature curves suggests that the test works well in finite samples.

## REFERENCES

ANTONIADIS, A. and SAPATINAS, T. (2003). Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes. *J. Multivariate Anal.* **87** 133–158. MR2007265 https://doi.org/10.1016/S0047-259X(03)00028-9

AUE, A. and HORVÁTH, L. (2013). Structural breaks in time series. *J. Time Series Anal.* **34** 1–16. MR3008012 https://doi.org/10.1111/j.1467-9892.2012.00819.x

AUE, A., DUBART NORINHO, D. D. and HÖRMANN, S. (2015). On the prediction of stationary functional time series. *J. Amer. Statist. Assoc.* **110** 378–392. MR3338510 https://doi.org/10.1080/01621459.2014.909317

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

AUE, A. and VAN DELFT, A. (2019). Supplement to "Testing for stationarity of functional time series in the frequency domain." https://doi.org/10.1214/19-AOS1895SUPP.

BANDYOPADHYAY, S., JENTSCH, C. and SUBBA RAO, S. (2017). A spectral domain test for stationarity of spatio-temporal data. *J. Time Series Anal.* **38** 326–351. MR3611747 https://doi.org/10.1111/jtsa.12222

BANDYOPADHYAY, S. and SUBBA RAO, S. (2017). A test for stationarity for irregularly spaced spatial data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 95–123. MR3597966 https://doi.org/10.1111/rssb.12161

BESSE, P., CARDOT, H. and STEPHENSON, D. (2000). Autoregressive forecasting of some functional climatic variations. *Scand. J. Stat.* **27** 673–687.

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

BRILLINGER, D. (1981). *Time Series: Data Analysis and Theory*. McGraw Hill, New York.

BRILLINGER, D. R. and ROSENBLATT, M. (1967). Asymptotic theory of estimates of $k$-th order spectra. In *Spectral Analysis Time Series (Proc. Advanced Sem., Madison, Wis.,* 1966) 153–188. John Wiley, New York. MR0211566

CREMERS, H. and KADELKA, D. (1986). On weak convergence of integral functionals of stochastic processes with applications to processes taking paths in $L_p^E$. *Stochastic Process. Appl.* **21** 305–317. MR0833957 https://doi.org/10.1016/0304-4149(86)90102-X

*MSC2020 subject classifications.* Primary 62G99, 62H99; secondary 62M10, 62M15, 91B84.
*Key words and phrases.* Frequency domain methods, functional data analysis, locally stationary processes, spectral analysis.

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

DETTE, H., PREUSS, P. and VETTER, M. (2011). A measure of stationarity in locally stationary processes with applications to testing. *J. Amer. Statist. Assoc.* **106** 1113–1124. MR2894768 https://doi.org/10.1198/jasa.2011.tm10811

DWIVEDI, Y. and SUBBA RAO, S. (2011). A test for second-order stationarity of a time series based on the discrete Fourier transform. *J. Time Series Anal.* **32** 68–91. MR2790673 https://doi.org/10.1111/j.1467-9892.2010.00685.x

FERRATY, F. and VIEU, P. (2010). *Nonparametric Functional Data Analysis*: *Theory and Practice*. *Springer Series in Statistics*. Springer, New York. MR2229687

FREMDT, S., HORVÁTH, L., KOKOSZKA, P. and STEINEBACH, J. G. (2014). Functional data analysis with increasing number of projections. *J. Multivariate Anal.* **124** 313–332. MR3147328 https://doi.org/10.1016/j.jmva.2013.11.009

HALL, P. and HOSSEINI-NASAB, M. (2006). On properties of functional principal components analysis. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 109–126. MR2212577 https://doi.org/10.1111/j.1467-9868.2005.00535.x

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

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

HORVÁTH, L. and KOKOSZKA, P. (2012). *Inference for Functional Data with Applications*. *Springer Series in Statistics*. Springer, New York. MR2920735 https://doi.org/10.1007/978-1-4614-3655-3

HORVÁTH, L., KOKOSZKA, P. and RICE, G. (2014). Testing stationarity of functional time series. *J. Econometrics* **179** 66–82. MR3153649 https://doi.org/10.1016/j.jeconom.2013.11.002

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

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

LEE, J. and RAO, S. S. (2017). A note on general quadratic forms of nonstationary stochastic processes. *Statistics* **51** 949–968. MR3698495 https://doi.org/10.1080/02331888.2017.1318880

LI, Y. and HSING, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist.* **38** 3321–3351. MR2766854 https://doi.org/10.1214/10-AOS813

MAS, A. and MENNETEAU, L. (2003). Perturbation approach applied to the asymptotic study of random operators. In *High Dimensional Probability*, *III* (*Sandjberg*, 2002). *Progress in Probability* **55** 127–134. Birkhäuser, Basel. MR2033885

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

PAPARODITIS, E. (2009). Testing temporal constancy of the spectral structure of a time series. *Bernoulli* **15** 1190–1221. MR2597589 https://doi.org/10.3150/08-BEJ179

PREUSS, P., VETTER, M. and DETTE, H. (2013). A test for stationarity based on empirical processes. *Bernoulli* **19** 2715–2749. MR3160569 https://doi.org/10.3150/12-BEJ472

PRIESTLEY, M. B. and SUBBA RAO, T. (1969). A test for non-stationarity of time-series. *J. Roy. Statist. Soc. Ser. B* **31** 140–149. MR0269062

RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2168993

SUBBA RAO, S. (2018). Orthogonal samples for estimators in time series. *J. Time Series Anal.* **39** 313–337. MR3796522 https://doi.org/10.1111/jtsa.12269

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

VAN DELFT, A. and EICHLER, M. (2019) A note on Herglotz's theorem for time series on function spaces. https://doi.org/10.1016/j.spa.2019.10.006, forthcoming.

VON SACHS, R. and NEUMANN, M. H. (2000). A wavelet-based test for stationarity. *J. Time Series Anal.* **21** 597–613. MR1794489 https://doi.org/10.1111/1467-9892.00200

WILKINSON, H. (1965). *The Algebraic Eigenvalue Problem*. Clarendon Press, Oxford.

*The Annals of Statistics*
2020, Vol. 48, No. 5, 2548–2574
https://doi.org/10.1214/19-AOS1897


# ON SPIKE AND SLAB EMPIRICAL BAYES MULTIPLE TESTING

BY ISMAËL CASTILLO* AND ÉTIENNE ROQUAIN**

*Laboratoire de Probabilités, Statistique et Modélisation, LPSM, Sorbonne Université,* *ismael.castillo@upmc.fr; **etienne.roquain@upmc.fr

This paper explores a connection between empirical Bayes posterior distributions and false discovery rate (FDR) control. In the Gaussian sequence model this work shows that empirical Bayes-calibrated spike and slab posterior distributions allow a correct FDR control under sparsity. Doing so, it offers a frequentist theoretical validation of empirical Bayes methods in the context of multiple testing. Our theoretical results are illustrated with numerical experiments.

## REFERENCES

[1] ABRAMOVICH, F. and ANGELINI, C. (2006). Bayesian maximum *a posteriori* multiple testing procedure. *Sankhyā* **68** 436–460. MR2322194

[2] ABRAMOVICH, F., BENJAMINI, Y., DONOHO, D. L. and JOHNSTONE, I. M. (2006). Adapting to unknown sparsity by controlling the false discovery rate. *Ann. Statist.* **34** 584–653. MR2281879 https://doi.org/10.1214/009053606000000074

[3] ABRAMOVICH, F., GRINSHTEIN, V. and PENSKY, M. (2007). On optimality of Bayesian testimation in the normal means problem. *Ann. Statist.* **35** 2261–2286. MR2363971 https://doi.org/10.1214/009053607000000226

[4] ARIAS-CASTRO, E. and CHEN, S. (2017). Distribution-free multiple testing. *Electron. J. Stat.* **11** 1983–2001. MR3651021 https://doi.org/10.1214/17-EJS1277

[5] BELITSER, E. and GHOSAL, S. (2019). Empirical Bayes oracle uncertainty quantification for regression. *Ann. Statist.* To appear.

[6] BELITSER, E. and NURUSHEV, N. (2019). Needles and straw in a haystack: Robust empirical Bayes confidence for possibly sparse sequences. *Bernoulli*. To appear.

[7] 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

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

[9] 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

[10] BOGDAN, M., CHAKRABARTI, A., FROMMLET, F. and GHOSH, J. K. (2011). Asymptotic Bayes-optimality under sparsity of some multiple testing procedures. *Ann. Statist.* **39** 1551–1579. MR2850212 https://doi.org/10.1214/10-AOS869

[11] BOGDAN, M., GHOSH, J. K. and TOKDAR, S. T. (2008). A comparison of the Benjamini–Hochberg procedure with some Bayesian rules for multiple testing. In *Beyond Parametrics in Interdisciplinary Research*: *Festschrift in Honor of Professor Pranab K. Sen*. *Inst. Math. Stat.* (*IMS*) *Collect.* **1** 211–230. IMS, Beachwood, OH. MR2462208 https://doi.org/10.1214/193940307000000158

[12] BOGDAN, M., VAN DEN BERG, E., SABATTI, C., SU, W. and CANDÈS, E. J. (2015). SLOPE—Adaptive variable selection via convex optimization. *Ann. Appl. Stat.* **9** 1103–1140. MR3418717 https://doi.org/10.1214/15-AOAS842

[13] CAI, T. T. and SUN, W. (2009). Simultaneous testing of grouped hypotheses: Finding needles in multiple haystacks. *J. Amer. Statist. Assoc.* **104** 1467–1481. MR2597000 https://doi.org/10.1198/jasa.2009.tm08415

[14] CAI, T. T., SUN, W. and WANG, W. (2019). Covariate-assisted ranking and screening for large-scale two-sample inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 187–234. MR3928141

*MSC2020 subject classifications.* 62C12, 62G10.

*Key words and phrases.* Frequentist properties of Bayesian procedures, false discovery rate, sparsity, multiple testing.

[15] CAO, J., XIE, X.-J., ZHANG, S., WHITEHURST, A. and WHITE, M. A. (2009). Bayesian optimal discovery procedure for simultaneous significance testing. *BMC Bioinform*. **10** Article ID 5.

[16] CASTILLO, I. and MISMER, R. (2018). Empirical Bayes analysis of spike and slab posterior distributions. *Electron. J. Stat*. **12** 3953–4001. MR3885271 https://doi.org/10.1214/18-EJS1494

[17] CASTILLO, I. and ROQUAIN, E. (2020). Supplement to "On spike and slab empirical Bayes multiple testing." https://doi.org/10.1214/19-AOS1897SUPP.

[18] CASTILLO, I. and SZABÓ, B. (2019). Spike and slab empirical Bayes sparse credible sets. *Bernoulli*. To appear.

[19] CASTILLO, I. and VAN DER VAART, A. (2012). Needles and straw in a haystack: Posterior concentration for possibly sparse sequences. *Ann. Statist*. **40** 2069–2101. MR3059077 https://doi.org/10.1214/12-AOS1029

[20] DONOHO, D. L., JOHNSTONE, I. M., HOCH, J. C. and STERN, A. S. (1992). Maximum entropy and the nearly black object. *J. Roy. Statist. Soc. Ser. B* **54** 41–81. MR1157714

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

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

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

[24] FINNER, H., DICKHAUS, T. and ROTERS, M. (2007). Dependency and false discovery rate: Asymptotics. *Ann. Statist*. **35** 1432–1455. MR2351092 https://doi.org/10.1214/009053607000000046

[25] GEORGE, E. I. and FOSTER, D. P. (2000). Calibration and empirical Bayes variable selection. *Biometrika* **87** 731–747. MR1813972 https://doi.org/10.1093/biomet/87.4.731

[26] GERARD, D. and STEPHENS, M. (2018). Empirical Bayes shrinkage and false discovery rate estimation, allowing for unwanted variation. *Biostatistics* To appear.

[27] GHOSAL, S., GHOSH, J. K. and VAN DER VAART, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist*. **28** 500–531. MR1790007 https://doi.org/10.1214/aos/1016218228

[28] GUINDANI, M., MÜLLER, P. and ZHANG, S. (2009). A Bayesian discovery procedure. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **71** 905–925. MR2750250 https://doi.org/10.1111/j.1467-9868.2009.00714.x

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

[30] JOHNSTONE, I. M. and SILVERMAN, B. W. (2004). Needles and straw in haystacks: Empirical Bayes estimates of possibly sparse sequences. *Ann. Statist*. **32** 1594–1649. MR2089135 https://doi.org/10.1214/009053604000000030

[31] JOHNSTONE, I. M. and SILVERMAN, B. W. (2005). EbayesThresh: R programs for empirical Bayes thresholding. *J. Stat. Softw*. **12** Issue 8.

[32] MARTIN, R. and TOKDAR, S. (2012). A nonparametric empirical Bayes framework for large-scale significance testing. *Biostatistics* **13** 427–439.

[33] MITCHELL, T. J. and BEAUCHAMP, J. J. (1988). Bayesian variable selection in linear regression. *J. Amer. Statist. Assoc*. **83** 1023–1036. MR0997578

[34] MÜLLER, P., PARMIGIANI, G., ROBERT, C. and ROUSSEAU, J. (2004). Optimal sample size for multiple testing: The case of gene expression microarrays. *J. Amer. Statist. Assoc*. **99** 990–1001. MR2109489 https://doi.org/10.1198/016214504000001646

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

[36] RUDIN, W. (1976). *Principles of Mathematical Analysis*, 3rd ed. *International Series in Pure and Applied Mathematics*. McGraw-Hill, New York. MR0385023

[37] SALOMOND, J.-B. (2017). Risk quantification for the thresholding rule for multiple testing using Gaussian scale mixtures. Preprint. Available at arXiv:1711.08705.

[38] SARKAR, S. K. (2007). Stepup procedures controlling generalized FWER and generalized FDR. *Ann. Statist*. **35** 2405–2420. MR2382652 https://doi.org/10.1214/009053607000000398

[39] SARKAR, S. K., ZHOU, T. and GHOSH, D. (2008). A general decision theoretic formulation of procedures controlling FDR and FNR from a Bayesian perspective. *Statist. Sinica* **18** 925–945. MR2440399

[40] SCOTT, J. G. and BERGER, J. O. (2010). Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *Ann. Statist*. **38** 2587–2619. MR2722450 https://doi.org/10.1214/10-AOS792

[41] SHEN, X. and WASSERMAN, L. (2001). Rates of convergence of posterior distributions. *Ann. Statist*. **29** 687–714. MR1865337 https://doi.org/10.1214/aos/1009210686

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

[43] STOREY, J. D. (2003). The positive false discovery rate: A Bayesian interpretation and the $q$-value. *Ann. Statist.* **31** 2013–2035. MR2036398 https://doi.org/10.1214/aos/1074290335

[44] SU, W. and CANDÈS, E. (2016). SLOPE is adaptive to unknown sparsity and asymptotically minimax. *Ann. Statist.* **44** 1038–1068. MR3485953 https://doi.org/10.1214/15-AOS1397

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

[46] SUN, W. and CAI, T. T. (2009). Large-scale multiple testing under dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 393–424. MR2649603 https://doi.org/10.1111/j.1467-9868.2008.00694.x

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

[48] VAN DER PAS, S., SZABÓ, B. and VAN DER VAART, A. (2017). Uncertainty quantification for the horseshoe (with discussion). *Bayesian Anal.* **12** 1221–1274. MR3724985 https://doi.org/10.1214/17-BA1065

[49] VAN DER PAS, S. L., KLEIJN, B. J. K. and VAN DER VAART, A. W. (2014). The horseshoe estimator: Posterior concentration around nearly black vectors. *Electron. J. Stat.* **8** 2585–2618. MR3285877 https://doi.org/10.1214/14-EJS962

*The Annals of Statistics*
2020, Vol. 48, No. 5, 2575–2598
https://doi.org/10.1214/19-AOS1898


# THEORETICAL AND COMPUTATIONAL GUARANTEES OF MEAN FIELD VARIATIONAL INFERENCE FOR COMMUNITY DETECTION

BY ANDERSON Y. ZHANG[1] AND HARRISON H. ZHOU[2]

[1]*Department of Statistics, The Wharton School, University of Pennsylvania, ayz@wharton.upenn.edu*
[2]*Department of Statistics, Yale University, huibin.zhou@yale.edu*

The mean field variational Bayes method is becoming increasingly popular in statistics and machine learning. Its iterative coordinate ascent variational inference algorithm has been widely applied to large scale Bayesian inference. See Blei et al. (2017) for a recent comprehensive review. Despite the popularity of the mean field method, there exist remarkably little fundamental theoretical justifications. To the best of our knowledge, the iterative algorithm has never been investigated for any high-dimensional and complex model. In this paper, we study the mean field method for community detection under the stochastic block model. For an iterative batch coordinate ascent variational inference algorithm, we show that it has a linear convergence rate and converges to the minimax rate within $\log n$ iterations. This complements the results of Bickel et al. (2013) which studied the global minimum of the mean field variational Bayes and obtained asymptotic normal estimation of global model parameters. In addition, we obtain similar optimality results for Gibbs sampling and an iterative procedure to calculate maximum likelihood estimation, which can be of independent interest.

## REFERENCES

[1] AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P. (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.

[2] BEAL, M. J. (2003). *Variational Algorithms for Approximate Bayesian Inference*. Univ. London, London.

[3] BICKEL, P., CHOI, D., CHANG, X. and ZHANG, H. (2013). Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *Ann. Statist.* **41** 1922–1943. MR3127853 https://doi.org/10.1214/13-AOS1124

[4] BICKEL, P. J. and CHEN, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA* **106** 21068–21073.

[5] BICKEL, P. J., CHEN, A. and LEVINA, E. (2011). The method of moments and degree distributions for network models. *Ann. Statist.* **39** 2280–2301. MR2906868 https://doi.org/10.1214/11-AOS904

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

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

[8] BLEI, D. M., NG, A. Y. and JORDAN, M. I. (2003). Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3** 993–1022.

[9] CELISSE, A., DAUDIN, J.-J. and PIERRE, L. (2012). Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electron. J. Stat.* **6** 1847–1899. MR2988467 https://doi.org/10.1214/12-EJS729

[10] CHIN, P., RAO, A. and VU, V. (2015). Stochastic block model and community detection in sparse graphs: A spectral algorithm with optimal rate of recovery. In *COLT* 391–423.

[11] FEI, Y. and CHEN, Y. (2019). Exponential error rates of SDP for block models: Beyond Grothendieck's inequality. *IEEE Trans. Inform. Theory* **65** 551–571. MR3901009 https://doi.org/10.1109/TIT.2018.2839677

*MSC2020 subject classifications.* 60G05.
*Key words and phrases.* Mean field, variational inference, Bayesian, community detection, stochastic block model.

[12] GAO, C., MA, Z., ZHANG, A. Y. and ZHOU, H. H. (2017). Achieving optimal misclassification proportion in stochastic block models. *J. Mach. Learn. Res.* **18** Art. ID 60. MR3687603

[13] GAO, C., VAN DER VAART, A. W. and ZHOU, H. H. (2015). A general framework for Bayes structured linear models. Preprint. Available at arXiv:1506.02174.

[14] GELFAND, A. E. and SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** 398–409. MR1141740

[15] GRABSKA-BARWIŃSKA, A., BARTHELMÉ, S., BECK, J., MAINEN, Z. F., POUGET, A. and LATHAM, P. E. (2017). A probabilistic approach to demixing odors. *Nat. Neurosci.* **20** 98–106. https://doi.org/10.1038/nn.4444

[16] GUÉDON, O. and VERSHYNIN, R. (2016). Community detection in sparse networks via Grothendieck's inequality. *Probab. Theory Related Fields* **165** 1025–1049. MR3520025 https://doi.org/10.1007/s00440-015-0659-z

[17] HOFMAN, J. M. and WIGGINS, C. H. (2008). Bayesian approach to network modularity. *Phys. Rev. Lett.* **100** Art. ID 258701. https://doi.org/10.1103/PhysRevLett.100.258701

[18] HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. MR0718088 https://doi.org/10.1016/0378-8733(83)90021-7

[19] JORDAN, M. I., GHAHRAMANI, Z., JAAKKOLA, T. S. and SAUL, L. K. (1999). An introduction to variational methods for graphical models. *Mach. Learn.* **37** 183–233.

[20] LEI, J. and RINALDO, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann. Statist.* **43** 215–237. MR3285605 https://doi.org/10.1214/14-AOS1274

[21] LIANG, P., PETROV, S., JORDAN, M. I. and KLEIN, D. (2007). The infinite PCFG using hierarchical Dirichlet processes. In *EMNLP-CoNLL* 688–697.

[22] LU, Y. and ZHOU, H. H. (2016). Statistical and computational guarantees of Lloyd's algorithm and its variants. Preprint. Available at arXiv:1612.02099.

[23] MOSSEL, E., NEEMAN, J. and SLY, A. (2012). Stochastic block models and reconstruction. Preprint. Available at arXiv:1202.1499.

[24] NEWMAN, M. E. J. (2006). Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA* **103** 8577–8582. https://doi.org/10.1073/pnas.0601602103

[25] PENNY, W. D., TRUJILLO-BARRETO, N. J. and FRISTON, K. J. (2005). Bayesian fMRI time series analysis with spatial priors. *NeuroImage* **24** 350–362.

[26] RAZAEE, Z. S., AMINI, A. A. and LI, J. J. (2019). Matched bipartite block model with covariates. *J. Mach. Learn. Res.* **20** Art. ID 34. MR3948074

[27] ROBERT, C. P. (2004). *Monte Carlo Methods*. Wiley, New York.

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

[29] WAINWRIGHT, M. J. and JORDAN, M. I. (2008). Graphical models, exponential families, and variational inference. *Found. Trends Mach. Learn.* **1** 1–305.

[30] WANG, B. and TITTERINGTON, D. M. (2006). Convergence properties of a general algorithm for calculating variational Bayesian estimates for a normal mixture model. *Bayesian Anal.* **1** 625–649. MR2221291 https://doi.org/10.1214/06-BA121

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

[32] WESTLING, T. and MCCORMICK, T. H. (2015). Establishing consistency and improving uncertainty estimates of variational inference through M-estimation. Preprint. Available at arXiv:1510.08151.

[33] YOU, C., ORMEROD, J. T. and MÜLLER, S. (2014). On variational Bayes estimation and variational information criteria for linear regression models. *Aust. N. Z. J. Stat.* **56** 73–87. MR3200293 https://doi.org/10.1111/anzs.12063

[34] ZHANG, A. Y. and ZHOU, H. H. (2016). Minimax rates of community detection in stochastic block models. *Ann. Statist.* **44** 2252–2280. MR3546450 https://doi.org/10.1214/15-AOS1428

[35] ZHANG, A. Y. and ZHOU, H. H. (2020). Supplement to "Theoretical and computational guarantees of mean field variational inference for community detection". https://doi.org/10.1214/19-AOS1898SUPP

*The Annals of Statistics*
2020, Vol. 48, No. 5, 2599–2621
https://doi.org/10.1214/19-AOS1899


# MINIMAX OPTIMAL SEQUENTIAL HYPOTHESIS TESTS FOR MARKOV PROCESSES

BY MICHAEL FAUSS[1,*], ABDELHAK M. ZOUBIR[1,**] AND H. VINCENT POOR[2]

[1]*Signal Processing Group, TU Darmstadt,* [*]*mfauss@princeton.edu;* [**]*zoubir@spg.tu-darmstadt.de*
[2]*Department of Electrical Engineering, Princeton University,* *poor@princeton.edu*

Under mild Markov assumptions, sufficient conditions for strict minimax optimality of sequential tests for multiple hypotheses under distributional uncertainty are derived. First, the design of optimal sequential tests for simple hypotheses is revisited, and it is shown that the partial derivatives of the corresponding cost function are closely related to the performance metrics of the underlying sequential test. Second, an implicit characterization of the least favorable distributions for a given testing policy is stated. By combining the results on optimal sequential tests and least favorable distributions, sufficient conditions for a sequential test to be minimax optimal under general distributional uncertainties are obtained. The cost function of the minimax optimal test is further identified as a generalized $f$-dissimilarity and the least favorable distributions as those that are most similar with respect to this dissimilarity. Numerical examples for minimax optimal sequential tests under different uncertainties illustrate the theoretical results.

## REFERENCES

[1] AVELLA MEDINA, M. and RONCHETTI, E. (2015). Robust statistics: A selective overview and new directions. *Wiley Interdiscip. Rev.: Comput. Stat.* **7** 372–393. MR3415420 https://doi.org/10.1002/wics.1363
[2] BANERJEE, T. and VEERAVALLI, V. V. (2015). Data-efficient minimax quickest change detection with composite post-change distribution. *IEEE Trans. Inform. Theory* **61** 5172–5184. MR3386508 https://doi.org/10.1109/TIT.2015.2458864
[3] BASU, A., HARRIS, I. R., HJORT, N. L. and JONES, M. C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika* **85** 549–559. MR1665873 https://doi.org/10.1093/biomet/85.3.549
[4] BREUER, T. and CSISZÁR, I. (2016). Measuring distribution model risk. *Math. Finance* **26** 395–411. MR3481309 https://doi.org/10.1111/mafi.12050
[5] BRODSKY, B. and DARKHOVSKY, B. (2008). Minimax methods for multihypothesis sequential testing and change-point detection problems. *Sequential Anal.* **27** 141–173. MR2439985 https://doi.org/10.1080/07474940801989111
[6] BRODSKY, B. E. and DARKHOVSKY, B. S. (2008). Minimax sequential tests for many composite hypotheses. I. *Theory Probab. Appl.* **52** 565–579.
[7] CAPASSO, V. and BAKSTEIN, D. (2015). *An Introduction to Continuous-Time Stochastic Processes: Theory, Models, and Applications to Finance, Biology, and Medicine*, 3rd ed. *Modeling and Simulation in Science, Engineering and Technology*. Birkhäuser/Springer, New York. MR3362672 https://doi.org/10.1007/978-1-4939-2757-9
[8] CHOQUET, G. (1954). Theory of capacities. *Ann. Inst. Fourier (Grenoble)* **5** 131–295. MR0080760
[9] DEGROOT, M. H. (1960). Minimax sequential tests of some composite hypotheses. *Ann. Math. Stat.* **31** 1193–1200. MR0120734 https://doi.org/10.1214/aoms/1177705690
[10] DRAGALIN, V. P. and NOVIKOV, A. A. (1987). Asymptotic solution of the Kiefer–Weiss problem for processes with independent increments. *Theory Probab. Appl.* **32** 617–627.
[11] DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1953). Sequential decision problems for processes with continuous time parameter. Testing hypotheses. *Ann. Math. Stat.* **24** 254–264. MR0054911 https://doi.org/10.1214/aoms/1177729031

*MSC2020 subject classifications.* Primary 62L10; secondary 62C20.
*Key words and phrases.* Sequential analysis, minimax procedures, distributional uncertainty, robust hypothesis testing, multiple hypothesis testing.

[12] EL-SAWY, A. H. and VANDELINDE, V. D. (1979). Robust sequential detection of signals in noise. *IEEE Trans. Inform. Theory* **25** 346–353. MR0528012 https://doi.org/10.1109/TIT.1979.1056049

[13] FAUSS, M. (2016). Design and analysis of optimal and minimax robust sequential hypthesis tests. Ph.D. thesis, TU Darmstadt, Institute of Communications.

[14] FAUSS, M. and ZOUBIR, A. M. (2015). A linear programming approach to sequential hypothesis testing. *Sequential Anal.* **34** 235–263. MR3348300 https://doi.org/10.1080/07474946.2015.1030981

[15] FAUSS, M. and ZOUBIR, A. M. (2016). Old bands, new tracks—Revisiting the band model for robust hypothesis testing. *IEEE Trans. Signal Process.* **64** 5875–5886. MR3554590 https://doi.org/10.1109/TSP.2016.2600505

[16] FAUSS, M., ZOUBIR, A. M. and POOR, V. H. (2018). On the equivalence of $f$-divergence balls and density bands in robust detection. In *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

[17] FAUSS, M., ZOUBIR, A. M. and POOR, H. V. (2020). Supplement to "Minimax optimal sequential hypothesis tests for Markov processes." https://doi.org/10.1214/19-AOS1899SUPP.

[18] FELLOURIS, G. and TARTAKOVSKY, A. G. (2012). Nearly minimax one-sided mixture-based sequential tests. *Sequential Anal.* **31** 297–325. MR2958638

[19] FELLOURIS, G. and TARTAKOVSKY, A. G. (2013). Almost optimal sequential tests of discrete composite hypotheses. *Statist. Sinica* **23** 1717–1741. MR3222817

[20] FERRARI, S. and STENGEL, R. F. (2005). Smooth function approximation using neural networks. *IEEE Trans. Neural Netw.* **16** 24–38.

[21] GAO, R., XIE, L., XIE, Y. and XU, H. (2018). Robust hypothesis testing using Wasserstein uncertainty sets. *Adv. Neural Inf. Process. Syst.* 7913–7923.

[22] GHOSH, B. K. and SEN, P. K., eds. (1991) *Handbook of Sequential Analysis. Statistics: Textbooks and Monographs* **118**. Dekker, New York. MR1174296

[23] GÜL, G. and ZOUBIR, A. M. (2016). Robust hypothesis testing with $\alpha$-divergence. *IEEE Trans. Signal Process.* **64** 4737–4750. MR3538377 https://doi.org/10.1109/TSP.2016.2569405

[24] GYÖRFI, L. and NEMETZ, T. (1975). On the dissimilarity of probability measures. Technical report, Mathematical Institute of the Hungarian Academy of Science.

[25] GYÖRFI, L. and NEMETZ, T. (1977). $f$-dissimilarity: A general class of separation measures of several probability measures. In *Topics in Information Theory (Second Colloq., Keszthely*, 1975). *Colloq. Math. Soc. János Bolyai* **16** 309–321. MR0459923

[26] GYÖRFI, L. and NEMETZ, T. (1978). $f$-dissimilarity: A generalization of the affinity of several distributions. *Ann. Inst. Statist. Math.* **30** 105–113. MR0507086 https://doi.org/10.1007/BF02480206

[27] HAFNER, R. (1993). Construction of minimax-tests for bounded families of probability-densities. *Metrika* **40** 1–23. MR1210451 https://doi.org/10.1007/BF02613659

[28] HAZEWINKEL, M., ed. (1994). Kolmogorov–Chapman equation. In *Encyclopaedia of Mathematics* **5** 292. Springer, Dordrecht, Netherlands.

[29] HUBER, P. J. (1965). A robust version of the probability ratio test. *Ann. Math. Stat.* **36** 1753–1758. MR0185747 https://doi.org/10.1214/aoms/1177699803

[30] HUBER, P. J. (1981). *Robust Statistics*. Wiley, Hoboken, NJ.

[31] HUBER, P. J. and STRASSEN, V. (1973). Minimax tests and the Neyman–Pearson lemma for capacities. *Ann. Statist.* **1** 251–263. MR0356306

[32] KASSAM, S. A. (1981). Robust hypothesis testing for bounded classes of probability densities. *IEEE Trans. Inform. Theory* **27** 242–247. MR0633425 https://doi.org/10.1109/TIT.1981.1056314

[33] KASSAM, S. A. and POOR, H. V. (1985). Robust techniques for signal processing: A survey. *Proc. IEEE* **73** 433–481.

[34] KHARIN, A. (2002). On robustifying of the sequential probability ratio test for a discrete model under "contaminations". *Aust. J. Stat.* **31** 267–277.

[35] KIEFER, J. and WEISS, L. (1957). Some properties of generalized sequential probability ratio tests. *Ann. Math. Stat.* **28** 57–74. MR0087290 https://doi.org/10.1214/aoms/1177707037

[36] KUNSCH, R. (2017). High-dimensional function approximation: Breaking the curse with Monte Carlo methods. Ph.D. thesis, Friedrich-Schiller-Universität Jena, Jena, Germany.

[37] LORDEN, G. (1976). 2-SPRT's and the modified Kiefer–Weiss problem of minimizing an expected sample size. *Ann. Statist.* **4** 281–291. MR0405750

[38] MARONNA, R. A., MARTIN, R. D. and YOHAI, V. J. (2006). *Robust Statistics: Theory and Methods. Wiley Series in Probability and Statistics*. Wiley, Chichester. MR2238141 https://doi.org/10.1002/0470010940

[39] MAURICE, R. J. (1957). A minimax procedure for choosing between two populations using sequential sampling. *J. Roy. Statist. Soc. Ser. B* **19** 255–261.

[40] NGUYEN, X., WAINWRIGHT, M. J. and JORDAN, M. I. (2009). On surrogate loss functions and $f$-divergences. *Ann. Statist.* **37** 876–904. MR2502654 https://doi.org/10.1214/08-AOS595

[41] NOVIKOV, A. (2009). Optimal sequential multiple hypothesis tests. *Kybernetika* (*Prague*) **45** 309–330. MR2518154

[42] ÖSTERREICHER, F. (1978). On the construction of least favourable pairs of distributions. *Z. Wahrsch. Verw. Gebiete* **43** 49–55. MR0493481 https://doi.org/10.1007/BF00535275

[43] PAPAGEORGIOU, N. S. (1997). Convex integral functionals. *Trans. Amer. Math. Soc.* **349** 1421–1436. MR1321585 https://doi.org/10.1090/S0002-9947-97-01478-5

[44] PARDO, L. (2006). *Statistical Inference Based on Divergence Measures*. *Statistics*: *Textbooks and Monographs* **185**. CRC Press/CRC, Boca Raton, FL. MR2183173

[45] PAVLOV, I. V. (1991). A sequential procedure for testing composite hypotheses with application to the Kiefer–Weiss problem. *Theory Probab. Appl.* **35** 280–292.

[46] POOR, H. V. (1980). Robust decision design using a distance criterion. *IEEE Trans. Inform. Theory* **26** 575–587. MR0583942 https://doi.org/10.1109/TIT.1980.1056249

[47] POOR, H. V. and HADJILIADIS, O. (2009). *Quickest Detection*. Cambridge Univ. Press, Cambridge. MR2482527

[48] REID, M. D. and WILLIAMSON, R. C. (2011). Information, divergence and risk for binary experiments. *J. Mach. Learn. Res.* **12** 731–817. MR2786911

[49] ROCKAFELLAR, R. T. (1968). Integrals which are convex functionals. *Pacific J. Math.* **24** 525–539. MR0236689

[50] ROCKAFELLAR, R. T. (1970). *Convex Analysis*. *Princeton Mathematical Series* **28**. Princeton Univ. Press, Princeton, NJ. MR0274683

[51] RUDIN, W. (1976). *Principles of Mathematical Analysis*, 3rd ed. *International Series in Pure and Applied Mathematics*. McGraw-Hill, New York–Auckland–Düsseldorf. MR0385023

[52] SCHMITZ, N. (1987). Minimax sequential tests of composite hypotheses on the drift of a Wiener process. *Stat. Hefte* **28** 247–261. MR0975919

[53] SIEGMUND, D. (1985). *Sequential Analysis*: *Tests and Confidence Intervals*. *Springer Series in Statistics*. Springer, New York. MR0799155 https://doi.org/10.1007/978-1-4757-1862-1

[54] SOCHMAN, J. and MATAS, J. (2005). WaldBoost—Learning for time constrained sequential detection. In *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (C. Schmid, S. Soatto and C. Tomasi, eds.) 150–156.

[55] TARTAKOVSKY, A., NIKIFOROV, I. and BASSEVILLE, M. (2015). *Sequential Analysis*: *Hypothesis Testing and Changepoint Detection*. *Monographs on Statistics and Applied Probability* **136**. CRC Press, Boca Raton, FL. MR3241619

[56] TARTAKOVSKY, A. G., LI, X. R. and YARALOV, G. (2003). Sequential detection of targets in multichannel systems. *IEEE Trans. Inform. Theory* **49** 425–445. MR1966790 https://doi.org/10.1109/TIT.2002.807288

[57] UNNIKRISHNAN, J., VEERAVALLI, V. V. and MEYN, S. P. (2011). Minimax robust quickest change detection. *IEEE Trans. Inform. Theory* **57** 1604–1614. MR2815837 https://doi.org/10.1109/TIT.2011.2104993

[58] VARSHNEY, K. R. (2011). Bayes risk error is a Bregman divergence. *IEEE Trans. Signal Process.* **59** 4470–4472. MR2865998 https://doi.org/10.1109/TSP.2011.2159500

[59] VERDÚ, S. and POOR, H. V. (1984). On minimax robustness: A general approach and applications. *IEEE Trans. Inform. Theory* **30** 328–340. MR0754862 https://doi.org/10.1109/TIT.1984.1056876

[60] VOS, H. J. (2001). A minimax procedure in the context of sequential testing problems in psychodiagnostics. *Br. J. Math. Stat. Psychol.* **54** 139–159.

[61] VOUDOURI, E. and KURZ, L. (1988). A robust approach to sequential detection. *IEEE Trans. Acoust. Speech Signal Process.* **36** 1200–1210.

[62] WALD, A. (1947). *Sequential Analysis*. Wiley, Hoboken, NJ. MR0020764

[63] WEISS, C. H. and KIM, H.-Y. (2013). Binomial AR(1) processes: Moments, cumulants, and estimation. *Statistics* **47** 494–510. MR3060925 https://doi.org/10.1080/02331888.2011.605893

[64] WEISS, C. H. and KIM, H.-Y. (2013). Parameter estimation for binomial AR(1) models with applications in finance and industry. *Statist. Papers* **54** 563–590. MR3072889 https://doi.org/10.1007/s00362-012-0449-y

[65] YILMAZ, E., GEMMEKE, J. F. and HAMME, H. V. (2016). Noise robust exemplar matching with alpha–beta divergence. *Speech Commun.* **76** 127–142.

[66] ZHITLUKHIN, M. V., MURAVLEV, A. A. and SHIRYAEV, A. N. (2013). The optimal decision rule in the Kiefer–Weiss problem for a Brownian motion. *Russian Math. Surveys* **68** 389–391.

[67] ZOUBIR, A. M., KOIVUNEN, V., CHAKHCHOUKH, Y. and MUMA, M. (2012). Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts. *IEEE Signal Process. Mag.* **29** 61–80.

[68] ZOUBIR, A. M., KOIVUNEN, V., OLLILA, E. and MUMA, M. (2018). *Robust Statistics for Signal Processing*. Cambridge Univ. Press, Cambridge, UK.

*The Annals of Statistics*
2020, Vol. 48, No. 5, 2622–2645
https://doi.org/10.1214/19-AOS1900


# TEST OF SIGNIFICANCE FOR HIGH-DIMENSIONAL LONGITUDINAL DATA

BY ETHAN X. FANG[1,*], YANG NING[2] AND RUNZE LI[1,**]

[1]*Department of Statistics, Pennsylvania State University,* [*]*xxf13@psu.edu;* [**]*rzli@psu.edu*
[2]*Department of Statistics and Data Science, Cornell University, yn265@cornell.edu*

This paper concerns statistical inference for longitudinal data with ultrahigh dimensional covariates. We first study the problem of constructing confidence intervals and hypothesis tests for a low-dimensional parameter of interest. The major challenge is how to construct a powerful test statistic in the presence of high-dimensional nuisance parameters and sophisticated within-subject correlation of longitudinal data. To deal with the challenge, we propose a new quadratic decorrelated inference function approach which simultaneously removes the impact of nuisance parameters and incorporates the correlation to enhance the efficiency of the estimation procedure. When the parameter of interest is of fixed dimension, we prove that the proposed estimator is asymptotically normal and attains the semiparametric information bound, based on which we can construct an optimal Wald test statistic. We further extend this result and establish the limiting distribution of the estimator under the setting with the dimension of the parameter of interest growing with the sample size at a polynomial rate. Finally, we study how to control the false discovery rate (FDR) when a vector of high-dimensional regression parameters is of interest. We prove that applying the Storey (*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** (2002) 479–498) procedure to the proposed test statistics for each regression parameter controls FDR asymptotically in longitudinal data. We conduct simulation studies to assess the finite sample performance of the proposed procedures. Our simulation results imply that the newly proposed procedure can control both Type I error for testing a low dimensional parameter of interest and the FDR in the multiple testing problem. We also apply the proposed procedure to a real data example.

## REFERENCES

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

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

BENJAMINI, Y. and 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

BENTKUS, V. (2003). On the dependence of the Berry–Esseen bound on dimension. *J. Statist. Plann. Inference* **113** 385–402. MR1965117 https://doi.org/10.1016/S0378-3758(02)00094-0

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

DONG, C., WANG, S., LI, W.-D., LI, D., ZHAO, H. and PRICE, R. A. (2003). Interacting genetic loci on chromosomes 20 and 10 influence extreme human obesity. *Am. J. Hum. Genet.* **72** 115–124.

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

FAN, J., LIU, H., SUN, Q. and ZHANG, T. (2018). I-LAMM for sparse learning: Simultaneous control of algorithmic complexity and statistical error. *Ann. Statist.* **46** 814–841. MR3782385 https://doi.org/10.1214/17-AOS1568

*MSC2020 subject classifications.* Primary 62F03; secondary 62F05.
*Key words and phrases.* False discovery rate, generalized estimating equation, quadratic inference function.

FANG, E. X., NING, Y. and LI, R. (2020). Supplement to "Test of significance for high-dimensional longitudinal data." https://doi.org/10.1214/19-AOS1900SUPP.

FANG, E. X., NING, Y. and LIU, H. (2017). Testing and confidence intervals for high dimensional proportional hazards models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1415–1437. MR3731669 https://doi.org/10.1111/rssb.12224

GRAZIER G'SELL, M., WAGER, S., CHOULDECHOVA, A. and TIBSHIRANI, R. (2016). Sequential selection procedures and false discovery rate control. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 423–444. MR3454203 https://doi.org/10.1111/rssb.12122

JAQUISH, C. E. (2007). The framingham heart study, on its way to becoming the gold standard for cardiovascular genetic epidemiology? *BMC Med. Genet.* **8** 63. https://doi.org/10.1186/1471-2350-8-63

JAVANMARD, A. and MONTANARI, A. (2013). Confidence intervals and hypothesis testing for high-dimensional statistical models. In *Advances in Neural Information Processing Systems* 1187–1195.

LIANG, K. Y. and ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73** 13–22. MR0836430 https://doi.org/10.1093/biomet/73.1.13

LIU, W. (2013). Gaussian graphical model estimation with false discovery rate control. *Ann. Statist.* **41** 2948–2978. MR3161453 https://doi.org/10.1214/13-AOS1169

LOH, P.-L. and WAINWRIGHT, M. J. (2013). Regularized M-estimators with nonconvexity: Statistical and algorithmic theory for local optima. In *Advances in Neural Information Processing Systems* 476–484.

MA, S., SONG, Q. and WANG, L. (2013). Simultaneous variable selection and estimation in semiparametric modeling of longitudinal/clustered data. *Bernoulli* **19** 252–274. MR3019494 https://doi.org/10.3150/11-BEJ386

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

QU, A., LINDSAY, B. G. and LI, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika* **87** 823–836. MR1813977 https://doi.org/10.1093/biomet/87.4.823

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

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

VAN DE GEER, S. and MÜLLER, P. (2012). Quasi-likelihood and/or robust estimation in high dimensions. *Statist. Sci.* **27** 469–480. MR3025129 https://doi.org/10.1214/12-STS397

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

WANG, L. (2011). GEE analysis of clustered binary data with diverging number of covariates. *Ann. Statist.* **39** 389–417. MR2797851 https://doi.org/10.1214/10-AOS846

WANG, L., KIM, Y. and LI, R. (2013). Calibrating nonconvex penalized regression in ultra-high dimension. *Ann. Statist.* **41** 2505–2536. MR3127873 https://doi.org/10.1214/13-AOS1159

WANG, L. and QU, A. (2009). Consistent model selection and data-driven smooth tests for longitudinal data in the estimating equations approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 177–190. MR2655529 https://doi.org/10.1111/j.1467-9868.2008.00679.x

WANG, L., ZHOU, J. and QU, A. (2012). Penalized generalized estimating equations for high-dimensional longitudinal data analysis. *Biometrics* **68** 353–360. MR2959601 https://doi.org/10.1111/j.1541-0420.2011.01678.x

WANG, L., XUE, L., QU, A. and LIANG, H. (2014). Estimation and model selection in generalized additive partial linear models for correlated data with diverging number of covariates. *Ann. Statist.* **42** 592–624. MR3210980 https://doi.org/10.1214/13-AOS1194

XUE, L., QU, A. and ZHOU, J. (2010). Consistent model selection for marginal generalized additive model for correlated data. *J. Amer. Statist. Assoc.* **105** 1518–1530. Supplementary materials available online. MR2796568 https://doi.org/10.1198/jasa.2010.tm10128

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

ZHAO, T., LIU, H. and ZHANG, T. (2018). Pathwise coordinate optimization for sparse learning: Algorithm and theory. *Ann. Statist.* **46** 180–218. MR3766950 https://doi.org/10.1214/17-AOS1547

*The Annals of Statistics*
2020, Vol. 48, No. 5, 2646–2670
https://doi.org/10.1214/19-AOS1901


# GEOMETRIZING RATES OF CONVERGENCE UNDER LOCAL DIFFERENTIAL PRIVACY CONSTRAINTS

BY ANGELIKA ROHDE[1] AND LUKAS STEINBERGER[2]

[1]*University of Freiburg, angelika.rohde@stochastik.uni-freiburg.de*
[2]*Department of Statistics and OR, University of Vienna, lukas.steinberger@univie.ac.at*

We study the problem of estimating a functional $\theta(\mathbb{P})$ of an unknown probability distribution $\mathbb{P} \in \mathcal{P}$ in which the original iid sample $X_1, \dots, X_n$ is kept private even from the statistician via an $\alpha$-local differential privacy constraint. Let $\omega_{\mathrm{TV}}$ denote the modulus of continuity of the functional $\theta$ over $\mathcal{P}$ with respect to total variation distance. For a large class of loss functions $l$ and a fixed privacy level $\alpha$, we prove that the privatized minimax risk is equivalent to $l(\omega_{\mathrm{TV}}(n^{-1/2}))$ to within constants, under regularity conditions that are satisfied, in particular, if $\theta$ is linear and $\mathcal{P}$ is convex. Our results complement the theory developed by Donoho and Liu (1991) with the nowadays highly relevant case of privatized data. Somewhat surprisingly, the difficulty of the estimation problem in the private case is characterized by $\omega_{\mathrm{TV}}$, whereas, it is characterized by the Hellinger modulus of continuity if the original data $X_1, \dots, X_n$ are available. We also find that for locally private estimation of linear functionals over a convex model a simple sample mean estimator, based on independently and binary privatized observations, always achieves the minimax rate. We further provide a general recipe for choosing the functional parameter in the optimal binary privatization mechanisms and illustrate the general theory in numerous examples. Our theory allows us to quantify the price to be paid for local differential privacy in a large class of estimation problems. This price appears to be highly problem specific.

## REFERENCES

AWAN, J. and SLAVKOVIĆ, A. (2018). Differentially private uniformly most powerful tests for binomial data. In *Advances in Neural Information Processing Systems* 4212–4222.

DINUR, I. and NISSIM, K. (2003). Revealing information while preserving privacy. In *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* 202–210. ACM, New York.

DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence. II. *Ann. Statist.* **19** 633–667. MR1105839 https://doi.org/10.1214/aos/1176348114

DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2013a). Local privacy and statistical minimax rates. In 2013 *IEEE* 54*th Annual Symposium on Foundations of Computer Science—FOCS* 2013 429–438. IEEE Computer Soc., Los Alamitos, CA. MR3246246 https://doi.org/10.1109/FOCS.2013.53

DUCHI, J., JORDAN, M. I. and WAINWRIGHT, M. J. (2013b). Local privacy and minimax bounds: Sharp rates for probability estimation. In *Adv. Neural Inf. Process. Syst.* 1529–1537.

DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2014). Local privacy, data processing inequalities, and statistical minimax rates. Preprint. Available at arXiv:1302.3203.

DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2018). Minimax optimal procedures for locally private estimation. *J. Amer. Statist. Assoc.* **113** 182–201. MR3803452 https://doi.org/10.1080/01621459.2017.1389735

DUCHI, J. and ROGERS, R. (2019). Lower bounds for locally private estimation via communication complexity. Preprint. Available at arXiv:1902.00582.

DWORK, C. (2008). Differential privacy: A survey of results. In *Theory and Applications of Models of Computation. Lecture Notes in Computer Science* **4978** 1–19. Springer, Berlin. MR2472670 https://doi.org/10.1007/978-3-540-79228-4_1

*MSC2020 subject classifications.* Primary 62G05; secondary 62C20.
*Key words and phrases.* Local differential privacy, minimax estimation, rate of convergence, moduli of continuity, nonparametric estimation.

DWORK, C. and NISSIM, K. (2004). Privacy-preserving datamining on vertically partitioned databases. In *Advances in Cryptology—CRYPTO* 2004. *Lecture Notes in Computer Science* **3152** 528–544. Springer, Berlin. MR2147523 https://doi.org/10.1007/978-3-540-28628-8_32

DWORK, C. and SMITH, A. (2010). Differential privacy for statistics: What we know and what we want to learn. *J. Priv. Confid.* **1** 2.

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

EVFIMIEVSKI, A., GEHRKE, J. and SRIKANT, R. (2003). Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* 211–222. ACM, New York.

GENG, Q. and VISWANATH, P. (2016). The optimal noise-adding mechanism in differential privacy. *IEEE Trans. Inform. Theory* **62** 925–951. MR3455907 https://doi.org/10.1109/TIT.2015.2504967

KRAFFT, O. and WITTING, H. (1967). Optimale Tests und ungünstigste Verteilungen. *Z. Wahrsch. Verw. Gebiete* **7** 289–302. MR0217929 https://doi.org/10.1007/BF01844447

KRAFT, C. (1955). Some conditions for consistency and uniform consistency of statistical procedures. *Univ. California Publ. Statist.* **2** 125–141. MR0073896

LECAM, L. (1973). Convergence of estimates under dimensionality restrictions. *Ann. Statist.* **1** 38–53. MR0334381

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

ROHDE, A. and STEINBERGER, L. (2020). Supplement to "Geometrizing rates of convergence under local differential privacy constraints." https://doi.org/10.1214/19-AOS1901SUPP.

SION, M. (1958). On general minimax theorems. *Pacific J. Math.* **8** 171–176. MR0097026

SMITH, A. (2008). Efficient, differentially private point estimators. Preprint. Available at arXiv:0809.4794.

SMITH, A. (2011). Privacy-preserving statistical estimation with optimal convergence rates [extended abstract]. In *STOC'*11*—Proceedings of the* 43*rd ACM Symposium on Theory of Computing* 813–821. ACM, New York. MR2932032 https://doi.org/10.1145/1993636.1993743

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

WARNER, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc.* **60** 63–69.

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

YE, M. and BARG, A. (2017). Asymptotically optimal private estimation under mean square loss. Preprint. Available at arXiv:1708.00059.

*The Annals of Statistics*
2020, Vol. 48, No. 5, 2671–2697
https://doi.org/10.1214/19-AOS1902


# ADDITIVE REGRESSION WITH HILBERTIAN RESPONSES

BY JEONG MIN JEON* AND BYEONG U. PARK**

*Department of Statistics, Seoul National University,* *jeongmin.jeon@kuleuven.be; **bupark@stats.snu.ac.kr

This paper develops a foundation of methodology and theory for the estimation of structured nonparametric regression models with Hilbertian responses. Our method and theory are focused on the additive model, while the main ideas may be adapted to other structured models. For this, the notion of Bochner integration is introduced for Banach-space-valued maps as a generalization of Lebesgue integration. Several statistical properties of Bochner integrals, relevant for our method and theory and also of importance in their own right, are presented for the first time. Our theory is complete. The existence of our estimators and the convergence of a practical algorithm that evaluates the estimators are established. These results are nonasymptotic as well as asymptotic. Furthermore, it is proved that the estimators achieve the univariate rates in pointwise, $L^2$ and uniform convergence, and that the estimators of the component maps converge jointly in distribution to Gaussian random elements. Our numerical examples include the cases of functional, density-valued and simplex-valued responses, demonstrating the validity of our approach.

## REFERENCES

[1] ANEIROS, G., CAO, R., FRAIMAN, R., GENEST, C. and VIEU, P. (2019). Recent advances in functional data analysis and high-dimensional statistics. *J. Multivariate Anal.* **170** 3–9. MR3913024 https://doi.org/10.1016/j.jmva.2018.11.007

[2] 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

[3] BISSANTZ, N., DETTE, H., HILDEBRANDT, T. and BISSANTZ, K. (2016). Smooth backfitting in additive inverse regression. *Ann. Inst. Statist. Math.* **68** 827–853. MR3520045 https://doi.org/10.1007/s10463-015-0517-x

[4] 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

[5] BUSBY, R. C., SCHOCHETMAN, I. and SMITH, H. A. (1972). Integral operators and the compactness of induced representations. *Trans. Amer. Math. Soc.* **164** 461–477. MR0295099 https://doi.org/10.2307/1995990

[6] CHIOU, J.-M., MÜLLER, H.-G. and WANG, J.-L. (2003). Functional quasi-likelihood regression models with smooth random effects. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 405–423. MR1983755 https://doi.org/10.1111/1467-9868.00393

[7] COHN, D. L. (2013). *Measure Theory*, 2nd ed. *Birkhäuser Advanced Texts*: *Basler Lehrbücher*. [*Birkhäuser Advanced Texts*: *Basel Textbooks*]. Birkhäuser/Springer, New York. MR3098996 https://doi.org/10.1007/978-1-4614-6956-8

[8] CONWAY, J. B. (1985). *A Course in Functional Analysis*. *Graduate Texts in Mathematics* **96**. Springer, New York. MR0768926 https://doi.org/10.1007/978-1-4757-3828-5

[9] DABO-NIANG, S. and RHOMARI, N. (2009). Kernel regression estimation in a Banach space. *J. Statist. Plann. Inference* **139** 1421–1434. MR2485136 https://doi.org/10.1016/j.jspi.2008.06.015

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

---

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

*Key words and phrases.* Additive models, smooth backfitting, Bochner integral, non-Euclidean data, infinite-dimensional spaces, Hilbert spaces, functional responses.

[11] DURRETT, R. (2019). *Probability—Theory and Examples*. *Cambridge Series in Statistical and Probabilistic Mathematics* **49**. Cambridge Univ. Press, Cambridge. Fifth edition of [MR1068527]. MR3930614 https://doi.org/10.1017/9781108591034

[12] FERRATY, F., LAKSACI, A., TADJ, A. and VIEU, P. (2011). Kernel regression with functional response. *Electron*. *J*. *Stat*. **5** 159–171. MR2786486 https://doi.org/10.1214/11-EJS600

[13] FERRATY, F., VAN KEILEGOM, I. and VIEU, P. (2012). Regression when both response and predictor are functions. *J*. *Multivariate Anal*. **109** 10–28. MR2922850 https://doi.org/10.1016/j.jmva.2012.02.008

[14] FERRATY, F. and VIEU, P. (2009). Additive prediction and boosting for functional data. *Comput*. *Statist*. *Data Anal*. **53** 1400–1413. MR2657100 https://doi.org/10.1016/j.csda.2008.11.023

[15] GOLDBERG, S. (1959). Some properties of the space of compact operators on a Hilbert space. *Math*. *Ann*. **138** 329–331. MR0121673 https://doi.org/10.1007/BF01344152

[16] HAN, K., MÜLLER, H.-G. and PARK, B. U. (2018). Smooth backfitting for additive modeling with small errors-in-variables, with an application to additive functional regression for multiple predictor functions. *Bernoulli* **24** 1233–1265. MR3706793 https://doi.org/10.3150/16-BEJ898

[17] HAN, K. and PARK, B. U. (2018). Smooth backfitting for errors-in-variables additive models. *Ann*. *Statist*. **46** 2216–2250. MR3845016 https://doi.org/10.1214/17-AOS1617

[18] JAYASUMANA, S., SALZMANN, M., LI, H. and HARANDI, M. (2013). A framework for shape analysis via Hilbert space embedding. In *Proceedings of the IEEE International Conference on Computer Vision* 1249–1256.

[19] JEON, J. M. and PARK, B. U. (2020). Supplement to "Additive regression with Hilbertian responses." https://doi.org/10.1214/19-AOS1902SUPP.

[20] JIANG, C.-R. and WANG, J.-L. (2011). Functional single index models for longitudinal data. *Ann*. *Statist*. **39** 362–388. MR2797850 https://doi.org/10.1214/10-AOS845

[21] KALLENBERG, O. (1997). *Foundations of Modern Probability*. *Probability and Its Applications* (*New York*). Springer, New York. MR1464694

[22] KARA-ZAITRI, L., LAKSACI, A., RACHDI, M. and VIEU, P. (2017). Uniform in bandwidth consistency for various kernel estimators involving functional data. *J*. *Nonparametr*. *Stat*. **29** 85–107. MR3597219 https://doi.org/10.1080/10485252.2016.1254780

[23] KUNDU, S., MAJUMDAR, S. and MUKHERJEE, K. (2000). Central limit theorems revisited. *Statist*. *Probab*. *Lett*. **47** 265–275. MR1747487 https://doi.org/10.1016/S0167-7152(99)00164-9

[24] LEADBETTER, R., CAMBANIS, S. and PIPIRAS, V. (2014). *A Basic Course in Measure and Probability*: *Theory for Applications*. Cambridge Univ. Press, Cambridge. MR3445362

[25] LEE, Y. K., MAMMEN, E. and PARK, B. U. (2010). Backfitting and smooth backfitting for additive quantile models. *Ann*. *Statist*. **38** 2857–2883. MR2722458 https://doi.org/10.1214/10-AOS808

[26] LEE, Y. K., MAMMEN, E. and PARK, B. U. (2012). Flexible generalized varying coefficient regression models. *Ann*. *Statist*. **40** 1906–1933. MR3015048 https://doi.org/10.1214/12-AOS1026

[27] LIAN, H. (2011). Convergence of functional $k$-nearest neighbor regression estimate with functional responses. *Electron*. *J*. *Stat*. **5** 31–40. MR2773606 https://doi.org/10.1214/11-EJS595

[28] LIAN, H. (2012). Convergence of nonparametric functional regression estimates with functional responses. *Electron*. *J*. *Stat*. **6** 1373–1391. MR2988451 https://doi.org/10.1214/12-EJS716

[29] LING, N. and VIEU, P. (2018). Nonparametric modelling for functional data: Selected survey and tracks for future. *Statistics* **52** 934–949. MR3827427 https://doi.org/10.1080/02331888.2018.1487120

[30] LINTON, O., SPERLICH, S. and VAN KEILEGOM, I. (2008). Estimation of a semiparametric transformation model. *Ann*. *Statist*. **36** 686–718. MR2396812 https://doi.org/10.1214/009053607000000848

[31] MAMMEN, E., LINTON, O. and NIELSEN, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Ann*. *Statist*. **27** 1443–1490. MR1742496 https://doi.org/10.1214/aos/1017939137

[32] MAMMEN, E. and PARK, B. U. (2005). Bandwidth selection for smooth backfitting in additive models. *Ann*. *Statist*. **33** 1260–1294. MR2195635 https://doi.org/10.1214/009053605000000101

[33] MAMMEN, E., PARK, B. U. and SCHIENLE, M. (2014). Additive models: Extensions and related models. In *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics* 176–211. Oxford Univ. Press, Oxford. MR3306926

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

[35] MUANDET, K., SRIPERUMBUDUR, B., FUKUMIZU, K., GRETTON, A. and SCHÖLKOPF, B. (2016). Kernel mean shrinkage estimators. *J*. *Mach*. *Learn*. *Res*. **17** Paper No. 48, 41. MR3504608

[36] MÜLLER, H.-G. and YAO, F. (2008). Functional additive models. *J*. *Amer*. *Statist*. *Assoc*. **103** 1534–1544. MR2504202 https://doi.org/10.1198/016214508000000751

[37] PARK, B. U., CHEN, C.-J., TAO, W. and MÜLLER, H.-G. (2018). Singular additive models for function to function regression. *Statist*. *Sinica* **28** 2497–2520. MR3839871

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

[39] PINI, A., STAMM, A. and VANTINI, S. (2017). Hotelling in Wonderland. In *Functional Statistics and Related Fields*. *Contrib. Stat.* 211–216. Springer, Cham. MR3823575

[40] SCHEIPL, F., STAICU, A.-M. and GREVEN, S. (2015). Functional additive mixed models. *J. Comput. Graph. Statist.* **24** 477–501. MR3357391 https://doi.org/10.1080/10618600.2014.901914

[41] TALSKÁ, R., MENAFOGLIO, A., MACHALOVÁ, J., HRON, K. and FIŠEROVÁ, E. (2018). Compositional regression with functional response. *Comput. Statist. Data Anal.* **123** 66–85. MR3777086 https://doi.org/10.1016/j.csda.2018.01.018

[42] TSAGRIS, M. (2015). Regression analysis with compositional data containing zero values. *Chil. J. Stat.* **6** 47–57. MR3407274

[43] VAN DEN BOOGAART, K. G., EGOZCUE, J. J. and PAWLOWSKY-GLAHN, V. (2014). Bayes Hilbert spaces. *Aust. N. Z. J. Stat.* **56** 171–194. MR3226435 https://doi.org/10.1111/anzs.12074

[44] VÄTH, M. (2000). *Volterra and Integral Equations of Vector Functions*. *Monographs and Textbooks in Pure and Applied Mathematics* **224**. Dekker, New York. MR1738341

[45] XU, J. and ZIKATANOV, L. (2002). The method of alternating projections and the method of subspace corrections in Hilbert space. *J. Amer. Math. Soc.* **15** 573–597. MR1896233 https://doi.org/10.1090/S0894-0347-02-00398-3

[46] YU, K., PARK, B. U. and MAMMEN, E. (2008). Smooth backfitting in generalized additive models. *Ann. Statist.* **36** 228–260. MR2387970 https://doi.org/10.1214/009053607000000596

[47] ZHANG, X., PARK, B. U. and WANG, J.-L. (2013). Time-varying additive models for longitudinal data. *J. Amer. Statist. Assoc.* **108** 983–998. MR3174678 https://doi.org/10.1080/01621459.2013.778776

[48] ZHU, H., LI, R. and KONG, L. (2012). Multivariate varying coefficient model for functional responses. *Ann. Statist.* **40** 2634–2666. MR3097615 https://doi.org/10.1214/12-AOS1045

*The Annals of Statistics*
2020, Vol. 48, No. 5, 2698–2727
https://doi.org/10.1214/19-AOS1903


# NONPARAMETRIC BAYESIAN ESTIMATION FOR MULTIVARIATE HAWKES PROCESSES

BY SOPHIE DONNET[1], VINCENT RIVOIRARD[2,*] AND JUDITH ROUSSEAU[2,**]

[1]*INRA, Université Paris Saclay, sophie.donnet@inra.fr*
[2]*CEREMADE, CNRS, UMR 7534, Université Paris–Dauphine, *Vincent.Rivoirard@dauphine.fr;*
*** *Judith.Rousseau@stats.ox.ac.uk*

This paper studies nonparametric estimation of parameters of multivariate Hawkes processes. We consider the Bayesian setting and derive posterior concentration rates. First, rates are derived for $\mathbb{L}_1$-metrics for stochastic intensities of the Hawkes process. We then deduce rates for the $\mathbb{L}_1$-norm of interactions functions of the process. Our results are exemplified by using priors based on piecewise constant functions, with regular or random partitions and priors based on mixtures of Betas distributions. We also present a simulation study to illustrate our results and to study empirically the inference on functional connectivity graphs of neurons

## REFERENCES

[1] AÏT-SAHALIA, Y., CACHO-DIAZ, J. and LAEVEN, R. J. (2015). Modeling financial contagion using mutually exciting jump processes. *J. Financ. Econ.* **117** 585–606.

[2] ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes. Springer Series in Statistics*. Springer, New York. MR1198884 https://doi.org/10.1007/978-1-4612-4348-9

[3] BACRY, E., DAYRI, K. and MUZY, J. F. (2012). Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data. *Eur. Phys. J. B* **85** 157.

[4] BACRY, E., DELATTRE, S., HOFFMANN, M. and MUZY, J. F. (2013). Modelling microstructure noise with mutually exciting point processes. *Quant. Finance* **13** 65–77. MR3005350 https://doi.org/10.1080/14697688.2011.647054

[5] BACRY, E., GAÏFFAS, S. and MUZY, J.-F. (2015). A generalization error bound for sparse and low-rank multivariate Hawkes processes. ArXiv e-prints.

[6] BACRY, E., JAISSON, T. and MUZY, J. (2016). Estimation of slowly decreasing Hawkes kernels: Application to high-frequency order book dynamics. *Quant. Finance* **16** 1179–1201. MR3520321 https://doi.org/10.1080/14697688.2015.1123287

[7] BACRY, E., MASTROMATTEO, I. and MUZY, J.-F. (2015). Hawkes processes in finance. *Mark. Microstruct. Liq.* **1** 1550005.

[8] BACRY, E. and MUZY, J.-F. (2016). First- and second-order statistics characterization of Hawkes processes and non-parametric estimation. *IEEE Trans. Inform. Theory* **62** 2184–2202. MR3480107 https://doi.org/10.1109/TIT.2016.2533397

[9] BLUNDELL, C., BECK, J. and HELLER, K. A. (2012). Modelling reciprocating relationships with Hawkes processes. In *Advances in Neural Information Processing Systems* (F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger, eds.) **25** 2600–2608. Curran Associates, Red Hook.

[10] BRÉMAUD, P. and MASSOULIÉ, L. (1996). Stability of nonlinear Hawkes processes. *Ann. Probab.* **24** 1563–1588. MR1411506 https://doi.org/10.1214/aop/1065725193

[11] BRILLINGER, D. R. (1988). Maximum likelihood analysis of spike trains of interacting nerve cells. *Biol. Cybernet.* **59** 189–200. https://doi.org/10.1007/bf00318010

[12] CARSTENSEN, L., SANDELIN, A., WINTHER, O. and HANSEN, N. R. (2010). Multivariate Hawkes process models of the occurrence of regulatory elements. *BMC Bioinform.* **11** 456. https://doi.org/10.1186/1471-2105-11-456

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

*MSC2020 subject classifications.* Primary 62G20, 60G55; secondary 62G05.
*Key words and phrases.* Multivariate counting process, Hawkes processes, nonparametric Bayesian estimation, posterior concentration rates.

[14] CHEN, S., SHOJAIE, A., SHEA-BROWN, E. and WITTEN, D. (2017). The multivariate Hawkes process in high dimensions: Beyond mutual excitation. ArXiv e-prints.

[15] CHEN, S., WITTEN, D. and SHOJAIE, A. (2017). Nearly assumptionless screening for the mutually-exciting multivariate Hawkes process. *Electron. J. Stat.* **11** 1207–1234. MR3634334 https://doi.org/10.1214/17-EJS1251

[16] CHORNOBOY, E. S., SCHRAMM, L. P. and KARR, A. F. (1988). Maximum likelihood identification of neural point process systems. *Biol. Cybernet.* **59** 265–275. MR0961117 https://doi.org/10.1007/BF00332915

[17] CRANE, R. and SORNETTE, D. (2008). Robust dynamic classes revealed by measuring the response function of a social system. *Proc. Natl. Acad. Sci. USA* **105** 15649–15653.

[18] DALEY, D. J. and VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes. Vol. I: Elementary Theory and Methods*, 2nd ed. *Probability and Its Applications* (*New York*). Springer, New York. MR1950431

[19] DONNET, S., RIVOIRARD, V. and ROUSSEAU, J. (2020). Supplement to "Nonparametric Bayesian estimation for multivariate Hawkes processes." https://doi.org/10.1214/19-AOS1903SUPP.

[20] EMBRECHTS, P., LINIGER, T. and LIN, L. (2011). Multivariate Hawkes processes: An application to financial data. *J. Appl. Probab.* **48A** 367–378. MR2865638 https://doi.org/10.1239/jap/1318940477

[21] GHOSAL, S., GHOSH, J. K. and VAN DER VAART, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531. MR1790007 https://doi.org/10.1214/aos/1016218228

[22] GHOSAL, S. and VAN DER VAART, A. (2007). Convergence rates of posterior distributions for non-i.i.d. observations. *Ann. Statist.* **35** 192–223. MR2332274 https://doi.org/10.1214/009053606000001172

[23] GHOSAL, S. and VAN DER VAART, A. (2007). Posterior convergence rates of Dirichlet mixtures at smooth densities. *Ann. Statist.* **35** 697–723. MR2336864 https://doi.org/10.1214/009053606000001271

[24] GREEN, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82** 711–732. MR1380810 https://doi.org/10.1093/biomet/82.4.711

[25] GUSTO, G. and SCHBATH, S. (2005). FADO: A statistical method to detect favored or avoided distances between occurrences of motifs using the Hawkes' model. *Stat. Appl. Genet. Mol. Biol.* **4** Art. 24, 28. MR2170440 https://doi.org/10.2202/1544-6115.1119

[26] HANSEN, N. R., REYNAUD-BOURET, P. and RIVOIRARD, V. (2015). Lasso and probabilistic inequalities for multivariate point processes. *Bernoulli* **21** 83–143. MR3322314 https://doi.org/10.3150/13-BEJ562

[27] HAWKES, A. G. (1971). Point spectra of some mutually exciting point processes. *J. Roy. Statist. Soc. Ser. B* **33** 438–443. MR0358976

[28] HAWKES, A. G. and OAKES, D. (1974). A cluster process representation of a self-exciting process. *J. Appl. Probab.* **11** 493–503. MR0378093 https://doi.org/10.2307/3212693

[29] LAMBERT, R. C., TULEAU-MALOT, C., BESSAIH, T., RIVOIRARD, V., BOURET, Y., LERESCHE, N. and REYNAUD-BOURET, P. (2018). Reconstructing the functional connectivity of multiple spike trains using Hawkes models. *J. Neurosci. Methods* **297** 9–21. https://doi.org/10.1016/j.jneumeth.2017.12.026

[30] LI, L. and ZHA, H. (2014). Learning parametric models for social infectivity in multi-dimensional Hawkes processes. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, *AAAI'*14 101–107. AAAI Press, Menlo Park.

[31] MITCHELL, L. and CATES, M. E. (2010). Hawkes process as a model of social interactions: A view on video dynamics. *J. Phys. A* **43** 045101, 11. MR2578723 https://doi.org/10.1088/1751-8113/43/4/045101

[32] MOHLER, G. O., SHORT, M. B., BRANTINGHAM, P. J., SCHOENBERG, F. P. and TITA, G. E. (2011). Self-exciting point process modeling of crime. *J. Amer. Statist. Assoc.* **106** 100–108. MR2816705 https://doi.org/10.1198/jasa.2011.ap09546

[33] OGATA, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.* **83** 9–27.

[34] OKATAN, M., WILSON, M. A. and BROWN, E. N. (2005). Analyzing functional connectivity using a network likelihood model of ensemble neural spiking activity. *Neural Comput.* **17** 1927–1961.

[35] PANINSKI, L., PILLOW, J. and LEWI, J. (2007). Statistical models for neural encoding, decoding, and optimal stimulus design. *Prog. Brain Res.* **165** 493–507. https://doi.org/10.1016/S0079-6123(06)65031-0

[36] PILLOW, J. W., SHLENS, J., PANINSKI, L., SHER, A., LITKE, A. M., CHICHILNISKY, E. and SIMONCELLI, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* **454** 995–999.

[37] PORTER, M. D. and WHITE, G. (2012). Self-exciting hurdle models for terrorist activity. *Ann. Appl. Stat.* **6** 106–124. MR2951531 https://doi.org/10.1214/11-AOAS513

[38] RASMUSSEN, J. G. (2013). Bayesian inference for Hawkes processes. *Methodol. Comput. Appl. Probab.* **15** 623–642. MR3085883 https://doi.org/10.1007/s11009-011-9272-5

[39] REYNAUD-BOURET, P., RIVOIRARD, V., GRAMMONT, F. and TULEAU-MALOT, C. (2014). Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *J. Math. Neurosci.* **4** Art. 3, 41. MR3197017 https://doi.org/10.1186/2190-8567-4-3

[40] REYNAUD-BOURET, P., RIVOIRARD, V. and TULEAU-MALOT, C. (2013). Inference of functional connectivity in neurosciences via Hawkes processes. In *Global Conference on Signal and Information Processing* (*GlobalSIP*), 2013 *IEEE* 317–320. IEEE.

[41] REYNAUD-BOURET, P. and ROY, E. (2006). Some non asymptotic tail estimates for Hawkes processes. *Bull. Belg. Math. Soc. Simon Stevin* **13** 883–896. MR2293215

[42] REYNAUD-BOURET, P. and SCHBATH, S. (2010). Adaptive estimation for Hawkes processes; application to genome analysis. *Ann. Statist.* **38** 2781–2822. MR2722456 https://doi.org/10.1214/10-AOS806

[43] ROUSSEAU, J. (2010). Rates of convergence for the posterior distributions of mixtures of betas and adaptive nonparametric estimation of the density. *Ann. Statist.* **38** 146–180. MR2589319 https://doi.org/10.1214/09-AOS703

[44] SIMMA, A. and JORDAN, M. I. (2012). Modeling events with cascades of Poisson processes. ArXiv e-prints.

[45] VERE-JONES, D. and OZAKI, T. (1982). Some examples of statistical estimation applied to earthquake data I: Cyclic Poisson and self-exciting models. *Ann. Inst. Statist. Math.* **34** 189–207.

[46] YANG, S.-H. and ZHA, H. (2013). Mixture of mutually exciting processes for viral diffusion. *ICML* (2) **28** 1–9.

[47] ZHOU, K., ZHA, H. and SONG, L. (2013). Learning triggering kernels for multi-dimensional Hawkes processes. In *Proceedings of the* 30*th International Conference on Machine Learning* (*ICML*-13) (S. Dasgupta and D. Mcallester, eds.). *JMLR Workshop and Conference Proceedings* **28** 1301–1309.

[48] ZHUANG, J., OGATA, Y. and VERE-JONES, D. (2002). Stochastic declustering of space–time earthquake occurrences. *J. Amer. Statist. Assoc.* **97** 369–380. MR1941459 https://doi.org/10.1198/016214502760046925

*The Annals of Statistics*
2020, Vol. 48, No. 5, 2728–2758
https://doi.org/10.1214/19-AOS1904


# HYPOTHESIS TESTING FOR HIGH-DIMENSIONAL TIME SERIES VIA SELF-NORMALIZATION

BY RUNMIN WANG* AND XIAOFENG SHAO**

*Department of Statistics, University of Illinois at Urbana-Champaign,* *rwang52@illinois.edu; **xshao@illinois.edu

Self-normalization has attracted considerable attention in the recent literature of time series analysis, but its scope of applicability has been limited to low-/fixed-dimensional parameters for low-dimensional time series. In this article, we propose a new formulation of self-normalization for inference about the mean of high-dimensional stationary processes. Our original test statistic is a U-statistic with a trimming parameter to remove the bias caused by weak dependence. Under the framework of nonlinear causal processes, we show the asymptotic normality of our U-statistic with the convergence rate dependent upon the order of the Frobenius norm of the long-run covariance matrix. The self-normalized test statistic is then constructed on the basis of recursive subsampled U-statistics and its limiting null distribution is shown to be a functional of time-changed Brownian motion, which differs from the pivotal limit used in the low-dimensional setting. An interesting phenomenon associated with self-normalization is that it works in the high-dimensional context even if the convergence rate of original test statistic is unknown. We also present applications to testing for bandedness of the covariance matrix and testing for white noise for high-dimensional stationary time series and compare the finite sample performance with existing methods in simulation studies. At the root of our theoretical arguments, we extend the martingale approximation to the high-dimensional setting, which could be of independent theoretical interest.

## REFERENCES

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

ANDREWS, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59** 817–858. MR1106513 https://doi.org/10.2307/2938229

AYYALA, D. N., PARK, J. and ROY, A. (2017). Mean vector testing for high-dimensional dependent observations. *J. Multivariate Anal*. **153** 136–155. MR3578843 https://doi.org/10.1016/j.jmva.2016.09.012

BAI, Z. and SARANADASA, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6** 311–329. MR1399305

BASU, S. and MICHAILIDIS, G. (2015). Regularized estimation in sparse high-dimensional time series models. *Ann. Statist*. **43** 1535–1567. MR3357870 https://doi.org/10.1214/15-AOS1315

BICKEL, P. J. and LEVINA, E. (2008). Regularized estimation of large covariance matrices. *Ann. Statist*. **36** 199–227. MR2387969 https://doi.org/10.1214/009053607000000758

BILLINGSLEY, P. (2008). *Probability and Measure*. *Wiley Series in Probability and Mathematical Statistics* Wiley, New York–Chichester–Brisbane. MR0534323

BRILLINGER, D. R. (2001). *Time Series*: *Data Analysis and Theory*. *Classics in Applied Mathematics* **36**. SIAM, Philadelphia, PA. MR1853554 https://doi.org/10.1137/1.9780898719246

CAI, T. T. and JIANG, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist*. **39** 1496–1525. MR2850210 https://doi.org/10.1214/11-AOS879

CAI, T. T., LIU, W. and XIA, Y. (2014). Two-sample test of high dimensional means under dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol*. **76** 349–372. MR3164870 https://doi.org/10.1111/rssb.12034

*MSC2020 subject classifications.* Primary 62H15, 60K35; secondary 62G10, 62G20.

*Key words and phrases.* High-dimensional inference, nonlinear time series, pivotal, self-normalization, U-statistic.

CAI, T. T., ZHANG, C.-H. and ZHOU, H. H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* **38** 2118–2144. MR2676885 https://doi.org/10.1214/09-AOS752

CHANG, J., YAO, Q. and ZHOU, W. (2017). Testing for high-dimensional white noise using maximum cross-correlations. *Biometrika* **104** 111–127. MR3626482 https://doi.org/10.1093/biomet/asw066

CHEN, S. X. and QIN, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. MR2604697 https://doi.org/10.1214/09-AOS716

CHEN, L. and WU, W. B. (2019). Testing for trends in high-dimensional time series. *J. Amer. Statist. Assoc.* **114** 869–881. MR3963187 https://doi.org/10.1080/01621459.2018.1456935

CHEN, S. X., ZHANG, L.-X. and ZHONG, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** 810–819. MR2724863 https://doi.org/10.1198/jasa.2010.tm09560

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

GORDIN, M. I. (1969). The central limit theorem for stationary processes. *Dokl. Akad. Nauk SSSR* **188** 739–741. MR0251785

GREGORY, K. B., CARROLL, R. J., BALADANDAYUTHAPANI, V. and LAHIRI, S. N. (2015). A two-sample test for equality of means in high dimension. *J. Amer. Statist. Assoc.* **110** 837–849. MR3367268 https://doi.org/10.1080/01621459.2014.934826

HALL, P. and HEYDE, C. C. (2014). *Martingale Limit Theory and Its Application: Probability and Mathematical Statistics*. Academic Press, New York–London. MR0624435

HE, Y., XU, G., WU, C. and PAN, W. (2018). Asymptotically independent $u$-statistics in high dimensional testing. Preprint. Available at arXiv:1809.00411.

HOSKING, J. R. M. (1980). The multivariate portmanteau statistic. *J. Amer. Statist. Assoc.* **75** 602–608. MR0590689

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

HUTCHISON, R. M., WOMELSDORF, T., ALLEN, E. A., BANDETTINI, P. A., CALHOUN, V. D., CORBETTA, M., DELLA PENNA, S., DUYN, J. H., GLOVER, G. H. et al. (2013). Dynamic functional connectivity: Promise, issues, and interpretations. *NeuroImage* **80** 360–378.

JENTSCH, C. and POLITIS, D. N. (2015). Covariance matrix estimation and linear process bootstrap for multivariate time series of possibly increasing dimension. *Ann. Statist.* **43** 1117–1140. MR3346699 https://doi.org/10.1214/14-AOS1301

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

KIEFER, N. M., VOGELSANG, T. J. and BUNZEL, H. (2000). Simple robust testing of regression hypotheses. *Econometrica* **68** 695–714. MR1769382 https://doi.org/10.1111/1468-0262.00128

KIM, S., ZHAO, Z. and SHAO, X. (2015). Nonparametric functional central limit theorem for time series regression with application to self-normalized confidence interval. *J. Multivariate Anal.* **133** 277–290. MR3282031 https://doi.org/10.1016/j.jmva.2014.09.017

KITAMURA, Y. (1997). Empirical likelihood methods with weakly dependent processes. *Ann. Statist.* **25** 2084–2102. MR1474084 https://doi.org/10.1214/aos/1069362388

KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217–1241. MR1015147 https://doi.org/10.1214/aos/1176347265

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

LEDOIT, O. and WOLF, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30** 1081–1102. MR1926169 https://doi.org/10.1214/aos/1031689018

LEE, A. J. (1990). *$U$-Statistics: Theory and Practice. Statistics: Textbooks and Monographs* **110**. Dekker, New York. MR1075417

LI, W. K. (2004). *Diagnostic Checks in Time Series*. Chapman and Hall/CRC, Boca Raton.

LI, W. K. and MCLEOD, A. I. (1981). Distribution of the residual autocorrelations in multivariate ARMA time series models. *J. Roy. Statist. Soc. Ser. B* **43** 231–239. MR0626770

LI, Z., LAM, C., YAO, J. and YAO, Q. (2019). On testing for high-dimensional white noise. *Ann. Statist.* **47** 3382–3412. MR4025746 https://doi.org/10.1214/18-AOS1782

LIU, R. Y. and SINGH, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap* (*East Lansing*, *MI*, 1990). *Wiley Ser. Probab. Math. Statist. Probab. Math. Statist.* 225–248. Wiley, New York. MR1197787

LIU, Y., TAMURA, Y. and TANIGUCHI, M. (2018). Asymptotic theory of test statistic for sphericity of high-dimensional time series. *J. Time Series Anal.* **39** 402–416. MR3796526 https://doi.org/10.1111/jtsa.12288

LOBATO, I. N. (2001). Testing that a dependent process is uncorrelated. *J. Amer. Statist. Assoc.* **96** 1066–1076. MR1947254 https://doi.org/10.1198/016214501753208726

LOPES, M., JACOB, L. and WAINWRIGHT, M. J. (2011). A more powerful two-sample test in high dimensions using random projection. In *Advances in Neural Information Processing Systems* 1206–1214.

LÜTKEPOHL, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer, Berlin. MR2172368 https://doi.org/10.1007/978-3-540-27752-1

MCMURRY, T. L. and POLITIS, D. N. (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *J. Time Series Anal.* **31** 471–482. MR2732601 https://doi.org/10.1111/j.1467-9892.2010.00679.x

NEWEY, W. K. and WEST, K. D. (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55** 703–708. MR0890864 https://doi.org/10.2307/1913610

PAUL, D. and WANG, L. (2016). Discussion of "Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation" [MR3466172]. *Electron. J. Stat.* **10** 74–80. MR3466176 https://doi.org/10.1214/15-EJS1019

PHILLIPS, P. C. B. and SOLO, V. (1992). Asymptotics for linear processes. *Ann. Statist.* **20** 971–1001. MR1165602 https://doi.org/10.1214/aos/1176348666

PINI, A., STAMM, A. and VANTINI, S. (2018). Hotelling's $T^2$ in separable Hilbert spaces. *J. Multivariate Anal.* **167** 284–305. MR3830647 https://doi.org/10.1016/j.jmva.2018.05.007

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

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

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

SECCHI, P., STAMM, A. and VANTINI, S. (2013). Inference for the mean of large $p$ small $n$ data: A finite-sample high-dimensional generalization of Hotelling's theorem. *Electron. J. Stat.* **7** 2005–2031. MR3085016 https://doi.org/10.1214/13-EJS833

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

SHAO, X. (2011). A simple test of changes in mean in the possible presence of long-range dependence. *J. Time Series Anal.* **32** 598–606. MR2846559 https://doi.org/10.1111/j.1467-9892.2010.00717.x

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

SHAO, X. and WU, W. B. (2007). Asymptotic spectral theory for nonlinear time series. *Ann. Statist.* **35** 1773–1801. MR2351105 https://doi.org/10.1214/009053606000001479

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

SRIVASTAVA, M. S. (2009). A test for the mean vector with fewer observations than the dimension under non-normality. *J. Multivariate Anal.* **100** 518–532. MR2483435 https://doi.org/10.1016/j.jmva.2008.06.006

SRIVASTAVA, M. S. and DU, M. (2008). A test for the mean vector with fewer observations than the dimension. *J. Multivariate Anal.* **99** 386–402. MR2396970 https://doi.org/10.1016/j.jmva.2006.11.002

SUN, Y. (2014). Let's fix it: Fixed-$b$ asymptotics versus small-$b$ asymptotics in heteroskedasticity and autocorrelation robust inference. *J. Econometrics* **178** 659–677. MR3132458 https://doi.org/10.1016/j.jeconom.2013.10.001

WANG, R. and SHAO, X. (2020). Supplement to "Hypothesis testing for high-dimensional time series via self-normalization." https://doi.org/10.1214/19-AOS1904SUPP.

WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. MR2172215 https://doi.org/10.1073/pnas.0506715102

WU, W. B. (2007). Strong invariance principles for dependent random variables. *Ann. Probab.* **35** 2294–2320. MR2353389 https://doi.org/10.1214/009117907000000060

WU, W. B. (2011). Asymptotic theory for stationary processes. *Stat. Interface* **4** 207–226. MR2812816 https://doi.org/10.4310/SII.2011.v4.n2.a15

WU, W. B. and MIN, W. (2005). On linear processes with dependent innovations. *Stochastic Process. Appl.* **115** 939–958. MR2138809 https://doi.org/10.1016/j.spa.2005.01.001

WU, W. B. and SHAO, X. (2004). Limit theorems for iterated random functions. *J. Appl. Probab.* **41** 425–436. MR2052582 https://doi.org/10.1239/jap/1082999076

WU, W. B. and WOODROOFE, M. (2004). Martingale approximations for sums of stationary processes. *Ann. Probab.* **32** 1674–1690. MR2060314 https://doi.org/10.1214/009117904000000351

XU, G., LIN, L., WEI, P. and PAN, W. (2016). An adaptive two-sample test for high-dimensional means. *Biometrika* **103** 609–624. MR3551787 https://doi.org/10.1093/biomet/asw029

ZHANG, X. (2017). Testing high dimensional mean under sparsity. Preprint. Available at http://www.stat.tamu.edu/~zhangxiany/LR-HD2017.pdf.

ZHANG, X. and CHENG, G. (2018). Gaussian approximation for high dimensional vector under physical dependence. *Bernoulli* **24** 2640–2675. MR3779697 https://doi.org/10.3150/17-BEJ939

ZHANG, D. and WU, W. B. (2017). Gaussian approximation for high dimensional time series. *Ann. Statist.* **45** 1895–1919. MR3718156 https://doi.org/10.1214/16-AOS1512

ZHANG, X., SHAO, X., HAYHOE, K. and WUEBBLES, D. J. (2011). Testing the structural stability of temporally dependent functional observations and application to climate projections. *Electron. J. Stat.* **5** 1765–1796. MR2870150 https://doi.org/10.1214/11-EJS655

ZHOU, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *J. Amer. Statist. Assoc.* **108** 726–740. MR3174655 https://doi.org/10.1080/01621459.2013.787184

ZHOU, Z. and SHAO, X. (2013). Inference for linear models with dependent errors. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 323–343. MR3021390 https://doi.org/10.1111/j.1467-9868.2012.01044.x

*The Annals of Statistics*
2020, Vol. 48, No. 5, 2759–2790
https://doi.org/10.1214/19-AOS1905


# VARIATIONAL ANALYSIS OF CONSTRAINED M-ESTIMATORS

BY JOHANNES O. ROYSET[1] AND ROGER J-B WETS[2]

[1] *Operations Research Department, Naval Postgraduate School, joroyset@nps.edu*
[2] *Department of Mathematics, University of California, Davis, rjwets@ucdavis.edu*

We propose a unified framework for establishing existence of nonparametric $M$-estimators, computing the corresponding estimates, and proving their strong consistency when the class of functions is exceptionally rich. In particular, the framework addresses situations where the class of functions is complex involving information and assumptions about shape, pointwise bounds, location of modes, height at modes, location of level-sets, values of moments, size of subgradients, continuity, distance to a "prior" function, multivariate total positivity and any combination of the above. The class might be engineered to perform well in a specific setting even in the presence of little data. The framework views the class of functions as a subset of a particular metric space of upper semicontinuous functions under the Attouch–Wets distance. In addition to allowing a systematic treatment of numerous $M$-estimators, the framework yields consistency of plug-in estimators of modes of densities, maximizers of regression functions, level-sets of classifiers and related quantities, and also enables computation by means of approximating parametric classes. We establish consistency through a one-sided law of large numbers, here extended to sieves, that relaxes assumptions of uniform laws, while ensuring global approximations even under model misspecification.

## REFERENCES

[1] ARTSTEIN, Z. and WETS, R. J.-B. (1995). Consistency of minimizers and the SLLN for stochastic programs. *J. Convex Anal.* **2** 1–17. MR1363357

[2] BEER, G., ROCKAFELLAR, R. T. and WETS, R. J.-B. (1992). A characterization of epi-convergence in terms of convergence of level sets. *Proc. Amer. Math. Soc.* **116** 753–761. MR1119262 https://doi.org/10.2307/2159443

[3] BIRGÉ, L. (1997). Estimation of unimodal densities without smoothness assumptions. *Ann. Statist.* **25** 970–981. MR1447736 https://doi.org/10.1214/aos/1069362733

[4] CANNON, A., HOWSE, J., HUSH, D. and SCOVEL, C. (2002). Learning with the Neyman–Pearson and minmax criteria. Technical Report LA-UR 02–2951, Los Alamos National Laboratory, Los Alamos, USA.

[5] CASASENT, D. and CHEN, X. (2003). Radial basis function neural networks for nonlinear Fisher discrimination and Neyman–Pearson classification. *Neural Netw.* **16** 529–535.

[6] CHEN, X. (2007). Large sample sieve estimation of semi-nonparametric models, Chapter 76. In *Handbook of Econometric* **6B** 5549–5632.

[7] CULE, M. and SAMWORTH, R. (2010). Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electron. J. Stat.* **4** 254–270. MR2645484 https://doi.org/10.1214/09-EJS505

[8] CULE, M., SAMWORTH, R. and STEWART, M. (2010). Maximum likelihood estimation of a multi-dimensional log-concave density. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 545–607. MR2758237 https://doi.org/10.1111/j.1467-9868.2010.00753.x

[9] DECHEVSKY, L. and PENEV, S. (1997). On shape-preserving probabilistic wavelet approximators. *Stoch. Anal. Appl.* **15** 187–215. MR1440149 https://doi.org/10.1080/07362999708809471

[10] DEVORE, R. A. (1977). Monotone approximation by polynomials. *SIAM J. Math. Anal.* **8** 906–921. MR0510582 https://doi.org/10.1137/0508069

[11] DEVORE, R. A. (1977). Monotone approximation by splines. *SIAM J. Math. Anal.* **8** 891–905. MR0510725 https://doi.org/10.1137/0508068

*MSC2020 subject classifications.* 62G07.
*Key words and phrases.* Shape-constrained estimation, variational approximations.

[12] DONG, M. X. and WETS, R. J.-B. (2000). Estimating density functions: A constrained maximum likelihood approach. *J. Nonparametr. Stat.* **12** 549–595. MR1785398 https://doi.org/10.1080/10485250008832822

[13] DOSS, C. R. and WELLNER, J. A. (2019). Univariate log-concave density estimation with symmetry or modal constraints. *Electron. J. Stat.* **13** 2391–2461. MR3983344 https://doi.org/10.1214/19-EJS1574

[14] DÜMBGEN, L., SAMWORTH, R. and SCHUHMACHER, D. (2011). Approximation by log-concave distributions, with applications to regression. *Ann. Statist.* **39** 702–730. MR2816336 https://doi.org/10.1214/10-AOS853

[15] DUPAČOVÁ, J. and WETS, R. (1988). Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *Ann. Statist.* **16** 1517–1549. MR0964937 https://doi.org/10.1214/aos/1176351052

[16] DURRETT, R. (1996). *Probability: Theory and Examples*, 2nd ed. Duxbury Press, Belmont, CA. MR1609153

[17] FALLAT, S., LAURITZEN, S., SADEGHI, K., UHLER, C., WERMUTH, N. and ZWIERNIK, P. (2017). Total positivity in Markov structures. *Ann. Statist.* **45** 1152–1184. MR3662451 https://doi.org/10.1214/16-AOS1478

[18] GEMAN, S. and HWANG, C.-R. (1982). Nonparametric maximum likelihood estimation by the method of sieves. *Ann. Statist.* **10** 401–414. MR0653512

[19] GRENANDER, U. (1981). *Abstract Inference. Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0599175

[20] HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability* (*Berkeley, Calif.*, 1965/66), *Vol. I: Statistics* 221–233. Univ. California Press, Berkeley, CA. MR0216620

[21] KOENKER, R. and MIZERA, I. (2010). Quasi-concave density estimation. *Ann. Statist.* **38** 2998–3027. MR2722462 https://doi.org/10.1214/10-AOS814

[22] KORF, L. A. and WETS, R. J.-B. (2001). Random-lsc functions: An ergodic theorem. *Math. Oper. Res.* **26** 421–445. MR1895837 https://doi.org/10.1287/moor.26.2.421.10548

[23] LAVRIČ, B. (1993). Continuity of monotone functions. *Arch. Math.* (*Brno*) **29** 1–4. MR1242622

[24] MEYER, M. C. (2012). Constrained penalized splines. *Canad. J. Statist.* **40** 190–206. MR2896937 https://doi.org/10.1002/cjs.10137

[25] MEYER, M. C. (2012). Nonparametric estimation of a smooth density with shape restrictions. *Statist. Sinica* **22** 681–701. MR2954357 https://doi.org/10.5705/ss.2010.309

[26] MEYER, M. C. and HABTZGHI, D. (2011). Nonparametric estimation of density and hazard rate functions with shape restrictions. *J. Nonparametr. Stat.* **23** 455–470. MR2801304 https://doi.org/10.1080/10485252.2010.531133

[27] PROBST, P., BOULESTEIX, A.-L. and BISCHL, B. (2019). Tunability: Importance of hyperparameters of machine learning algorithms. *J. Mach. Learn. Res.* **20** Paper No. 53, 32. MR3948093

[28] ROCKAFELLAR, R. T. and WETS, R. J.-B. (1998). *Variational Analysis. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **317**. Springer, Berlin. MR1491362 https://doi.org/10.1007/978-3-642-02431-3

[29] ROYSET, J. O. (2018). Approximations and solution estimates in optimization. *Math. Program.* **170** 479–506. MR3820483 https://doi.org/10.1007/s10107-017-1165-0

[30] ROYSET, J. O. (2019). Approximations of semicontinuous functions with applications to stochastic optimization and statistical estimation. *Mathematical Programming*, OnlineFirst.

[31] ROYSET, J. O. and WETS, R. J.-B. (2015). Fusion of hard and soft information in nonparametric density estimation. *European J. Oper. Res.* **247** 532–547. MR3375840 https://doi.org/10.1016/j.ejor.2015.06.034

[32] ROYSET, J. O. and WETS, R. J.-B. (2016). Multivariate epi-splines and evolving function identification problems. *Set-Valued Var. Anal.* **24** 517–545. Erratum: pp. 547–549. MR3570341 https://doi.org/10.1007/s11228-015-0329-4

[33] ROYSET, J. O. and WETS, R. J.-B. (2017). Variational theory for optimization under stochastic ambiguity. *SIAM J. Optim.* **27** 1118–1149. MR3663008 https://doi.org/10.1137/16M1060704

[34] ROYSET, J. O. and WETS, R. J.-B. (2014). From data to assessments and decisions: Epi-spline technology. In *INFORMS Tutorials* (A. Newman, ed.) INFORMS, Catonsville.

[35] ROYSET, J. O. and WETS, R. J.-B. (2019). Lopsided convergence: An extension and its quantification. *Math. Program.* **177** 395–423. MR3987205 https://doi.org/10.1007/s10107-018-1275-3

[36] SALINETTI, G. and WETS, R. (1986). On the hypoconvergence of probability measures. In *Optimization and Related Fields* (*Erice*, 1984). *Lecture Notes in Math.* **1190** 371–395. Springer, Berlin. MR0858359 https://doi.org/10.1007/BFb0076715

[37] SALINETTI, G. and WETS, R. J.-B. (1986). On the convergence in distribution of measurable multifunctions (random sets), normal integrands, stochastic processes and stochastic infima. *Math. Oper. Res.* **11** 385–419. MR0852332 https://doi.org/10.1287/moor.11.3.385

[38] SASABUCHI, S., INUTSUKA, M. and KULATUNGA, D. D. S. (1983). A multivariate version of isotonic regression. *Biometrika* **70** 465–472. MR0712034 https://doi.org/10.1093/biomet/70.2.465

[39] SCOTT, C. and NOWAK, R. (2005). A Neyman–Pearson approach to statistical learning. *IEEE Trans. Inform. Theory* **51** 3806–3819. MR2239000 https://doi.org/10.1109/TIT.2005.856955

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

[41] SEREGIN, A. and WELLNER, J. A. (2010). Nonparametric estimation of multivariate convex-transformed densities. *Ann. Statist.* **38** 3751–3781. MR2766867 https://doi.org/10.1214/10-AOS840

[42] STEINWART, I. (2005). Consistency of support vector machines and other regularized kernel classifiers. *IEEE Trans. Inform. Theory* **51** 128–142. MR2234577 https://doi.org/10.1109/TIT.2004.839514

[43] VAN DE GEER, S. (2000). *Empirical Processes in M-Estimation*. Cambridge Univ. Press, Cambridge.

[44] VAN DE GEER, S. and WEGKAMP, M. (1996). Consistency for the least squares estimator in nonparametric regression. *Ann. Statist.* **24** 2513–2523. MR1425964 https://doi.org/10.1214/aos/1032181165

[45] 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

[46] VAN DER VAART, A. W. (2011). Empirical Processes and Statistical Learning. Lecture Notes, Vrije Universiteit, Amsterdam, Netherland.

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

[48] WAECHTER, A. (2018). Ipopt interior point optimizer. Available at http://projects.coin-or.org/Ipopt.

[49] WALD, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Stat.* **20** 595–601. MR0032169 https://doi.org/10.1214/aoms/1177729952

[50] WANG, J. (1996). Asymptotics of least-squares estimators for constrained nonlinear regression. *Ann. Statist.* **24** 1316–1326. MR1401852 https://doi.org/10.1214/aos/1032526971

*The Annals of Statistics*
2020, Vol. 48, No. 5, 2791–2823
https://doi.org/10.1214/19-AOS1906


# WHICH BRIDGE ESTIMATOR IS THE BEST FOR VARIABLE SELECTION?

BY SHUAIWEN WANG[1,*], HAOLEI WENG[2] AND ARIAN MALEKI[1,†]

[1]*Department of Statistics, Columbia University,* [*]*sw2853@columbia.edu;* [†]*arian@stat.columbia.edu*
[2]*Department of Statistics and Probability, Michigan State University, wenghaol@msu.edu*

We study the problem of variable selection for linear models under the high-dimensional asymptotic setting, where the number of observations $n$ grows at the same rate as the number of predictors $p$. We consider two-stage variable selection techniques (TVS) in which the first stage uses bridge estimators to obtain an estimate of the regression coefficients, and the second stage simply thresholds this estimate to select the "important" predictors. The asymptotic false discovery proportion (AFDP) and true positive proportion (ATPP) of these TVS are evaluated. We prove that for a fixed ATPP, in order to obtain a smaller AFDP, one should pick a bridge estimator with smaller asymptotic mean square error in the first stage of TVS. Based on such principled discovery, we present a sharp comparison of different TVS, via an in-depth investigation of the estimation properties of bridge estimators. Rather than "orderwise" error bounds with loose constants, our analysis focuses on precise error characterization. Various interesting signal-to-noise ratio and sparsity settings are studied. Our results offer new and thorough insights into high-dimensional variable selection. For instance, we prove that a TVS with Ridge in its first stage outperforms TVS with other bridge estimators in large noise settings; two-stage LASSO becomes inferior when the signal is rare and weak. As a by-product, we show that two-stage methods outperform some standard variable selection techniques, such as LASSO and Sure Independence Screening, under certain conditions.

## REFERENCES

[1] AERON, S., SALIGRAMA, V. and ZHAO, M. (2010). Information theoretic bounds for compressed sensing. *IEEE Trans. Inf. Theory* **56** 5111–5130. MR2808668 https://doi.org/10.1109/TIT.2010.2059891
[2] 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
[3] 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
[4] 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
[5] 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
[6] BOGDAN, M., VAN DEN BERG, E., SU, W. and CANDES, E. J. (2013). Supplementary materials for "Statistical estimation and testing via the sorted l1 norm". *Ann. Statist.*
[7] BUTUCEA, C., NDAOUD, M., STEPANOVA, N. A. and TSYBAKOV, A. B. (2018). Variable selection with Hamming loss. *Ann. Statist.* **46** 1837–1875. MR3845003 https://doi.org/10.1214/17-AOS1572
[8] 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
[9] CHO, H. and FRYZLEWICZ, P. (2012). High dimensional variable selection via tilting. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 593–622. MR2925375 https://doi.org/10.1111/j.1467-9868.2011.01023.x

---

*MSC2020 subject classifications.* 62J05, 62J07.
*Key words and phrases.* Variable selection, high dimension, bridge regression, two-stage methods, false discovery proportion, true positive proportion, rare signal, large noise, large sample, debiasing.

[10] DAVID, G. and ILIAS, Z. (2017). High dimensional regression with binary coefficients. Estimating squared error and a phase transition. In *Conference on Learning Theory* 948–953.

[11] 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

[12] 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

[13] 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

[14] DONOHO, D. L., JOHNSTONE, I. M., HOCH, J. C. and STERN, A. S. (1992). Maximum entropy and the nearly black object. *J. Roy. Statist. Soc. Ser. B* **54** 41–81. With discussion and a reply by the authors. MR1157714

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

[16] DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2011). The noise-sensitivity phase transition in compressed sensing. *IEEE Trans. Inf. Theory* **57** 6920–6941. MR2882271 https://doi.org/10.1109/TIT.2011.2165823

[17] DONOHO, D. L. and TANNER, J. (2005). Sparse nonnegative solution of underdetermined linear equations by linear programming. *Proc. Natl. Acad. Sci. USA* **102** 9446–9451. MR2168715 https://doi.org/10.1073/pnas.0502269102

[18] EL KAROUI, N. (2010). High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: Risk underestimation. *Ann. Statist.* **38** 3487–3566. MR2766860 https://doi.org/10.1214/10-AOS795

[19] 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.

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

[21] 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

[22] FLETCHER, A. K., RANGAN, S. and GOYAL, V. K. (2009). Necessary and sufficient conditions for sparsity pattern recovery. *IEEE Trans. Inf. Theory* **55** 5758–5772. MR2597192 https://doi.org/10.1109/TIT.2009.2032726

[23] FRANK, L. E. and FRIEDMAN, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics* **35** 109–135.

[24] GENOVESE, C. R., JIN, J., WASSERMAN, L. and YAO, Z. (2012). A comparison of the lasso and marginal regression. *J. Mach. Learn. Res.* **13** 2107–2143. MR2956354

[25] HASTIE, T., TIBSHIRANI, R. and TIBSHIRANI, R. J. (2017). Extended comparisons of best subset selection, forward stepwise selection, and the lasso. arXiv preprint arXiv:1707.08692.

[26] HUANG, J., HOROWITZ, J. L. and MA, S. (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *Ann. Statist.* **36** 587–613. MR2396808 https://doi.org/10.1214/009053607000000875

[27] HUBER, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821. MR0356373

[28] JI, P. and JIN, J. (2012). UPS delivers optimal phase diagram in high-dimensional variable selection. *Ann. Statist.* **40** 73–103. MR3013180 https://doi.org/10.1214/11-AOS947

[29] JIN, J., ZHANG, C.-H. and ZHANG, Q. (2014). Optimality of graphlet screening in high dimensional variable selection. *J. Mach. Learn. Res.* **15** 2723–2772. MR3270749 https://doi.org/10.1631/jzus.a1400233

[30] KE, Z. T., JIN, J. and FAN, J. (2014). Covariate assisted screening and estimation. *Ann. Statist.* **42** 2202–2242. MR3269978 https://doi.org/10.1214/14-AOS1243

[31] KNIGHT, K. and FU, W. (2000). Asymptotics for lasso-type estimators. *Ann. Statist.* **28** 1356–1378. MR1805787 https://doi.org/10.1214/aos/1015957397

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

[33] LUO, S. and CHEN, Z. (2014). Sequential lasso cum EBIC for feature selection with ultra-high dimensional feature space. *J. Amer. Statist. Assoc.* **109** 1229–1240. MR3265693 https://doi.org/10.1080/01621459.2013.877275

[34] MALEKI, A., ANITORI, L., YANG, Z. and BARANIUK, R. G. (2013). Asymptotic analysis of complex LASSO via complex approximate message passing (CAMP). *IEEE Trans. Inf. Theory* **59** 4290–4308. MR3071330 https://doi.org/10.1109/TIT.2013.2252232

[35] MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. MR2278363 https://doi.org/10.1214/009053606000000281

[36] MEINSHAUSEN, N. and YU, B. (2009). Lasso-type recovery of sparse representations for high-dimensional data. *Ann. Statist.* **37** 246–270. MR2488351 https://doi.org/10.1214/07-AOS582

[37] MILLER, A. (2002). *Subset Selection in Regression*, 2nd ed. *Monographs on Statistics and Applied Probability* **95**. CRC Press/CRC, Boca Raton, FL. MR2001193 https://doi.org/10.1201/9781420035933

[38] MOUSAVI, A., MALEKI, A. and BARANIUK, R. G. (2018). Consistent parameter estimation for LASSO and approximate message passing. *Ann. Statist.* **46** 119–148. MR3766948 https://doi.org/10.1214/17-AOS1544

[39] NDAOUD, M. and TSYBAKOV, A. B. (2018). Optimal variable selection and adaptive noisy compressed sensing. arXiv preprint arXiv:1809.03145.

[40] OYMAK, S. and HASSIBI, B. (2016). Sharp MSE bounds for proximal denoising. *Found. Comput. Math.* **16** 965–1029. MR3529131 https://doi.org/10.1007/s10208-015-9278-4

[41] RAD, K. R. (2011). Nearly sharp sufficient conditions on exact sparsity pattern recovery. *IEEE Trans. Inf. Theory* **57** 4672–4679. MR2840483 https://doi.org/10.1109/TIT.2011.2145670

[42] RANGAN, S., GOYAL, V. and FLETCHER, A. K. (2009). Asymptotic analysis of map estimation via the replica method and compressed sensing. In *Advances in Neural Information Processing Systems* 1545–1553.

[43] REEVES, G. and GASTPAR, M. C. (2013). Approximate sparsity pattern recovery: Information-theoretic lower bounds. *IEEE Trans. Inf. Theory* **59** 3451–3465. MR3061258 https://doi.org/10.1109/TIT.2013.2253852

[44] SU, W., BOGDAN, M. and CANDÈS, E. (2017). False discoveries occur early on the Lasso path. *Ann. Statist.* **45** 2133–2150. MR3718164 https://doi.org/10.1214/16-AOS1521

[45] 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

[46] 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

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

[48] WAINWRIGHT, M. J. (2009). Sharp thresholds for high-dimensional and noisy sparsity recovery using $\ell_1$-constrained quadratic programming (Lasso). *IEEE Trans. Inf. Theory* **55** 2183–2202. MR2729873 https://doi.org/10.1109/TIT.2009.2016018

[49] WAINWRIGHT, M. J. (2009). Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting. *IEEE Trans. Inf. Theory* **55** 5728–5741. MR2597190 https://doi.org/10.1109/TIT.2009.2032816

[50] WANG, S., WENG, H. and MALEKI, A. (2020). Which bridge estimator is optimal for variable selection? arXiv preprint arXiv:1705.08617.

[51] WANG, S., WENG, H. and MALEKI, A. (2020). Supplement to "Which bridge estimator is the best for variable selection?" https://doi.org/10.1214/19-AOS1906SUPP

[52] WANG, W., WAINWRIGHT, M. J. and RAMCHANDRAN, K. (2010). Information-theoretic limits on sparse signal recovery: Dense versus sparse measurement matrices. *IEEE Trans. Inf. Theory* **56** 2967–2979. MR2683451 https://doi.org/10.1109/TIT.2010.2046199

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

[54] WENG, H., FENG, Y. and QIAO, X. (2019). Regularization after retention in ultrahigh dimensional linear regression models. *Statist. Sinica* **29** 387–407. MR3889373

[55] WENG, H., MALEKI, A. and ZHENG, L. (2018). Overcoming the limitations of phase transition by higher order analysis of regularization techniques. *Ann. Statist.* **46** 3099–3129. MR3851766 https://doi.org/10.1214/17-AOS1651

[56] YANG, E., LOZANO, A. and RAVIKUMAR, P. (2014). Elementary estimators for high-dimensional linear regression. In *International Conference on Machine Learning* 388–396.

[57] ZHANG, C.-H. and HUANG, J. (2008). The sparsity and bias of the LASSO selection in high-dimensional linear regression. *Ann. Statist.* **36** 1567–1594. MR2435448 https://doi.org/10.1214/07-AOS520

[58] ZHANG, T. (2009). Some sharp performance bounds for least squares regression with $L_1$ regularization. *Ann. Statist.* **37** 2109–2144. MR2543687 https://doi.org/10.1214/08-AOS659

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

[60] ZHENG, L., MALEKI, A., WENG, H., WANG, X. and LONG, T. (2017). Does $\ell_p$-minimization outperform $\ell_1$-minimization? *IEEE Trans. Inf. Theory* **63** 6896–6935. MR3724407 https://doi.org/10.1109/TIT.2017.2717585

[61] ZHOU, S. (2009). Thresholding procedures for high dimensional variable selection and statistical estimation. In *Advances in Neural Information Processing Systems* 2304–2312.

[62] ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279469 https://doi.org/10.1198/016214506000000735

*The Annals of Statistics*
2020, Vol. 48, No. 5, 2824–2847
https://doi.org/10.1214/19-AOS1907


# PERMUTATION METHODS FOR FACTOR ANALYSIS AND PCA

BY EDGAR DOBRIBAN

*Department of Statistics, The Wharton School, University of Pennsylvania, dobriban@wharton.upenn.edu*

Researchers often have datasets measuring features $x_{ij}$ of samples, such as test scores of students. In factor analysis and PCA, these features are thought to be influenced by unobserved factors, such as skills. Can we determine how many components affect the data? This is an important problem, because decisions made here have a large impact on all downstream data analysis. Consequently, many approaches have been developed. *Parallel Analysis* is a popular permutation method: it randomly scrambles each feature of the data. It selects components if their singular values are larger than those of the permuted data. Despite widespread use, as well as empirical evidence for its accuracy, it currently has no theoretical justification.

In this paper, we show that parallel analysis (or permutation methods) consistently select the large components in certain high-dimensional factor models. However, when the signals are too large, the smaller components are not selected. The intuition is that permutations keep the noise invariant, while "destroying" the low-rank signal. This provides justification for permutation methods. Our work also uncovers drawbacks of permutation methods, and paves the way to improvements.

## REFERENCES

ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. *Wiley Publications in Statistics*. Wiley, New York; CRC Press, London. MR0091588

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

BAI, J. and NG, S. (2008). *Large Dimensional Factor Analysis*. Now Publishers, Hanover.

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

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

BARTLETT, M. S. (1950). Tests of significance in factor analysis. *Br*. *J*. *Math*. *Stat*. *Psychol*. **3** 77–85.

BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2012). The singular values and vectors of low rank perturbations of large rectangular random matrices. *J*. *Multivariate Anal*. **111** 120–135. MR2944410 https://doi.org/10.1016/j.jmva.2012.04.019

BROWN, T. A. (2014). *Confirmatory Factor Analysis for Applied Research*. Guilford, New York.

BUJA, A. and EYUBOGLU, N. (1992). Remarks on parallel analysis. *Multivar. Behav. Res*. **27** 509–540.

CATTELL, R. B. (1966). The scree test for the number of factors. *Multivar. Behav. Res*. **1** 245–276.

CHURCHILL, G. A. JR. (1979). A paradigm for developing better measures of marketing constructs. *J*. *Mark*. *Res*. 64–73.

COSTELLO, A. B. and OSBORNE, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Pract*. *Assess*., *Res*. *Eval*. **10** 1–9.

DOBRIBAN, E., LEEB, W. and SINGER, A. (2017). Optimal prediction in the linearly transformed spiked model. Preprint. Available at arXiv:1709.03393. To appear in the Annals of Statistics.

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

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

*MSC2020 subject classifications.* Primary 62H25; secondary 62H12.

*Key words and phrases.* Factor analysis, PCA, parallel analysis, permutation methods, high-dimensional asymptotics.

FABRIGAR, L. R., WEGENER, D. T., MACCALLUM, R. C. and STRAHAN, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychol. Methods* **4** 272.

GASKIN, C. J. and HAPPELL, B. (2014). On exploratory factor analysis: A review of recent evidence, an assessment of current practice, and recommendations for future use. *Int. J. Nurs. Stud.* **51** 511–521.

GERARD, D. and STEPHENS, M. (2017). Unifying and generalizing methods for removing unwanted variation based on negative controls. Preprint. Available at arXiv:1705.08393.

GLORFELD, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educ. Psychol. Meas.* **55** 377–393.

GREEN, S. B., LEVY, R., THOMPSON, M. S., LU, M. and LO, W.-J. (2012). A proposed solution to the problem with using completely random data to assess the number of factors with parallel analysis. *Educ. Psychol. Meas.* **72** 357–374.

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

HAYTON, J. C., ALLEN, D. G. and SCARPELLO, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organ. Res. Methods* **7** 191–205.

HORN, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika* **30** 179–185.

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

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

KAISER, H. F. (1960). The application of electronic computers to factor analysis. *Educ. Psychol. Meas.* **20** 141–151.

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

LEEK, J. T. and STOREY, J. D. (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3** e161.

LEEK, J. T. and STOREY, J. D. (2008). A general framework for multiple testing dependence. *Proc. Natl. Acad. Sci. USA* **105** 18718–18723.

LIN, Z., YANG, C., ZHU, Y. et al. (2016). Simultaneous dimension reduction and adjustment for confounding variation. *Proc. Natl. Acad. Sci. USA* **113** 14662–14667. MR3600520 https://doi.org/10.1073/pnas.1617317113

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

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

ONATSKI, A. (2009). Testing hypotheses about the numbers of factors in large factor models. *Econometrica* **77** 1447–1479. MR2561070 https://doi.org/10.3982/ECTA6964

ONATSKI, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *J. Econometrics* **168** 244–258. MR2923766 https://doi.org/10.1016/j.jeconom.2012.01.034

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

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

PAUL, D. and AUE, A. (2014). Random matrix theory in statistics: A review. *J. Statist. Plann. Inference* **150** 1–29. MR3206718 https://doi.org/10.1016/j.jspi.2013.09.005

PERES-NETO, P. R., JACKSON, D. A. and SOMERS, K. M. (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Comput. Statist. Data Anal.* **49** 974–997. MR2143053 https://doi.org/10.1016/j.csda.2004.06.015

QUADEER, A. A., LOUIE, R. H., SHEKHAR, K., CHAKRABORTY, A. K., HSING, I.-M. and MCKAY, M. R. (2014). Statistical linkage analysis of substitutions in patient-derived sequences of genotype 1a hepatitis C virus nonstructural protein 3 exposes targets for immunogen design. *J. Virol.* **88** 7628–7644.

RAICHE, G., MAGIS, D. and RAICHE, M. G. Package 'nfactors'. 2010.

SACCENTI, E. and TIMMERMAN, M. E. (2017). Considering Horn's parallel analysis from a random matrix theory point of view. *Psychometrika* **82** 186–209. MR3614813 https://doi.org/10.1007/s11336-016-9515-z

SPEARMAN, C. (1904). "General intelligence", objectively determined and measured. *Am. J. Psychol.* **15** 201–292.

STEWART, D. W. (1981). The application and misapplication of factor analysis in marketing research. *J. Mark. Res.* 51–62.

THURSTONE, L. L. (1947). *Multiple-factor analysis*. University of Chicago Press, Chicago.

VELICER, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika* **41** 321–327.

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

YAO, J., ZHENG, S. and BAI, Z. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis*. *Cambridge Series in Statistical and Probabilistic Mathematics* **39**. Cambridge Univ. Press, New York. MR3468554 https://doi.org/10.1017/CBO9781107588080

ZHOU, Y.-H., MARRON, J. S. and WRIGHT, F. A. (2018). Eigenvalue significance testing for genetic association. *Biometrics* **74** 439–447. MR3825330 https://doi.org/10.1111/biom.12767

ZWICK, W. R. and VELICER, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychol*. *Bull*. **99** 432.

*The Annals of Statistics*
2020, Vol. 48, No. 5, 2848–2878
https://doi.org/10.1214/19-AOS1909


# A GENERAL FRAMEWORK FOR BAYES STRUCTURED LINEAR MODELS

BY CHAO GAO[1], AAD W. VAN DER VAART[2] AND HARRISON H. ZHOU[3]

[1]*Department of Statistics, University of Chicago, chaogao@galton.uchicago.edu*
[2]*Mathematical Institute, Faculty of Science, Leiden University, avdvaart@math.leidenuniv.nl*
[3]*Department of Statistics, Yale University, huibin.zhou@yale.edu*

High dimensional statistics deals with the challenge of extracting structured information from complex model settings. Compared with a large number of frequentist methodologies, there are rather few theoretically optimal Bayes methods for high dimensional models. This paper provides a unified approach to both Bayes high dimensional statistics and Bayes nonparametrics in a general framework of structured linear models. With a proposed two-step prior, we prove a general oracle inequality for posterior contraction under an abstract setting that allows model misspecification. The general result can be used to derive new results on optimal posterior contraction under many complex model settings including recent works for stochastic block model, graphon estimation and dictionary learning. It can also be used to improve upon posterior contraction results in literature including sparse linear regression and nonparametric aggregation. The key of the success lies in the novel two-step prior distribution: one for model structure, that is, model selection, and the other one for model parameters. The prior on the parameters of a model is an elliptical Laplace distribution that is capable of modeling signals with large magnitude, and the prior on the model structure involves a factor that compensates the effect of the normalizing constant of the elliptical Laplace distribution, which is important to attain rate-optimal posterior contraction.

## REFERENCES

[1] AGARWAL, A., ANANDKUMAR, A. and NETRAPALLI, P. (2013). Exact recovery of sparsely used overcomplete dictionaries. ArXiv preprint. Available at arXiv:1309.1952.

[2] ALDOUS, D. J. (1981). Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* **11** 581–598. MR0637937 https://doi.org/10.1016/0047-259X(81)90099-3

[3] BAKIN, S. (1999). Adaptive regression and model selection in data mining problems.

[4] BANERJEE, S. and GHOSAL, S. (2014). Posterior convergence rates for estimating large precision matrices using graphical models. *Electron. J. Stat.* **8** 2111–2137. MR3273620 https://doi.org/10.1214/14-EJS945

[5] BARRON, A., BIRGÉ, L. and MASSART, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113** 301–413. MR1679028 https://doi.org/10.1007/s004400050210

[6] BARRON, A., SCHERVISH, M. J. and WASSERMAN, L. (1999). The consistency of posterior distributions in nonparametric problems. *Ann. Statist.* **27** 536–561. MR1714718 https://doi.org/10.1214/aos/1018031206

[7] BARRON, A. R. (1988). The exponential convergence of posterior probabilities with implications for Bayes estimators of density functions. Univ. of Illinois.

[8] BARRON, A. R. and COVER, T. M. (1991). Minimum complexity density estimation. *IEEE Trans. Inform. Theory* **37** 1034–1054. MR1111806 https://doi.org/10.1109/18.86996

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

[10] BIRGÉ, L. and MASSART, P. (2001). Gaussian model selection. *J. Eur. Math. Soc.* (*JEMS*) **3** 203–268. MR1848946 https://doi.org/10.1007/s100970100031

*MSC2020 subject classifications.* Primary 62C10; secondary 62F15.
*Key words and phrases.* Oracle inequality, stochastic block model, graphon, sparse linear regression, aggregation, dictionary learning, posterior contraction.

[11] 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

[12] 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

[13] Bunea, F. (2008). Consistent selection via the Lasso for high dimensional approximating regression models. In *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh. Inst. Math. Stat. (IMS) Collect.* **3** 122–137. IMS, Beachwood, OH. MR2459221 https://doi.org/10.1214/074921708000000101

[14] Candes, E. J. and Tao, T. (2005). Decoding by linear programming. *IEEE Trans. Inform. Theory* **51** 4203–4215. MR2243152 https://doi.org/10.1109/TIT.2005.858979

[15] Castillo, I. (2014). On Bayesian supremum norm contraction rates. *Ann. Statist.* **42** 2058–2091. MR3262477 https://doi.org/10.1214/14-AOS1253

[16] Castillo, I., Schmidt-Hieber, J. and van der Vaart, A. (2015). Bayesian linear regression with sparse priors. *Ann. Statist.* **43** 1986–2018. MR3375874 https://doi.org/10.1214/15-AOS1334

[17] Castillo, I. and van der Vaart, A. (2012). Needles and straw in a haystack: Posterior concentration for possibly sparse sequences. *Ann. Statist.* **40** 2069–2101. MR3059077 https://doi.org/10.1214/12-AOS1029

[18] Catoni, O. (2004). *Statistical Learning Theory and Stochastic Optimization. Lecture Notes in Math.* **1851**. Springer, Berlin. MR2163920 https://doi.org/10.1007/b99352

[19] Diaconis, P. and Janson, S. (2008). Graph limits and exchangeable random graphs. *Rend. Mat. Appl.* (7) **28** 33–61. MR2463439

[20] Donoho, D. L., Elad, M. and Temlyakov, V. N. (2006). Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inform. Theory* **52** 6–18. MR2237332 https://doi.org/10.1109/TIT.2005.860430

[21] Fang, K. T., Kotz, S. and Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions. Monographs on Statistics and Applied Probability* **36**. CRC Press, London. MR1071174 https://doi.org/10.1007/978-1-4899-2937-2

[22] Gao, C., Lu, Y. and Zhou, H. H. (2015). Rate-optimal graphon estimation. *Ann. Statist.* **43** 2624–2652. MR3405606 https://doi.org/10.1214/15-AOS1354

[23] Gao, C., van der Vaart, A. W and Zhou, H. H (2020). Supplement to "A General Framework for Bayes Structured Linear Models." https://doi.org/10.1214/19-AOS1909SUPP.

[24] Gao, C. and Zhou, H. H. (2015). Rate-optimal posterior contraction for sparse PCA. *Ann. Statist.* **43** 785–818. MR3325710 https://doi.org/10.1214/14-AOS1268

[25] Gao, C. and Zhou, H. H. (2016). Rate exact Bayesian adaptation with modified block priors. *Ann. Statist.* **44** 318–345. MR3449770 https://doi.org/10.1214/15-AOS1368

[26] Ghosal, S., Ghosh, J. K. and Ramamoorthi, R. V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *Ann. Statist.* **27** 143–158. MR1701105 https://doi.org/10.1214/aos/1018031105

[27] Ghosal, S., Ghosh, J. K. and van der Vaart, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531. MR1790007 https://doi.org/10.1214/aos/1016218228

[28] Hartigan, J. A. (1972). Direct clustering of a data matrix. *J. Amer. Statist. Assoc.* **67** 123–129.

[29] Hoffmann, M., Rousseau, J. and Schmidt-Hieber, J. (2015). On adaptive posterior concentration rates. *Ann. Statist.* **43** 2259–2295. MR3396985 https://doi.org/10.1214/15-AOS1341

[30] Holland, P. W., Laskey, K. B. and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. MR0718088 https://doi.org/10.1016/0378-8733(83)90021-7

[31] Hoover, D. N. (1979). Relations on probability spaces and arrays of random variables 2. Institute for Advanced Study, Princeton, NJ. Preprint.

[32] Johnstone, I. M. (2017). Gaussian estimation: Sequence and wavelet models.

[33] Kallenberg, O. (1989). On the representation theorem for exchangeable arrays. *J. Multivariate Anal.* **30** 137–154. MR1003713 https://doi.org/10.1016/0047-259X(89)90092-4

[34] Kleijn, B. J. K. and van der Vaart, A. W. (2006). Misspecification in infinite-dimensional Bayesian statistics. *Ann. Statist.* **34** 837–877. MR2283395 https://doi.org/10.1214/009053606000000029

[35] Klopp, O., Lu, Y., Tsybakov, A. B. and Zhou, H. H. (2019). Structured matrix estimation and completion. *Bernoulli* **25** 3883–3911. MR4010976 https://doi.org/10.3150/19-bej1114

[36] Leung, G. and Barron, A. R. (2006). Information theory and mixing least-squares regressions. *IEEE Trans. Inform. Theory* **52** 3396–3410. MR2242356 https://doi.org/10.1109/TIT.2006.878172

[37] Lounici, K. (2008). Sup-norm convergence rate and sign concentration property of Lasso and Dantzig estimators. *Electron. J. Stat.* **2** 90–102. MR2386087 https://doi.org/10.1214/08-EJS177

[38] LOUNICI, K., PONTIL, M., VAN DE GEER, S. and TSYBAKOV, A. B. (2011). Oracle inequalities and optimal inference under group sparsity. *Ann. Statist.* **39** 2164–2204. MR2893865 https://doi.org/10.1214/11-AOS896
[39] LOVÁSZ, L. (2012). *Large Networks and Graph Limits*. *American Mathematical Society Colloquium Publications* **60**. Amer. Math. Soc., Providence, RI. MR3012035 https://doi.org/10.1090/coll/060
[40] LOVÁSZ, L. and SZEGEDY, B. (2006). Limits of dense graph sequences. *J. Combin. Theory Ser. B* **96** 933–957. MR2274085 https://doi.org/10.1016/j.jctb.2006.05.002
[41] MA, Z. and WU, Y. (2015). Volume ratio, sparsity, and minimaxity under unitarily invariant norms. *IEEE Trans. Inform. Theory* **61** 6939–6956. MR3430731 https://doi.org/10.1109/TIT.2015.2487541
[42] MARTIN, R., MESS, R. and WALKER, S. G. (2017). Empirical Bayes posterior concentration in sparse high-dimensional linear models. *Bernoulli* **23** 1822–1847. MR3624879 https://doi.org/10.3150/15-BEJ797
[43] NEMIROVSKI, A. (2000). Topics in non-parametric statistics. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1998). *Lecture Notes in Math.* **1738** 85–277. Springer, Berlin. MR1775640
[44] 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
[45] OLSHAUSEN, B. A. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381** 607–609.
[46] PATI, D. and BHATTACHARYA, A. (2015). Optimal Bayesian estimation in stochastic block models. ArXiv preprint. Available at arXiv:1505.06794.
[47] RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2011). Minimax rates of estimation for high-dimensional linear regression over $\ell_q$-balls. *IEEE Trans. Inform. Theory* **57** 6976–6994. MR2882274 https://doi.org/10.1109/TIT.2011.2165799
[48] RIGOLLET, P. and TSYBAKOV, A. (2011). Exponential screening and optimal rates of sparse estimation. *Ann. Statist.* **39** 731–771. MR2816337 https://doi.org/10.1214/10-AOS854
[49] RIGOLLET, P. and TSYBAKOV, A. B. (2012). Sparse estimation by exponential weighting. *Statist. Sci.* **27** 558–575. MR3025134 https://doi.org/10.1214/12-STS393
[50] RIVOIRARD, V. and ROUSSEAU, J. (2012). Posterior concentration rates for infinite dimensional exponential families. *Bayesian Anal.* **7** 311–333. MR2934953 https://doi.org/10.1214/12-BA710
[51] SHEN, X. and WASSERMAN, L. (2001). Rates of convergence of posterior distributions. *Ann. Statist.* **29** 687–714. MR1865337 https://doi.org/10.1214/aos/1009210686
[52] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242
[53] TSYBAKOV, A. B. (2003). Optimal rates of aggregation. In *Learning Theory and Kernel Machines* 303–313. Springer, Berlin.
[54] TSYBAKOV, A. B. (2014). Aggregation and minimax optimality in high-dimensional estimation. In *Proceedings of the International Congress of Mathematicians—Seoul* 2014. *Vol. IV* 225–246. Kyung Moon Sa, Seoul. MR3727610
[55] VAN DE GEER, S. A. and BÜHLMANN, P. (2009). On the conditions used to prove oracle results for the Lasso. *Electron. J. Stat.* **3** 1360–1392. MR2576316 https://doi.org/10.1214/09-EJS506
[56] VAN DER PAS, S. L. and VAN DER VAART, A. W. (2018). Bayesian community detection. *Bayesian Anal.* **13** 767–796. MR3807866 https://doi.org/10.1214/17-BA1078
[57] VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2008). Rates of contraction of posterior distributions based on Gaussian process priors. *Ann. Statist.* **36** 1435–1463. MR2418663 https://doi.org/10.1214/009053607000000613
[58] WANG, Z., PATERLINI, S., GAO, F. and YANG, Y. (2014). Adaptive minimax regression estimation over sparse $\ell_q$-hulls. *J. Mach. Learn. Res.* **15** 1675–1711. MR3225246
[59] YANG, Y. (1999). Model selection for nonparametric regression. *Statist. Sinica* **9** 475–499. MR1707850
[60] YANG, Y. (2000). Combining different procedures for adaptive regression. *J. Multivariate Anal.* **74** 135–161. MR1790617 https://doi.org/10.1006/jmva.1999.1884
[61] YANG, Y. (2004). Aggregating regression procedures to improve performance. *Bernoulli* **10** 25–47. MR2044592 https://doi.org/10.3150/bj/1077544602
[62] YANG, Y. and BARRON, A. R. (1998). An asymptotic property of model selection criteria. *IEEE Trans. Inform. Theory* **44** 95–116. MR1486651 https://doi.org/10.1109/18.650993
[63] YANG, Y. and DUNSON, D. B. (2014). Minimax optimal bayesian aggregation. ArXiv preprint. Available at arXiv:1403.1345.
[64] YANG, Y., WAINWRIGHT, M. J. and JORDAN, M. I. (2016). On the computational complexity of high-dimensional Bayesian variable selection. *Ann. Statist.* **44** 2497–2532. MR3576552 https://doi.org/10.1214/15-AOS1417

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

[66] ZHANG, C.-H. and HUANG, J. (2008). The sparsity and bias of the LASSO selection in high-dimensional linear regression. *Ann. Statist.* **36** 1567–1594. MR2435448 https://doi.org/10.1214/07-AOS520

*The Annals of Statistics*
2020, Vol. 48, No. 5, 2879–2903
https://doi.org/10.1214/19-AOS1913


# ASYMPTOTIC DISTRIBUTION AND DETECTION THRESHOLDS FOR TWO-SAMPLE TESTS BASED ON GEOMETRIC GRAPHS

BY BHASWAR B. BHATTACHARYA

*Department of Statistics, The Wharton School, University of Pennsylvania, bhaswar@wharton.upenn.edu*

In this paper, we consider the problem of testing the equality of two multivariate distributions based on geometric graphs constructed using the interpoint distances between the observations. These include the tests based on the minimum spanning tree and the $K$-nearest neighbor (NN) graphs, among others. These tests are asymptotically distribution-free, universally consistent and computationally efficient, making them particularly useful in modern applications. However, very little is known about the power properties of these tests. In this paper, using the theory of stabilizing geometric graphs, we derive the asymptotic distribution of these tests under general alternatives, in the Poissonized setting. Using this, the detection threshold and the limiting local power of the test based on the $K$-NN graph are obtained, where interesting exponents depending on dimension emerge. This provides a way to compare and justify the performance of these tests in different examples.

## REFERENCES

[1] ALDOUS, D. and STEELE, J. M. (1992). Asymptotics for Euclidean minimal spanning trees on random points. *Probab. Theory Related Fields* **92** 247–258. MR1161188 https://doi.org/10.1007/BF01194923

[2] ARIAS-CASTRO, E. and PELLETIER, B. (2016). On the consistency of the crossmatch test. *J. Statist. Plann. Inference* **171** 184–190. MR3458077 https://doi.org/10.1016/j.jspi.2015.10.003

[3] ASLAN, B. and ZECH, G. (2005). New test for the multivariate two-sample problem based on the concept of minimum energy. *J. Stat. Comput. Simul.* **75** 109–119. MR2117010 https://doi.org/10.1080/00949650410001661440

[4] BARINGHAUS, L. and FRANZ, C. (2004). On a new multivariate two-sample test. *J. Multivariate Anal.* **88** 190–206. MR2021870 https://doi.org/10.1016/S0047-259X(03)00079-4

[5] BATU, T., FORTNOW, L., RUBINFELD, R., SMITH, W. D. and WHITE, P. (2013). Testing closeness of discrete distributions. *J. ACM* **60** Art. 4, 25. MR3033221 https://doi.org/10.1145/2432622.2432626

[6] BHATTACHARYA, B. B. (2019). A general asymptotic framework for distribution-free graph-based two-sample tests. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 575–602. MR3961499

[7] BHATTACHARYA, B. B. (2020). Supplement to "Asymptotic distribution and detection thresholds for two-sample tests based on geometric graphs." https://doi.org/10.1214/19-AOS1913SUPP.

[8] BICKEL, P. J. (1968). A distribution free version of the Smirnov two sample test in the $p$-variate case. *Ann. Math. Stat.* **40** 1–23. MR0256519 https://doi.org/10.1214/aoms/1177697800

[9] BICKEL, P. J. and BREIMAN, L. (1983). Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. *Ann. Probab.* **11** 185–214. MR0682809

[10] BISWAS, M., MUKHOPADHYAY, M. and GHOSH, A. K. (2014). A distribution-free two-sample run test applicable to high-dimensional data. *Biometrika* **101** 913–926. MR3286925 https://doi.org/10.1093/biomet/asu045

[11] CHAN, S.-O., DIAKONIKOLAS, I., VALIANT, G. and VALIANT, P. (2014). Optimal algorithms for testing closeness of discrete distributions. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms* 1193–1203. ACM, New York. MR3376448 https://doi.org/10.1137/1.9781611973402.88

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

---

*MSC2020 subject classifications.* 62F07, 62G10, 60D05, 60F05, 60C05.

*Key words and phrases.* Efficiency, local power, geometric probability, nearest-neighbor graphs, nonparametric hypothesis testing.

[13] EBNER, B., HENZE, N. and YUKICH, J. E. (2018). Multivariate goodness-of-fit on flat and curved spaces via nearest neighbor distances. *J. Multivariate Anal.* **165** 231–242. MR3768763 https://doi.org/10.1016/j.jmva.2017.12.009

[14] FRIEDMAN, J. H. and RAFSKY, L. C. (1979). Multivariate generalizations of the Wald–Wolfowitz and Smirnov two-sample tests. *Ann. Statist.* **7** 697–717. MR0532236

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

[16] GYŐRFI, L. and NEMETZ, T. (1977). $f$-dissimilarity: A general class of separation measures of several probability measures. In *Topics in Information Theory. Colloq. Math. Soc. János Bolyai*, **16** 309–321. MR0459923

[17] HALL, P. and TAJVIDI, N. (2002). Permutation tests for equality of distributions in high-dimensional settings. *Biometrika* **89** 359–374. MR1913964 https://doi.org/10.1093/biomet/89.2.359

[18] HENZE, N. (1988). A multivariate two-sample test based on the number of nearest neighbor type coincidences. *Ann. Statist.* **16** 772–783. MR0947577 https://doi.org/10.1214/aos/1176350835

[19] HENZE, N. and PENROSE, M. D. (1999). On the multivariate runs test. *Ann. Statist.* **27** 290–298. MR1701112 https://doi.org/10.1214/aos/1018031112

[20] LIU, R. Y. and SINGH, K. (1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.* **88** 252–260. MR1212489

[21] MAA, J.-F., PEARL, D. K. and BARTOSZYŃSKI, R. (1996). Reducing multidimensional two-sample data to one-dimensional interpoint comparisons. *Ann. Statist.* **24** 1069–1074. MR1401837 https://doi.org/10.1214/aos/1032526956

[22] MANN, H. B. and WHITNEY, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **18** 50–60. MR0022058 https://doi.org/10.1214/aoms/1177730491

[23] PENROSE, M. (2003). *Random Geometric Graphs. Oxford Studies in Probability* **5**. Oxford Univ. Press, Oxford. MR1986198 https://doi.org/10.1093/acprof:oso/9780198506263.001.0001

[24] PENROSE, M. D. (2007). Gaussian limits for random geometric measures. *Electron. J. Probab.* **12** 989–1035. MR2336596 https://doi.org/10.1214/EJP.v12-429

[25] PENROSE, M. D. and YUKICH, J. E. (2003). Weak laws of large numbers in geometric probability. *Ann. Appl. Probab.* **13** 277–303. MR1952000 https://doi.org/10.1214/aoap/1042765669

[26] ROSENBAUM, P. R. (2005). An exact distribution-free test comparing two multivariate distributions based on adjacency. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 515–530. MR2168202 https://doi.org/10.1111/j.1467-9868.2005.00513.x

[27] ROUSSON, V. (2002). On distribution-free tests for the multivariate two-sample location-scale model. *J. Multivariate Anal.* **80** 43–57. MR1889832 https://doi.org/10.1006/jmva.2000.1981

[28] SCHILLING, M. F. (1986). Multivariate two-sample tests based on nearest neighbors. *J. Amer. Statist. Assoc.* **81** 799–806. MR0860514

[29] SMIRNOFF, N. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Moscow Univ. Math. Bull.* **2** 16. MR0002062

[30] TUKEY, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians* (*Vancouver*, *B. C.*, 1974), *Vol.* 2 523–531. MR0426989

[31] WALD, A. and WOLFOWITZ, J. (1940). On a test whether two samples are from the same population. *Ann. Math. Stat.* **11** 147–162. MR0002083 https://doi.org/10.1214/aoms/1177731909

[32] WEISS, L. (1960). Two-sample tests for multivariate distributions. *Ann. Math. Stat.* **31** 159–164. MR0119305 https://doi.org/10.1214/aoms/1177705995

[33] YUKICH, J. (2013). Limit theorems in discrete stochastic geometry. In *Stochastic Geometry, Spatial Statistics and Random Fields. Lecture Notes in Math.* **2068** 239–275. Springer, Heidelberg. MR3059650 https://doi.org/10.1007/978-3-642-33305-7_8

*The Annals of Statistics*
2020, Vol. 48, No. 5, 2904–2929
https://doi.org/10.1214/19-AOS1914


# CONTROLLED SEQUENTIAL MONTE CARLO

BY JEREMY HENG[1], ADRIAN N. BISHOP[2], GEORGE DELIGIANNIDIS[3,*] AND ARNAUD DOUCET[3,**]

[1]*ESSEC Business School, heng@essec.edu*
[2]*CSIRO, adrian.bishop@uts.edu.au*
[3]*University of Oxford, [*]deligian@stats.ox.ac.uk; [**]doucet@stats.ox.ac.uk*

Sequential Monte Carlo methods, also known as particle methods, are a popular set of techniques for approximating high-dimensional probability distributions and their normalizing constants. These methods have found numerous applications in statistics and related fields; for example, for inference in nonlinear non-Gaussian state space models, and in complex static models. Like many Monte Carlo sampling schemes, they rely on proposal distributions which crucially impact their performance. We introduce here a class of controlled sequential Monte Carlo algorithms, where the proposal distributions are determined by approximating the solution to an associated optimal control problem using an iterative scheme. This method builds upon a number of existing algorithms in econometrics, physics and statistics for inference in state space models, and generalizes these methods so as to accommodate complex static models. We provide a theoretical analysis concerning the fluctuation and stability of this methodology that also provides insight into the properties of related algorithms. We demonstrate significant gains over state-of-the-art methods at a fixed computational complexity on a variety of applications.

## REFERENCES

[1] ANDRIEU, C., DOUCET, A. and HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 269–342. MR2758115 https://doi.org/10.1111/j.1467-9868.2009.00736.x

[2] BERTSEKAS, D. P. and TSITSIKLIS, J. N. (1996). *Neuro-dynamic Programming*. Athena Scientific, Belmont, MA.

[3] BESKOS, A., JASRA, A., KANTAS, N. and THIERY, A. (2016). On the convergence of adaptive sequential Monte Carlo methods. *Ann. Appl. Probab.* **26** 1111–1146. MR3476634 https://doi.org/10.1214/15-AAP1113

[4] BICHTELER, K. (2002). *Stochastic Integration with Jumps. Encyclopedia of Mathematics and Its Applications* **89**. Cambridge Univ. Press, Cambridge. MR1906715 https://doi.org/10.1017/CBO9780511549878

[5] BRESLER, Y. (1986). Two-filter formula for discrete-time non-linear Bayesian smoothing. *Internat. J. Control* **43** 629–641.

[6] BRIERS, M., DOUCET, A. and MASKELL, S. (2010). Smoothing algorithms for state-space models. *Ann. Inst. Statist. Math.* **62** 61–89. MR2577439 https://doi.org/10.1007/s10463-009-0236-2

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

[8] CHOPIN, N. (2002). A sequential particle filter method for static models. *Biometrika* **89** 539–552. MR1929161 https://doi.org/10.1093/biomet/89.3.539

[9] CHOPIN, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.* **32** 2385–2411. MR2153989 https://doi.org/10.1214/009053604000000698

*MSC2020 subject classifications.* Primary 62M05; secondary 62F12, 62M10.
*Key words and phrases.* State space models, annealed importance sampling, normalizing constants, optimal control, approximate dynamic programming, reinforcement learning.

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

[11] Del Moral, P. (2004). *Feynman–Kac Formulae. Probability and Its Applications* (*New York*). Springer, New York. MR2044973 https://doi.org/10.1007/978-1-4684-9393-1

[12] Del Moral, P., Doucet, A. and Jasra, A. (2006). Sequential Monte Carlo samplers. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 411–436. MR2278333 https://doi.org/10.1111/j.1467-9868.2006.00553.x

[13] Del Moral, P. and Guionnet, A. (1999). Central limit theorem for nonlinear filtering and interacting particle systems. *Ann. Appl. Probab.* **9** 275–297. MR1687359 https://doi.org/10.1214/aoap/1029962742

[14] Diaconis, P. and Freedman, D. (1999). Iterated random functions. *SIAM Rev.* **41** 45–76. MR1669737 https://doi.org/10.1137/S0036144598338446

[15] Doucet, A., Godsill, S. J. and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat. Comput.* **10** 197–208.

[16] Doucet, A. and Johansen, A. M. (2011). A tutorial on particle filtering and smoothing: Fifteen years later. In *The Oxford Handbook of Nonlinear Filtering* (D. Crisan and B. L. Rozovsky, eds.) 656–704. Oxford Univ. Press, Oxford. MR2884612

[17] Gelman, A. and Meng, X.-L. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statist. Sci.* **13** 163–185. MR1647507 https://doi.org/10.1214/ss/1028905934

[18] Gerber, M., Chopin, N. and Whiteley, N. (2019). Negative association, ordering and convergence of resampling methods. *Ann. Statist.* **47** 2236–2260. MR3953450 https://doi.org/10.1214/18-AOS1746

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

[20] Gobet, E. (2016). *Monte-Carlo Methods and Stochastic Processes*: *From Linear to Non-linear*. CRC Press, Boca Raton, FL. MR3558839

[21] Gordon, N. J., Salmond, D. and Smith, A. F. M. (1993). A novel approach to non-linear/non-Gaussian Bayesian state estimation. *IEE Proc. Radar Signal Process.* **140** 107–113.

[22] Guarniero, P., Johansen, A. M. and Lee, A. (2017). The iterated auxiliary particle filter. *J. Amer. Statist. Assoc.* **112** 1636–1647. MR3750887 https://doi.org/10.1080/01621459.2016.1222291

[23] Gupta, A., Jain, R. and Glynn, P. (2018). A fixed point theorem for iterative random contraction operators over Banach spaces. Available at arXiv:1804.01195.

[24] Heng, J., Bishop, A. N, Deligiannidis, G. and Doucet, A. (2020). Supplement to "Controlled sequential Monte Carlo." https://doi.org/10.1214/19-AOS1914SUPP.

[25] Higham, N. J. (1988). Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra Appl.* **103** 103–118. MR0943997 https://doi.org/10.1016/0024-3795(88)90223-6

[26] Jacob, P. E., Murray, L. M. and Rubenthaler, S. (2015). Path storage in the particle filter. *Stat. Comput.* **25** 487–496. MR3306720 https://doi.org/10.1007/s11222-013-9445-x

[27] Jasra, A., Stephens, D. A., Doucet, A. and Tsagaris, T. (2011). Inference for Lévy-driven stochastic volatility models via adaptive sequential Monte Carlo. *Scand. J. Stat.* **38** 1–22. MR2760137 https://doi.org/10.1111/j.1467-9469.2010.00723.x

[28] Kappen, H. J. and Ruiz, H.-C. (2016). Adaptive importance sampling for control and inference. *J. Stat. Phys.* **162** 1244–1266. MR3462019 https://doi.org/10.1007/s10955-016-1446-7

[29] Künsch, H. R. (2005). Recursive Monte Carlo filters: Algorithms and theoretical analysis. *Ann. Statist.* **33** 1983–2021. MR2211077 https://doi.org/10.1214/009053605000000426

[30] Künsch, H. R. (2013). Particle filters. *Bernoulli* **19** 1391–1403. MR3102556 https://doi.org/10.3150/12-BEJSP07

[31] Lin, M., Chen, R. and Liu, J. S. (2013). Lookahead strategies for sequential Monte Carlo. *Statist. Sci.* **28** 69–94. MR3075339 https://doi.org/10.1214/12-STS401

[32] Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing. Springer Series in Statistics*. Springer, New York. MR1842342

[33] Liu, J. S. and Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems. *J. Amer. Statist. Assoc.* **93** 1032–1044. MR1649198 https://doi.org/10.2307/2669847

[34] Lorenz, E. N. (1996). Predictability: A problem partly solved. In *Proc. Seminar on Predictability* **1**.

[35] Mattingly, J. C., Stuart, A. M. and Higham, D. J. (2002). Ergodicity for SDEs and approximations: Locally Lipschitz vector fields and degenerate noise. *Stochastic Process. Appl.* **101** 185–232. MR1931266 https://doi.org/10.1016/S0304-4149(02)00150-3

[36] Milstein, G. N. and Tretyakov, M. V. (2004). *Stochastic Numerics for Mathematical Physics. Scientific Computation*. Springer, Berlin. MR2069903 https://doi.org/10.1007/978-3-662-10063-9

[37] MØLLER, J., SYVERSVEEN, A. R. and WAAGEPETERSEN, R. P. (1998). Log Gaussian Cox processes. *Scand. J. Stat.* **25** 451–482. MR1650019 https://doi.org/10.1111/1467-9469.00115

[38] MURRAY, L. M., SINGH, S., JACOB, P. E. and LEE, A. (2016). Anytime Monte Carlo. Available at arXiv:612.03319.

[39] NEAL, R. M. (2001). Annealed importance sampling. *Stat. Comput.* **11** 125–139. MR1837132 https://doi.org/10.1023/A:1008923215028

[40] NEMOTO, T., BOUCHET, F., JACK, R. L. and LECOMTE, V. (2016). Population-dynamics method with a multicanonical feedback control. *Phys. Rev. E* **93** 062123. MR3713187 https://doi.org/10.1103/physreve.93.062123

[41] PITT, M. K. and SHEPHARD, N. (1999). Filtering via simulation: Auxiliary particle filters. *J. Amer. Statist. Assoc.* **94** 590–599. MR1702328 https://doi.org/10.2307/2670179

[42] RICHARD, J.-F. and ZHANG, W. (2007). Efficient high-dimensional importance sampling. *J. Econometrics* **141** 1385–1411. MR2413506 https://doi.org/10.1016/j.jeconom.2007.02.007

[43] ROBERTS, G. O. and STRAMER, O. (2002). Langevin diffusions and Metropolis–Hastings algorithms. *Methodol. Comput. Appl. Probab.* **4** 337–357. MR2002247 https://doi.org/10.1023/A:1023562417138

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

[45] RUIZ, H.-C. and KAPPEN, H. J. (2017). Particle smoothing for hidden diffusion processes: Adaptive path integral smoother. *IEEE Trans. Signal Process.* **65** 3191–3203. MR3650283 https://doi.org/10.1109/TSP.2017.2686340

[46] SCHÄFER, C. and CHOPIN, N. (2013). Sequential Monte Carlo on large binary sampling spaces. *Stat. Comput.* **23** 163–184. MR3016936 https://doi.org/10.1007/s11222-011-9299-z

[47] SCHARTH, M. and KOHN, R. (2016). Particle efficient importance sampling. *J. Econometrics* **190** 133–147. MR3425276 https://doi.org/10.1016/j.jeconom.2015.03.047

[48] TEMEREANCA, S., BROWN, E. N. and SIMONS, D. J. (2008). Rapid changes in thalamic firing synchrony during repetitive whisker stimulation. *J. Neurosci.* **28** 11153–11164.

[49] THEODOROU, E. A. and TODOROV, E. (2012). Relative entropy and free energy dualities: Connections to path integral and KL control. In *Proceedings* 51*st IEEE Conference on Decision and Control* (*CDC*), 1466–1473. Oxford Univ. Press, Oxford.

[50] THIJSSEN, S. and KAPPEN, H. J. (2015). Path integral control and state-dependent feedback. *Phys. Rev. E* **91** 032104. MR3460947 https://doi.org/10.1103/PhysRevE.91.032104

[51] ZHOU, Y., JOHANSEN, A. M. and ASTON, J. A. D. (2016). Toward automatic model comparison: An adaptive sequential Monte Carlo approach. *J. Comput. Graph. Statist.* **25** 701–726. MR3533634 https://doi.org/10.1080/10618600.2015.1060885

*The Annals of Statistics*
2020, Vol. 48, No. 5, 2930–2952
https://doi.org/10.1214/19-AOS1916


# A FRAMEWORK FOR ADAPTIVE MCMC TARGETING MULTIMODAL DISTRIBUTIONS

BY EMILIA POMPE[1,*], CHRIS HOLMES[1,†] AND KRZYSZTOF ŁATUSZYŃSKI[2]

[1] *Department of Statistics, University of Oxford,* [*] *emilia.pompe@stats.ox.ac.uk;* [†] *cholmes@stats.ox.ac.uk*
[2] *Department of Statistics, University of Warwick, K.G.Latuszynski@warwick.ac.uk*

We propose a new Monte Carlo method for sampling from multimodal distributions. The idea of this technique is based on splitting the task into two: finding the modes of a target distribution $\pi$ and sampling, given the knowledge of the locations of the modes. The sampling algorithm relies on steps of two types: local ones, preserving the mode; and jumps to regions associated with different modes. Besides, the method learns the optimal parameters of the algorithm, while it runs, without requiring user intervention. Our technique should be considered as a flexible framework, in which the design of moves can follow various strategies known from the broad MCMC literature.

In order to design an adaptive scheme that facilitates both local and jump moves, we introduce an auxiliary variable representing each mode, and we define a new target distribution $\tilde{\pi}$ on an augmented state space $\mathcal{X} \times \mathcal{I}$, where $\mathcal{X}$ is the original state space of $\pi$ and $\mathcal{I}$ is the set of the modes. As the algorithm runs and updates its parameters, the target distribution $\tilde{\pi}$ also keeps being modified. This motivates a new class of algorithms, Auxiliary Variable Adaptive MCMC. We prove general ergodic results for the whole class before specialising to the case of our algorithm.

## REFERENCES

[1] AHN, S., CHEN, Y. and WELLING, M. (2013). Distributed and adaptive darting Monte Carlo through regenerations. In *Artificial Intelligence and Statistics* 108–116.

[2] ANDRICIOAEI, I., STRAUB, J. E. and VOTER, A. F. (2001). Smart darting Monte Carlo. *J. Chem. Phys.* **114** 6994–7000.

[3] ANDRIEU, C. and MOULINES, É. (2006). On the ergodicity properties of some adaptive MCMC algorithms. *Ann. Appl. Probab.* **16** 1462–1505. MR2260070 https://doi.org/10.1214/105051606000000286

[4] ANDRIEU, C. and ROBERTS, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* **37** 697–725. MR2502648 https://doi.org/10.1214/07-AOS574

[5] ANDRIEU, C. and THOMS, J. (2008). A tutorial on adaptive MCMC. *Stat. Comput.* **18** 343–373. MR2461882 https://doi.org/10.1007/s11222-008-9110-y

[6] ANDRIEU, C. and VIHOLA, M. (2015). Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms. *Ann. Appl. Probab.* **25** 1030–1077. MR3313762 https://doi.org/10.1214/14-AAP1022

[7] ATCHADÉ, Y. and FORT, G. (2010). Limit theorems for some adaptive MCMC algorithms with subgeometric kernels. *Bernoulli* **16** 116–154. MR2648752 https://doi.org/10.3150/09-BEJ199

[8] ATCHADÉ, Y. F., ROBERTS, G. O. and ROSENTHAL, J. S. (2011). Towards optimal scaling of Metropolis-coupled Markov chain Monte Carlo. *Stat. Comput.* **21** 555–568. MR2826692 https://doi.org/10.1007/s11222-010-9192-1

[9] BAI, Y., ROBERTS, G. O. and ROSENTHAL, J. S. (2011). On the containment condition for adaptive Markov chain Monte Carlo algorithms. *Adv. Appl. Stat.* **21** 1–54. MR2849670

[10] BORNN, L., JACOB, P. E., DEL MORAL, P. and DOUCET, A. (2013). An adaptive interacting Wang–Landau algorithm for automatic density exploration. *J. Comput. Graph. Statist.* **22** 749–773. MR3173740 https://doi.org/10.1080/10618600.2012.723569

[11] CHIMISOV, C., ŁATUSZYŃSKI, K. and ROBERTS, G. (2018). Air Markov chain Monte Carlo. Preprint. Available at arXiv:1801.09309.

*MSC2020 subject classifications.* Primary 60J05, 65C05; secondary 62F15.
*Key words and phrases.* Multimodal distribution, adaptive MCMC, ergodicity.

[12] CRAIU, R. V., ROSENTHAL, J. and YANG, C. (2009). Learn from thy neighbor: Parallel-chain and regional adaptive MCMC. *J. Amer. Statist. Assoc.* **104** 1454–1466. MR2750572 https://doi.org/10.1198/jasa.2009.tm08393

[13] DRTON, M. and RICHARDSON, T. S. (2004). Multimodality of the likelihood in the bivariate seemingly unrelated regressions model. *Biometrika* **91** 383–392. MR2081308 https://doi.org/10.1093/biomet/91.2.383

[14] FEROZ, F., HOBSON, M. P. and BRIDGES, M. (2009). MultiNest: An efficient and robust Bayesian inference tool for cosmology and particle physics. *Mon. Not. R. Astron. Soc.* **398** 1601–1614.

[15] FEROZ, F., HOBSON, M. P., CAMERON, E. and PETTITT, A. N. (2013). Importance nested sampling and the MultiNest algorithm. Preprint. Available at arXiv:1306.2144.

[16] FORT, G., MOULINES, E. and PRIOURET, P. (2011). Convergence of adaptive and interacting Markov chain Monte Carlo algorithms. *Ann. Statist.* **39** 3262–3289. MR3012408 https://doi.org/10.1214/11-AOS938

[17] GÅSEMYR, J. (2006). The spectrum of the independent Metropolis–Hastings algorithm. *J. Theoret. Probab.* **19** 152–165. MR2256484 https://doi.org/10.1007/s10959-006-0009-2

[18] GEYER, C. J. (1991). Markov chain Monte Carlo maximum likelihood.

[19] GREEN, P. J., ŁATUSZYŃSKI, K., PEREYRA, M. and ROBERT, C. P. (2015). Bayesian computation: A summary of the current state, and samples backwards and forwards. *Stat. Comput.* **25** 835–862. MR3360496 https://doi.org/10.1007/s11222-015-9574-5

[20] HAARIO, H., SAKSMAN, E. and TAMMINEN, J. (2001). An adaptive Metropolis algorithm. *Bernoulli* **7** 223–242. MR1828504 https://doi.org/10.2307/3318737

[21] IHLER, A. T., FISHER, J. W., MOSES, R. L. and WILLSKY, A. S. (2005). Nonparametric belief propagation for self-localization of sensor networks. *IEEE J. Sel. Areas Commun.* **23** 809–819.

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

[23] JASRA, A., HOLMES, C. C. and STEPHENS, D. A. (2005). Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statist. Sci.* **20** 50–67. MR2182987 https://doi.org/10.1214/088342305000000016

[24] KOU, S. C., ZHOU, Q. and WONG, W. H. (2006). Equi-energy sampler with applications in statistical inference and statistical mechanics. *Ann. Statist.* **34** 1581–1619. MR2283711 https://doi.org/10.1214/009053606000000515

[25] LAN, S., STREETS, J. and SHAHBABA, B. (2014). Wormhole Hamiltonian Monte Carlo. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence* 1953–1959.

[26] MARINARI, E. and PARISI, G. (1992). Simulated tempering: A new Monte Carlo scheme. *Europhys. Lett.* **19** 451–458.

[27] MIASOJEDOW, B., MOULINES, E. and VIHOLA, M. (2013). An adaptive parallel tempering algorithm. *J. Comput. Graph. Statist.* **22** 649–664. MR3173735 https://doi.org/10.1080/10618600.2013.778779

[28] MYKLAND, P., TIERNEY, L. and YU, B. (1995). Regeneration in Markov chain samplers. *J. Amer. Statist. Assoc.* **90** 233–241. MR1325131

[29] NEAL, R. M. (1996). Sampling from multimodal distributions using tempered transitions. *Stat. Comput.* **6** 353–366.

[30] NOCEDAL, J. and WRIGHT, S. J. (2006). *Numerical Optimization*, 2nd ed. *Springer Series in Operations Research and Financial Engineering*. Springer, New York. MR2244940

[31] PAULIN, D., JASRA, A. and THIERY, A. (2019). Error bounds for sequential Monte Carlo samplers for multimodal distributions. *Bernoulli* **25** 310–340. MR3892321 https://doi.org/10.3150/17-bej988

[32] POMPE, E., HOLMES, C. and ŁATUSZYŃSKI, K. (2020). Supplement to "A framework for adaptive MCMC targeting multimodal distributions." https://doi.org/10.1214/19-AOS1916SUPP

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

[34] ROBERTS, G. O. and ROSENTHAL, J. S. (2001). Optimal scaling for various Metropolis–Hastings algorithms. *Statist. Sci.* **16** 351–367. MR1888450 https://doi.org/10.1214/ss/1015346320

[35] ROBERTS, G. O. and ROSENTHAL, J. S. (2007). Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *J. Appl. Probab.* **44** 458–475. MR2340211 https://doi.org/10.1239/jap/1183667414

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

[37] ROSENTHAL, J. S. (2011). Optimal proposal distributions and adaptive MCMC. In *Handbook of Markov Chain Monte Carlo*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 93–111. CRC Press, Boca Raton, FL. MR2858446

[38] SMINCHISESCU, C. and WELLING, M. (2011). Generalized darting Monte Carlo. *Pattern Recognit.* **44** 2738–2748.

[39] TAK, H., MENG, X.-L. and VAN DYK, D. A. (2018). A repelling-attracting Metropolis algorithm for multimodality. *J. Comput. Graph. Statist.* **27** 479–490. MR3863751 https://doi.org/10.1080/10618600.2017.1415911

[40] TAWN, N. G., ROBERTS, G. O. and ROSENTHAL, J. S. (2020). Weight-preserving simulated tempering. *Stat. Comput.* **30** 27–41. MR4057469 https://doi.org/10.1007/s11222-019-09863-3

[41] TJELMELAND, H. and HEGSTAD, B. K. (2001). Mode jumping proposals in MCMC. *Scand. J. Stat.* **28** 205–223. MR1844357 https://doi.org/10.1111/1467-9469.00232

[42] VIHOLA, M. (2011). On the stability and ergodicity of adaptive scaling Metropolis algorithms. *Stochastic Process. Appl.* **121** 2839–2860. MR2844543 https://doi.org/10.1016/j.spa.2011.08.006

[43] WANG, F. and LANDAU, D. P. (2001). Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram. *Phys. Rev. E* **64** Art. ID 056101.

[44] WANG, F. and LANDAU, D. P. (2001). Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* **86** 2050–2053. https://doi.org/10.1103/PhysRevLett.86.2050

[45] WOODARD, D. B., SCHMIDLER, S. C. and HUBER, M. (2009). Sufficient conditions for torpid mixing of parallel and simulated tempering. *Electron. J. Probab.* **14** 780–804. MR2495560 https://doi.org/10.1214/EJP.v14-638

[46] ZHOU, Q. (2011). Multi-domain sampling with applications to structural inference of Bayesian networks. *J. Amer. Statist. Assoc.* **106** 1317–1330. MR2896838 https://doi.org/10.1198/jasa.2011.ap10346

*The Annals of Statistics*
2020, Vol. 48, No. 5, 2953–2981
https://doi.org/10.1214/19-AOS1917


# VALID POST-SELECTION INFERENCE IN MODEL-FREE LINEAR REGRESSION

BY ARUN K. KUCHIBHOTLA*, LAWRENCE D. BROWN, ANDREAS BUJA, JUNHUI CAI, EDWARD I. GEORGE AND LINDA H. ZHAO

*Department of Statistics, The Wharton School, University of Pennsylvania,* *arunku@wharton.upenn.edu

Modern data-driven approaches to modeling make extensive use of covariate/model selection. Such selection incurs a cost: it invalidates classical statistical inference. A conservative remedy to the problem was proposed by Berk et al. (*Ann. Statist.* **41** (2013) 802–837) and further extended by Bachoc, Preinerstorfer and Steinberger (2016). These proposals, labeled "PoSI methods," provide valid inference after arbitrary model selection. They are computationally NP-hard and have limitations in their theoretical justifications. We therefore propose computationally efficient confidence regions, named "UPoSI"[1] and prove large-$p$ asymptotics for them. We do this for linear OLS regression allowing misspecification of the normal linear model, for both fixed and random covariates, and for independent as well as some types of dependent data. We start by proving a general equivalence result for the post-selection inference problem and a simultaneous inference problem in a setting that strips inessential features still present in a related result of Berk et al. (*Ann. Statist.* **41** (2013) 802–837). We then construct valid PoSI confidence regions that are the first to have vastly improved computational efficiency in that the required computation times grow only quadratically rather than exponentially with the total number $p$ of covariates. These are also the first PoSI confidence regions with guaranteed asymptotic validity when the total number of covariates $p$ diverges (almost exponentially) with the sample size $n$. Under standard tail assumptions, we only require $(\log p)^7 = o(n)$ and $k = o(\sqrt{n/\log p})$ where $k$ ($\le p$) is the largest number of covariates (model size) considered for selection. We study various properties of these confidence regions, including their Lebesgue measures, and compare them theoretically with those proposed previously.

## REFERENCES

BACHOC, F., BLANCHARD, G. and NEUVIAL, P. (2018). On the post selection inference constant under restricted isometry properties. *Electron. J. Stat.* **12** 3736–3757. MR3878579 https://doi.org/10.1214/18-ejs1490

BACHOC, F., LEEB, H. and PÖTSCHER, B. M. (2019). Valid confidence intervals for post-model-selection predictors. *Ann. Statist.* **47** 1475–1504. MR3911119 https://doi.org/10.1214/18-AOS1721

BACHOC, F., PREINERSTORFER, D. and STEINBERGER, L. (2016). Uniformly valid confidence intervals post-model-selection. Preprint. Available at arXiv:1611.01043.

BELLONI, A., ROSENBAUM, M. and TSYBAKOV, A. B. (2017). Linear and conic programming estimators in high dimensional errors-in-variables models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 939–956. MR3641415 https://doi.org/10.1111/rssb.12196

BERK, R., BROWN, L., BUJA, A., ZHANG, K. and ZHAO, L. (2013). Valid post-selection inference. *Ann. Statist.* **41** 802–837. MR3099122 https://doi.org/10.1214/12-AOS1077

BUEHLER, R. J. and FEDDERSEN, A. P. (1963). Note on a conditional property of Student's $t$. *Ann. Math. Stat.* **34** 1098–1100. MR0150864 https://doi.org/10.1214/aoms/1177704034

BUJA, A., BROWN, L. D., BERK, R. A., GEORGE, E. I., TRASKIN, M., PITKIN, E., ZHAO, L. H. and ZHANG, K. (2019). Models as approximations, part I: Consequences illustrated with linear regression. *Statist. Sci.* To appear. Available at arXiv:1404.1578.

*MSC2020 subject classifications.* 62J05, 62F40, 62F25, 62F12.

*Key words and phrases.* Simultaneous inference, multiplier bootstrap, uniform consistency, high-dimensional linear regression, concentration inequalities, Orlicz norms, model selection.

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

CLAESKENS, G. and CARROLL, R. J. (2007). An asymptotic theory for model selection inference in general semiparametric problems. *Biometrika* **94** 249–265. MR2331485 https://doi.org/10.1093/biomet/asm034

CUI, Y., LENG, C. and SUN, D. (2016). Sparse estimation of high-dimensional correlation matrices. *Comput. Statist. Data Anal.* **93** 390–403. MR3406221 https://doi.org/10.1016/j.csda.2014.10.001

DENG, H. and ZHANG, C.-H. (2017). Beyond Gaussian approximation: Bootstrap for maxima of sums of independent random vectors. Preprint. Available at arXiv:1705.09528.

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

FREEDMAN, D. A. (1983). A note on screening regression equations. *Amer. Statist.* **37** 152–155. MR0702208 https://doi.org/10.2307/2685877

HJORT, N. L. and CLAESKENS, G. (2003). Frequentist model average estimators. *J. Amer. Statist. Assoc.* **98** 879–899. MR2041481 https://doi.org/10.1198/016214503000000828

KIVARANOVIC, D. and LEEB, H. (2018). Expected length of post-model-selection confidence intervals conditional on polyhedral constraints. Preprint. Available at arXiv:1803.01665.

KUCHIBHOTLA, A. K., BROWN, L. D. and BUJA, A. (2018). Model-free study of ordinary least squares linear regression. Preprint. Available at arXiv:1809.10538.

KUCHIBHOTLA, A. K. and CHAKRABORTTY, A. (2018). Moving beyond sub-Gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression. Preprint. Available at arXiv:1804.02605.

KUCHIBHOTLA, A. K., BROWN, L. D., BUJA, A., GEORGE, E. I. and ZHAO, L. (2018). A model free perspective for linear regression: Uniform-in-model bounds for post selection inference. Preprint. Available at arXiv:1802.05801.

KUCHIBHOTLA, A. K., BROWN, L. D., BUJA, A., CAI, J., GEORGE, E. I. and ZHAO, L. H. (2020). Supplement to "Valid post-selection inference in model-free linear regression." https://doi.org/10.1214/19-AOS1917SUPP.

LEE, J. D., SUN, D. L., SUN, Y. and TAYLOR, J. E. (2016). Exact post-selection inference, with application to the lasso. *Ann. Statist.* **44** 907–927. MR3485948 https://doi.org/10.1214/15-AOS1371

LIU, R. Y. and SINGH, K. (1995). Using i.i.d. bootstrap inference for general non-i.i.d. models. *J. Statist. Plann. Inference* **43** 67–75. MR1314128 https://doi.org/10.1016/0378-3758(94)00008-J

OLSHEN, R. A. (1973). The conditional level of the $F$-test. *J. Amer. Statist. Assoc.* **68** 692–698. MR0359198

RENCHER, A. C. and PUN, F. C. (1980). Inflation of $R^2$ in best subset regression. *Technometrics* **22** 49–53.

RINALDO, A., WASSERMAN, L., G'SELL, M.,and LEI, J. (2019). Bootstrapping and sample splitting for high-dimensional, assumption-free inference. *Ann. Statist.* To appear. Available at arXiv:1611.05401.

SIMMONS, J. P., NELSON, L. D. and SIMONSOHN, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* **22** 1359–1366.

TIAN, X., BI, N. and TAYLOR, J. (2016). MAGIC: A general, powerful and tractable method for selective inference. Preprint. Available at arXiv:1607.02630.

TIBSHIRANI, R. J., TAYLOR, J., LOCKHART, R. and TIBSHIRANI, R. (2016). Exact post-selection inference for sequential regression procedures. *J. Amer. Statist. Assoc.* **111** 600–620. MR3538689 https://doi.org/10.1080/01621459.2015.1108848

TIBSHIRANI, R. J., RINALDO, A., TIBSHIRANI, R. and WASSERMAN, L. (2018). Uniform asymptotic inference and the bootstrap after model selection. *Ann. Statist.* **46** 1255–1287. MR3798003 https://doi.org/10.1214/17-AOS1584

ZHANG, X. and CHENG, G. (2014). Bootstrapping high dimensional time series. Preprint. Available at arXiv:1406.1037.

*The Annals of Statistics*
2020, Vol. 48, No. 5, 2982–2998
https://doi.org/10.1214/19-AOS1918


# INFERENCE FOR SPHERICAL LOCATION UNDER HIGH CONCENTRATION

BY DAVY PAINDAVEINE[1,2,*] AND THOMAS VERDEBOUT[1,**]

[1]*Université libre de Bruxelles.* [*]*dpaindav@ulb.ac.be;* [**]*tverdebo@ulb.ac.be*
[2]*Université Toulouse 1 Capitole*

Motivated by the fact that circular or spherical data are often much concentrated around a location $\boldsymbol{\theta}$, we consider inference about $\boldsymbol{\theta}$ under *high concentration* asymptotic scenarios for which the probability of any fixed spherical cap centered at $\boldsymbol{\theta}$ converges to one as the sample size $n$ diverges to infinity. Rather than restricting to Fisher–von Mises–Langevin distributions, we consider a much broader, semiparametric, class of rotationally symmetric distributions indexed by the location parameter $\boldsymbol{\theta}$, a scalar concentration parameter $\kappa$ and a functional nuisance $f$. We determine the class of distributions for which high concentration is obtained as $\kappa$ diverges to infinity. For such distributions, we then consider inference (point estimation, confidence zone estimation, hypothesis testing) on $\boldsymbol{\theta}$ in asymptotic scenarios where $\kappa_n$ diverges to infinity at an arbitrary rate with the sample size $n$. Our asymptotic investigation reveals that, interestingly, optimal inference procedures on $\boldsymbol{\theta}$ show consistency rates that depend on $f$. Using asymptotics "à la Le Cam," we show that the spherical mean is, at any $f$, a parametrically superefficient estimator of $\boldsymbol{\theta}$ and that the Watson and Wald tests for $\mathcal{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ enjoy similar, nonstandard, optimality properties. We illustrate our results through simulations and treat a real data example. On a technical point of view, our asymptotic derivations require challenging expansions of rotationally symmetric functionals for large arguments of $f$.

## REFERENCES

ARNOLD, R. and JUPP, P. E. (2013). Statistics of orthogonal axial frames. *Biometrika* **100** 571–586. MR3094438 https://doi.org/10.1093/biomet/ast017

ARNOLD, R., JUPP, P. E. and SCHAEBEN, H. (2018). Statistics of ambiguous rotations. *J. Multivariate Anal.* **165** 73–85. MR3768753 https://doi.org/10.1016/j.jmva.2017.10.007

CHIKUSE, Y. (2003a). Concentrated matrix Langevin distributions. *J. Multivariate Anal.* **85** 375–394. MR1983803 https://doi.org/10.1016/S0047-259X(02)00065-9

CHIKUSE, Y. (2003b). *Statistics on Special Manifolds. Lecture Notes in Statistics* **174**. Springer, New York. MR1960435 https://doi.org/10.1007/978-0-387-21540-2

DAI, X. and MÜLLER, H.-G. (2018). Principal component analysis for functional data on Riemannian manifolds and spheres. *Ann. Statist.* **46** 3334–3361. MR3852654 https://doi.org/10.1214/17-AOS1660

DOWNS, T. D. (2003). Spherical regression. *Biometrika* **90** 655–668. MR2006842 https://doi.org/10.1093/biomet/90.3.655

DOWNS, T. D. and MARDIA, K. V. (2002). Circular regression. *Biometrika* **89** 683–697. MR1929172 https://doi.org/10.1093/biomet/89.3.683

FISHER, N. I., LEWIS, T. and EMBLETON, B. J. J. (1987). *Statistical Analysis of Spherical Data*. Cambridge Univ. Press, Cambridge. MR0899958 https://doi.org/10.1017/CBO9780511623059

FUJIKOSHI, Y. and WATAMORI, Y. (1992). Tests for the mean direction of the Langevin distribution with large concentration parameter. *J. Multivariate Anal.* **42** 210–225. MR1183843 https://doi.org/10.1016/0047-259X(92)90044-G

GARCÍA-PORTUGUÉS, E., PAINDAVEINE, D. and VERDEBOUT, T. (2019). On optimal tests for rotational symmetry against new classes of hyperspherical distributions. *J. Amer. Statist. Assoc.*, to appear.

*MSC2020 subject classifications.* Primary 62E20, 62F30; secondary 62F05, 62F12.
*Key words and phrases.* Concentrated distributions, directional statistics, Le Cam's asymptotic theory of statistical experiments, local asymptotic normality, superefficiency.

HALLIN, M. and PAINDAVEINE, D. (2002). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist.* **30** 1103–1133. MR1926170 https://doi.org/10.1214/aos/1031689019

HAYAKAWA, T. (1990). On tests for the mean direction of the Langevin distribution. *Ann. Inst. Statist. Math.* **42** 359–373. MR1064794 https://doi.org/10.1007/BF00050842

HAYAKAWA, T. and PURI, M. L. (1985). Asymptotic expansions of the distributions of some test statistics. *Ann. Inst. Statist. Math.* **37** 95–108. MR0790378 https://doi.org/10.1007/BF02481083

JUPP, P. E. (2015). Copulae on products of compact Riemannian manifolds. *J. Multivariate Anal.* **140** 92–98. MR3372555 https://doi.org/10.1016/j.jmva.2015.04.008

LEY, C. and VERDEBOUT, T. (2017). *Modern Directional Statistics. Chapman & Hall/CRC Interdisciplinary Statistics Series*. CRC Press, Boca Raton, FL. MR3752655

MARDIA, K. V. and JUPP, P. E. (2000). *Directional Statistics. Wiley Series in Probability and Statistics*. Wiley, Chichester. Revised reprint of *Statistics of directional data* by Mardia [MR0336854 (49 #1627)]. MR1828667

PAINDAVEINE, D. and VERDEBOUT, T. (2017). Inference on the mode of weak directional signals: A Le Cam perspective on hypothesis testing near singularities. *Ann. Statist.* **45** 800–832. MR3650401 https://doi.org/10.1214/16-AOS1468

PAINDAVEINE, D. and VERDEBOUT, T. (2020). Supplement to "Inference for spherical location under high concentration." https://doi.org/10.1214/19-AOS1918SUPP.

RIVEST, L.-P. (1986). Modified Kent's statistics for testing goodness of fit for the Fisher distribution in small concentrated samples. *Statist. Probab. Lett.* **4** 1–4. MR0822715 https://doi.org/10.1016/0167-7152(86)90028-3

RIVEST, L.-P. (1989). Spherical regression for concentrated Fisher–von Mises distributions. *Ann. Statist.* **17** 307–317. MR0981452 https://doi.org/10.1214/aos/1176347018

RIVEST, L.-P., DUCHESNE, T., NICOSIA, A. and FORTIN, D. (2016). A general angular regression model for the analysis of data on animal movement in ecology. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **65** 445–463. MR3470586 https://doi.org/10.1111/rssc.12124

ROSENTHAL, M., WU, W., KLASSEN, E. and SRIVASTAVA, A. (2014). Spherical regression models using projective linear transformations. *J. Amer. Statist. Assoc.* **109** 1615–1624. MR3293614 https://doi.org/10.1080/01621459.2014.892881

SENGUPTA, A., KIM, S. and ARNOLD, B. C. (2013). Inverse circular-circular regression. *J. Multivariate Anal.* **119** 200–208. MR3061424 https://doi.org/10.1016/j.jmva.2013.04.011

SHORACK, G. R. (2000). *Probability for Statisticians. Springer Texts in Statistics*. Springer, New York. MR1762415

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

WATAMORI, Y. (1996). Statistical inference of Langevin distribution for directional data. *Hiroshima Math. J.* **26** 25–74. MR1380423

WATSON, G. S. (1983). *Statistics on Spheres. University of Arkansas Lecture Notes in the Mathematical Sciences* **6**. Wiley, New York. A Wiley-Interscience Publication. MR0709262

WATSON, G. S. (1984). The theory of concentrated Langevin distributions. *J. Multivariate Anal.* **14** 74–82. MR0734099 https://doi.org/10.1016/0047-259X(84)90047-2

*The Annals of Statistics*
2020, Vol. 48, No. 5, 2999–3020
https://doi.org/10.1214/19-AOS1919


# SEMIPARAMETRIC BAYESIAN CAUSAL INFERENCE

BY KOLYAN RAY[1] AND AAD VAN DER VAART[2]

[1]*Department of Mathematics, Imperial College London, kolyan.ray@imperial.ac.uk*
[2]*Mathematical Institute, Leiden University, avdvaart@math.leidenuniv.nl*

We develop a semiparametric Bayesian approach for estimating the mean response in a missing data model with binary outcomes and a nonparametrically modelled propensity score. Equivalently, we estimate the causal effect of a treatment, correcting nonparametrically for confounding. We show that standard Gaussian process priors satisfy a semiparametric Bernstein–von Mises theorem under smoothness conditions. We further propose a novel propensity score-dependent prior that provides efficient inference under strictly weaker conditions. We also show that it is theoretically preferable to model the covariate distribution with a Dirichlet process or Bayesian bootstrap, rather than modelling its density.

## REFERENCES

[1] ALAA, A. and VAN DER SCHAAR, M. (2018). Limits of estimating heterogeneous treatment effects: Guidelines for practical algorithm design. In *Proceedings of the* 35*th International Conference on Machine Learning* 129–138.

[2] ALAA, A. M. and VAN DER SCHAAR, M. (2017). Deep multi-task Gaussian processes for survival analysis with competing risks. In *Advances in Neural Information Processing Systems* 30 2329–2337.

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

[4] ATHEY, S., IMBENS, G., PHAM, T. and WAGER, S. (2017). Estimating average treatment effects: Supplementary analyses and remaining challenges. *Am. Econ. Rev.* **107** 278–281. https://doi.org/10.1257/aer.p20171042

[5] BEGUN, J. M., HALL, W. J., HUANG, W.-M. and WELLNER, J. A. (1983). Information and asymptotic efficiency in parametric–nonparametric models. *Ann. Statist.* **11** 432–452. MR0696057 https://doi.org/10.1214/aos/1176346151

[6] BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York. MR1623559

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

[8] BICKEL, P. J. and KWON, J. (2001). Inference for semiparametric models: Some questions and an answer. *Statist. Sinica* **11** 863–960. MR1867326

[9] CASTILLO, I. (2008). Lower bounds for posterior rates with Gaussian process priors. *Electron. J. Stat.* **2** 1281–1299. MR2471287 https://doi.org/10.1214/08-EJS273

[10] CASTILLO, I. (2012). A semiparametric Bernstein–von Mises theorem for Gaussian process priors. *Probab. Theory Related Fields* **152** 53–99. MR2875753 https://doi.org/10.1007/s00440-010-0316-5

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

[12] DUDLEY, R. M. (2002). *Real Analysis and Probability*. *Cambridge Studies in Advanced Mathematics* **74**. Cambridge Univ. Press, Cambridge. MR1932358 https://doi.org/10.1017/CBO9780511755347

[13] FERGUSON, T. S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2** 615–629. MR0438568

[14] FREEDMAN, D. (1999). On the Bernstein–von Mises theorem with infinite-dimensional parameters. *Ann. Statist.* **27** 1119–1140. MR1740119 https://doi.org/10.1214/aos/1017938917

*MSC2020 subject classifications.* Primary 62G20; secondary 62G15, 62G08.
*Key words and phrases.* Bernstein–von Mises, Gaussian processes, propensity score-dependent priors, causal inference, Dirichlet process.

[15] FUTOMA, J., HARIHARAN, S. and HELLER, K. (2017). Learning to detect sepsis with a multitask Gaussian process RNN classifier. In *Proceedings of the* 34*th International Conference on Machine Learning* 1174–1182.

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

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

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

[19] HAHN, P. R., CARVALHO, C. M., PUELZ, D. and HE, J. (2018). Regularization and confounding in linear regression for treatment effect estimation. *Bayesian Anal.* **13** 163–182. MR3737947 https://doi.org/10.1214/16-BA1044

[20] HAHN, P. R., MURRAY, J. S. and CARVALHO, C. (2017). Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects. Available at arXiv:1706.09523.

[21] HECKMAN, J. J., LOPES, H. F. and PIATEK, R. (2014). Treatment effects: A Bayesian perspective. *Econometric Rev.* **33** 36–67. MR3170840 https://doi.org/10.1080/07474938.2013.807103

[22] HILL, J. L. (2011). Bayesian nonparametric modeling for causal inference. *J. Comput. Graph. Statist.* **20** 217–240. MR2816546 https://doi.org/10.1198/jcgs.2010.08162

[23] 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

[24] MURPHY, S. A. and VAN DER VAART, A. W. (2000). On profile likelihood. *J. Amer. Statist. Assoc.* **95** 449–485. MR1803168 https://doi.org/10.2307/2669386

[25] PRÆSTGAARD, J. and WELLNER, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.* **21** 2053–2086. MR1245301

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

[27] RAY, K. and SZABÓ, B. T. (2019). Debiased Bayesian inference for average treatment effects. In *Advances in Neural Information Processing Systems* 33 11952–11962.

[28] RAY, K. and VAN DER VAART, A. (2020). Supplement to "Semiparametric Bayesian causal inference." https://doi.org/10.1214/19-AOS1919SUPP

[29] RAY, K. and VAN DER VAART, A. W. (2020). On the Bernstein–von Mises theorem for the Dirichlet process. Available at arXiv:2008.01130.

[30] RITOV, Y., BICKEL, P. J., GAMST, A. C. and KLEIJN, B. J. K. (2014). The Bayesian analysis of complex, high-dimensional models: Can it be CODA? *Statist. Sci.* **29** 619–639. MR3300362 https://doi.org/10.1214/14-STS483

[31] RIVOIRARD, V. and ROUSSEAU, J. (2012). Bernstein–von Mises theorem for linear functionals of the density. *Ann. Statist.* **40** 1489–1523. MR3015033 https://doi.org/10.1214/12-AOS1004

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

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

[34] ROBINS, J., TCHETGEN TCHETGEN, E., LI, L. and VAN DER VAART, A. (2009). Semiparametric minimax rates. *Electron. J. Stat.* **3** 1305–1321. MR2566189 https://doi.org/10.1214/09-EJS479

[35] ROBINS, J. M., LI, L., MUKHERJEE, R., TCHETGEN, E. T. and VAN DER VAART, A. (2017). Minimax estimation of a functional on a structured high-dimensional model. *Ann. Statist.* **45** 1951–1987. MR3718158 https://doi.org/10.1214/16-AOS1515

[36] ROBINS, J. M. and RITOV, Y. (1997). Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models. *Stat. Med.* **16** 285–319.

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

[38] 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

[39] ROTNITZKY, A. and ROBINS, J. M. (1995). Semi-parametric estimation of models for means and covariances in the presence of missing data. *Scand. J. Stat.* **22** 323–333. MR1363216

[40] RUBIN, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Ann. Statist.* **6** 34–58. MR0472152

[41] SEAMAN, S. R. and VANSTEELANDT, S. (2018). Introduction to double robust methods for incomplete data. *Statist. Sci.* **33** 184–197. MR3797709 https://doi.org/10.1214/18-STS647

[42] TADDY, M., GARDNER, M., CHEN, L. and DRAPER, D. (2016). A nonparametric Bayesian analysis of heterogenous treatment effects in digital experimentation. *J. Bus. Econom. Statist.* **34** 661–672. MR3548002 https://doi.org/10.1080/07350015.2016.1172013

[43] VAN DER VAART, A. (1991). On differentiable functionals. *Ann. Statist.* **19** 178–204. MR1091845 https://doi.org/10.1214/aos/1176347976

[44] VAN DER VAART, A. (2014). Higher order tangent spaces and influence functions. *Statist. Sci.* **29** 679–686. MR3300365 https://doi.org/10.1214/14-STS478

[45] VAN DER VAART, A. and VAN ZANTEN, H. (2007). Bayesian inference with rescaled Gaussian process priors. *Electron. J. Stat.* **1** 433–448. MR2357712 https://doi.org/10.1214/07-EJS098

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

[47] VAN DER VAART, A. and WELLNER, J. A. (2000). Preservation theorems for Glivenko–Cantelli and uniform Glivenko–Cantelli classes. In *High Dimensional Probability, II* (*Seattle, WA*, 1999). *Progress in Probability* **47** 115–133. Birkhäuser, Boston, MA. MR1857319

[48] 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

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

[50] VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2008). Reproducing kernel Hilbert spaces of Gaussian priors. In *Pushing the Limits of Contemporary Statistics*: *Contributions in Honor of Jayanta K. Ghosh*. *Inst. Math. Stat.* (*IMS*) *Collect.* **3** 200–222. IMS, Beachwood, OH. MR2459226 https://doi.org/10.1214/074921708000000156

[51] VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2009). Adaptive Bayesian estimation using a Gaussian random field with inverse gamma bandwidth. *Ann. Statist.* **37** 2655–2675. MR2541442 https://doi.org/10.1214/08-AOS678

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

[53] 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

[54] ZIGLER, C. M. and DOMINICI, F. (2014). Uncertainty in propensity score estimation: Bayesian methods for variable selection and model-averaged causal effects. *J. Amer. Statist. Assoc.* **109** 95–107. MR3180549 https://doi.org/10.1080/01621459.2013.869498

*The Annals of Statistics*
2020, Vol. 48, No. 5, 3021–3042
https://doi.org/10.1214/19-AOS1920


# RELAXING THE ASSUMPTIONS OF KNOCKOFFS BY CONDITIONING

BY DONGMING HUANG[*] AND LUCAS JANSON[**]

*Department of Statistics, Harvard University, [*]dhuang01@g.harvard.edu; [**]ljanson@fas.harvard.edu*

The recent paper Candès et al. (*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** (2018) 551–577) introduced model-X knockoffs, a method for variable selection that provably and nonasymptotically controls the false discovery rate with no restrictions or assumptions on the dimensionality of the data or the conditional distribution of the response given the covariates. The one requirement for the procedure is that the covariate samples are drawn independently and identically from a precisely-known (but arbitrary) distribution. The present paper shows that the exact same guarantees can be made *without* knowing the covariate distribution fully, but instead knowing it only up to a parametric model with as many as $\Omega(n^* p)$ parameters, where $p$ is the dimension and $n^*$ is the number of covariate samples (which may exceed the usual sample size $n$ of labeled samples when unlabeled samples are also available). The key is to treat the covariates as if they are drawn conditionally on their observed value for a sufficient statistic of the model. Although this idea is simple, even in Gaussian models conditioning on a sufficient statistic leads to a distribution supported on a set of zero Lebesgue measure, requiring techniques from topological measure theory to establish valid algorithms. We demonstrate how to do this for three models of interest, with simulations showing the new approach remains powerful under the weaker assumptions.

## REFERENCES

[1] 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

[2] BARBER, R. F., CANDÈS, E. J. and SAMWORTH, R. J. (2020). Robust inference with knockoffs. *Ann. Statist.* **48** 1409–1431.

[3] BATES, S., CANDÈS, E., JANSON, L. and WANG, W. (2020). Metropolized knockoff sampling. *J. Amer. Statist. Assoc.* https://doi.org/10.1080/01621459.2020.1729163

[4] 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

[5] BELLONI, A., CHERNOZHUKOV, V. and KATO, K. (2015). Uniform post-selection inference for least absolute deviation regression and other Z-estimation problems. *Biometrika* **102** 77–94. MR3335097 https://doi.org/10.1093/biomet/asu056

[6] 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

[7] 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

[8] BERRETT, T. B., WANG, Y., BARBER, R. F. and SAMWORTH, R. J. (2018). The conditional permutation test. Preprint. Available at arXiv:1807.05405.

[9] 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

[10] 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

---

*MSC2020 subject classifications.* Primary 62G10; secondary 62B05, 62J02.

*Key words and phrases.* High-dimensional inference, knockoffs, model-X, sufficient statistic, false discovery rate (FDR), topological measure, graphical model.

[11] CHATTERJEE, A. and LAHIRI, S. N. (2013). Rates of convergence of the adaptive LASSO estimators to the oracle distribution and higher order refinements by the bootstrap. *Ann. Statist.* **41** 1232–1259. MR3113809 https://doi.org/10.1214/13-AOS1106

[12] CHERNOZHUKOV, V., HANSEN, C. and SPINDLER, M. (2015). Valid post-selection and post-regularization inference: An elementary, general approach. *Ann. Rev. Econ.* **7** 649–688. https://doi.org/10.1146/annurev-economics-012315-015826

[13] DIACONIS, P., HOLMES, S. and SHAHSHAHANI, M. (2013). Sampling from a manifold. In *Advances in Modern Statistical Theory and Applications*: *A Festschrift in Honor of Morris L. Eaton*. *Inst. Math. Stat.* (*IMS*) *Collect.* **10** 102–125. IMS, Beachwood, OH. MR3586941

[14] EDWARDS, D. (2000). *Introduction to Graphical Modelling*, 2nd ed. *Springer Texts in Statistics*. Springer, New York. MR1880319 https://doi.org/10.1007/978-1-4612-0493-0

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

[16] GAREY, M. R. and JOHNSON, D. S. (1979). *Computers and Intractability*: *A Guide to the Theory of NP-Completeness*. W. H. Freeman and Co., San Francisco, CA. MR0519066

[17] GIMENEZ, J. R., GHORBANI, A. and ZOU, J. (2019). Knockoffs for the mass: New feature importance statistics with false discovery guarantees. In *Proceedings of Machine Learning Research* (K. Chaudhuri and M. Sugiyama, eds.). *Proceedings of Machine Learning Research* **89** 2125–2133.

[18] HUANG, D. and JANSON, L. (2020). Supplement to "Relaxing the assumptions of knockoffs by conditioning." https://doi.org/10.1214/19-AOS1920SUPP.

[19] HUBER, P. J. (1972). The 1972 Wald lecture. Robust statistics: A review. *Ann. Math. Stat.* **43** 1041–1067. MR0314180 https://doi.org/10.1214/aoms/1177692459

[20] JANSON, L. (2017). A model-free approach to high-dimensional inference. Ph.D. thesis, Stanford Univ.

[21] 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

[22] JAVANMARD, A. and JAVADI, H. (2019). False discovery rate control via debiased lasso. *Electron. J. Stat.* **13** 1212–1253. MR3935848 https://doi.org/10.1214/19-ejs1554

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

[24] JORDON, J., YOON, J. and VAN DER SCHAAR, M. (2019). KnockoffGAN: Generating knockoffs for feature selection using generative adversarial networks. In *International Conference on Learning Representations*.

[25] LEE, J. D., SUN, D. L., SUN, Y. and TAYLOR, J. E. (2016). Exact post-selection inference, with application to the lasso. *Ann. Statist.* **44** 907–927. MR3485948 https://doi.org/10.1214/15-AOS1371

[26] LEWIS, R. M. R. (2016). *A Guide to Graph Colouring*: *Algorithms and Applications*. Springer, Cham. MR3408102 https://doi.org/10.1007/978-3-319-25730-3

[27] LIU, Y. and ZHENG, C. (2019). Deep latent variable models for generating knockoffs. *Stat* **8** e260. MR4072122 https://doi.org/10.1002/sta4.260

[28] MALAGUTI, E. and TOTH, P. (2010). A survey on vertex coloring problems. *Int. Trans. Oper. Res.* **17** 1–34. MR2598219 https://doi.org/10.1111/j.1475-3995.2009.00696.x

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

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

[31] ROMANO, Y., SESIA, M. and CANDÈS, E. J. (2019). Deep knockoffs. *J. Amer. Statist. Assoc.* https://doi.org/10.1080/01621459.2019.1660174

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

[33] 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

[34] 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

[35] 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

[36] 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

[37] ZHU, Y. and BRADIC, J. (2018). Significance testing in non-sparse high-dimensional linear models. *Electron. J. Stat.* **12** 3312–3364. MR3861831 https://doi.org/10.1214/18-EJS1443

*The Annals of Statistics*
2020, Vol. 48, No. 5, 3043–3065
https://doi.org/10.1214/19-AOS1921


# ANALYTICAL NONLINEAR SHRINKAGE OF LARGE-DIMENSIONAL COVARIANCE MATRICES

BY OLIVIER LEDOIT* AND MICHAEL WOLF**

*Department of Economics, University of Zurich,* *olivier.ledoit@econ.uzh.ch; **michael.wolf@econ.uzh.ch

This paper establishes the first analytical formula for nonlinear shrinkage estimation of large-dimensional covariance matrices. We achieve this by identifying and mathematically exploiting a deep connection between nonlinear shrinkage and nonparametric estimation of the Hilbert transform of the sample spectral density. Previous nonlinear shrinkage methods were of numerical nature: QuEST requires numerical inversion of a complex equation from random matrix theory whereas NERCOME is based on a sample-splitting scheme. The new analytical method is more elegant and also has more potential to accommodate future variations or extensions. Immediate benefits are (i) that it is typically 1000 times faster with basically the same accuracy as QuEST and (ii) that it accommodates covariance matrices of dimension up to 10,000 and more. The difficult case where the matrix dimension exceeds the sample size is also covered.

## REFERENCES

ABADIR, K. M., DISTASO, W. and ŽIKEŠ, F. (2014). Design-free estimation of variance matrices. *J. Econometrics* **181** 165–180. MR3209862 https://doi.org/10.1016/j.jeconom.2014.03.010

BAI, Z. D. and SILVERSTEIN, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.* **26** 316–345. MR1617051 https://doi.org/10.1214/aop/1022855421

CAPON, J. (1969). High-resolution frequency-wavenumber spectrum analysis. *Proc. IEEE* **57** 1408–1418.

DAWSON, H. G. (1898). On the Numerical Value of $\int_0^h e^{x^2}\,dx$. *Proc. Lond. Math. Soc.* **29** 519–522. MR1576451 https://doi.org/10.1112/plms/s1-29.1.519

ENGLE, R. and COLACITO, R. (2006). Testing and valuing dynamic correlations for asset allocation. *J. Bus. Econom. Statist.* **24** 238–253. MR2234449 https://doi.org/10.1198/073500106000000017

ENGLE, R. F., LEDOIT, O. and WOLF, M. (2019). Large dynamic covariance matrices. *J. Bus. Econom. Statist.* **37** 363–375. MR3948411 https://doi.org/10.1080/07350015.2017.1345683

EPANECHNIKOV, V. A. (1969). Nonparametric estimation of a multidimensional probability density. *Theory Probab. Appl.* **14** 153–158. MR0250422

ERDÉLYI, A., MAGNUS, W., OBERHETTINGER, F. and TRICOMI, F. G. (1954). *Tables of Integral Transforms. Vol. II*. McGraw-Hill Book Company, Inc., New York–Toronto–London. MR0065685

GIRSHICK, M. A. (1939). On the sampling theory of roots of determinantal equations. *Ann. Math. Stat.* **10** 203–224. MR0000127 https://doi.org/10.1214/aoms/1177732180

JING, B.-Y., PAN, G., SHAO, Q.-M. and ZHOU, W. (2010). Nonparametric estimate of spectral density functions of sample covariance matrices: A first step. *Ann. Statist.* **38** 3724–3750. MR2766866 https://doi.org/10.1214/10-AOS833

KRANTZ, S. G. (2009). *Explorations in Harmonic Analysis. Applied and Numerical Harmonic Analysis*. Birkhäuser, Inc., Boston, MA. MR2508404 https://doi.org/10.1007/978-0-8176-4669-1

LAM, C. (2016). Nonparametric eigenvalue-regularized precision or covariance matrix estimator. *Ann. Statist.* **44** 928–953. MR3485949 https://doi.org/10.1214/15-AOS1393

LEDOIT, O. and PÉCHÉ, S. (2011). Eigenvectors of some large sample covariance matrix ensembles. *Probab. Theory Related Fields* **151** 233–264. MR2834718 https://doi.org/10.1007/s00440-010-0298-3

LEDOIT, O. and WOLF, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivariate Anal.* **88** 365–411. MR2026339 https://doi.org/10.1016/S0047-259X(03)00096-4

---

*MSC2020 subject classifications.* Primary 62H12; secondary 62G20, 15A52.
*Key words and phrases.* Hilbert transform, kernel estimation, rotation equivariance.

LEDOIT, O. and WOLF, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Ann. Statist.* **40** 1024–1060. MR2985942 https://doi.org/10.1214/12-AOS989

LEDOIT, O. and WOLF, M. (2015). Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. *J. Multivariate Anal.* **139** 360–384. MR3349498 https://doi.org/10.1016/j.jmva.2015.04.006

LEDOIT, O. and WOLF, M. (2017a). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *Rev. Financ. Stud.* **30** 4349–4388.

LEDOIT, O. and WOLF, M. (2017b). Numerical implementation of the QuEST function. *Comput. Statist. Data Anal.* **115** 199–223. MR3683138 https://doi.org/10.1016/j.csda.2017.06.004

LEDOIT, O. and WOLF, M. (2018). Optimal estimation of a large-dimensional covariance matrix under Stein's loss. *Bernoulli* **24** 3791–3832. MR3788189 https://doi.org/10.3150/17-BEJ979

LEDOIT, O. and WOLF, M. (2020). Supplement to "Analytical nonlinear shrinkage of large-dimensional covariance matrices." https://doi.org/10.1214/19-AOS1921SUPP.

MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Sb. Math.* **1** 457–483.

MARKOWITZ, H. M. (1952). Portfolio selection. *J. Finance* **7** 77–91.

RIBES, A., AZAÏS, J.-M. and PLANTON, S. (2009). Adaptation of the optimal fingerprint method for climate change detection using a well-conditioned covariance matrix estimate. *Clim. Dyn.* **33** 707–722.

SILVERSTEIN, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.* **55** 331–339. MR1370408 https://doi.org/10.1006/jmva.1995.1083

SILVERSTEIN, J. W. and BAI, Z. D. (1995). On the empirical distribution of eigenvalues of a class of large-dimensional random matrices. *J. Multivariate Anal.* **54** 175–192. MR1345534 https://doi.org/10.1006/jmva.1995.1051

SILVERSTEIN, J. W. and CHOI, S.-I. (1995). Analysis of the limiting spectral distribution of large-dimensional random matrices. *J. Multivariate Anal.* **54** 295–309. MR1345541 https://doi.org/10.1006/jmva.1995.1058

STEIN, C. (1986). Lectures on the theory of estimation of many parameters. *J. Math. Sci.* **34** 1373–1403.

STIELTJES, T.-J. (1894). Recherches sur les fractions continues. *Ann. Fac. Sci. Univ. Toulouse Sci. Math. Sci. Phys.* **8** J1–J122. MR1508159

*The Annals of Statistics*
2020, Vol. 48, No. 5, 3066–3089
https://doi.org/10.1214/19-AOS1922


# COUPLED CONDITIONAL BACKWARD SAMPLING PARTICLE FILTER

BY ANTHONY LEE[1], SUMEETPAL S. SINGH[2] AND MATTI VIHOLA[3]

[1]*School of Mathematics, University of Bristol, anthony.lee@bristol.ac.uk*
[2]*Department of Engineering, University of Cambridge*
[3]*Department of Mathematics and Statistics, University of Jyväskylä*

The conditional particle filter (CPF) is a promising algorithm for general hidden Markov model smoothing. Empirical evidence suggests that the variant of CPF with backward sampling (CBPF) performs well even with long time series. Previous theoretical results have not been able to demonstrate the improvement brought by backward sampling, whereas we provide rates showing that CBPF can remain effective with a fixed number of particles independent of the time horizon. Our result is based on analysis of a new coupling of two CBPFs, the coupled conditional backward sampling particle filter (CCBPF). We show that CCBPF has good stability properties in the sense that with fixed number of particles, the coupling time in terms of iterations increases only linearly with respect to the time horizon under a general (strong mixing) condition. The CCBPF is useful not only as a theoretical tool, but also as a practical method that allows for unbiased estimation of smoothing expectations, following the recent developments by Jacob, Lindsten and Schön (2020). Unbiased estimation has many advantages, such as enabling the construction of asymptotically exact confidence intervals and straightforward parallelisation.

## REFERENCES

ANDRIEU, C., DOUCET, A. and HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 269–342. MR2758115 https://doi.org/10.1111/j.1467-9868.2009.00736.x

ANDRIEU, C., LEE, A. and VIHOLA, M. (2018). Uniform ergodicity of the iterated conditional SMC and geometric ergodicity of particle Gibbs samplers. *Bernoulli* **24** 842–872. MR3706778 https://doi.org/10.3150/15-BEJ785

CHOPIN, N. and SINGH, S. S. (2015). On particle Gibbs sampling. *Bernoulli* **21** 1855–1883. MR3352064 https://doi.org/10.3150/14-BEJ629

DEL MORAL, P. (2004). *Feynman–Kac Formulae*: *Genealogical and Interacting Particle Systems with Applications*. *Probability and Its Applications* (*New York*). Springer, New York. MR2044973 https://doi.org/10.1007/978-1-4684-9393-1

DEL MORAL, P. and GUIONNET, A. (2001). On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann. Inst. Henri Poincaré Probab. Stat.* **37** 155–194. MR1819122 https://doi.org/10.1016/S0246-0203(00)01064-5

DELYON, B., LAVIELLE, M. and MOULINES, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *Ann. Statist.* **27** 94–128. MR1701103 https://doi.org/10.1214/aos/1018031103

DOUC, R., GARIVIER, A., MOULINES, E. and OLSSON, J. (2011). Sequential Monte Carlo smoothing for general state space hidden Markov models. *Ann. Appl. Probab.* **21** 2109–2145. MR2895411 https://doi.org/10.1214/10-AAP735

FEARNHEAD, P. and KÜNSCH, H. R. (2018). Particle filters and data assimilation. *Annu. Rev. Stat. Appl.* **5** 421–452. MR3774754 https://doi.org/10.1146/annurev-statistics-031017-100232

GLYNN, P. W. and RHEE, C.-H. (2014). Exact estimation for Markov chain equilibrium expectations. *J. Appl. Probab.* **51A** 377–389. MR3317370 https://doi.org/10.1239/jap/1417528487

JACOB, P. E., LINDSTEN, F. and SCHÖN, T. B. Smoothing with couplings of conditional particle filters. *J. Amer. Statist. Assoc.* To appear.

JACOB, P. E., O'LEARY, J. and ATCHADÉ, Y. F. Unbiased Markov chain Monte Carlo with couplings. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* To appear.

*MSC2020 subject classifications.* Primary 65C40; secondary 65C05, 65C35, 65C60.
*Key words and phrases.* Backward sampling, convergence rate, coupling, conditional particle filter, unbiased.

LINDSTEN, F., DOUC, R. and MOULINES, E. (2015). Uniform ergodicity of the particle Gibbs sampler. *Scand. J. Stat.* **42** 775–797. MR3391692 https://doi.org/10.1111/sjos.12136

LINDSTEN, F., JORDAN, M. I. and SCHÖN, T. B. (2014). Particle Gibbs with ancestor sampling. *J. Mach. Learn. Res.* **15** 2145–2184. MR3231604

SHAKED, M. and SHANTHIKUMAR, J. G. (2007). *Stochastic Orders*. *Springer Series in Statistics*. Springer, New York. MR2265633 https://doi.org/10.1007/978-0-387-34675-5

SINGH, S. S., LINDSTEN, F. and MOULINES, E. (2017). Blocking strategies and stability of particle Gibbs samplers. *Biometrika* **104** 953–969. MR3737314 https://doi.org/10.1093/biomet/asx051

WHITELEY, N. (2010). Discussion on Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 306–307.

WHITELEY, N. (2013). Stability properties of some particle filters. *Ann. Appl. Probab.* **23** 2500–2537. MR3127943 https://doi.org/10.1214/12-AAP909

*The Annals of Statistics*
2020, Vol. 48, No. 5, 3090–3111
https://doi.org/10.1214/19-AOS1923


# ASYMPTOTIC RISK AND PHASE TRANSITION OF $l_1$-PENALIZED ROBUST ESTIMATOR

BY HANWEN HUANG

*Department of Epidemiology and Biostatistics, University of Georgia, huanghw@uga.edu*

Mean square error (MSE) of the estimator can be used to evaluate the performance of a regression model. In this paper, we derive the asymptotic MSE of $l_1$-penalized robust estimators in the limit of both sample size $n$ and dimension $p$ going to infinity with fixed ratio $n/p \to \delta$. We focus on the $l_1$-penalized least absolute deviation and $l_1$-penalized Huber's regressions. Our analytic study shows the appearance of a sharp phase transition in the two-dimensional sparsity-undersampling phase space. We derive the explicit formula of the phase boundary. Remarkably, the phase boundary is identical to the phase transition curve of LASSO which is also identical to the previously known Donoho–Tanner phase transition for sparse recovery. Our derivation is based on the asymptotic analysis of the generalized approximation passing (GAMP) algorithm. We establish the asymptotic MSE of the $l_1$-penalized robust estimator by connecting it to the asymptotic MSE of the corresponding GAMP estimator. Our results provide some theoretical insight into the high-dimensional regression methods. Extensive computational experiments have been conducted to validate the correctness of our analytic results. We obtain fairly good agreement between theoretical prediction and numerical simulations on finite-size systems.

## REFERENCES

[1] BAYATI, M., ERDOGDU, M. A. and MONTANARI, A. (2013). Estimating LASSO risk and noise level. In *Advances in Neural Information Processing Systems* 26 (C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani and K. Q. Weinberger, eds.) 944–952. Curran Associates.

[2] 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

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

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

[5] 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. https://doi.org/10.1073/pnas.1307845110

[6] BERTHIER, R., MONTANARI, A. and NGUYEN, P. (2017). State evolution for approximate message passing with non-separable functions. CoRR abs/1708.03950.

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

[8] BRADIC, J. (2016). Robustness in sparse high-dimensional linear models: Relative efficiency and robust approximate message passing. *Electron. J. Stat.* **10** 3894–3944. MR3581957 https://doi.org/10.1214/16-EJS1212

[9] 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

[10] CAI, T. T. and GUO, Z. (2018). Accuracy assessment for high-dimensional linear regression. *Ann. Statist.* **46** 1807–1836. MR3819118 https://doi.org/10.1214/17-AOS1604

*MSC2020 subject classifications.* Primary 62J05, 62J07; secondary 62H12.
*Key words and phrases.* Mean square error, minimax, penalized, phase transition, robust.

[11] CANDÈS, E. J., ROMBERG, J. K. and TAO, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.* **59** 1207–1223. MR2230846 https://doi.org/10.1002/cpa.20124

[12] 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

[13] DONOHO, D. and TANNER, J. (2009). Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **367** 4273–4293. MR2546388 https://doi.org/10.1098/rsta.2009.0152

[14] 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. Inform. Theory* **59** 3396–3433. MR3061255 https://doi.org/10.1109/TIT.2013.2239356

[15] DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455. MR1311089 https://doi.org/10.1093/biomet/81.3.425

[16] DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2009). Message-passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci. USA* **106** 18914–18919. https://doi.org/10.1073/pnas.0909892106

[17] DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2011). The noise-sensitivity phase transition in compressed sensing. *IEEE Trans. Inform. Theory* **57** 6920–6941. MR2882271 https://doi.org/10.1109/TIT.2011.2165823

[18] 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. https://doi.org/10.1073/pnas.1307842110

[19] HUANG, H. (2020). Supplement to "Asymptotic risk and phase transition of $l_1$-penalized robust estimator." https://doi.org/10.1214/19-AOS1923SUPP.

[20] JAVANMARD, A. and MONTANARI, A. (2014). Hypothesis testing in high-dimensional regression under the Gaussian random design model: Asymptotic theory. *IEEE Trans. Inform. Theory* **60** 6522–6554. MR3265038 https://doi.org/10.1109/TIT.2014.2343629

[21] KRZAKALA, F., MÉZARD, M., SAUSSET, F., SUN, Y. F. and ZDEBOROVÁ, L. (2012). Statistical-physics-based reconstruction in compressed sensing. *Phys. Rev. X* **2** 021005. https://doi.org/10.1103/PhysRevX.2.021005

[22] LAMBERT-LACROIX, S. and ZWALD, L. (2011). Robust regression through the Huber's criterion and adaptive lasso penalty. *Electron. J. Stat.* **5** 1015–1053. MR2836768 https://doi.org/10.1214/11-EJS635

[23] MOUSAVI, A., MALEKI, A. and BARANIUK, R. G. (2017). Consistent parameter estimation for LASSO and approximate message passing. *Ann. Statist.* **45** 2427–2454. MR3737897 https://doi.org/10.1214/16-AOS1529

[24] RANGAN, S. (2011). Generalized approximate message passing for estimation with random linear mixing. In *Information Theory Proceedings* (*ISIT*), 2011 *IEEE International Symposium on* 2168–2172. https://doi.org/10.1109/ISIT.2011.6033942

[25] RANGAN, S., SCHNITER, P. and FLETCHER, A. (2014). On the convergence of approximate message passing with arbitrary matrices. In 2014 *IEEE International Symposium on Information Theory* 236–240. https://doi.org/10.1109/ISIT.2014.6874830

[26] RANGAN, S., SCHNITER, P. and FLETCHER, A. K. (2017). Vector approximate message passing. In 2017 *IEEE International Symposium on Information Theory* (*ISIT*) 1588–1592.

[27] RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2011). Minimax rates of estimation for high-dimensional linear regression over $\ell_q$-balls. *IEEE Trans. Inform. Theory* **57** 6976–6994. MR2882274 https://doi.org/10.1109/TIT.2011.2165799

[28] SCHNITER, P., RANGAN, S. and FLETCHER, A. K. (2016). Vector approximate message passing for the generalized linear model. In 2016 50*th Asilomar Conference on Signals*, *Systems and Computers* 1525–1529. https://doi.org/10.1109/ACSSC.2016.7869633

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

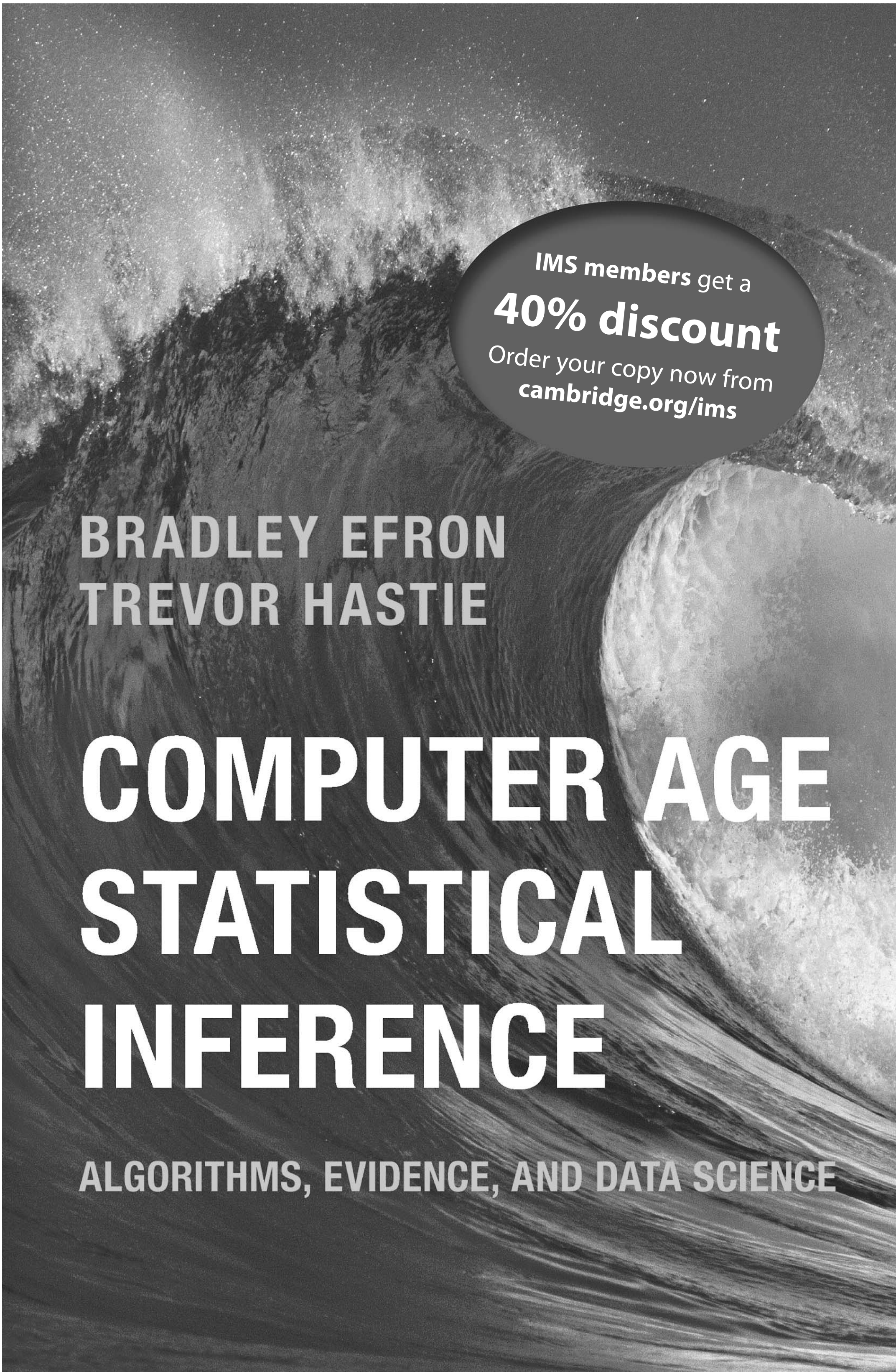
[30] VERZELEN, N. (2012). Minimax risks for sparse regressions: Ultra-high dimensional phenomenons. *Electron. J. Stat.* **6** 38–90. MR2879672 https://doi.org/10.1214/12-EJS666

[31] VILA, J., SCHNITER, P., RANGAN, S., KRZAKALA, F. and ZDEBOROVÁ, L. (2015). Adaptive damping and mean removal for the generalized approximate message passing algorithm. In 2015 *IEEE International Conference on Acoustics*, *Speech and Signal Processing* (*ICASSP*) 2021–2025. https://doi.org/10.1109/ICASSP.2015.7178325

[32] 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

[33] ZHANG, C.-H. (2012). Minimax $\ell_q$ risk in $\ell_p$ balls. In *Contemporary Developments in Bayesian Analysis and Statistical Decision Theory*: *A Festschrift for William E. Strawderman*. *Inst*. *Math*. *Stat*. (*IMS*) *Collect*. **8** 78–89. IMS, Beachwood, OH. MR3202504 https://doi.org/10.1214/11-IMSCOLL806

[34] ZHENG, L., MALEKI, A., WENG, H., WANG, X. and LONG, T. (2017). Does $\ell_p$-minimization outperform $\ell_1$-minimization? *IEEE Trans*. *Inform*. *Theory* **63** 6896–6935. MR3724407 https://doi.org/10.1109/TIT.2017.2717585


IMS members get a
40% discount
Order your copy now from
cambridge.org/ims
BRADLEY EFRON
TREVOR HASTIE
COMPUTER AGE
STATISTICAL
INFERENCE
ALGORITHMS, EVIDENCE, AND DATA SCIENCE