BERNOULLI

Official Journal of the Bernoulli Society for Mathematical Statistics and Probability

Volume Twenty Eight Number Three August 2022 ISSN: 1350-7265

CONTENTS

(continued)

A list of forthcoming papers can be found online at http://www.bernoullisociety.org/index.php/publications/bernoulli-journal/bernoulli-journal-papers

BERNOULLI

Official Journal of the Bernoulli Society for Mathematical Statistics and Probability

Volume Twenty Eight Number Three August 2022 ISSN: 1350-7265

CONTENTS

(continued)

BERNOULLI Volume 28 Number 3 August 2022 Pages 1519–2121 ISI/BS

Volume 28 Number 3   August 2022   ISSN 1350-7265

BERNOULLI

Official Journal of the Bernoulli Society for Mathematical Statistics and Probability

Aims and Scope

Issued four times per year, BERNOULLI is the flagship journal of the Bernoulli Society for Mathematical Statistics and Probability. The journal aims at publishing original research contributions of the highest quality in all subfields of Mathematical Statistics and Probability. The main emphasis of Bernoulli is on theoretical work, yet discussion of interesting applications in relation to the proposed methodology is also welcome.

Bernoulli Society for Mathematical Statistics and Probability

The Bernoulli Society was founded in 1973. It is an autonomous Association of the International Statistical Institute, ISI. According to its statutes, the object of the Bernoulli Society is the advancement, through international contacts, of the sciences of probability (including the theory of stochastic processes) and mathematical statistics and of their applications to all those aspects of human endeavour which are directed towards the increase of natural knowledge and the welfare of mankind.

Meetings: http://www.bernoulli-society.org/index.php/meetings

The Society holds a World Congress every four years; more frequent meetings, coordinated by the Society's standing committees and often organised in collaboration with other organisations, are the European Meeting of Statisticians, the Conference on Stochastic Processes and their Applications, the CLAPEM meeting (Latin-American Congress on Probability and Mathematical Statistics), the European Young Statisticians Meeting, and various meetings on special topics – in the physical sciences in particular. The Society, as an association of the ISI, also collaborates with other ISI associations in the organization of the biennial ISI World Statistics Congresses (formerly ISI Sessions).

Executive Committee

Detailed information about the members of the Executive Committee can be found on http://www.bernoulli-society.org/who-is-who

---

The papers published in Bernoulli are indexed or abstracted in Mathematical Reviews (*MathSciNet*), Zentralblatt MATH (*zbMATH Open*), Science Citation Index Expanded (*Web of Science*), *SCOPUS* and *Google Scholar*.

In 2022 Bernoulli consists of 4 issues published in February, May, August and November.

Bernoulli Society
for Mathematical Statistics
and Probability

Bernoulli **28**(3), 2022, 1519–1547
https://doi.org/10.3150/20-BEJ1315

Quadratic shrinkage for large covariance matrices

OLIVIER LEDOIT[1,2,a,c] and MICHAEL WOLF[1,b]

[1]*Department of Economics, University of Zurich, 8032 Zurich, Switzerland.* [a]*olivier.ledoit@econ.uzh.ch,* [b]*michael.wolf@econ.uzh.ch*
[2]*AlphaCrest Capital Management, New York, NY 10036, USA.* [c]*olivier.ledoit@alphacrestcapital.com*

This paper constructs a new estimator for large covariance matrices by drawing a bridge between the classic (Stein (1975)) estimator in finite samples and recent progress under large-dimensional asymptotics. The estimator keeps the eigenvectors of the sample covariance matrix and applies shrinkage to the *inverse* sample eigenvalues. The corresponding formula is *quadratic*: it has two shrinkage targets weighted by quadratic functions of the concentration (that is, matrix dimension divided by sample size). The first target dominates mid-level concentrations and the second one higher levels. This extra degree of freedom enables us to outperform linear shrinkage when the optimal shrinkage is not linear, which is the general case. Both of our targets are based on what we term the "Stein shrinker", a local attraction operator that pulls sample covariance matrix eigenvalues towards their nearest neighbors, but whose force diminishes with distance (like gravitation). We prove that no cubic or higher-order nonlinearities beat quadratic with respect to Frobenius loss under large-dimensional asymptotics. Non-normality and the case where the matrix dimension exceeds the sample size are accommodated. Monte Carlo simulations confirm state-of-the-art performance in terms of accuracy, speed, and scalability.

Keywords: Inverse shrinkage; kernel estimation; large-dimensional asymptotics; signal amplitude; Stein shrinkage

References

Abrahamsson, R., Selen, Y. and Stoica, P. (2007). Enhanced covariance matrix estimators in adaptive beamforming. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing – ICASSP 2007* **II** 969–972.

Arsigny, V., Fillard, P., Pennec, X. and Ayache, N. (2006). Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magn. Reson. Med.* **56** 411–421.

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

Bickel, P.J. and Levina, E. (2008a). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. MR2485008 https://doi.org/10.1214/08-AOS600

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

Bodnar, T., Gupta, A.K. and Parolya, N. (2016). Direct shrinkage estimation of large dimensional precision matrix. *J. Multivariate Anal.* **146** 223–236. MR3477661 https://doi.org/10.1016/j.jmva.2015.09.010

Capon, J. (1969). High-resolution frequency-wavenumber spectrum analysis. *Proc. IEEE* **57** 1408–1418.

Chen, Y., Wiesel, A. and Hero, A. O. (2009). Shrinkage estimation of high dimensional covariance matrices. IEEE International Conference on Acoustics, Speech, and Signal Processing, Taiwan.

Chen, Y., Wiesel, A. and Hero, A.O. III (2011). Robust shrinkage estimation of high-dimensional covariance matrices. *IEEE Trans. Signal Process.* **59** 4097–4107. MR2865971 https://doi.org/10.1109/TSP.2011.2138698

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

Daniels, M.J. and Kass, R.E. (2001). Shrinkage estimators for covariance matrices. *Biometrics* **57** 1173–1184. MR1950425 https://doi.org/10.1111/j.0006-341X.2001.01173.x

Deligianni, F., Centeno, M., Carmichael, D.W. and Clayden, J.D. (2014). Relating resting-state fMRI and EEG whole-brain connectomes across frequency bands. *Front. Neurosci.* **8** 258. https://doi.org/10.3389/fnins.2014.00258

Dey, D.K. and Srinivasan, C. (1985). Estimation of a covariance matrix under Stein's loss. *Ann. Statist.* **13** 1581–1591. MR0811511 https://doi.org/10.1214/aos/1176349756

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

Efron, B. and Morris, C. (1976). Multivariate empirical Bayes and estimation of covariance matrices. *Ann. Statist.* **4** 22–32. MR0394960

El Karoui, N. (2008). Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Ann. Statist.* **36** 2717–2756. MR2485011 https://doi.org/10.1214/07-AOS559

Elsheikh, A.H., Wheeler, M.F. and Hoteit, I. (2013). An iterative stochastic ensemble method for parameter estimation of subsurface flow models. *J. Comput. Phys.* **242** 696–714. MR3062055 https://doi.org/10.1016/j.jcp.2013.01.047

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

Förstner, W. and Moonen, B. (1999). A metric for covariance matrices. In *Quo Vadis Geodesia . . . ? Festschrift for Erik W. Grafarend on the Occasion of His 60th Birthday* (F. Krumm and V.S. Schwarze, eds.). *Technical Reports of the Department of Geodesy and Geoinformatics* **1999.6** 113–128. Univ. Stuttgart, Institute of Geodesy.

Gabor, D. (1946). Theory of communication. Part 1: The analysis of information. *J. Inst. Electr. Eng.* **93** 429–441.

Guo, S.M.A., He, J.A., Monnier, N.A., Sun, G.B., Wohland, T.B. and Bathe, M.A. (2012). Bayesian approach to the analysis of fluorescence correlation spectroscopy data II: Application to simulated and in vitro data. *Anal. Chem.* **84** 3880–3888.

Haff, L.R. (1979). Estimation of the inverse covariance matrix: Random mixtures of the inverse Wishart matrix and the identity. *Ann. Statist.* **7** 1264–1276. MR0550149

Haff, L.R. (1980). Empirical Bayes estimation of the multivariate normal covariance matrix. *Ann. Statist.* **8** 586–597. MR0568722

Haff, L.R. (1991). The variational form of certain Bayes estimators. *Ann. Statist.* **19** 1163–1190. MR1126320 https://doi.org/10.1214/aos/1176348244

Hasselmann, K. (1993). Optimal fingerprints for the detection of time-dependent climate change. *J. Climate* **6** 1957–1971.

Stein, C. (1977). Lectures on the theory of estimation of many parameters (in Russian). In *Studies in the Statistical Theory of Estimation, Part I* (I.A. Ibragimov and M.S. Nikulin, eds.). *Proceedings of Scientific Seminars of the Steklov Institute, Leningrad Division* **74** 4–65.

IPCC (2007). Climate change 2007: The physical science basis. In *Working Group I Contribution to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (S. Solomon, D. Qin, M. Manning, M. Marquis, K. Averyt, M.M.B. Tignor, H.L. Miller and Z. Chen, eds.) **4**. Cambridge and New York: Cambridge Univ. Press. p. 996.

James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I* 361–379. Berkeley, Calif.: Univ. California Press. MR0133191

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

John, S. (1971). Some optimal multivariate tests. *Biometrika* **58** 123–127. MR0275568 https://doi.org/10.1093/biomet/58.1.123

Korniotis, G.M. (2008). Habit formation, incomplete markets, and the significance of regional risk for expected returns. *Rev. Financ. Stud.* **21** 2139–2172.

Krishnamoorthy, K. and Gupta, A.K. (1989). Improved minimax estimation of a normal precision matrix. *Canad. J. Statist.* **17** 91–102. MR1014094 https://doi.org/10.2307/3314766

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. (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

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

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. (2017). 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). Analytical nonlinear shrinkage of large-dimensional covariance matrices. *Ann. Statist.* **48** 3043–3065. MR4152634 https://doi.org/10.1214/19-AOS1921

Ledoit, O. and Wolf, M.(2022). Supplement to "Quadratic shrinkage for large covariance matrices." https://doi.org/10.3150/20-BEJ1315SUPP

Lin, S.P. and Perlman, M.D. (1985). A Monte Carlo comparison of four estimators of a covariance matrix. In *Multivariate Analysis VI (Pittsburgh, Pa., 1983)* 411–429. Amsterdam: North-Holland. MR0822310

Loh, W.-L. (1991). Estimating covariance matrices. *Ann. Statist.* **19** 283–296. MR1091851 https://doi.org/10.1214/aos/1176347982

Marčenko, V.A. and Pastur, L.A. (1967). Distribution of eigenvalues for some sets of random matrices. *Sb. Math.* **1** 457–483.

Markon, K.E. (2010). Modeling psychopathology structure: A symptom-level analysis of axis I and II disorders. *Psychol. Med.* **40** 273–288.

Markowitz, H. (1952). Portfolio Selection. *Journal of Finance* **7** 77–91.

Moakher, M. and Batchelor, P.G. (2006). Symmetric positive-definite matrices: From geometry to applications and visualization. In *Visualization and Processing of Tensor Fields*. *Math. Vis.* 285–298, 452. Berlin: Springer. MR2210524 https://doi.org/10.1007/3-540-31272-2_17

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

Pal, N. (1993). Estimating the normal dispersion matrix and the precision matrix from a decision-theoretic point of view: A review. *Statist. Papers* **34** 1–26. MR1221520 https://doi.org/10.1007/BF02925524

Pyeon, D., Newton, M.A., Lambert, P.F., Den Boon, J.A., Sengupta, S., Marsit, C.J., Woodworth, C.D., Connor, J.P., Haugen, T.H., Smith, E.M., Kelsey, K.T., Turek, L.P. and Ahlquist, P. (2007). Fundamental differences in cell cycle deregulation in human papillomavirus-positive and human papillomavirus-negative head/neck and cervical cancers. *Cancer Res.* **67** 4605–4619.

Rajaratnam, B. and Vincenzi, D. (2016a). A theoretical study of Stein's covariance estimator. *Biometrika* **103** 653–666. MR3551790 https://doi.org/10.1093/biomet/asw030

Rajaratnam, B. and Vincenzi, D. (2016b). A note on covariance estimation in the unbiased estimator of risk framework. *J. Statist. Plann. Inference* **175** 25–39. MR3492762 https://doi.org/10.1016/j.jspi.2016.02.004

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.

Schäfer, J. and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat. Appl. Genet. Mol. Biol.* **4** Art. 32, 28. MR2183942 https://doi.org/10.2202/1544-6115.1175

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

Stein, C. (1975). Estimation of a covariance matrix. Rietz lecture, 39th Annual Meeting IMS. Atlanta, Georgia.

Stein, C. (1986). Lectures on the theory of estimation of many parameters. *J. Math. Sci.* **34** 1373–1403.

Stoica, P., Li, J., Zhu, X. and Guerci, J.R. (2008). On using a priori knowledge in space-time adaptive processing. *IEEE Trans. Signal Process.* **56** 2598–2602. MR2516660 https://doi.org/10.1109/TSP.2007.914347

Tenenhaus, A. and Tenenhaus, M. (2011). Regularized generalized canonical correlation analysis. *Psychometrika* **76** 257–284. MR2788885 https://doi.org/10.1007/s11336-011-9206-8

Tsukuma, H. (2005). Estimating the inverse matrix of scale parameters in an elliptically contoured distribution. *J. Japan Statist. Soc.* **35** 21–39. MR2183498 https://doi.org/10.14490/jjss.35.21

Vakman, D. (1996). On the analytic signal, the Teager-Kaiser energy algorithm, and other methods for defining amplitude and frequency. *IEEE Transactions on Signal Processing* **44** 791–797.

Vidaurre, C., Krämer, N., Blankertz, B. and Schlögl, A. (2009). Time domain parameters as a feature for EEG-based brain–computer interfaces. *Neural Netw.* **22** 1313–1319.

Wigner, E.P. (1955). Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math. (2)* **62** 548–564. MR0077805 https://doi.org/10.2307/1970079

Won, J.-H., Lim, J., Kim, S.-J. and Rajaratnam, B. (2013). Condition-number-regularized covariance estimation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 427–450. MR3065474 https://doi.org/10.1111/j.1467-9868.2012.01049.x

Yang, R. and Berger, J.O. (1994). Estimation of a covariance matrix using the reference prior. *Ann. Statist.* **22** 1195–1211. MR1311972 https://doi.org/10.1214/aos/1176325625

Zhang, Y., Sun, D. and Zhang, D. (2009). Robust adaptive acoustic vector sensor beamforming using automated diagonal loading. *Appl. Acoust.* **70** 1029–1033.

Bernoulli **28**(3), 2022, 1548–1576
https://doi.org/10.3150/21-BEJ1336

Berry–Esseen bounds for multivariate nonlinear statistics with applications to M-estimators and stochastic gradient descent algorithms

QI-MAN SHAO[1,2,a] and ZHUO-SONG ZHANG[2,3,b]

[1]*Department of Statistics and Data Scinece, Southern University of Science and Technology, Shenzhen, Guangdong, P.R. China.* [a]*shaoqm@sustech.edu.cn*
[2]*Department of Statistics, The Chinese University of Hong Kong, Shatin, N.T. Hong Kong*
[3]*Department of Statistics and Applied Probability, National University of Singapore, Singapore 117546.* [b]*zszhang.stat@gmail.com*

We establish a Berry–Esseen bound for general multivariate nonlinear statistics by developing a new multivariate-type randomized concentration inequality. The bound is the best possible for many known statistics. As applications, Berry–Esseen bounds for M-estimators and averaged stochastic gradient descent algorithms are obtained.

Keywords: Berry–Esseen bound; multivariate normal approximation; randomized concentration inequality; Stein's method; M-estimators; averaged stochastic gradient descent algorithms

References

[1] Anastasiou, A., Balasubramanian, K. and Erdogdu, M.A. (2019). Normal approximation for stochastic gradient descent via non-asymptotic rates of martingale CLT. In *Proceedings of the Thirty-Second Conference on Learning Theory*. *PMLR* **99** 115–137.
[2] Bach, F. and Moulines, E. (2011). Non-asymptotic analysis of stochastic approximation algorithms for machine learning. In *Neural Information Processing Systems (NIPS)*. Spain.
[3] Barbour, A.D. (1990). Stein's method for diffusion approximations. *Probab. Theory Related Fields* **84** 297–322. MR1035659 https://doi.org/10.1007/BF01197887
[4] 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
[5] Bentkus, V. (2004). A Lyapunov type bound in $\mathbf{R}^d$. *Teor. Veroyatn. Primen.* **49** 400–410. MR2144310 https://doi.org/10.1137/S0040585X97981123
[6] Bentkus, V., Bloznelis, M. and Götze, F. (1997). A Berry-Esséen bound for M-estimators. *Scand. J. Stat.* **24** 485–502. MR1615335 https://doi.org/10.1111/1467-9469.00076
[7] Bhattacharya, R. and Holmes, S. (2010). An exposition of Götze's estimation of the rate of convergence in the multivariate central limit theorem. *Technical report* Stanford University.
[8] Chatterjee, S. and Meckes, E. (2008). Multivariate normal approximation using exchangeable pairs. *ALEA Lat. Am. J. Probab. Math. Stat.* **4** 257–283. MR2453473
[9] Chen, L.H.Y., Goldstein, L. and Shao, Q.-M. (2011). *Normal Approximation by Stein's Method*. *Probability and Its Applications (New York)*. Heidelberg: Springer. MR2732624 https://doi.org/10.1007/978-3-642-15007-4
[10] Chen, L.H.Y. and Shao, Q.-M. (2007). Normal approximation for nonlinear statistics using a concentration inequality approach. *Bernoulli* **13** 581–599. MR2331265 https://doi.org/10.3150/07-BEJ5164
[11] Chen, L.H.Y. and Fang, X. (2015). Multivariate normal approximation by Stein's method: The concentration inequality approach. Available at arXiv:1111.4073.
[12] Goldstein, L. and Rinott, Y. (1996). Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.* **33** 1–17. MR1371949 https://doi.org/10.1017/s0021900200103675

[13] Götze, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19** 724–739. MR1106283
[14] 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. Berkeley, Calif.: Univ. California Press. MR0216620
[15] Nagaev, S.V. (1976). An estimate of the remainder term in the multidimensional central limit theorem. In *Proceedings of the Third Japan-USSR Symposium on Probability Theory (Tashkent, 1975)* 419–438. *Lecture Notes in Math.*, Vol. **550**. MR0443043
[16] Nemirovski, A., Juditsky, A., Lan, G. and Shapiro, A. (2008). Robust stochastic approximation approach to stochastic programming. *SIAM J. Optim.* **19** 1574–1609. MR2486041 https://doi.org/10.1137/070704277
[17] Paulauskas, V. (1996). Rates of convergence in the asymptotic normality for some local maximum estimators. *Liet. Mat. Rink.* **36** 85–114. MR1429819 https://doi.org/10.1007/BF02473976
[18] Pfanzagl, J. (1971). The Berry-Esseen bound for minimum contrast estimates. *Metrika* **17** 82–91. MR0295467 https://doi.org/10.1007/BF02613813
[19] Pfanzagl, J. (1972). Further results on asymptotic normality. I. *Metrika* **18** 174–198.
[20] Pfanzagl, J. (1973). The accuracy of the normal approximation for estimates of vector parameters. *Z. Wahrsch. Verw. Gebiete* **25** 171–198. MR0329093 https://doi.org/10.1007/BF00535890
[21] Pollard, D. (1985). New ways to prove central limit theorems. *Econom. Theory* **1** 295–313.
[22] Polyak, B.T. (1990). A new method of stochastic approximation type. *Avtomat. i Telemekh.* **7** 98–107. MR1071220
[23] Polyak, B.T. and Juditsky, A.B. (1992). Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.* **30** 838–855. MR1167814 https://doi.org/10.1137/0330046
[24] Raič, M. (2019). A multivariate Berry-Esseen theorem with explicit constants. *Bernoulli* **25** 2824–2853. MR4003566 https://doi.org/10.3150/18-BEJ1072
[25] Reinert, G. and Röllin, A. (2009). Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *Ann. Probab.* **37** 2150–2173. MR2573554 https://doi.org/10.1214/09-AOP467
[26] Ruppert, D. (1988). Efficient estimations from a slowly convergent Robbins–Monro process Technical Report, Cornell Univ. Operations Research and Industrial Engineering.
[27] Senatov, V.V. (1980). Some uniform estimates of the convergence rate in the multidimensional central limit theorem. *Teor. Veroyatn. Primen.* **25** 757–770. MR0595137
[28] Shao, Q.-M. and Zhang, Z.-S. (2022). Supplement to "Berry–Esseen bounds for multivariate nonlinear statistics with applications to M-estimators and stochastic gradient descent algorithms." https://doi.org/10.3150/21-BEJ1336SUPP
[29] Shao, Q.-M. and Zhou, W.-X. (2016). Cramér type moderate deviation theorems for self-normalized processes. *Bernoulli* **22** 2029–2079. MR3498022 https://doi.org/10.3150/15-BEJ719
[30] van der Vaart, A.W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. MR1652247 https://doi.org/10.1017/CBO9780511802256
[31] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. *Springer Series in Statistics*. New York: Springer. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2
[32] Vershynin, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge: Cambridge Univ. Press. MR2963170

Bernoulli **28**(3), 2022, 1577–1601
https://doi.org/10.3150/21-BEJ1343

Improved bounds for discretization of Langevin diffusions: Near-optimal rates without convexity

WENLONG MOU[1,a], NICOLAS FLAMMARION[3,d], MARTIN J. WAINWRIGHT[1,2,b] and PETER L. BARTLETT[1,2,c]

[1]*Department of EECS, UC Berkeley, Cory Hall, Berkeley, CA, 94720, USA.* [a]*wmou@berkeley.edu,* [b]*wainwrig@berkeley.edu,* [c]*peter@berkeley.edu*
[2]*Department of Statistics, UC Berkeley, Evans Hall, Berkeley, CA, 94720, USA*
[3]*School of Computer and Communication Sciences, EPFL, INJ 336, Station 14, CH-1015 Lausanne, Switzerland.* [d]*nicolas.flammarion@epfl.ch*

Discretizations of the Langevin diffusion have been proven very useful for developing and analyzing algorithms for sampling and stochastic optimization. We present an improved non-asymptotic analysis of the Euler-Maruyama discretization of the Langevin diffusion. Our analysis does not require global contractivity, and yields polynomial dependence on the time horizon. Compared to existing approaches, we make an additional smoothness assumption, and improve the existing rate in discretization step size from $O(\eta)$ to $O(\eta^2)$ in terms of the KL divergence. This result matches the correct order for numerical SDEs, without suffering from exponential time dependence. When applied to MCMC, this result simultaneously improves on the analyses of a range of sampling algorithms that are based on Dalalyan's approach.

Keywords: Langevin diffusion; Markov chain Monte Carlo; Euler-Maruyama discretization; KL divergence; non-asymptotic bound

References

[1] Alfonsi, A., Jourdain, B. and Kohatsu-Higa, A. (2015). Optimal transport bounds between the time-marginals of a multidimensional diffusion and its Euler scheme. *Electron. J. Probab.* **20** no. 70, 31. MR3361258 https://doi.org/10.1214/EJP.v20-4195

[2] Bakry, D., Barthe, F., Cattiaux, P. and Guillin, A. (2008). A simple proof of the Poincaré inequality for a large class of probability measures including the log-concave case. *Electron. Commun. Probab.* **13** 60–66. MR2386063 https://doi.org/10.1214/ECP.v13-1352

[3] Bakry, D. and Émery, M. (1985). Diffusions hypercontractives. In *Séminaire de Probabilités, XIX, 1983/84*. *Lecture Notes in Math.* **1123** 177–206. Berlin: Springer. MR0889476 https://doi.org/10.1007/BFb0075847

[4] Bernton, E. (2018) Langevin Monte Carlo and JKO splitting. In *Proceedings of the 31st Conference on Learning Theory*. *Proceedings of Machine Learning Research* **75** 1777–1798. PMLR.

[5] Bolley, F. and Villani, C. (2005). Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities. *Ann. Fac. Sci. Toulouse Math. (6)* **14** 331–352. MR2172583

[6] Bou-Rabee, N., Eberle, A. and Zimmer, R. (2020). Coupling and convergence for Hamiltonian Monte Carlo. *Ann. Appl. Probab.* **30** 1209–1250. MR4133372 https://doi.org/10.1214/19-AAP1528

[7] Bou-Rabee, N. and Hairer, M. (2013). Nonasymptotic mixing of the MALA algorithm. *IMA J. Numer. Anal.* **33** 80–110. MR3020951 https://doi.org/10.1093/imanum/drs003

[8] Bou-Rabee, N. and Vanden-Eijnden, E. (2010). Pathwise accuracy and ergodicity of metropolized integrators for SDEs. *Comm. Pure Appl. Math.* **63** 655–696. MR2583309 https://doi.org/10.1002/cpa.20306

[9] Bovier, A., Eckhoff, M., Gayrard, V. and Klein, M. (2004). Metastability in reversible diffusion processes. I. Sharp asymptotics for capacities and exit times. *J. Eur. Math. Soc. (JEMS)* **6** 399–424. MR2094397 https://doi.org/10.4171/JEMS/14

1350-7265

[10] Burkholder, D.L., Davis, B.J. and Gundy, R.F. (1972). Integral inequalities for convex functions of operators on martingales. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability Theory* 223–240. MR0400380

[11] Chen, Y., Chen, J., Dong, J., Peng, J. and Wang, Z. (2019). Accelerating nonconvex learning via replica exchange Langevin diffusion. In *International Conference on Learning Representations*.

[12] Cheng, X. and Bartlett, P. (2018). Convergence of Langevin MCMC in KL-divergence. In *Algorithmic Learning Theory 2018*. *Proc. Mach. Learn. Res. (PMLR)* **83** 26. PMLR. MR3857306

[13] Cheng, X., Chatterji, N.S., Abbasi-Yadkori, Y., Bartlett, P.L. and Jordan, M.I. (2018). Sharp convergence rates for Langevin dynamics in the nonconvex setting. Preprint, arXiv:1805.01648.

[14] Cheng, X., Chatterji, N.S., Bartlett, P.L. and Jordan, M.I. (2018). Underdamped Langevin MCMC: A non-asymptotic analysis. In *Proceedings of the 31st Conference on Learning Theory*. *Proceedings of Machine Learning Research* **75** 300–323. PMLR.

[15] Dalalyan, A.S. (2017). Theoretical guarantees for approximate sampling from smooth and log-concave densities. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 651–676. MR3641401 https://doi.org/10.1111/rssb.12183

[16] Dalalyan, A.S. and Karagulyan, A. (2019). User-friendly guarantees for the Langevin Monte Carlo with inaccurate gradient. *Stochastic Process. Appl.* **129** 5278–5311. MR4025705 https://doi.org/10.1016/j.spa.2019.02.016

[17] Dalalyan, A.S. and Riou-Durand, L. (2020). On sampling from a log-concave density using kinetic Langevin diffusions. *Bernoulli* **26** 1956–1988. MR4091098 https://doi.org/10.3150/19-BEJ1178

[18] Dalalyan, A.S., Riou-Durand, L. and Karagulyan, A. (2019). Bounding the error of discretized Langevin algorithms for non-strongly log-concave targets. Preprint, arXiv:1906.08530.

[19] De Bortoli, V. and Durmus, A. (2019). Convergence of diffusions and their discretizations: From continuous to discrete processes and back. Preprint, arXiv:1904.09808.

[20] Durmus, A. and Moulines, É. (2017). Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *Ann. Appl. Probab.* **27** 1551–1587. MR3678479 https://doi.org/10.1214/16-AAP1238

[21] Durmus, A. and Moulines, É. (2019). High-dimensional Bayesian inference via the unadjusted Langevin algorithm. *Bernoulli* **25** 2854–2882. MR4003567 https://doi.org/10.3150/18-BEJ1073

[22] Dwivedi, R., Chen, Y., Wainwright, M.J. and Yu, B. (2019). Log-concave sampling: Metropolis-Hastings algorithms are fast. *J. Mach. Learn. Res.* **20** Paper No. 183, 42. MR4048994

[23] Eberle, A. (2016). Reflection couplings and contraction rates for diffusions. *Probab. Theory Related Fields* **166** 851–886. MR3568041 https://doi.org/10.1007/s00440-015-0673-1

[24] Eberle, A. and Majka, M.B. (2019). Quantitative contraction rates for Markov chains on general state spaces. *Electron. J. Probab.* **24** Paper No. 26, 36. MR3933205 https://doi.org/10.1214/19-EJP287

[25] Erdogdu, M.A., Mackey, L. and Shamir, O. (2018). Global non-convex optimization with discretized diffusions. In *Advances in Neural Information Processing Systems* 9694–9703.

[26] Fang, W. and Giles, M.B. (2019). Multilevel Monte Carlo method for ergodic SDEs without contractivity. *J. Math. Anal. Appl.* **476** 149–176. MR3944422 https://doi.org/10.1016/j.jmaa.2018.12.032

[27] Gorham, J., Duncan, A.B., Vollmer, S.J. and Mackey, L. (2019). Measuring sample quality with diffusions. *Ann. Appl. Probab.* **29** 2884–2928. MR4019878 https://doi.org/10.1214/19-AAP1467

[28] Gross, L. (1975). Logarithmic Sobolev inequalities. *Amer. J. Math.* **97** 1061–1083. MR0420249 https://doi.org/10.2307/2373688

[29] Higham, D.J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Rev.* **43** 525–546. MR1872387 https://doi.org/10.1137/S0036144500378302

[30] Holley, R. and Stroock, D. (1987). Logarithmic Sobolev inequalities and stochastic Ising models. *J. Stat. Phys.* **46** 1159–1194. MR0893137 https://doi.org/10.1007/BF01011161

[31] Iserles, A. (2009). *A First Course in the Numerical Analysis of Differential Equations*, 2nd ed. *Cambridge Texts in Applied Mathematics*. Cambridge: Cambridge Univ. Press. MR2478556

[32] Khasminskii, R. (2012). *Stochastic Stability of Differential Equations*, 2nd ed. *Stochastic Modelling and Applied Probability* **66**. Heidelberg: Springer. MR2894052 https://doi.org/10.1007/978-3-642-23280-0

[33] Kloeden, P.E. and Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*. *Applications of Mathematics (New York)* **23**. Berlin: Springer. MR1214374 https://doi.org/10.1007/978-3-662-12616-5

[34] Lee, H., Risteski, A. and Ge, R. (2018). Beyond log-concavity: Provable guarantees for sampling multi-modal distributions using simulated tempering Langevin Monte Carlo. In *Advances in Neural Information Processing Systems* 7858–7867.
[35] Lee, Y.T., Song, Z. and Vempala, S.S. (2018). Algorithmic theory of ODEs and sampling from well-conditioned log-concave densities. Preprint, arXiv:1812.06243.
[36] Liang, T. and Su, W.J. (2019). Statistical inference for the population landscape via moment-adjusted stochastic gradients. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 431–456. MR3928149
[37] Ma, Y.-A., Chen, Y., Jin, C., Flammarion, N. and Jordan, M.I. (2019). Sampling can be faster than optimization. *Proc. Natl. Acad. Sci. USA* **116** 20881–20885. MR4025861 https://doi.org/10.1073/pnas.1820003116
[38] Majka, M.B., Mijatović, A. and Szpruch, Ł. (2020). Nonasymptotic bounds for sampling algorithms without log-concavity. *Ann. Appl. Probab.* **30** 1534–1581. MR4132634 https://doi.org/10.1214/19-AAP1535
[39] Mangoubi, O. and Vishnoi, N. (2018). Dimensionally tight bounds for second-order Hamiltonian Monte Carlo. In *Advances in Neural Information Processing Systems 31* 6030–6040. Curran Associates.
[40] Markowich, P.A. and Villani, C. (2000). On the trend to equilibrium for the Fokker-Planck equation: An interplay between physics and functional analysis **19** 1–29. Mat. Contemp., VI Workshop on Partial Differential Equations, Part II (Rio de Janeiro, 1999). MR1812873
[41] 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
[42] Mattingly, J.C., Stuart, A.M. and Tretyakov, M.V. (2010). Convergence of numerical time-averaging and stationary measures via Poisson equations. *SIAM J. Numer. Anal.* **48** 552–577. MR2669996 https://doi.org/10.1137/090770527
[43] Milstein, G.N. and Tretyakov, M.V. (2004). *Stochastic Numerics for Mathematical Physics*. *Scientific Computation*. Berlin: Springer. MR2069903 https://doi.org/10.1007/978-3-662-10063-9
[44] Mou, W., Flammarion, N., Wainwright, M.J. and Bartlett, P.L. (2022). Supplement to "Improved Bounds for Discretization of Langevin Diffusions: Near-Optimal Rates without Convexity." https://doi.org/10.3150/21-BEJ1343SUPP
[45] Pavliotis, G.A. (2014). *Stochastic Processes and Applications*. *Texts in Applied Mathematics* **60**. New York: Springer. Diffusion processes, the Fokker-Planck and Langevin equations. MR3288096 https://doi.org/10.1007/978-1-4939-1323-7
[46] Raginsky, M., Rakhlin, A. and Telgarsky, M. (2017). Non-convex learning via stochastic gradient Langevin dynamics: A nonasymptotic analysis. In *Proceedings of the 2017 Conference on Learning Theory*. *Proceedings of Machine Learning Research* **65** 1674–1703. PMLR.
[47] Revuz, D. and Yor, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **293**. Berlin: Springer. MR1725357 https://doi.org/10.1007/978-3-662-06400-9
[48] 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
[49] Rudin, W. (1991). *Functional Analysis*, 2nd ed. *International Series in Pure and Applied Mathematics*. New York: McGraw-Hill, Inc. MR1157815
[50] Stam, A.J. (1959). Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Inf. Control* **2** 101–112. MR0109101
[51] Talay, D. (1990). Simulation and numerical analysis of stochastic differential systems: A review Research Report RR-1313, INRIA.
[52] Talay, D. and Tubaro, L. (1990). Expansion of the global error for numerical schemes solving stochastic differential equations. *Stoch. Anal. Appl.* **8** 483–509. MR1091544 https://doi.org/10.1080/07362999008809220
[53] Toscani, G. (1999). Entropy production and the rate of convergence to equilibrium for the Fokker-Planck equation. *Quart. Appl. Math.* **57** 521–541. MR1704435 https://doi.org/10.1090/qam/1704435
[54] Tzen, B., Liang, T. and Raginsky, M. (2018). Local optimality and generalization guarantees for the Langevin algorithm via empirical metastability. In *Proceedings of the 31st Conference on Learning Theory*. *Proceedings of Machine Learning Research* **75** 857–875. PMLR.
[55] Vempala, S. and Wibisono, A. (2019). Rapid convergence of the unadjusted Langevin algorithm: Isoperimetry suffices. In *Advances in Neural Information Processing Systems* 8094–8106.

[56] Veretennikov, A.Yu. (1997). On polynomial mixing bounds for stochastic differential equations. *Stochastic Process. Appl.* **70** 115–127. MR1472961 https://doi.org/10.1016/S0304-4149(97)00056-2

[57] Wibisono, A. (2018). Sampling as optimization in the space of measures: The Langevin dynamics as a composite optimization problem. In *Proceedings of the 31st Conference on Learning Theory*. *Proceedings of Machine Learning Research* **75** 2093–3027. PMLR.

[58] Zhang, Y., Liang, P. and Charikar, M. (2017). A hitting time analysis of stochastic gradient Langevin dynamics. In *Proceedings of the 2017 Conference on Learning Theory*. *Proceedings of Machine Learning Research* **65** 1980–2022. PMLR.

Bernoulli **28**(3), 2022, 1602–1624
https://doi.org/10.3150/21-BEJ1349

Limit theorems for supercritical branching processes in random environment

DARIUSZ BURACZEWSKI[a] and EWA DAMEK[b]

Mathematical Institute, University of Wroclaw, Pl. Grunwaldzki 2/4 50-384 Wroclaw, Poland.
[a]*dariusz.buraczewski@uwr.edu.pl*, [b]*edamek@math.uni.wroc.pl*

The branching process in random environment $\{Z_n\}_{n\geq 0}$ is a population growth process where individuals reproduce independently of each other with the reproduction law randomly picked at each generation. We focus on the supercritical case, when the process survives with positive probability and grows exponentially fast on the nonextinction set. Using Fourier techniques we improve existing central limit theorem as well as we obtain Edgeworth expansions and the renewal theorem for the sequence $\{\log Z_n\}_{n\geq 0}$. The strategy is to compare $\log Z_n$ with partial sums of i.i.d. random variables in order to obtain precise estimates.

Keywords: Branching process; random environment; central limit theorem; Berry Esseen bound; Edgeworth expansions; renewal theorem; Fourier transform; characteristic function

References

[1] Athreya, K.B. and Karlin, S. (1971). On branching processes with random environments. I. Extinction probabilities. *Ann. Math. Stat.* **42** 1499–1520. MR0298780 https://doi.org/10.1214/aoms/1177693150
[2] Athreya, K.B. and Ney, P.E. (1972). *Branching Processes*. *Die Grundlehren der Mathematischen Wissenschaften, Band 196*. New York: Springer. MR0373040
[3] Bansaye, V. and Böinghoff, C. (2013). Lower large deviations for supercritical branching processes in random environment. *Tr. Mat. Inst. Steklova* **282** 22–41. MR3308579 https://doi.org/10.1134/s0081543813060035
[4] Bansaye, V. and Böinghoff, C. (2014). Small positive values for supercritical branching processes in random environment. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** 770–805. MR3224289 https://doi.org/10.1214/13-AIHP538
[5] Biggins, J.D. (1992). Uniform convergence of martingales in the branching random walk. *Ann. Probab.* **20** 137–151. MR1143415
[6] Breiman, L. (1968). *Probability*. Reading, Mass.-London-Don Mills, Ont.: Addison-Wesley Co. MR0229267
[7] Buraczewski, D. and Dyszewski, P. (2017). Precise large deviation estimates for branching process in random environment. arXiv preprint. Available at arXiv:1706.03874.
[8] Damek, E., Gantert, N. and Kolesko, K. (2019). Absolute continuity of the martingale limit in branching processes in random environment. *Electron. Commun. Probab.* **24** 42. MR3978691 https://doi.org/10.1214/19-ECP229
[9] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications. Vol. II*, 2nd ed. New York: Wiley. MR0270403
[10] Féray, V., Méliot, P.L. and Nikeghbali, A. (2016). *Mod-phi Convergence I: Normality Zones and Precise Deviations*. Berlin: Springer.
[11] Grama, I., Liu, Q. and Miqueu, E. (2017). Berry-Esseen's bound and Cramér's large deviation expansion for a supercritical branching process in a random environment. *Stochastic Process. Appl.* **127** 1255–1281. MR3619270 https://doi.org/10.1016/j.spa.2016.07.014
[12] Gut, A. (1988). Stopped random walks. Limit theorems and applications. In *Applied Probability. A Series of the Applied Probability Trust* **5** x+199 pp. New York: Springer.
[13] Hervé, L. and Pène, F. (2010). The Nagaev-Guivarc'h method via the Keller-Liverani theorem. *Bull. Soc. Math. France* **138** 415–489. MR2729019 https://doi.org/10.24033/bsmf.2594

[14] Huang, C. and Liu, Q. (2012). Moments, moderate and large deviations for a branching process in a random environment. *Stochastic Process. Appl.* **122** 522–545. MR2868929 https://doi.org/10.1016/j.spa.2011.09.001

[15] Huang, C. and Liu, Q. (2014). Convergence in L^p and its exponential rate for a branching process in a random environment. *Electron. J. Probab.* **19** 104. MR3275856 https://doi.org/10.1214/EJP.v19-3388

[16] Kabluchko, Z., Marynych, A. and Sulzbach, H. (2017). General Edgeworth expansions with applications to profiles of random trees. *Ann. Appl. Probab.* **27** 3478–3524. MR3737930 https://doi.org/10.1214/17-AAP1285

[17] Kersting, G. and Vatutin, V. (2017). *Discrete Time Branching Processes in Random Environment*. New York: Wiley.

[18] Smith, W.L. and Wilkinson, W.E. (1969). On branching processes in random environments. *Ann. Math. Stat.* **40** 814–827. MR0246380 https://doi.org/10.1214/aoms/1177697589

[19] Tanny, D. (1977). Limit theorems for branching processes in a random environment. *Ann. Probab.* **5** 100–116. MR0426189 https://doi.org/10.1214/aop/1176995894

[20] Tanny, D. (1988). A necessary and sufficient condition for a branching process in a random environment to grow like the product of its means. *Stochastic Process. Appl.* **28** 123–139. MR0936379 https://doi.org/10.1016/0304-4149(88)90070-1

[21] Tao, T. (2006). *Analysis II*. Hindustan Book Agency.

Bernoulli **28**(3), 2022, 1625–1647
https://doi.org/10.3150/21-BEJ1363

Two-step Bayesian methods for generalized regression driven by partial differential equations

PRITHWISH BHAUMIK[a], WENLI SHI[b] and SUBHASHIS GHOSAL[c]

Department of Statistics, North Carolina State University, 2311 Stinson Drive, Raleigh, North Carolina 27695-8203, USA. [a]*pbhaumi@ncsu.edu,* [b]*wenli_shi@ncsu.edu,* [c]*sghosal@ncsu.edu*

In certain non-linear regression models, the functional form of the regression function is not explicitly available, but is only described by a set of differential equations. For regression models described by a set of ordinary differential equations (ODEs), both Bayesian and non-Bayesian methods for inference were developed in the literature. In this paper, we consider a Bayesian approach to non-linear regression with respect to a multidimensional predictor variable given by a set of partial differential equations (PDEs). We consider a computationally convenient two-step approach by first representing the functions nonparametrically, constructing a finite random series prior using tensor products of B-splines and directly inducing a posterior distribution on parameter space through an appropriate projection map. By considering three different choices of the projection map, we propose three different approaches with their merits. We allow generalized non-linear regression with the response variable following an exponential family of distributions, extending the method beyond regression with additive normal errors. We establish Bernstein-von Mises type theorems which show $\sqrt{n}$-consistency and asymptotically correct frequentist coverage of Bayesian credible regions. We also conduct a simulation study to evaluate the finite sample performances of the proposed methods.

Keywords: Partial differential equation; generalized regression; two-step method; projection posterior; Bernstein-von Mises theorem; contiguity; B-splines; tensor products

References

[1] Baake, E., Baake, M., Bock, H. and Briggs, K. (1992). Fitting ordinary differential equations to chaotic data. *Phys. Rev. A* **45** 5524–5529.
[2] Bhaumik, P. and Ghosal, S. (2015). Bayesian two-step estimation in differential equation models. *Electron. J. Stat.* **9** 3124–3154. MR3453972 https://doi.org/10.1214/15-EJS1099
[3] Bhaumik, P. and Ghosal, S. (2017). Bayesian inference for higher-order ordinary differential equation models. *J. Multivariate Anal.* **157** 103–114. MR3641739 https://doi.org/10.1016/j.jmva.2017.03.003
[4] Bhaumik, P. and Ghosal, S. (2017). Efficient Bayesian estimation and uncertainty quantification in ordinary differential equation models. *Bernoulli* **23** 3537–3570. MR3654815 https://doi.org/10.3150/16-BEJ856
[5] Brunel, N.J.-B. (2008). Parameter estimation of ODE's via nonparametric estimators. *Electron. J. Stat.* **2** 1242–1267. MR2471285 https://doi.org/10.1214/07-EJS132
[6] Campbell, D. and Steele, R.J. (2012). Smooth functional tempering for nonlinear differential equation models. *Stat. Comput.* **22** 429–443. MR2865027 https://doi.org/10.1007/s11222-011-9234-3
[7] Chkrebtii, O.A., Campbell, D.A., Calderhead, B. and Girolami, M.A. (2016). Bayesian solution uncertainty quantification for differential equations. *Bayesian Anal.* **11** 1239–1267. MR3577378 https://doi.org/10.1214/16-BA1017
[8] Dattner, I. and Gugushvili, S. (2018). Application of one-step method to parameter estimation in ODE models. *Stat. Neerl.* **72** 126–156. MR3790757 https://doi.org/10.1111/stan.12124
[9] Gelman, A., Bois, F. and Jiang, J. (1996). Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *J. Amer. Statist. Assoc.* **91** 1400–1412.

[10] Ghosal, S. (1997). Normal approximation to the posterior distribution for generalized linear models with many covariates. *Math. Methods Statist.* **6** 332–348. MR1475901
[11] Girolami, M. (2008). Bayesian inference for differential equations. *Theoret. Comput. Sci.* **408** 4–16. MR2460604 https://doi.org/10.1016/j.tcs.2008.07.005
[12] Gugushvili, S. and Klaassen, C.A.J. (2012). $\sqrt{n}$-consistent parameter estimation for systems of ordinary differential equations: Bypassing numerical integration via smoothing. *Bernoulli* **18** 1061–1098. MR2948913 https://doi.org/10.3150/11-BEJ362
[13] Hairer, E., Nørsett, S.P. and Wanner, G. (1993). *Solving Ordinary Differential Equations. I: Nonstiff Problems*, 2nd ed. *Springer Series in Computational Mathematics* **8**. Berlin: Springer. MR1227985
[14] Johnson, C. (2012). Numerical Solution of Partial Differential Equations by the Finite Element Method. Courier Corporation.
[15] Kleijn, B.J.K. and van der Vaart, A.W. (2012). The Bernstein-Von-Mises theorem under misspecification. *Electron. J. Stat.* **6** 354–381. MR2988412 https://doi.org/10.1214/12-EJS675
[16] Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quart. Appl. Math.* **2** 164–168. MR0010666 https://doi.org/10.1090/qam/10666
[17] Marquardt, D.W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Indust. Appl. Math.* **11** 431–441. MR0153071
[18] Mattheij, R. and Molenaar, J. (2002). *Ordinary Differential Equations in Theory and Practice*. *Classics in Applied Mathematics* **43**. Philadelphia, PA: SIAM. MR1946758 https://doi.org/10.1137/1.9780898719178
[19] Morton, K.W. and Mayers, D.F. (2005). *Numerical Solution of Partial Differential Equations: An Introduction*, 2nd ed. Cambridge: Cambridge Univ. Press. MR2153063 https://doi.org/10.1017/CBO9780511812248
[20] Müller, T.G. and Timmer, J. (2004). Parameter identification techniques for partial differential equations. *Internat. J. Bifur. Chaos Appl. Sci. Engrg.* **14** 2052–2060. MR2076181 https://doi.org/10.1142/S0218127404010424
[21] Rogers, S., Khanin, R. and Girolami, M. (2007). Bayesian model-based inference of transcription factor activity. *BMC Bioinform.* **8** S2.
[22] Schumaker, L.L. (2007). *Spline Functions: Basic Theory*, 3rd ed. *Cambridge Mathematical Library*. Cambridge: Cambridge Univ. Press. MR2348176 https://doi.org/10.1017/CBO9780511618994
[23] Tadmor, E. (2012). A review of numerical methods for nonlinear partial differential equations. *Bull. Amer. Math. Soc. (N.S.)* **49** 507–554. MR2958929 https://doi.org/10.1090/S0273-0979-2012-01379-4
[24] van der Vaart, A.W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. MR1652247 https://doi.org/10.1017/CBO9780511802256
[25] Varah, J.M. (1982). A spline least squares method for numerical parameter estimation in differential equations. *SIAM J. Sci. Statist. Comput.* **3** 28–46. MR0651865 https://doi.org/10.1137/0903003
[26] Voit, E.O. and Almeida, J. (2004). Decoupling dynamical systems for pathway identification from metabolic profiles. *Bioinformatics* **20** 1670–1681.
[27] Wu, H., Xue, H. and Kumar, A. (2012). Numerical discretization-based estimation methods for ordinary differential equation models via penalized spline smoothing with applications in biomedical research. *Biometrics* **68** 344–352. MR2959600 https://doi.org/10.1111/j.1541-0420.2012.01752.x
[28] Xue, H., Miao, H. and Wu, H. (2010). Sieve estimation of constant and time-varying coefficients in nonlinear ordinary differential equation models by considering both numerical error and measurement error. *Ann. Statist.* **38** 2351–2387. MR2676892 https://doi.org/10.1214/09-AOS784
[29] Xun, X., Cao, J., Mallick, B., Maity, A. and Carroll, R.J. (2013). Parameter estimation of partial differential equation models. *J. Amer. Statist. Assoc.* **108** 1009–1020. MR3174680 https://doi.org/10.1080/01621459.2013.794730
[30] Yoo, W.W. and Ghosal, S. (2016). Supremum norm posterior contraction and credible sets for nonparametric multivariate regression. *Ann. Statist.* **44** 1069–1102. MR3485954 https://doi.org/10.1214/15-AOS1398

Bernoulli **28**(3), 2022, 1648–1674
https://doi.org/10.3150/21-BEJ1365

Homogenization of nonlocal partial differential equations related to stochastic differential equations with Lévy noise

QIAO HUANG[1,a], JINQIAO DUAN[2,b] and RENMING SONG[3,c]

[1]*Center for Mathematical Sciences, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China.* [a]*hq932309@alumni.hust.edu.cn*
[2]*Department of Applied Mathematics, Illinois Institute of Technology, Chicago, IL 60616, USA.* [b]*duan@iit.edu*
[3]*Department of Mathematics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.* [c]*rsong@illinois.edu*

We study the "periodic homogenization" for a class of nonlocal partial differential equations of parabolic-type with rapidly oscillating coefficients, related to stochastic differential equations driven by multiplicative isotropic α-stable Lévy noise ($1 < \alpha < 2$) which is nonlinear in the noise component. Our homogenization method is probabilistic. It turns out that, under suitable regularity assumptions, the limit of the solutions satisfies a nonlocal partial differential equation with constant coefficients, which are associated to a symmetric α-stable Lévy process.

Keywords: Homogenization; nonlocal parabolic PDEs; SDEs with jumps; Zvonkin's transform; strong well-posedness; Feller semigroups; Feynman-Kac formula

References

Allaire, G. (2002). *Shape Optimization by the Homogenization Method*. *Applied Mathematical Sciences* **146**. New York: Springer. MR1859696 https://doi.org/10.1007/978-1-4684-9286-6

Applebaum, D. (2009). *Lévy Processes and Stochastic Calculus*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **116**. Cambridge: Cambridge Univ. Press. MR2512800 https://doi.org/10.1017/CBO9780511809781

Arisawa, M. (2009). Homogenization of a class of integro-differential equations with Lévy operators. *Comm. Partial Differential Equations* **34** 617–624. MR2560294 https://doi.org/10.1080/03605300902963518

Bal, G. and Jing, W. (2016). Fluctuations in the homogenization of semilinear equations with random potentials. *Comm. Partial Differential Equations* **41** 1839–1859. MR3572560 https://doi.org/10.1080/03605302.2016.1238482

Barczy, M., Li, Z. and Pap, G. (2015). Yamada-Watanabe results for stochastic differential equations with jumps. *Int. J. Stoch. Anal.* **2015** 1–23.

Bass, R.F. (2009). Regularity results for stable-like operators. *J. Funct. Anal.* **257** 2693–2722. MR2555009 https://doi.org/10.1016/j.jfa.2009.05.012

Bella, P., Fehrman, B., Fischer, J. and Otto, F. (2017). Stochastic homogenization of linear elliptic equations: Higher-order error estimates in weak norms via second-order correctors. *SIAM J. Math. Anal.* **49** 4658–4703. MR3725278 https://doi.org/10.1137/16M110229X

Benchérif-Madani, A. and Pardoux, É. (2007). Homogenization of a semilinear parabolic PDE with locally periodic coefficients: A probabilistic approach. *ESAIM Probab. Stat.* **11** 385–411. MR2339300 https://doi.org/10.1051/ps:2007026

Bensoussan, A., Lions, J.-L. and Papanicolaou, G. (1978). *Asymptotic Analysis for Periodic Structures*. *Studies in Mathematics and Its Applications* **5**. Amsterdam: North-Holland. MR0503330

Bessaih, H., Hausenblas, E. and Razafimandimby, P.A. (2015). Strong solutions to stochastic hydrodynamical systems with multiplicative noise of jump type. *NoDEA Nonlinear Differential Equations Appl.* **22** 1661–1697. MR3415017 https://doi.org/10.1007/s00030-015-0339-9

1350-7265 © 2022 ISI/BS

Chen, Z.-Q. and Zhang, X. (2018). Heat kernels for time-dependent non-symmetric stable-like operators. *J. Math. Anal. Appl.* **465** 1–21. MR3806688 https://doi.org/10.1016/j.jmaa.2018.03.054

Cioranescu, D. and Saint Jean Paulin, J. (1999). *Homogenization of Reticulated Structures*. *Applied Mathematical Sciences* **136**. New York: Springer. MR1676922 https://doi.org/10.1007/978-1-4612-2158-6

Engel, K.-J. and Nagel, R. (2000). *One-Parameter Semigroups for Linear Evolution Equations*. *Graduate Texts in Mathematics* **194**. New York: Springer.

Ethier, S.N. and Kurtz, T.G. (1986). *Markov Processes: Characterization and Convergence*. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley.

Franke, B. (2006). The scaling limit behaviour of periodic stable-like processes. *Bernoulli* **12** 551–570. MR2232732 https://doi.org/10.3150/bj/1151525136

Franke, B. (2007). A functional non-central limit theorem for jump-diffusions with periodic coefficients driven by stable Lévy-noise. *J. Theoret. Probab.* **20** 1087–1100. MR2362589 https://doi.org/10.1007/s10959-007-0099-5

Gilbarg, D. and Trudinger, N.S. (2001). *Elliptic Partial Differential Equations of Second Order*. Reprint of the 1998 edition. *Classics in Mathematics* **224**. Berlin: Springer.

Hairer, M. and Pardoux, E. (2008). Homogenization of periodic linear degenerate PDEs. *J. Funct. Anal.* **255** 2462–2487. MR2473263 https://doi.org/10.1016/j.jfa.2008.04.014

Horie, M., Inuzuka, T. and Tanaka, H. (1977). Homogenization of certain one-dimensional discontinuous Markov processes. *Hiroshima Math. J.* **7** 629–641. MR0488335

Huang, Q., Duan, J. Song, R. (2022). Supplement to "Homogenization of Nonlocal Partial Differential Equations Related to Stochastic Differential Equations with Lévy Noise." https://doi.org/10.3150/21-BEJ1365SUPP

Ikeda, N. and Watanabe, S. (1981). *Stochastic Differential Equations and Diffusion Processes*. *North-Holland Mathematical Library* **24**. Amsterdam: Elsevier.

Jacod, J. and Shiryaev, A. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. *Grundlehren der mathematischen Wissenschaften* **288**. Berlin: Springer.

Kallenberg, O. (2002). *Foundations of Modern Probability*, 2nd ed. *Probability and Its Applications*. New York: Springer.

Kassmann, M., Piatnitski, A. and Zhizhina, E. (2019). Homogenization of Lévy-type operators with oscillating coefficients. *SIAM J. Math. Anal.* **51** 3641–3665. MR4000213 https://doi.org/10.1137/18M1200038

Krantz, S.G. and Parks, H.R. (2013). *The Implicit Function Theorem: History, Theory, and Applications*. Reprint of the 2003 edition. *Modern Birkhäuser Classics*. New York: Springer.

Kurtz, T.G. (2011). Equivalence of stochastic equations and martingale problems. In *Stochastic Analysis 2010* 113–130. Heidelberg: Springer. MR2789081 https://doi.org/10.1007/978-3-642-15358-7_6

Le Bris, C., Legoll, F. and Madiot, F. (2019). Multiscale finite element methods for advection-dominated problems in perforated domains. *Multiscale Model. Simul.* **17** 773–825. MR3961974 https://doi.org/10.1137/17M1152048

Masuda, H. (2007). Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps. *Stochastic Process. Appl.* **117** 35–56. MR2287102 https://doi.org/10.1016/j.spa.2006.04.010

Oleĭnik, O.A., Shamaev, A.S. and Yosifian, G.A. (1992). *Mathematical Problems in Elasticity and Homogenization*. *Studies in Mathematics and Its Applications* **26**. Amsterdam: North-Holland. MR1195131

Pardoux, É. (1999). Homogenization of linear and semilinear second order parabolic PDEs with periodic coefficients: A probabilistic approach. *J. Funct. Anal.* **167** 498–520. MR1716206 https://doi.org/10.1006/jfan.1999.3441

Pardoux, É. and Piatnitski, A. (2012). Homogenization of a singular random one-dimensional PDE with time-varying coefficients. *Ann. Probab.* **40** 1316–1356. MR2962093 https://doi.org/10.1214/11-AOP650

Pardoux, E., Răşcanu, A. and Răşcanu, A. (2014). *Stochastic Differential Equations, Backward SDEs, Partial Differential Equations*. *Stochastic Modelling and Applied Probability* **69**. Cham: Springer. MR3308895 https://doi.org/10.1007/978-3-319-05714-9

Pardoux, E. and Veretennikov, A.Y. (2001). On the Poisson equation and diffusion approximation. I. *Ann. Probab.* **29** 1061–1085. MR1872736 https://doi.org/10.1214/aop/1015345596

Piatnitski, A. and Zhizhina, E. (2017). Periodic homogenization of nonlocal operators with a convolution-type kernel. *SIAM J. Math. Anal.* **49** 64–81. MR3595876 https://doi.org/10.1137/16M1072292

Priola, E. (2012). Pathwise uniqueness for singular SDEs driven by stable processes. *Osaka J. Math.* **49** 421–447. MR2945756

Rhodes, R. and Sow, A.B. (2011). Critical homogenization of SDEs driven by a Levy process in random medium. *Stoch. Anal. Appl.* **29** 838–859. MR2827434 https://doi.org/10.1080/07362994.2011.598795

Sakamoto, Y. and Yoshida, N. (2004). Asymptotic expansion formulas for functionals of ϵ-Markov processes with a mixing property. *Ann. Inst. Statist. Math.* **56** 545–597. MR2095018 https://doi.org/10.1007/BF02530541

Sandrić, N. (2016). Homogenization of periodic diffusion with small jumps. *J. Math. Anal. Appl.* **435** 551–577. MR3423413 https://doi.org/10.1016/j.jmaa.2015.10.054

Schäffer, J.J. (1970). Norms and determinants of linear mappings. *Math. Z.* **118** 331–339. MR0281730 https://doi.org/10.1007/BF01109869

Schwab, R.W. (2010). Periodic homogenization for nonlinear integro-differential equations. *SIAM J. Math. Anal.* **42** 2652–2680. MR2733264 https://doi.org/10.1137/080737897

Tomisaki, M. (1992). Homogenization of càdlàg processes. *J. Math. Soc. Japan* **44** 281–305. MR1154844 https://doi.org/10.2969/jmsj/04420281

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

Xie, L. and Zhang, X. (2014). Heat kernel estimates for critical fractional diffusion operators. *Studia Math.* **224** 221–263. MR3294616 https://doi.org/10.4064/sm224-3-3

Yosida, K. (1995). *Functional Analysis*, Reprint of the sixth (1980) edition. *Classics in Mathematics* **123**. Berlin: Springer.

Zvonkin, A.K. (1974). A transformation of the phase space of a diffusion process that removes the drift. *Mat. Sb.* **93** 129–149.

Bernoulli **28**(3), 2022, 1675–1703
https://doi.org/10.3150/21-BEJ1378

Doubly robust semiparametric inference using regularized calibrated estimation with high-dimensional data

SATYAJIT GHOSH and ZHIQIANG TAN[a]

Department of Statistics, Rutgers University, Piscataway, New Jersey 08854, USA. [a]*ztan@stat.rutgers.edu*

Consider semiparametric estimation where a doubly robust estimating function for a low-dimensional parameter is available, depending on two working models. With high-dimensional data, we develop regularized calibrated estimation as a general method for estimating the parameters in the two working models, such that valid Wald confidence intervals can be obtained for the parameter of interest under suitable sparsity conditions if either of the two working models is correctly specified. We propose a computationally tractable two-step algorithm and provide rigorous theoretical analysis which justifies sufficiently fast rates of convergence for the regularized calibrated estimators in spite of sequential construction and establishes a desired asymptotic expansion for the doubly robust estimator. As concrete examples, we discuss applications to partially linear, log-linear, and logistic models and estimation of average treatment effects. Numerical studies in the former three examples demonstrate superior performance of our method, compared with debiased Lasso.

Keywords: Average treatment effect; calibration estimation; debiased Lasso; double robustness; high-dimensional data; Lasso penalty; partially linear model; semiparametric estimation

References

[1] Avagyan, V. and Vansteelandt, S. (2017). Honest data-adaptive inference for the average treatment effect under model misspecification using penalised bias-reduced double-robust estimation. Available at arXiv:1708.03787.

[2] 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

[3] Bickel, P.J., Klaassen, C.A., Ritov, Y. and Wellner, J.A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins Univ. Press. MR1245941

[4] Bradic, J., Wager, S. and Zhu, Y. (2019). Sparsity double robust inference of average treatment effects. Available at arXiv:1905.00744.

[5] Bühlmann, P. and Van de Geer, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Berlin: Springer. MR2807761 https://doi.org/10.1007/978-3-642-20192-9

[6] Bühlmann, P. and Van de Geer, S. (2015). High-dimensional inference in misspecified linear models. *Electron. J. Stat.* **9** 1449–1473.

[7] Candes, E. and Tao, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n (with discussion). *Ann. Statist.* **35** 2313–2351. MR2382644 https://doi.org/10.1214/009053606000001523

[8] Chen, H.Y. (2007). A semiparametric odds ratio model for measuring association. *Biometrics* **63** 413–421. MR2370799 https://doi.org/10.1111/j.1541-0420.2006.00701.x

[9] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. and Robins, J.M. (2018). Double/debiased machine learning for treatment and structural parameters. *Econom. J.* **21** C1–C68. MR3769544 https://doi.org/10.1111/ectj.12097

[10] Dukes, O., Avagyan, V. and Vansteelandt, S. (2020). Doubly robust tests of exposure effects under high-dimensional confounding. *Biometrics* **76** 1190–1200. https://doi.org/10.1111/biom.13231

1350-7265

[11] Dukes, O. and Vansteelandt, S. (2020). Inference for treatment effect parameters in potentially misspecified high-dimensional models. *Biometrika* **2** 321–334. MR4259134 https://doi.org/10.1093/biomet/asaa071
[12] Farrell, M.H. (2015). Robust inference on average treatment effects with possibly more covariates than observations. *J. Econometrics* **189** 1–23. MR3397349 https://doi.org/10.1016/j.jeconom.2015.06.017
[13] Ferguson, T.S. (1996). *A Course in Large Sample Theory*. Boca Raton: CRC Press.
[14] Folsom, R.E. (1991). Exponential and logistic weight adjustments for sampling and nonresponse error reduction. In *Proceedings of the American Statistical Association, Social Statistics Section* **101** 197–202.
[15] Ghosh, S. and Tan, Z. (2022). Supplement to "Doubly Robust Semiparametric Inference Using Regularized Calibrated Estimation with High-dimensional Data." https://doi.org/10.3150/21-BEJ1378SUPPA, https://doi.org/10.3150/21-BEJ1378SUPPB
[16] Javanmard, A. and Montanari, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15** 2869–2909. MR3277152
[17] Kang, J.D. and Schafer, J.L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data (with discussion). *Statist. Sci.* **22** 523–539. MR2420458 https://doi.org/10.1214/07-STS227
[18] Kim, J.K. and Haziza, D. (2014). Doubly robust inference with missing data in survey sampling. *Statist. Sinica* **24** 375–394. MR3183689
[19] Ma, Y. and Zhu, L. (2012). A semiparametric approach to dimension reduction. *J. Amer. Statist. Assoc.* **107** 168–179. MR2949349 https://doi.org/10.1080/01621459.2011.646925
[20] Manski, C.F. (1988). *Analog Estimation Methods in Econometrics*. Boca Raton: CRC Press. MR0996421
[21] McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, 2nd ed. Boca Raton: CRC Press. MR3223057 https://doi.org/10.1007/978-1-4899-3242-6
[22] Negahban, S.N., Ravikumar, P., Wainwright, M.J. and Yu, B. (2012). A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers. *Statist. Sci.* **27** 538–557. MR3025133 https://doi.org/10.1214/12-STS400
[23] Neykov, M., Ning, Y., Liu, J.S. and Liu, H. (2018). A unified theory of confidence regions and testing for high-dimensional estimating equations. *Statist. Sci.* **33** 427–443. MR3843384 https://doi.org/10.1214/18-STS661
[24] 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
[25] Ning, Y., Peng, S. and Imai, K. (2020). Robust estimation of causal effects via a high-dimensional covariate balancing propensity score. *Biometrika* **107** 533–554. MR4138975 https://doi.org/10.1093/biomet/asaa020
[26] Okui, R., Small, D., Tan, Z. and Robins, J.M. (2012). Doubly robust instrumental variables regression. *Statist. Sinica* **22** 173–205. MR2933172 https://doi.org/10.5705/ss.2009.265
[27] Radchenko, P. and James, G.M. (2011). Improved variable selection with forward-Lasso adaptive shrinkage. *Ann. Appl. Stat.* **5** 427–448. MR2810404 https://doi.org/10.1214/10-AOAS375
[28] Robins, J.M., Mark, S.D. and Newey, W.K. (1992). Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics* **48** 479–495. MR1173493 https://doi.org/10.2307/2532304
[29] Robins, J.M. and Rotnitzky, A. (2001). Comment on "Inference for semiparametric models: Some questions and an answer" by Bickel and Kwon. *Statist. Sinica* **11** 920–936.
[30] Robins, J.M., Rotnitzky, A. and Zhao, L.P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89** 846–866. MR1294730
[31] Rubin, D.B. (1976). Inference and missing data. *Biometrika* **63** 581–592. MR0455196 https://doi.org/10.1093/biomet/63.3.581
[32] Scharfstein, D.O., Rotnitzky, A. and Robins, J.M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion). *J. Amer. Statist. Assoc.* **94** 1096–1120. MR1731478 https://doi.org/10.2307/2669923
[33] Small, C.G., Wang, J. and Yang, Z. (2000). Eliminating multiple root problems in estimation (with discussion). *Statist. Sci.* **15** 313–341.
[34] Smucler, E., Rotnitzky, A. and Robins, J.M. (2019). A unifying approach for doubly-robust ℓ_1 regularized estimation of causal contrasts. Available at arXiv:1904.03737.
[35] Tan, Z. (2006). A distributional approach for causal inference using propensity scores. *J. Amer. Statist. Assoc.* **101** 1619–1637.

[36] Tan, Z. (2006). Regression and weighting methods for causal inference using instrumental variables. *J. Amer. Statist. Assoc.* **101** 1607–1618. MR2279483 https://doi.org/10.1198/016214505000001366

[37] Tan, Z. (2010). Bounded, efficient and doubly robust estimation with inverse weighting. *Biometrika* **97** 661–682.

[38] Tan, Z. (2019). On doubly robust estimation for logistic partially linear models. *Statist. Probab. Lett.* **155** 108577. MR3994482 https://doi.org/10.1016/j.spl.2019.108577

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

[40] Tan, Z. (2020). Regularized calibrated estimation of propensity scores with model misspecification and high-dimensional data. *Biometrika* **107** 137–158.

[41] Tchetgen Tchetgen, E.J., Robins, J.M. and Rotnitzky, A. (2010). On doubly robust estimation in a semiparametric odds ratio model. *Biometrika* **97** 171–180.

[42] Tchetgen Tchetgen, E.J. and Shpitser, I. (2012). Semiparametric theory for causal mediation analysis: Efficiency bounds, multipleobustness, and sensitivity analysis. *Ann. Statist.* **40** 1816.

[43] 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.

[44] Van der Vaart, A.W. (2000). *Asymptotic Statistics*. Cambridge: Cambridge Univ. Press.

[45] Vermeulen, K. and Vansteelandt, S. (2015). Bias-reduced doubly robust estimation. *J. Amer. Statist. Assoc.* **110** 1024–1036. MR3420681 https://doi.org/10.1080/01621459.2014.958155

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

[47] White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50** 1–25.

[48] Zhang, C.-H. and Zhang, S.S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. Roy. Statist. Soc. Ser. B* **76** 217–242. MR3153940 https://doi.org/10.1111/rssb.12026

Bernoulli **28**(3), 2022, 1704–1728
https://doi.org/10.3150/21-BEJ1388

Sequential estimation of quantiles with applications to A/B testing and best-arm identification

STEVEN R. HOWARD[1,a] and AADITYA RAMDAS[2,b]

[1]*The Voleon Group, 2150 Dwight Way, Berkley, CA 94704.* [a]*steve@stevehoward.org*
[2]*Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA 15213.*
[b]*aramdas@stat.cmu.edu*

We design confidence sequences—sequences of confidence intervals which are valid uniformly over time—for quantiles of any distribution over a complete, fully-ordered set, based on a stream of i.i.d. observations. We give methods both for tracking a fixed quantile and for tracking all quantiles simultaneously. Specifically, we provide explicit expressions with small constants for intervals whose widths shrink at the fastest possible $\sqrt{t^{-1}\log\log t}$ rate, along with a nonasymptotic concentration inequality for the empirical distribution function which holds uniformly over time with the same rate. The latter strengthens Smirnov's empirical process law of the iterated logarithm and extends the Dvoretzky-Kiefer-Wolfowitz inequality to hold uniformly over time. We give a new algorithm and sample complexity bound for selecting an arm with an approximately best quantile in a multi-armed bandit framework. In simulations, our method needs fewer samples than existing methods by a factor of five to fifty.

Keywords: Quantile estimation; confidence sequences; empirical process; Dvoretzky-Kiefer-Wolfowitz inequality; best-arm identification

References

Anderson, C.W. (1984). Large deviations of extremes. In *Statistical Extremes and Applications (Vimeiro, 1983). NATO Adv. Sci. Inst. Ser. C: Math. Phys. Sci.* **131** 325–340. Dordrecht: Reidel. MR0784827

Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N. (2008). *A First Course in Order Statistics*. *Classics in Applied Mathematics* **54**. Philadelphia, PA: SIAM. MR2399836 https://doi.org/10.1137/1.9780898719062

Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford: Oxford Univ. Press. MR3185193 https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

Boucheron, S. and Thomas, M. (2012). Concentration inequalities for order statistics. *Electron. Commun. Probab.* **17** no. 51, 12. MR2994876 https://doi.org/10.1214/ECP.v17-2210

Bubeck, S., Munos, R. and Stoltz, G. (2009). Pure exploration in multi-armed bandits problems. In *Algorithmic Learning Theory*. *Lecture Notes in Computer Science* **5809** 23–37. Berlin: Springer. MR2564216 https://doi.org/10.1007/978-3-642-04414-4_7

Corless, R.M., Gonnet, G.H., Hare, D.E.G., Jeffrey, D.J. and Knuth, D.E. (1996). On the Lambert W function. *Adv. Comput. Math.* **5** 329–359. MR1414285 https://doi.org/10.1007/BF02124750

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

Darling, D.A. and Robbins, H. (1967b). Iterated logarithm inequalities. *Proc. Natl. Acad. Sci. USA* **57** 1188–1192. MR0211441 https://doi.org/10.1073/pnas.57.5.1188

Darling, D.A. and Robbins, H. (1968). Some nonparametric sequential tests with power one. *Proc. Natl. Acad. Sci. USA* **61** 804–809. MR0238437 https://doi.org/10.1073/pnas.61.3.804

David, Y. and Shimkin, N. (2016). Pure exploration for max-quantile bandits. In *Machine Learning and Knowledge Discovery in Databases*. *Lecture Notes in Computer Science* 556–571. Springer.

Dekkers, A.L.M. and de Haan, L. (1989). On the estimation of the extreme-value index and large quantile estimation. *Ann. Statist.* **17** 1795–1832. MR1026314 https://doi.org/10.1214/aos/1176347396

Drees, H. (1998). On smooth statistical tail functionals. *Scand. J. Stat.* **25** 187–210. MR1614276 https://doi.org/10.1111/1467-9469.00097

Drees, H., de Haan, L. and Li, D. (2003). On large deviation for extremes. *Statist. Probab. Lett.* **64** 51–62. MR1995809 https://doi.org/10.1016/S0167-7152(03)00130-5

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

Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Stat.* **27** 642–669. MR0083864 https://doi.org/10.1214/aoms/1177728174

Even-Dar, E., Mannor, S. and Mansour, Y. (2002). PAC bounds for multi-armed bandit and Markov decision processes. In *Computational Learning Theory (Sydney, 2002)*. *Lecture Notes in Computer Science* **2375** 255–270. Berlin: Springer. MR2040418 https://doi.org/10.1007/3-540-45435-7_18

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

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

Howard, S.R. and Ramdas, A. (2022). Supplement to "Sequential estimation of quantiles with applications to A/B testing and best-arm identification." https://doi.org/10.3150/21-BEJ1388SUPP

Howard, S.R., Ramdas, A., McAuliffe, J. and Sekhon, J. (2020). Time-uniform Chernoff bounds via nonnegative supermartingales. *Probab. Surv.* **17** 257–317. MR4100718 https://doi.org/10.1214/18-PS321

Howard, S.R., Ramdas, A., McAuliffe, J. and Sekhon, J. (2021). Time-uniform, nonparametric, nonasymptotic confidence sequences. *Ann. Statist.* **49** 1055–1080. MR4255119 https://doi.org/10.1214/20-aos1991

James, B.R. (1975). A functional law of the iterated logarithm for weighted empirical distributions. *Ann. Probab.* **3** 762–772. MR0402881 https://doi.org/10.1214/aop/1176996263

Jamieson, K. and Nowak, R. (2014). Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting. In *48th Annual Conference on Information Sciences and Systems (CISS)* 1–6.

Jamieson, K., Malloy, M., Nowak, R. and Bubeck, S. (2014). Lil' UCB: An optimal exploration algorithm for multi-armed bandits. In *Proceedings of the 27th Conference on Learning Theory*. *Proceedings of Machine Learning Research* **35** 423–439.

Kalogerias, D.S., Nikolakakis, K.E., Sarwate, A.D. and Sheffet, O. (2020). Best-Arm Identification for Quantile Bandits with Privacy. arXiv:2006.06792.

Kalyanakrishnan, S., Tewari, A., Auer, P. and Stone, P. (2012). PAC subset selection in stochastic multi-armed bandits. In *Proceedings of the 29th International Conference on Machine Learning* 655–662.

Kaufmann, E., Cappé, O. and Garivier, A. (2016). On the complexity of best-arm identification in multi-armed bandit models. *J. Mach. Learn. Res.* **17** Paper No. 1, 42. MR3482921

Mannor, S. and Tsitsiklis, J.N. (2003/04). The sample complexity of exploration in the multi-armed bandit problem. *J. Mach. Learn. Res.* **5** 623–648. MR2247994 https://doi.org/10.1007/978-3-540-45167-9_31

Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Probab.* **18** 1269–1283. MR1062069

Schreuder, N., Brunel, V.-E. and Dalalyan, A. (2020). A nonasymptotic law of iterated logarithm for general M-estimators. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*. *Proceedings of Machine Learning Research, PMLR* **108** 1331–1341.

Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics*. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. New York: Wiley. MR0838963

Smirnov, N.V. (1944). Approximate laws of distribution of random variables from empirical data. *Uspekhi Mat. Nauk* **10** 179–206. MR0012387

Stephanou, M., Varughese, M. and Macdonald, I. (2017). Sequential quantiles via Hermite series density estimation. *Electron. J. Stat.* **11** 570–607. MR3619317 https://doi.org/10.1214/17-EJS1245

Szörényi, B., Busa-Fekete, R., Weng, E. and Hüllermeier, P. (2015). Qualitative multi-armed bandits: A quantile-based approach. In *Proceedings of the 32nd International Conference on Machine Learning* 1660–1668.

Talagrand, M. (2005). *The Generic Chaining: Upper and Lower Bounds of Stochastic Processes*. *Springer Monographs in Mathematics*. Berlin: Springer. MR2133757

Torossian, L., Garivier, A. and Picheny, V. (2019). X-armed bandits: Optimizing quantiles, CVaR and other risks. In *'Asian Conference on Machine Learning', PMLR* 252–267.

Khintchine, A. (1924). Über einen Satz der Wahrscheinlichkeitsrechnung. *Fund. Math.* **6** 9–20.

Ville, J. (1939). *Étude Critique de la Notion de Collectif*. Gauthier-Villars, Paris. MR3533075

Yu, J.Y. and Nikolova, E. (2013). Sample complexity of risk-averse bandit-arm selection. In *Twenty-Third International Joint Conference on Artificial Intelligence*.

Zhao, S., Zhou, E., Sabharwal, A. and Ermon, S. (2016). Adaptive concentration inequalities for sequential decision problems. In *30th Conference on Neural Information Processing Systems*.

Bernoulli **28**(3), 2022, 1729–1756
https://doi.org/10.3150/21-BEJ1391

Spectral statistics of high dimensional sample covariance matrix with unbounded population spectral norm

YANQING YIN[a]

School of Mathematics and Statistics, Chongqing University, Chongqing 401331, P. R. China.
[a]*yinyq799@nenu.edu.cn*

In this paper, we establish some new central limit theorems for certain spectral statistics of a high-dimensional sample covariance matrix under a divergent spectral norm population model. This model covers the divergent spiked population model as a special case. Meanwhile, the number of the spiked eigenvalues can either be fixed or grow to infinity. It is seen from our theorems that the divergence of population spectral norm affects the fluctuations of the linear spectral statistics in a fickle way, depending on the divergence rate.

Keywords: Large covariance matrix; unbounded spectral norm; linear spectral statistics; spiked eigenvalue

References

Bai, Z.D. and Silverstein, J.W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** 553–605. MR2040792 https://doi.org/10.1214/aop/1078415845

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

Baltagi, B.H., Kao, C. and Wang, F. (2017). Asymptotic power of the sphericity test under weak and strong factors in a fixed effects panel data model. *Econometric Rev.* **36** 853–882. MR3680747 https://doi.org/10.1080/07474938.2017.1307580

Billingsley, P. (1995). *Probability and Measure*. New York: Wiley.

Gao, J., Han, X., Pan, G. and Yang, Y. (2017). High dimensional correlation matrices: The central limit theorem and its applications. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 677–693. MR3641402 https://doi.org/10.1111/rssb.12189

Hu, J., Li, W., Liu, Z. and Zhou, W. (2019). High-dimensional covariance matrices in elliptical distributions with application to spherical test. *Ann. Statist.* **47** 527–555. MR3909941 https://doi.org/10.1214/18-AOS1699

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

Jonsson, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.* **12** 1–38. MR0650926 https://doi.org/10.1016/0047-259X(82)90080-X

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

Pan, G.M. and Zhou, W. (2008). Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. *Ann. Appl. Probab.* **18** 1232–1270. MR2418244 https://doi.org/10.1214/07-AAP477

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

Srivastava, M.S. (2005). Some tests concerning the covariance matrix in high dimensional data. *J. Japan Statist. Soc.* **35** 251–272. MR2328427 https://doi.org/10.14490/jjss.35.251

Yang, Y. and Pan, G. (2015). Independence test for high dimensional data based on regularized canonical correlation coefficients. *Ann. Statist.* **43** 467–500. MR3316187 https://doi.org/10.1214/14-AOS1284

Zheng, S., Bai, Z. and Yao, J. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *Ann. Statist.* **43** 546–591. MR3316190 https://doi.org/10.1214/14-AOS1292

Bernoulli **28**(3), 2022, 1757–1783
https://doi.org/10.3150/21-BEJ1394

Multidimensional SDE with distributional drift and Lévy noise

HELENA KREMP[a] and NICOLAS PERKOWSKI[b]

Freie Universität Berlin, Arnimallee 7, 14195 Berlin, Germany. [a]*helena.kremp@fu-berlin.de*, [b]*perkowski@math.fu-berlin.de*

We solve multidimensional SDEs with distributional drift driven by symmetric, α-stable Lévy processes for $\alpha \in (1,2]$ by studying the associated (singular) martingale problem and by solving the Kolmogorov backward equation. We allow for drifts of regularity $(2-2\alpha)/3$, and in particular we go beyond the by now well understood "Young regime", where the drift must have better regularity than $(1-\alpha)/2$. The analysis of the Kolmogorov backward equation in the low regularity regime is based on paracontrolled distributions. As an application of our results we construct a Brox diffusion with Lévy noise.

Keywords: Singular diffusions; stable Lévy noise; distributional drift; paracontrolled distributions; Brox diffusion

References

[1] Alberts, T., Khanin, K. and Quastel, J. (2014). The intermediate disorder regime for directed polymers in dimension $1+1$. *Ann. Probab.* **42** 1212–1256. MR3189070 https://doi.org/10.1214/13-AOP858
[2] Athreya, S., Butkovsky, O. and Mytnik, L. (2020). Strong existence and uniqueness for stable stochastic differential equations with distributional drift. *Ann. Probab.* **48** 178–210. MR4079434 https://doi.org/10.1214/19-AOP1358
[3] Bahouri, H., Chemin, J.-Y. and Danchin, R. (2011). *Fourier Analysis and Nonlinear Partial Differential Equations. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **343**. Heidelberg: Springer. MR2768550 https://doi.org/10.1007/978-3-642-16830-7
[4] Bailleul, I. and Bernicot, F. (2019). High order paracontrolled calculus. *Forum Math. Sigma* **7** e44, 94. MR4061967 https://doi.org/10.1017/fms.2019.44
[5] Bass, R.F. and Chen, Z.-Q. (2001). Stochastic differential equations for Dirichlet processes. *Probab. Theory Related Fields* **121** 422–446. MR1867429 https://doi.org/10.1007/s004400100151
[6] Brox, Th. (1986). A one-dimensional diffusion process in a Wiener medium. *Ann. Probab.* **14** 1206–1218. MR0866343
[7] Cannizzaro, G. and Chouk, K. (2018). Multidimensional SDEs with singular drift and universal construction of the polymer measure with white noise potential. *Ann. Probab.* **46** 1710–1763. MR3785598 https://doi.org/10.1214/17-AOP1213
[8] Caravenna, F., Sun, R. and Zygouras, N. (2017). Polynomial chaos and scaling limits of disordered systems. *J. Eur. Math. Soc. (JEMS)* **19** 1–65. MR3584558 https://doi.org/10.4171/JEMS/660
[9] Carmona, P. (1997). The mean velocity of a Brownian motion in a random Lévy potential. *Ann. Probab.* **25** 1774–1788. MR1487435 https://doi.org/10.1214/aop/1023481110
[10] Catellier, R. and Gubinelli, M. (2016). Averaging along irregular curves and regularisation of ODEs. *Stochastic Process. Appl.* **126** 2323–2366. MR3505229 https://doi.org/10.1016/j.spa.2016.02.002
[11] Coutin, L. and Qian, Z. (2002). Stochastic analysis, rough path analysis and fractional Brownian motions. *Probab. Theory Related Fields* **122** 108–140. MR1883719 https://doi.org/10.1007/s004400100158
[12] de Raynal, P.-E.C. and Menozzi, S. (2019). On multidimensional stable-driven stochastic differential equations with Besov drift. Preprint arXiv:1907.12263.
[13] Delarue, F. and Diel, R. (2016). Rough paths and 1d SDE with a time dependent distributional drift: Application to polymers. *Probab. Theory Related Fields* **165** 1–63. MR3500267 https://doi.org/10.1007/s00440-015-0626-8

1350-7265

[14] Ethier, S.N. and Kurtz, T.G. (1986). *Markov Processes: Characterization and Convergence*. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. New York: Wiley. MR0838085 https://doi.org/10.1002/9780470316658

[15] Flandoli, F., Gubinelli, M. and Priola, E. (2010). Well-posedness of the transport equation by stochastic perturbation. *Invent. Math.* **180** 1–53. MR2593276 https://doi.org/10.1007/s00222-009-0224-4

[16] Flandoli, F., Issoglio, E. and Russo, F. (2017). Multidimensional stochastic differential equations with distributional drift. *Trans. Amer. Math. Soc.* **369** 1665–1688. MR3581216 https://doi.org/10.1090/tran/6729

[17] Flandoli, F., Russo, F. and Wolf, J. (2003). Some SDEs with distributional drift. I. General calculus. *Osaka J. Math.* **40** 493–542. MR1988703

[18] Gubinelli, M., Imkeller, P. and Perkowski, N. (2015). Paracontrolled distributions and singular PDEs. *Forum Math. Pi* **3** e6, 75. MR3406823 https://doi.org/10.1017/fmp.2015.2

[19] Gubinelli, M. and Perkowski, N. (2015). *Lectures on Singular Stochastic PDEs*. *Ensaios Matemáticos [Mathematical Surveys]* **29**. Rio de Janeiro: Sociedade Brasileira de Matemática. MR3445609

[20] Gubinelli, M. and Perkowski, N. (2017). KPZ reloaded. *Comm. Math. Phys.* **349** 165–269. MR3592748 https://doi.org/10.1007/s00220-016-2788-3

[21] Hairer, M. (2014). A theory of regularity structures. *Invent. Math.* **198** 269–504. MR3274562 https://doi.org/10.1007/s00222-014-0505-4

[22] Hairer, M. and Labbé, C. (2015). A simple construction of the continuum parabolic Anderson model on $\mathbf{R}^2$. *Electron. Commun. Probab.* **20** no. 43, 11. MR3358965 https://doi.org/10.1214/ECP.v20-4038

[23] Harang, F.A. and Ling, C. (2020). Regularity of local times associated to Volterra-Lévy processes and pathwise regularization of stochastic differential equations. Preprint arXiv:2007.01093.

[24] Harang, F.A. and Perkowski, N. (2021). C^∞-regularization of ODEs perturbed by noise. *Stoch. Dyn.* **21** Paper No. 2140010, 29. MR4342752 https://doi.org/10.1142/S0219493721400104

[25] Jacod, J. and Shiryaev, A.N. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **288**. Berlin: Springer. MR1943877 https://doi.org/10.1007/978-3-662-05265-5

[26] Janson, S. (1997). *Gaussian Hilbert Spaces*. *Cambridge Tracts in Mathematics* **129**. Cambridge: Cambridge Univ. Press. MR1474726 https://doi.org/10.1017/CBO9780511526169

[27] Kingman, J.F.C. (1993). *Poisson Processes*. *Oxford Studies in Probability* **3**. Oxford University Press: New York, The Clarendon Press. Oxford Science Publications. MR1207584

[28] Krylov, N.V. and Röckner, M. (2005). Strong solutions of stochastic equations with singular time dependent drift. *Probab. Theory Related Fields* **131** 154–196. MR2117951 https://doi.org/10.1007/s00440-004-0361-z

[29] Kusuoka, S., Takahashi, H. and Tamura, Y. (2017). Recurrence and transience properties of multidimensional diffusion processes in selfsimilar and semi-selfsimilar random environments. *Electron. Commun. Probab.* **22** Paper No. 4, 11. MR3607799 https://doi.org/10.1214/16-ECP36

[30] Ling, C. and Zhao, G. (2019). Nonlocal elliptic equation in Hölder space and the martingale problem. Preprint arXiv:1907.00588.

[31] Martin, J. and Perkowski, N. (2019). Paracontrolled distributions on Bravais lattices and weak universality of the 2d parabolic Anderson model. *Ann. Inst. Henri Poincaré Probab. Stat.* **55** 2058–2110. MR4029148 https://doi.org/10.1214/18-AIHP942

[32] Mathieu, P. (1994). Zero white noise limit through Dirichlet forms, with application to diffusions in a random medium. *Probab. Theory Related Fields* **99** 549–580. MR1288070 https://doi.org/10.1007/BF01206232

[33] Mourrat, J.-C. and Weber, H. (2017). The dynamic Φ^4_3 model comes down from infinity. *Comm. Math. Phys.* **356** 673–753. MR3719541 https://doi.org/10.1007/s00220-017-2997-4

[34] Perkowski, N. (2014). Studies of robustness in stochastic analysis and mathematical finance Ph.D. thesis Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II. https://doi.org/10.18452/16895

[35] Peszat, S. and Zabczyk, J. (2007). *Stochastic Partial Differential Equations with Lévy Noise: An Evolution Equation Approach*. *Encyclopedia of Mathematics and Its Applications* **113**. Cambridge: Cambridge Univ. Press. MR2356959 https://doi.org/10.1017/CBO9780511721373

[36] Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. *Cambridge Studies in Advanced Mathematics* **68**. Cambridge: Cambridge Univ. Press. MR1739520

[37] Tanaka, H. (1987). Limit distributions for one-dimensional diffusion processes in self-similar random environments. In *Hydrodynamic Behavior and Interacting Particle Systems (Minneapolis, Minn., 1986)*. *IMA Vol. Math. Appl.* **9** 189–210. New York: Springer. MR0914995 https://doi.org/10.1007/978-1-4684-6347-7_15
[38] Veretennikov, A. Y. (1981). On strong solution and explicit formulas for solutions of stochastic integral equations. *Math. USSR, Sb.* **39** 387–403.
[39] Zvonkin, A.K. (1974). A transformation of the phase space of a diffusion process that will remove the drift. *Mat. Sb. (N.S.)* **93(135)** 129–149, 152. MR0336813

Bernoulli **28**(3), 2022, 1784–1809
https://doi.org/10.3150/21-BEJ1397

At the edge of a one-dimensional jellium

DJALIL CHAFAÏ[1,a], DAVID GARCÍA-ZELADA[2,b] and PAUL JUNG[3,c]

[1] *CEREMADE, Université Paris-Dauphine, PSL University, France.* [a] *djalil@chafai.net*
[2] *Aix-Marseille University, Institut de Mathématiques de Marseille (I2M), France.*
[b] *david.garcia-zelada@univ-amu.fr*
[3] *KAIST, Daejon, Korea.* [c] *pauljung@kaist.ac.kr*

We consider a one-dimensional classical Wigner jellium, not necessarily charge neutral, for which the electrons are allowed to exist beyond the support of the background charge. The model can be seen as a one-dimensional Coulomb gas in which the external field is generated by a smeared background on an interval. It is a true one-dimensional Coulomb gas and not a one-dimensional log-gas. The system exists if and only if the total background charge is greater than the number of electrons minus one. For various backgrounds, we show convergence to point processes, at the edge of the support of the background. In particular, this provides asymptotic analysis of the fluctuations of the right-most particle. Our analysis reveals that these fluctuations are not universal, in the sense that depending on the background, the tails range anywhere from exponential to Gaussian-like behavior, including for instance Tracy – Widom-like behavior. We also obtain a Rényi-type probabilistic representation for the order statistics of the particle system beyond the support of the background.

Keywords: Jellium; Coulomb gas; edge statistics; one-dimensional model

References

[1] Agarwal, S., Dhar, A., Kulkarni, M., Kundu, A., Majumdar, S.N., Mukamel, D. and Schehr, G. (2019). Harmonically confined particles with long-range repulsive interactions. *Phys. Rev. Lett.* **123** 100603, 6. MR4008477 https://doi.org/10.1103/PhysRevLett.123.100603

[2] Aizenman, M., Goldstein, S. and Lebowitz, J.L. (2001). Bounded fluctuations and translation symmetry breaking in one-dimensional particle systems. *J. Stat. Phys.* **103** 601–618. MR1827927 https://doi.org/10.1023/A:1010397401128

[3] Aizenman, M., Jansen, S. and Jung, P. (2010). Symmetry breaking in quasi-1D Coulomb systems. *Ann. Henri Poincaré* **11** 1453–1485. MR2769702 https://doi.org/10.1007/s00023-010-0067-y

[4] Aizenman, M. and Martin, P.A. (1980/81). Structure of Gibbs states of one-dimensional Coulomb systems. *Comm. Math. Phys.* **78** 99–116. MR0597033

[5] Baxter, R.J. (1963). Statistical mechanics of a one-dimensional Coulomb system with a uniform charge background. *Proc. Camb. Philos. Soc.* **59** 779–787. MR0155636

[6] Borot, G. and Nadal, C. (2012). Right tail asymptotic expansion of Tracy-Widom beta laws. *Random Matrices Theory Appl.* **1** 1250006, 23. MR2967965 https://doi.org/10.1142/S2010326312500062

[7] Brascamp, H.J. and Lieb, E.H. (2002). Some inequalities for Gaussian measures and the long-range order of the one-dimensional plasma. In *Inequalities: Selecta of Elliott H. Lieb* (M. Loss and M.B. Ruskai, eds.) 403–416. Berlin: Springer. https://doi.org/10.1007/978-3-642-55925-9_34

[8] Butez, R. and García-Zelada, D. (2018). Extremal particles of two-dimensional Coulomb gases and random polynomials on a positive background. Preprint.

[9] Butez, R., García-Zelada, D., Nishry, A. and Wennman, A. (2020). Universality of the outliers for two dimensional determinantal jellium. Work in progress.

[10] Chafaï, D., García-Zelada, D. and Jung, P. (2020). Macroscopic and edge behavior of a planar jellium. *J. Math. Phys.* **61** 033304, 18. MR4079754 https://doi.org/10.1063/1.5126724

[11] Chafaï, D., García-Zelada, D. and Jung, P. (2020). At the edge of a one-dimensional jellium (extended version). Available online arXiv:2012.04633v1.

[12] Chafaï, D. and Péché, S. (2014). A note on the second order universality at the edge of Coulomb gases on the plane. *J. Stat. Phys.* **156** 368–383. MR3215627 https://doi.org/10.1007/s10955-014-1007-x

[13] Dhar, A., Kundu, A., Majumdar, S.N., Sabhapandit, S. and Schehr, G. (2017). Exact extremal statistics in the classical 1D Coulomb gas. *Phys. Rev. Lett.* **119** 060601. https://doi.org/10.1103/PhysRevLett.119.060601

[14] Dhar, A., Kundu, A., Majumdar, S.N., Sabhapandit, S. and Schehr, G. (2018). Extreme statistics and index distribution in the classical 1d Coulomb gas. *J. Phys. A* **51** 295001, 32. MR3817495 https://doi.org/10.1088/1751-8121/aac75f

[15] Dumaz, L. and Virág, B. (2013). The right tail exponent of the Tracy-Widom β distribution. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 915–933. MR3127907 https://doi.org/10.1214/11-AIHP475

[16] García-Zelada, D. (2021). Edge fluctuations for random normal matrix ensembles. Preprint.

[17] Helms, L.L. (2014). *Potential Theory*, 2nd ed. *Universitext*. London: Springer. MR3308615 https://doi.org/10.1007/978-1-4471-6422-7

[18] Hough, J.B., Krishnapur, M., Peres, Y. and Virág, B. (2009). *Zeros of Gaussian Analytic Functions and Determinantal Point Processes*. *University Lecture Series* **51**. Providence, RI: Amer. Math. Soc. MR2552864 https://doi.org/10.1090/ulect/051

[19] Jansen, S. and Jung, P. (2014). Wigner crystallization in the quantum 1D jellium at all densities. *Comm. Math. Phys.* **331** 1133–1154. MR3248061 https://doi.org/10.1007/s00220-014-2032-y

[20] Kostlan, E. (1992). On the spectra of Gaussian matrices. *Linear Algebra Appl.* **162/164** 385–388. MR1148410 https://doi.org/10.1016/0024-3795(92)90386-O

[21] Kunz, H. (1974). The one-dimensional classical electron gas. *Ann. Physics* **85** 303–335. MR0426742 https://doi.org/10.1016/0003-4916(74)90413-8

[22] Kylänpää, I., Cavaliere, F., Ziani, Traverso, N., Sassetti, M. and Räsänen, E. (2016). Thermal effects on the Wigner localization and Friedel oscillations in many-electron nanowires. *Phys. Rev. B* **94** 115417. https://doi.org/10.1103/PhysRevB.94.115417

[23] Landkof, N.S. (1972). *Foundations of Modern Potential Theory*. *Die Grundlehren der Mathematischen Wissenschaften* **180**. New York: Springer. Translated from the Russian by A.P. Doohovskoy. MR0350027

[24] Lewin, M., Lieb, E.H. and Seiringer, R. (2018). Statistical mechanics of the uniform electron gas. *J. Éc. Polytech. Math.* **5** 79–116. MR3732693 https://doi.org/10.5802/jep.64

[25] Ramírez, J.A., Rider, B. and Virág, B. (2011). Beta ensembles, stochastic Airy spectrum, and a diffusion. *J. Amer. Math. Soc.* **24** 919–944. MR2813333 https://doi.org/10.1090/S0894-0347-2011-00703-0

[26] Rényi, A. (1953). On the theory of order statistics. *Acta Math. Acad. Sci. Hung.* **4** 191–231. MR0061792 https://doi.org/10.1007/BF02127580

[27] Wigner, E.P. (1934). On the interaction of electrons in metals. *Phys. Rev.* **46** 1002.

[28] Wigner, E.P. (1938). Effects of the electron interaction on the energy levels of electrons in metals. *Trans. Faraday Soc.* **34** 678–685. https://doi.org/10.1039/TF9383400678

Bernoulli **28**(3), 2022, 1810–1834
https://doi.org/10.3150/21-BEJ1400

Stochastic zeroth-order discretizations of Langevin diffusions for Bayesian inference

ABHISHEK ROY[1,a], LINGQING SHEN[2,c],
KRISHNAKUMAR BALASUBRAMANIAN[1,b] and SAEED GHADIMI[3,d]

[1]*Department of Statistics, University of California, Davis, Davis, CA 95616, USA.* [a]*abroy@ucdavis.edu,* [b]*kbala@ucdavis.edu*
[2]*Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213.* [c]*lingqins@andrew.cmu.edu*
[3]*Department of Management Sciences, University of Waterloo, Waterloo, ON N2L 3G1, Canada.* [d]*sghadimi@uwaterloo.ca*

Discretizations of Langevin diffusions provide a powerful method for sampling and Bayesian inference. However, such discretizations require evaluation of the gradient of the potential function. In several real-world scenarios, obtaining gradient evaluations might either be computationally expensive, or simply impossible. In this work, we propose and analyze stochastic zeroth-order sampling algorithms for discretizing overdamped and underdamped Langevin diffusions. Our approach is based on estimating the gradients, based on Gaussian Stein's identities, widely used in the stochastic optimization literature. We provide a comprehensive oracle complexity analysis – number noisy function evaluations to be made to obtain an ϵ-approximate sample in Wasserstein distance – of stochastic zeroth-order discretizations of both overdamped and underdamped Langevin diffusions, under various noise models. Our theoretical contributions extend the applicability of sampling algorithms to the noisy black-box settings arising in practice.

Keywords: Monte Carlo sampling; Langevin diffusion; stochastic MCMC; derivative-free or zeroth-order sampling; Bayesian inference

References

[1] Agarwal, A., Dekel, O. and Xiao, L. (2010). Optimal algorithms for online convex optimization with multi-point bandit feedback. In *Conference on Learning Theory* 28–40.
[2] Akhavan, A., Pontil, M. and Tsybakov, A. (2020). Exploiting higher order smoothness in derivative-free optimization and continuous bandits. *Adv. Neural Inf. Process. Syst.* **33**.
[3] Alenlöv, J., Doucet, A. and Lindsten, F. (2021). Pseudo-marginal Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **22** 141. MR4318497
[4] 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
[5] Roy, A., Shen, L. Balasubramanian, K., Ghadimi, S. (2022). Supplement to "Stochastic Zeroth-order Discretizations of Langevin Diffusions for Bayesian Inference." https://doi.org/10.3150/21-BEJ1400SUPP
[6] 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
[7] Audet, C. and Hare, W. (2017). *Derivative-Free and Blackbox Optimization*. *Springer Series in Operations Research and Financial Engineering*. Cham: Springer. MR3729457 https://doi.org/10.1007/978-3-319-68913-5
[8] Bach, F. and Perchet, V. (2016). Highly-smooth zero-th order online optimization. In *Conference on Learning Theory* 257–283.
[9] Bakry, D., Gentil, I. and Ledoux, M. (2014). *Analysis and Geometry of Markov Diffusion Operators*. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **348**. Cham: Springer. MR3155209 https://doi.org/10.1007/978-3-319-00227-9

[10] Balasubramanian, K. and Ghadimi, S. (2022). Zeroth-Order Nonconvex Stochastic Optimization: Handling Constraints, High Dimensionality, and Saddle Points. *Found. Comput. Math.* **22** 35–76. MR4376588 https://doi.org/10.1007/s10208-021-09499-8

[11] Barp, A., Briol, F.-X., Kennedy, A.D. and Girolami, M. (2018). Geometry and dynamics for Markov chain Monte Carlo. *Annu. Rev. Stat. Appl.* **5** 451–474. MR3774755 https://doi.org/10.1146/annurev-statistics-031017-100141

[12] Beaumont, M.A. (2003). Estimation of population growth or decline in genetically monitored populations. *Genetics* **164** 1139–1160. https://doi.org/10.1093/genetics/164.3.1139

[13] Bélisle, C.J.P., Romeijn, H.E. and Smith, R.L. (1993). Hit-and-run algorithms for generating multivariate distributions. *Math. Oper. Res.* **18** 255–266. MR1250117 https://doi.org/10.1287/moor.18.2.255

[14] Belloni, A., Liang, T., Narayanan, H. and Rakhlin, A. (2015). Escaping the local minima via simulated annealing: Optimization of approximately convex functions. In *Conference on Learning Theory*.

[15] Blum, J.R. (1954). Multidimensional stochastic approximation methods. *Ann. Math. Stat.* **25** 737–744. MR0065092 https://doi.org/10.1214/aoms/1177728659

[16] Brazitikos, S., Giannopoulos, A., Valettas, P. and Vritsiou, B.-H. (2014). *Geometry of Isotropic Convex Bodies*. *Mathematical Surveys and Monographs* **196**. Providence, RI: Amer. Math. Soc. MR3185453 https://doi.org/10.1090/surv/196

[17] Brosse, N., Durmus, A., Moulines, É. and Pereyra, M. (2017). Sampling from a log-concave distribution with compact support with proximal Langevin Monte Carlo. In *Conference on Learning Theory*.

[18] Bubeck, S., Eldan, R. and Lehec, J. (2018). Sampling from a log-concave distribution with projected Langevin Monte Carlo. *Discrete Comput. Geom.* **59** 757–783. MR3802303 https://doi.org/10.1007/s00454-018-9992-1

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

[20] Cao, Y., Lu, J. and Wang, L. (2021). Complexity of randomized algorithms for underdamped Langevin dynamics. *Commun. Math. Sci.* **19** 1827–1853. MR4312746 https://doi.org/10.4310/CMS.2021.v19.n7.a4

[21] Chatterji, N.S., Bartlett, P.L. and Long, P.M. (2020). Oracle lower bounds for stochastic gradient sampling algorithms. arXiv preprint. Available at arXiv:2002.00291.

[22] Cheng, X., Chatterji, N.S., Abbasi-Yadkori, Y., Bartlett, P.L. and Sharp, M.I.J. (2018). Convergence Rates for Langevin Dynamics in the Nonconvex Setting. arXiv preprint. Available at arXiv:1805.01648.

[23] Cheng, X., Chatterji, N.S., Bartlett, P.L. and Jordan, M.I. (2018). Underdamped Langevin mcmc: A non-asymptotic analysis. In *Conference on Learning Theory* 300–323. PMLR.

[24] Conn, A.R., Scheinberg, K. and Vicente, L.N. (2009). *Introduction to Derivative-Free Optimization*. *MPS/SIAM Series on Optimization* **8**. Philadelphia, PA: SIAM. MR2487816 https://doi.org/10.1137/1.9780898718768

[25] Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. *Adv. Neural Inf. Process. Syst.* **26** 2292–2300.

[26] Dalalyan, A.S. (2017). Theoretical guarantees for approximate sampling from smooth and log-concave densities. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 651–676. MR3641401 https://doi.org/10.1111/rssb.12183

[27] Dalalyan, A.S. (2017). Theoretical guarantees for approximate sampling from smooth and log-concave densities. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 651–676. MR3641401 https://doi.org/10.1111/rssb.12183

[28] Dalalyan, A.S. and Karagulyan, A. (2019). User-friendly guarantees for the Langevin Monte Carlo with inaccurate gradient. *Stochastic Process. Appl.* **129** 5278–5311. MR4025705 https://doi.org/10.1016/j.spa.2019.02.016

[29] Dalalyan, A.S. and Riou-Durand, L. (2020). On sampling from a log-concave density using kinetic Langevin diffusions. *Bernoulli* **26** 1956–1988. MR4091098 https://doi.org/10.3150/19-BEJ1178

[30] Dippon, J. (2003). Accelerated randomized stochastic optimization. *Ann. Statist.* **31** 1260–1281. MR2001650 https://doi.org/10.1214/aos/1059655913

[31] Duchi, J.C., Jordan, M.I., Wainwright, M.J. and Wibisono, A. (2015). Optimal rates for zero-order convex optimization: The power of two function evaluations. *IEEE Trans. Inf. Theory* **61** 2788–2806. MR3342307 https://doi.org/10.1109/TIT.2015.2409256

[32] Dunson, D.B. and Johndrow, J.E. (2020). The Hastings algorithm at fifty. *Biometrika* **107** 1–23. MR4064137 https://doi.org/10.1093/biomet/asz066

[33] Durmus, A. and Moulines, É. (2017). Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *Ann. Appl. Probab.* **27** 1551–1587. MR3678479 https://doi.org/10.1214/16-AAP1238

[34] Durmus, A. and Moulines, É. (2019). High-dimensional Bayesian inference via the unadjusted Langevin algorithm. *Bernoulli* **25** 2854–2882. MR4003567 https://doi.org/10.3150/18-BEJ1073

[35] Dwivedi, R., Chen, Y., Wainwright, M.J. and Yu, B. (2019). Log-concave sampling: Metropolis-Hastings algorithms are fast. *J. Mach. Learn. Res.* **20** 183. MR4048994

[36] Eberle, A., Guillin, A. and Zimmer, R. (2019). Couplings and quantitative contraction rates for Langevin dynamics. *Ann. Probab.* **47** 1982–2010. MR3980913 https://doi.org/10.1214/18-AOP1299

[37] Erdogdu, M.A. and Hosseinzadeh, R. (2020). On the convergence of langevin monte carlo: The interplay between tail growth and smoothness. arXiv preprint. Available at arXiv:2005.13097.

[38] Gasnikov, A.V., Krymova, E.A., Lagunovskaya, A.A., Usmanova, I.N. and Fedorenko, F.A. (2017). Stochastic online optimization. Single-point and multi-point non-linear multi-armed bandits. Convex and strongly-convex case. *Autom. Remote Control* **78** 224–234.

[39] George, E.I. and McCulloch, R.E. (1997). Approaches for Bayesian variable selection. *Statist. Sinica* 339–373.

[40] Ghadimi, S. and Lan, G. (2013). Stochastic first- and zeroth-order methods for nonconvex stochastic programming. *SIAM J. Optim.* **23** 2341–2368. MR3134439 https://doi.org/10.1137/120880811

[41] 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

[42] Golightly, A. and Wilkinson, D.J. (2011). Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. *Interface Focus* **1** 807–820.

[43] Gorham, J. and Mackey, L. (2015). Measuring sample quality with Stein's method. In *Advances in Neural Information Processing Systems* 226–234.

[44] Griewank, A. and Walther, A. (2008). *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, 2nd ed. Philadelphia, PA: SIAM. MR2454953 https://doi.org/10.1137/1.9780898717761

[45] He, Y., Balasubramanian, K. and Erdogdu, M.A. (2020). On the ergodicity, bias and asymptotic normality of randomized midpoint sampling method. *Adv. Neural Inf. Process. Syst.* **33**.

[46] Ishwaran, H. and Rao, J.S. (2005). Spike and slab variable selection: Frequentist and Bayesian strategies. *Ann. Statist.* **33** 730–773. MR2163158 https://doi.org/10.1214/009053604000001147

[47] Jamieson, K.G., Nowak, R. and Recht, B. (2012). Query complexity of derivative-free optimization. *Adv. Neural Inf. Process. Syst.* **25** 2672–2680.

[48] Jensen, S. and Foster, D. (2014). A level-set hit-and-run sampler for quasi-concave distributions. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics* (S. Kaski and J. Corander, eds.). *Proceedings of Machine Learning Research* **33** 439–447. Reykjavik, Iceland.

[49] Kannan, R., Lovász, L. and Simonovits, M. (1995). Isoperimetric problems for convex bodies and a localization lemma. *Discrete Comput. Geom.* **13** 541–559. MR1318794 https://doi.org/10.1007/BF02574061

[50] Kiefer, J. and Wolfowitz, J. (1952). Stochastic estimation of the maximum of a regression function. *Ann. Math. Stat.* **23** 462–466. MR0050243 https://doi.org/10.1214/aoms/1177729392

[51] Knape, J. and De Valpine, P. (2012). Fitting complex population models by combining particle filters with Markov chain Monte Carlo. *Ecology* **93** 256–263.

[52] Korpelevič, G.M. (1976). An extragradient method for finding saddle points and for other problems. *Èkon. Mat. Metody* **12** 747–756. MR0451121

[53] Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A. and Blei, D.M. (2017). Automatic differentiation variational inference. *J. Mach. Learn. Res.* **18** 14. MR3634881

[54] Larson, J., Menickelly, M. and Wild, S.M. (2019). Derivative-free optimization methods. *Acta Numer.* **28** 287–404. MR3963507 https://doi.org/10.1017/s0962492919000060

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

[56] Ledoux, M. (2001). Logarithmic Sobolev inequalities for unbounded spin systems revisited. In *Séminaire de Probabilités, XXXV. Lecture Notes in Math.* **1755** 167–194. Berlin: Springer. MR1837286 https://doi.org/10.1007/978-3-540-44671-2_13
[57] Li, J. (2020). Krishnakumar Balasubramanian, and Shiqian Ma. Zeroth-order optimization on riemannian manifolds. arXiv preprint. Available at arXiv:2003.11238.
[58] Lovász, L. and Simonovits, M. (1990). The mixing rate of Markov chains, an isoperimetric inequality, and computing the volume. In *Foundations of Computer Science, 1990. Proceedings., 31st Annual Symposium on*.
[59] Lovász, L. and Vempala, S. (2007). The geometry of logconcave functions and sampling algorithms. *Random Structures Algorithms* **30** 307–358. MR2309621 https://doi.org/10.1002/rsa.20135
[60] Martin, G.M., Frazier, D.T. and Robert, C.P. (2020). Computing bayes: Bayesian computation from 1763 to the 21st century. arXiv preprint. Available at arXiv:2004.06425.
[61] Mengersen, K.L. and Tweedie, R.L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24** 101–121. MR1389882 https://doi.org/10.1214/aos/1033066201
[62] Mokkadem, A. and Pelletier, M. (2007). A companion for the Kiefer-Wolfowitz-Blum stochastic approximation algorithm. *Ann. Statist.* **35** 1749–1772. MR2351104 https://doi.org/10.1214/009053606000001451
[63] Nemirovsky, A.S. and Yudin, D.B. (1983). *Problem Complexity and Method Efficiency in Optimization*. *Wiley-Interscience Series in Discrete Mathematics*. New York: Wiley. MR0702836
[64] Nesterov, Y. and Spokoiny, V. (2017). Random gradient-free minimization of convex functions. *Found. Comput. Math.* **17** 527–566. MR3627456 https://doi.org/10.1007/s10208-015-9296-2
[65] Peyré, G. and Cuturi, M. (2019). Computational optimal transport: With applications to data science. *Found. Trends Mach. Learn.* **11** 355–607.
[66] Radford, M. and Neal (2011). MCMC using Hamiltonian dynamics. In Handbook of Markov Chain Monte Carlo 2 2. 11.
[67] Ranganath, R., Gerrish, S. and Blei, D. (2014). Black box variational inference. In *Artificial Intelligence and Statistics* 814–822.
[68] Ranganath, R., Tang, L., Charlin, L. and Blei, D. (2015). Deep exponential families. In *Artificial Intelligence and Statistics* 762–771.
[69] Roberts, G.O. and Rosenthal, J.S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 255–268. MR1625691 https://doi.org/10.1111/1467-9868.00123
[70] Sanz-Serna, J.M. (1992). Symplectic integrators for Hamiltonian problems: An overview. In *Acta Numerica, 1992*. *Acta Numer.* 243–286. Cambridge: Cambridge Univ. Press. MR1165727 https://doi.org/10.1017/s0962492900002282
[71] Shamir, O. (2013). On the complexity of bandit and derivative-free stochastic convex optimization. In *Conference on Learning Theory* 3–24.
[72] Shen, R. and Tat Lee, Y. (2019). The randomized midpoint method for log-concave sampling. In *Advances in Neural Information Processing Systems* 2100–2111.
[73] Sherlock, C., Thiery, A.H., Roberts, G.O. and Rosenthal, J.S. (2015). On the efficiency of pseudo-marginal random walk Metropolis algorithms. *Ann. Statist.* **43** 238–275. MR3285606 https://doi.org/10.1214/14-AOS1278
[74] Spall, J.C. (2003). *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. *Wiley-Interscience Series in Discrete Mathematics and Optimization*. Hoboken, NJ: Wiley Interscience. MR1968388 https://doi.org/10.1002/0471722138
[75] Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory. the Regents of the* Univ. California.
[76] Stramer, O. and Tweedie, R.L. (1999). Langevin-type models. I. Diffusions with given stationary distributions and their discretizations. *Methodol. Comput. Appl. Probab.* **1** 283–306. MR1730651 https://doi.org/10.1023/A:1010086427957
[77] Stramer, O. and Tweedie, R.L. (1999). Langevin-type models. II. Self-targeting candidates for MCMC algorithms. *Methodol. Comput. Appl. Probab.* **1** 307–328. MR1730652 https://doi.org/10.1023/A:1010090512027
[78] Vempala, S. and Wibisono, A. (2019). Rapid convergence of the unadjusted Langevin algorithm: Isoperimetry suffices. In *Advances in Neural Information Processing Systems* 8094–8106.

[79] Villani, C. (2009). *Hypocoercivity. Number 949-951*. American Mathematical Soc.

[80] Zhang, Y., Liang, P. and Charikar, M. (2017). A hitting time analysis of stochastic gradient Langevin dynamics. In *Conference on Learning Theory* 1980–2022.

Bernoulli **28**(3), 2022, 1835–1861
https://doi.org/10.3150/21-BEJ1401

The asymptotic distribution of the MLE in high-dimensional logistic models: Arbitrary covariance

QIAN ZHAO[1,a], PRAGYA SUR[2,b] and EMMANUEL J. CANDÈS[3,c]

[1]*Department of Statistics, Stanford University, Sequoia Hall, 390 Jane Stanford Way, Stanford, CA 94305-4020, USA.* [a]*qzhao1@stanford.edu*
[2]*Department of Statistics, Harvard University, Science Center 712, One Oxford Street, Cambridge, MA 02138, USA.* [b]*pragya@fas.harvard.edu*
[3]*Department of Mathematics and of Statistics, Stanford University, Sequoia Hall, 390 Jane Stanford Way, Stanford, CA 94305-4020, USA.* [c]*candes@stanford.edu*

We study the distribution of the maximum likelihood estimate (MLE) in high-dimensional logistic models, where covariates are Gaussian with an arbitrary covariance structure. We prove that in the limit of large problems holding the ratio between the number p of covariates and the sample size n constant, every finite list of MLE coordinates follows a multivariate normal distribution. Concretely, the jth coordinate $\hat{\beta}_j$ of the MLE is asymptotically normally distributed with mean $\alpha_\star \beta_j$ and standard deviation $\sigma_\star/\tau_j$; here, β_j is the value of the true regression coefficient, and τ_j the standard deviation of the jth predictor conditional on all the others. The numerical parameters $\alpha_\star > 1$ and $\sigma_\star$ only depend upon the problem dimensionality p/n and the overall signal strength, and can be accurately estimated. Our results imply that the MLE's magnitude is biased upwards and that the MLE's standard deviation is greater than that predicted by classical theory. We present a series of experiments on simulated and real data showing excellent agreement with the theory.

Keywords: High-dimensional inference; logistic regression; maximum likelihood estimation; Gaussian covariates

References

[1] Anastasiou, A. and Reinert, G. (2020). Bounds for the asymptotic distribution of the likelihood ratio. *Ann. Appl. Probab.* **30** 608–643. MR4108117 https://doi.org/10.1214/19-AAP1510

[2] Àngela, N., Francisco, M. and Antoni, E. (2015). Modeling wine preferences from physicochemical properties using fuzzy techniques. In *Proceedings of the 5th International Conference on Simulation and Modeling Methodologies, Technologies and Applications. SciTePress* 501–507.

[3] Barbier, J., Krzakala, F., Macris, N., Miolane, L. and Zdeborová, L. (2019). Optimal errors and phase transitions in high-dimensional generalized linear models. *Proc. Natl. Acad. Sci. USA* **116** 5451–5460. MR3939767 https://doi.org/10.1073/pnas.1802705116

[4] Bayati, M. and Montanari, A. (2011). The lasso risk for Gaussian matrices. *IEEE Trans. Inf. Theory* **58** 1997–2017.

[5] Bellec, P.C. and Zhang, C.-H. (2019). Second order Poincaré inequalities and de-biasing arbitrary convex regularizers when $p/n \to \gamma$. arXiv preprint. Available at arXiv:1912.11943.

[6] Bühlmann, P. and van de Geer, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. *Springer Series in Statistics*. Heidelberg: Springer. MR2807761 https://doi.org/10.1007/978-3-642-20192-9

[7] Cai, T.T. and Guo, Z. (2020). Semisupervised inference for explained variance in high dimensional linear regression and its applications. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 391–419. MR4084169

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

[9] Celentano, M., Montanari, A. and Wei, Y. (2020). The Lasso with general Gaussian designs with applications to hypothesis testing. arXiv preprint. Available at arXiv:2007.13716.

[10] Chatterjee, S. (2009). Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields* **143** 1–40. MR2449121 https://doi.org/10.1007/s00440-007-0118-6

[11] Donoho, D., Maleki, A. and Montanari, A. (2009). Message-passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci. USA* **106** 18914–18919.

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

[14] 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.

[15] Fan, Y., Demirkaya, E. and Lv, J. (2019). Nonuniformity of p-values can occur early in diverging dimensions. *J. Mach. Learn. Res.* **20** 77. MR3960931

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

[17] He, X. and Shao, Q.-M. (2000). On parameters of increasing dimensions. *J. Multivariate Anal.* **73** 120–135. MR1766124 https://doi.org/10.1006/jmva.1999.1873

[18] Janková, J. and van de Geer, S. (2021). De-biased sparse PCA: Inference for eigenstructure of large covariance matrices. *IEEE Trans. Inf. Theory* **67** 2507–2527. MR4282369 https://doi.org/10.1109/TIT.2021.3059765

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

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

[21] Kimmel, G. and Shamir, R. (2005). A block-free hidden Markov model for genotypes and its application to disease association. *J. Comput. Biol.* **12** 1243–1260.

[22] Ma, R., Cai, T.T. and Li, H. (2021). Global and simultaneous hypothesis testing for high-dimensional logistic regression models. *J. Amer. Statist. Assoc.* **116** 984–998. MR4270038 https://doi.org/10.1080/01621459.2019.1699421

[23] McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, 2nd ed. London; New York: CRC Press.

[24] Mézard, M., Parisi, G. and Virasoro, M.A. (1987). *Spin Glass Theory and Beyond. World Scientific Lecture Notes in Physics* **9**. Teaneck, NJ: World Scientific Co., Inc. MR1026102

[25] Montanari, A. (2018). Mean field asymptotics in high-dimensional statistics: From exact results to efficient algorithms. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. IV. Invited Lectures* 2973–2994. Hackensack, NJ: World Sci. Publ. MR3966519

[26] Portnoy, S. (1984). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I. Consistency. *Ann. Statist.* **12** 1298–1309. MR0760690 https://doi.org/10.1214/aos/1176346793

[27] 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

[28] Portnoy, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Ann. Statist.* **16** 356–366. MR0924876 https://doi.org/10.1214/aos/1176350710

[29] Rangan, S. (2011). Generalized approximate message passing for estimation with random linear mixing. In *2011 IEEE International Symposium on Information Theory Proceedings*.

[30] Rastas, P., Koivisto, M., Mannila, H. and Ukkonen, E. (2005). A hidden Markov technique for haplotype reconstruction. In *Algorithms in Bioinformatics*. *Lecture Notes in Computer Science* **3692** 140–151. Berlin: Springer. MR2226832 https://doi.org/10.1007/11557067_12

[31] Sabatti, C., Candès, E.J. and Sesia, M. (2018). Gene hunting with hidden Markov model knockoffs. *Biometrika* **106** 1–18.

[32] Salehi, F., Abbasi, E. and Hassibi, B. (2019). The impact of regularization on high-dimensional logistic regression. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*.

[33] Scheet, P. and Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* **78** 629–644.

[34] Sesia, M. (2019). Using SNPknock with genetic data. https://msesia.github.io/snpknock/articles/genotypes.html, [online; accessed 30-October-2019].

[35] Stojnic, M. (2013). A framework to characterize performance of Lasso algorithms. arXiv preprint. Available at arXiv:1303.7291.

[36] Sur, P. (2019). *A Modern Maximum Likelihood Theory for High-Dimensional Logistic Regression*. Ann Arbor, MI: ProQuest LLC. Thesis (Ph.D.)—Stanford University. Available at purl.stanford.edu/jw604jq1260. MR4197622

[37] 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

[38] 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

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

[40] Thrampoulidis, C., Oymak, S. and Hassibi, B. (2015). Regularized linear regression: A precise analysis of the estimation error. *Proc. Mach. Learn. Res.* **40** 1683–1709.

[41] 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

[42] van der Vaart, A.W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. MR1652247 https://doi.org/10.1017/CBO9780511802256

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

[44] Wilks, S.S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Statist.* **9** 60–62.

[45] 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

[46] Zhao, Q. (2020). Glmhd: Statistical inference in high-dimensional binary regression. R package version 0.0.0.9000. https://github.com/zq00/glmhd.

[47] Zhao, Q., Sur, P. and Candès, E.J. (2022). Supplement to "The Asymptotic Distribution of the MLE in High-Dimensional Logistic Models: Arbitrary Covariance." https://doi.org/10.3150/21-BEJ1401SUPPA, https://doi.org/10.3150/21-BEJ1401SUPPB

Bernoulli **28**(3), 2022, 1862–1891
https://doi.org/10.3150/21-BEJ1402

Conditional variance estimator for sufficient dimension reduction

LUKAS FERTL[a] and EFSTATHIA BURA[b]

Institute of Statistics and Mathematical Methods in Economics, Faculty of Mathematics and Geoinformation, TU Wien, Vienna, Austria. [a]*lukas.fertl@tuwien.ac.at,* [b]*efstathia.bura@tuwien.ac.at*

Conditional Variance Estimation (CVE) is a novel sufficient dimension reduction (SDR) method for additive error regressions with continuous predictors and link function. It operates under the assumption that the predictors can be replaced by a lower dimensional projection without loss of information. Conditional Variance Estimation is fully data driven, does not require the restrictive linearity and constant variance conditions, and is not based on inverse regression as the majority of moment and likelihood based sufficient dimension reduction methods. CVE is shown to be consistent and its objective function to be uniformly convergent. CVE outperforms the mean average variance estimation, (MAVE), its main competitor, in several simulation settings, remains on par under others, while it always outperforms inverse regression based linear SDR methods, such as Sliced Inverse Regression.

Keywords: Regression; nonparametric; mean subspace; minimum average variance estimation; dimension reduction

References

[1] Absil, P.-A., Mahony, R. and Sepulchre, R. (2008). *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ: Princeton Univ. Press. MR2364186 https://doi.org/10.1515/9781400830244

[2] Adragni, K.P. and Cook, R.D. (2009). Sufficient dimension reduction and prediction in regression. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **367** 4385–4405. MR2546393 https://doi.org/10.1098/rsta.2009.0110

[3] Amemiya, T. (1985). *Advanced Econometrics*. Cambridge: Harvard University Press.

[4] Bernstein, S.N. (1927). *Theory of Probability*. Moscow.

[5] Boothby, W.M. (2002). *An Introduction to Differentiable Manifolds and Riemannian Geometry*. San Diego: Academic Press.

[6] Bura, E., Duarte, S. and Forzani, L. (2016). Sufficient reductions in regressions with exponential family inverse predictors. *J. Amer. Statist. Assoc.* **111** 1313–1329. MR3561952 https://doi.org/10.1080/01621459.2015.1093944

[7] Bura, E. and Forzani, L. (2015). Sufficient reductions in regressions with elliptically contoured inverse predictors. *J. Amer. Statist. Assoc.* **110** 420–434. MR3338513 https://doi.org/10.1080/01621459.2014.914440

[8] Chikuse, Y. (1994). Invariant measures on Stiefel manifolds with applications to multivariate analysis. In *Multivariate Analysis and Its Applications (Hong Kong, 1992). Institute of Mathematical Statistics Lecture Notes—Monograph Series* **24** 177–193. Hayward, CA: IMS. MR1479466 https://doi.org/10.1214/lnms/1215463795

[9] Chikuse, Y. (2003). *Statistics on Special Manifolds*. *Lecture Notes in Statistics* **174**. New York: Springer. MR1960435 https://doi.org/10.1007/978-0-387-21540-2

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

[11] Cook, R.D. (2007). Fisher lecture: Dimension reduction in regression. *Statist. Sci.* **22** 1–26. MR2408655 https://doi.org/10.1214/088342306000000682

[12] Cook, R.D. and Forzani, L. (2008). Principal fitted components for dimension reduction in regression. *Statist. Sci.* **23** 485–501. MR2530547 https://doi.org/10.1214/08-STS275

[13] Cook, R.D. and Forzani, L. (2009). Likelihood-based sufficient dimension reduction. *J. Amer. Statist. Assoc.* **104** 197–208. MR2504373 https://doi.org/10.1198/jasa.2009.0106

[14] Cook, R.D. and Li, B. (2002). Dimension reduction for conditional mean in regression. *Ann. Statist.* **30** 455–474. MR1902895 https://doi.org/10.1214/aos/1021379861

[15] Cook, R.D. and Li, B. (2004). Determining the dimension of iterative Hessian transformation. *Ann. Statist.* **32** 2501–2531. MR2153993 https://doi.org/10.1214/009053604000000661

[16] Cook, R.D. and Weisberg, S. (1991). Sliced inverse regression for dimension reduction: Comment. *J. Amer. Statist. Assoc.* **86** 328–332.

[17] Faden, A.M. (1985). The existence of regular conditional probabilities: Necessary and sufficient conditions. *Ann. Probab.* **13** 288–298. MR0770643

[18] Fertl, L. and Bura, E. (2022). Supplement to "Conditional variance estimator for sufficient dimension reduction." https://doi.org/10.3150/21-BEJ1402SUPP

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

[20] Hang, W. and Xia, Y. (2019). MAVE: Methods for Dimension Reduction. R package version 1.3.10.

[21] Hansen, B.E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* **24** 726–748. MR2409261 https://doi.org/10.1017/S0266466608080304

[22] Heuser, H. (1995). *Analysis 2, 9 Auflage*. Leipzig: Teubner.

[23] James, I.M. (1976). *The Topology of Stiefel Manifolds*. *London Mathematical Society Lecture Note Series, No. 24*. Cambridge-Melbourne: Cambridge Univ. Press. MR0431239

[24] Kapla, D. and Fertl, L. (2021). CVarE: Conditional Variance Estimator for Sufficient Dimension Reduction. R package version 1.1.

[25] Karr, A.F. (1993). *Probability*. *Springer Texts in Statistics*. New York: Springer. MR1231974 https://doi.org/10.1007/978-1-4612-0891-4

[26] Leao, D. Jr., Fragoso, M. and Ruffino, P. (2004). Regular conditional probability, disintegration of probability and Radon spaces. *Proyecciones* **23** 15–29. MR2060837 https://doi.org/10.4067/S0716-09172004000100002

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

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

[29] Li, K.-C. (1992). On principal Hessian directions for data visualization and dimension reduction: Another application of Stein's lemma. *J. Amer. Statist. Assoc.* **87** 1025–1039. MR1209564

[30] Ma, Y. and Zhu, L. (2013). A review on dimension reduction. *Int. Stat. Rev.* **81** 134–150. MR3047506 https://doi.org/10.1111/j.1751-5823.2012.00182.x

[31] Nadarajah, S. (2005). A generalized normal distribution. *J. Appl. Stat.* **32** 685–694. MR2119411 https://doi.org/10.1080/02664760500079464

[32] Nocedal, J. and Wright, S.J. (2006). *Numerical Optimization*, 2nd ed. *Springer Series in Operations Research and Financial Engineering*. New York: Springer. MR2244940

[33] Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Stat.* **33** 1065–1076. MR0143282 https://doi.org/10.1214/aoms/1177704472

[34] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. *Monographs on Statistics and Applied Probability*. London: CRC Press. MR0848134 https://doi.org/10.1007/978-1-4899-3324-9

[35] Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *J. Roy. Statist. Soc. Ser. B* **36** 111–147. MR0356377

[36] Wen, Z. and Yin, W. (2013). A feasible method for optimization with orthogonality constraints. *Math. Program.* **142** 397–434. MR3127080 https://doi.org/10.1007/s10107-012-0584-1

[37] Xia, Y. (2008). A multiple-index model and dimension reduction. *J. Amer. Statist. Assoc.* **103** 1631–1640. MR2504209 https://doi.org/10.1198/016214508000000805

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

[39] Yin, X., Li, B. and Cook, R.D. (2008). Successive direction extraction for estimating the central subspace in a multiple-index regression. *J. Multivariate Anal.* **99** 1733–1757. MR2444817 https://doi.org/10.1016/j.jmva.2008.01.006

Bernoulli **28**(3), 2022, 1892–1911
https://doi.org/10.3150/21-BEJ1403

An optimal uniform concentration inequality for discrete entropies on finite alphabets in the high-dimensional setting

YUNPENG ZHAO[a]

School of Mathematical and Natural Sciences, Arizona State University, AZ, 85306. [a]*yunpeng.zhao@asu.edu*

We prove an exponential decay concentration inequality to bound the tail probability of the difference between the log-likelihood of discrete random variables on a finite alphabet and the negative entropy. The concentration bound we derive holds uniformly over all parameter values. The new result improves the convergence rate in an earlier result of Zhao (2020), from $(K^2 \log K)/n = o(1)$ to $(\log K)^2/n = o(1)$, where n is the sample size and K is the size of the alphabet. We further prove that the rate $(\log K)^2/n = o(1)$ is optimal. The result is extended to misspecified log-likelihoods for grouped random variables. We give applications of the new result in information theory.

Keywords: Concentration inequality; log-likelihood; entropy; typical set; source coding theorem; non-convex optimization

References

[1] Baccetti, V. and Visser, M. (2013). Infinite Shannon entropy. *J. Stat. Mech. Theory Exp.* **4** P04010, 12. MR3077811 https://doi.org/10.1088/1742-5468/2013/04/p04010
[2] Bernstein, S. (1924). On a modification of Chebyshev's inequality and of the error formula of Laplace. *Ann. Sci. Inst. Savantes Ukraine, Sect. Math.* 38–49.
[3] Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford: Oxford Univ. Press. MR3185193 https://doi.org/10.1093/acprof:oso/9780199535255.001.0001
[4] Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge: Cambridge Univ. Press. MR2061575 https://doi.org/10.1017/CBO9780511804441
[5] Breiman, L. (1957). The individual ergodic theorem of information theory. *Ann. Math. Stat.* **28** 809–811. MR0092710 https://doi.org/10.1214/aoms/1177706899
[6] Choi, D.S., Wolfe, P.J. and Airoldi, E.M. (2012). Stochastic blockmodels with a growing number of classes. *Biometrika* **99** 273–284. MR2931253 https://doi.org/10.1093/biomet/asr053
[7] Cover, T.M. and Thomas, J.A. (2006). *Elements of Information Theory*, 2nd ed. Hoboken, NJ: Wiley-Interscience. MR2239987
[8] Csiszár, I. and Körner, J. (2011). *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge: Cambridge Univ. Press. MR2839250 https://doi.org/10.1017/CBO9780511921889
[9] Dubhashi, D.P. and Panconesi, A. (2009). *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge: Cambridge Univ. Press. MR2547432 https://doi.org/10.1017/CBO9780511581274
[10] Durrett, R. (2019). *Probability—Theory and Examples*. *Cambridge Series in Statistical and Probabilistic Mathematics* **49**. Cambridge: Cambridge Univ. Press. MR3930614 https://doi.org/10.1017/9781108591034
[11] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. MR0144363
[12] Ledoux, M. (2001). *The Concentration of Measure Phenomenon*. *Mathematical Surveys and Monographs* **89**. Providence, RI: Amer. Math. Soc. MR1849347 https://doi.org/10.1090/surv/089
[13] MacKay, D.J.C. (2003). *Information Theory, Inference and Learning Algorithms*. New York: Cambridge Univ. Press. MR2012999

1350-7265

[14] McMillan, B. (1953). The basic theorems of information theory. *Ann. Math. Stat.* **24** 196–219. MR0055621 https://doi.org/10.1214/aoms/1177729028

[15] Paul, S. and Chen, Y. (2016). Consistent community detection in multi-relational data through restricted multi-layer stochastic blockmodel. *Electron. J. Stat.* **10** 3807–3870. MR3579677 https://doi.org/10.1214/16-EJS1211

[16] Raginsky, M. and Sason, I. (2018). *Concentration of Measure Inequalities in Information Theory, Communications, and Coding*, 3rd ed. Hanover: Now Publishers.

[17] Shannon, C.E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* **27** 379–423, 623–656. MR0026286 https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

[18] Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. *Cambridge Series in Statistical and Probabilistic Mathematics* **47**. Cambridge: Cambridge Univ. Press. MR3837109 https://doi.org/10.1017/9781108231596

[19] Vershynin, R. (2020). Concentration inequalities for random tensors. *Bernoulli* **26** 3139–3162. MR4140540 https://doi.org/10.3150/20-BEJ1218

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

[21] Zhao, Y. (2020). A note on new Bernstein-type inequalities for the log-likelihood function of Bernoulli variables. *Statist. Probab. Lett.* **163** 108779, 5. MR4083856 https://doi.org/10.1016/j.spl.2020.108779

[22] Zhao, Y., Bickel, P. and Weko, C. (2020). Identifiability and consistency of network inference using the hub model and variants: A restricted class of bernoulli mixture models. Preprint. Available at arXiv:2004.09709.

Bernoulli **28**(3), 2022, 1912–1934
https://doi.org/10.3150/21-BEJ1404

Multivariate ρ-quantiles: A spatial approach

DIMITRI KONEN[1] and DAVY PAINDAVEINE[1,2,a]

[1] *Université libre de Bruxelles, ECARES and Département de Mathématique, Avenue F.D. Roosevelt, 50, ECARES, CP114/04, B-1050, Brussels, Belgium.* [a]*Davy.Paindaveine@ulb.be*
[2] *Université Toulouse Capitole, Toulouse School of Economics, 1, Esplanade de l'Université, 31080 Toulouse Cedex 06, France*

By substituting an L_p loss function for the L_1 loss function in the optimization problem defining quantiles, one obtains L_p-quantiles that, as shown recently, dominate their classical L_1-counterparts in financial risk assessment. In this work, we propose a concept of multivariate L_p-quantiles generalizing the spatial (L_1-)quantiles introduced by Probal Chaudhuri (*J. Amer. Statist. Assoc.* **91** (1996) 862–872). Rather than restricting to power loss functions, we actually allow for a large class of convex loss functions ρ. We carefully study existence and uniqueness of the resulting ρ-quantiles, both for a general probability measure over $\mathbb{R}^d$ and for a spherically symmetric one. Interestingly, the results crucially depend on ρ and on the nature of the underlying probability measure. Building on an investigation of the differentiability properties of the objective function defining ρ-quantiles, we introduce a companion concept of spatial ρ-depth, that generalizes the classical spatial depth. We study extreme ρ-quantiles and show in particular that extreme L_p-quantiles behave in fundamentally different ways for $p \leq 2$ and $p > 2$. Finally, we establish Bahadur representation results for sample ρ-quantiles and derive their asymptotic distributions. Throughout, we impose only very mild assumptions on the underlying probability measure, and in particular we never assume absolute continuity with respect to the Lebesgue measure.

Keywords: Bahadur representation results; convex objective functions; M-estimation; multivariate quantiles; spatial depth; spatial quantiles

References

Azzalini, A. (2014). *The Skew-normal and Related Families*. *Institute of Mathematical Statistics (IMS) Monographs* **3**. Cambridge: Cambridge Univ. Press. MR3468021

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge: Cambridge Univ. Press. MR2061575 https://doi.org/10.1017/CBO9780511804441

Breckling, J. and Chambers, R. (1988). *M*-quantiles. *Biometrika* **75** 761–771. MR0995118 https://doi.org/10.2307/2336317

Brown, B.M. (1983). Statistical uses of the spatial median. *J. Roy. Statist. Soc. Ser. B* **45** 25–30. MR0701072

Cardot, H., Cénac, P. and Godichon-Baggioni, A. (2017). Online estimation of the geometric median in Hilbert spaces: Nonasymptotic confidence balls. *Ann. Statist.* **45** 591–614. MR3650394 https://doi.org/10.1214/16-AOS1460

Cardot, H., Cénac, P. and Zitt, P.-A. (2013). Efficient and fast estimation of the geometric median in Hilbert spaces with an averaged stochastic gradient algorithm. *Bernoulli* **19** 18–43. MR3019484 https://doi.org/10.3150/11-BEJ390

Chakraborty, B. (2003). On multivariate quantile regression. *J. Statist. Plann. Inference* **110** 109–132. MR1944636 https://doi.org/10.1016/S0378-3758(01)00277-4

Chakraborty, A. and Chaudhuri, P. (2014). The spatial distribution in infinite dimensional spaces and related quantiles and depths. *Ann. Statist.* **42** 1203–1231. MR3224286 https://doi.org/10.1214/14-AOS1226

Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.* **91** 862–872. MR1395753 https://doi.org/10.2307/2291681

Chen, Z. (1996). Conditional L_p-quantiles and their application to the testing of symmetry in non-parametric regression. *Statist. Probab. Lett.* **29** 107–115. MR1411408 https://doi.org/10.1016/0167-7152(95)00163-8

1350-7265

Cheng, Y. and De Gooijer, J.G. (2007). On the uth geometric conditional quantile. *J. Statist. Plann. Inference* **137** 1914–1930. MR2323873 https://doi.org/10.1016/j.jspi.2006.02.014

Chowdhury, J. and Chaudhuri, P. (2019). Nonparametric depth and quantile regression for functional data. *Bernoulli* **25** 395–423. MR3892324 https://doi.org/10.3150/17-bej991

Daouia, A., Girard, S. and Stupfler, G. (2018). Estimation of tail risk based on extreme expectiles. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 263–292. MR3763692 https://doi.org/10.1111/rssb.12254

Daouia, A., Girard, S. and Stupfler, G. (2019). Extreme M-quantiles as risk measures: From L^1 to L^p optimization. *Bernoulli* **25** 264–309. MR3892320 https://doi.org/10.3150/17-bej987

Francisci, G., Nieto-Reyes, A. and Agostinelli, C. (2019). Generalization of the simplicial depth: No vanishment outside the convex hull of the distribution support. Preprint arXiv:1909.02739.

Gardes, L., Girard, S. and Stupfler, G. (2020). Beyond tail median and conditional tail expectation: Extreme risk estimation using tail L^p-optimization. *Scand. J. Stat.* **47** 922–949. MR4157164 https://doi.org/10.1111/sjos.12433

Ghosh, A.K. and Chaudhuri, P. (2005). On maximum depth and related classifiers. *Scand. J. Statist.* **32** 327–350. MR2188677 https://doi.org/10.1111/j.1467-9469.2005.00423.x

Girard, S. and Stupfler, G. (2015). Extreme geometric quantiles in a multivariate regular variation framework. *Extremes* **18** 629–663. MR3418771 https://doi.org/10.1007/s10687-015-0226-0

Girard, S. and Stupfler, G. (2017). Intriguing properties of extreme geometric quantiles. *REVSTAT* **15** 107–139. MR3614901 https://doi.org/10.4310/jsg.2017.v15.n1.a4

Haberman, S.J. (1989). Concavity and estimation. *Ann. Statist.* **17** 1631–1661. MR1026303 https://doi.org/10.1214/aos/1176347385

Hallin, M., Paindaveine, D. and Šiman, M. (2010). Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth. *Ann. Statist.* **38** 635–669. MR2604670 https://doi.org/10.1214/09-AOS723

Hallin, M., del Barrio, E., Cuesta-Albertos, J. and Matrán, C. (2021). Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach. *Ann. Statist.* **49** 1139–1165. MR4255122 https://doi.org/10.1214/20-aos1996

Herrmann, K., Hofert, M. and Mailhot, M. (2018). Multivariate geometric expectiles. *Scand. Actuar. J.* **7** 629–659. MR3836392 https://doi.org/10.1080/03461238.2018.1426038

Koenker, R. and Bassett, G. Jr. (1978). Regression quantiles. *Econometrica* **46** 33–50. MR0474644 https://doi.org/10.2307/1913643

Koltchinskii, V.I. (1997). M-estimation, convexity and quantiles. *Ann. Statist.* **25** 435–477. MR1439309 https://doi.org/10.1214/aos/1031833659

Konen, D. and Paindaveine, D. (2022). Supplement to "Multivariate ρ-quantiles: A spatial approach." https://doi.org/10.3150/21-BEJ1404SUPP.

Kuan, C.-M., Yeh, J.-H. and Hsu, Y.-C. (2009). Assessing value at risk with CARE, the conditional autoregressive expectile models. *J. Econometrics* **150** 261–270. MR2535521 https://doi.org/10.1016/j.jeconom.2008.12.002

Li, J., Cuesta-Albertos, J.A. and Liu, R.Y. (2012). DD-classifier: Nonparametric classification procedure based on DD-plot. *J. Amer. Statist. Assoc.* **107** 737–753. MR2980081 https://doi.org/10.1080/01621459.2012.688462

Magyar, A. and Tyler, D.E. (2011). The asymptotic efficiency of the spatial median for elliptically symmetric distributions. *Sankhya B* **73** 165–192. MR2924190 https://doi.org/10.1007/s13571-011-0032-x

Massé, J.-C. (2002). Asymptotics for the Tukey median. *J. Multivariate Anal.* **81** 286–300. MR1906382 https://doi.org/10.1006/jmva.2001.2008

Mosler, K. (2002). *Multivariate Dispersion, Central Regions and Depth: The lift zonoid approach*. *Lecture Notes in Statistics* **165**. Berlin: Springer. MR1913862 https://doi.org/10.1007/978-1-4613-0045-8

Möttönen, J., Nordhausen, K. and Oja, H. (2010). Asymptotic theory of the spatial median. In *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in Honor of Professor Jana Jurečková*. *Inst. Math. Stat. (IMS) Collect.* **7** 182–193. Beachwood, OH: IMS. MR2808378

Mukhopadhyay, N.D. and Chatterjee, S. (2011). High dimensional data analysis using multivariate generalized spatial quantiles. *J. Multivariate Anal.* **102** 768–780. MR2772334 https://doi.org/10.1016/j.jmva.2010.12.002

Newey, W.K. and Powell, J.L. (1987). Asymmetric least squares estimation and testing. *Econometrica* **55** 819–847. MR0906565 https://doi.org/10.2307/1911031

Paindaveine, D. and Virta, J. (2021). On the behavior of extreme d-dimensional spatial quantiles under minimal assumptions. In *Advances in Contemporary Statistics and Econometrics* 243–259. Cham: Springer. MR4299283 https://doi.org/10.1007/978-3-030-73249-3_13

Pokotylo, O., Mozharovskyi, P. and Dyckerhoff, R. (2019). Depth and depth-based classification with R-package ddalpha. *J. Statist. Softw.* **91** 1–46.

Serfling, R. (2002). Quantile functions for multivariate analysis: approaches and applications. *Stat. Neerl.* **56** 214–232.

Serfling, R. (2010). Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardisation. *J. Nonparametr. Stat.* **22** 915–936. MR2738875 https://doi.org/10.1080/10485250903431710

Serfling, R. (2019). Depth functions on general data spaces, I. Perspectives, with consideration of "density" and "local" depths. Submitted.

Serfling, R. and Wijesuriya, U. (2017). Depth-based nonparametric description of functional data, with emphasis on use of spatial depth. *Comput. Statist. Data Anal.* **105** 24–45. MR3552186 https://doi.org/10.1016/j.csda.2016.07.007

Serfling, R. and Zuo, Y. (2010). Discussion [MR2604670]. *Ann. Statist.* **38** 676–684. MR2604672 https://doi.org/10.1214/09-AOS723B

Taylor, J. (2008). Estimating value at risk and expected shortfall using expectiles. *J. Financ. Econometrics* **6** 231–252.

Usseglio-Carleve, A. (2018). Estimation of conditional extreme risk measures from heavy-tailed elliptical random vectors. *Electron. J. Stat.* **12** 4057–4093. MR3887178 https://doi.org/10.1214/18-EJS1499

Vardi, Y. and Zhang, C.-H. (2000). The multivariate L_1-median and associated data depth. *Proc. Natl. Acad. Sci. USA* **97** 1423–1426. MR1740461 https://doi.org/10.1073/pnas.97.4.1423

Zhou, W. and Serfling, R. (2008). Multivariate spatial U-quantiles: A Bahadur-Kiefer representation, a Theil-Sen estimator for multiple regression, and a robust dispersion estimator. *J. Statist. Plann. Inference* **138** 1660–1678. MR2400473 https://doi.org/10.1016/j.jspi.2007.05.043

Bernoulli **28**(3), 2022, 1935–1960
https://doi.org/10.3150/21-BEJ1405

On the theoretical properties of the exchange algorithm

GUANYANG WANG[a]

Department of Statistics, Rutgers University, Piscataway, NJ 08854, USA. [a]*guanyang.wang@rutgers.edu*

The exchange algorithm is one of the most popular extensions of the Metropolis–Hastings algorithm to sample from doubly-intractable distributions. However, the theoretical exploration of the exchange algorithm is very limited. For example, natural questions like 'Does exchange algorithm converge at a geometric rate?' or 'Does the exchange algorithm admit a Central Limit Theorem?' have not been answered yet. In this paper, we study the theoretical properties of the exchange algorithm, in terms of asymptotic variance and convergence speed. We compare the exchange algorithm with the original Metropolis–Hastings algorithm and provide both necessary and sufficient conditions for the geometric ergodicity of the exchange algorithm. Moreover, we prove that our results can be applied to various practical applications such as location models, Gaussian models, Poisson models, and a large class of exponential families, which includes most of the practical applications of the exchange algorithm. A central limit theorem for the exchange algorithm is also established. Our results justify the theoretical usefulness of the exchange algorithm.

Keywords: Markov chain Monte Carlo; convergence; geometrically ergodic

References

[1] Andrieu, C., Doucet, A., Yıldırım, S. and Chopin, N. (2018). On the utility of Metropolis-Hastings with asymmetric acceptance ratio. Preprint. Available at arXiv:1803.09527.

[2] 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

[3] 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

[4] Andrieu, C. and Vihola, M. (2016). Establishing some order amongst exact approximations of MCMCs. *Ann. Appl. Probab.* **26** 2661–2696. MR3563190 https://doi.org/10.1214/15-AAP1158

[5] Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B* **36** 192–236. MR0373208

[6] Besag, J.E. (1972). Nearest-neighbour systems and the auto-logistic model for binary data. *J. Roy. Statist. Soc. Ser. B* **34** 75–83. MR0323276

[7] Caimo, A. and Friel, N. (2011). Bayesian inference for exponential random graph models. *Soc. Netw.* **33** 41–55.

[8] Chan, K.S. and Geyer, C.J. (1994). Discussion: Markov chains for exploring posterior distributions. *Ann. Statist.* **22** 1747–1758.

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

[10] Cucala, L., Marin, J.-M., Robert, C.P. and Titterington, D.M. (2009). A Bayesian reassessment of nearest-neighbor classification. *J. Amer. Statist. Assoc.* **104** 263–273. MR2663042 https://doi.org/10.1198/jasa.2009.0125

[11] Diaconis, P. and Wang, G. (2018). Bayesian goodness of fit tests: A conversation for David Mumford. *Ann. Math. Sci. Appl.* **3** 287–308. MR3781267 https://doi.org/10.4310/AMSA.2018.v3.n1.a9

[12] Geyer, C.J. (1991). Markov chain Monte Carlo maximum likelihood.

[13] Geyer, C.J. (1992). Practical Markov chain Monte Carlo. *Statist. Sci.* 473–483.

[14] Habeck, M., Rudolf, D. and Sprungk, B. (2020). Stability of doubly-intractable distributions. *Electron. Commun. Probab.* **25** Paper No. 62, 13. MR4151879 https://doi.org/10.3390/mca25010013

[15] Hsieh, C.-S. and Lee, L.F. (2016). A social interactions model with endogenous friendship formation and selectivity. *J. Appl. Econometrics* **31** 301–319. MR3481364 https://doi.org/10.1002/jae.2426

[16] Hunter, J.K. and Nachtergaele, B. (2001). *Applied Analysis*. River Edge, NJ: World Scientific Co., Inc. MR1829589 https://doi.org/10.1142/4319

[17] Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. *Z. Phys.* **31** 253–258.

[18] 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

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

[20] Kipnis, C. and Varadhan, S.R.S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.* **104** 1–19. MR0834478

[21] Medina-Aguayo, F., Rudolf, D. and Schweizer, N. (2020). Perturbation bounds for Monte Carlo within Metropolis via restricted approximations. *Stochastic Process. Appl.* **130** 2200–2227. MR4074698 https://doi.org/10.1016/j.spa.2019.06.015

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

[23] Meyn, S.P. and Tweedie, R.L. (1993). *Markov Chains and Stochastic Stability*. *Communications and Control Engineering Series*. London: Springer. MR1287609 https://doi.org/10.1007/978-1-4471-3267-7

[24] Mira, A. (2001). Ordering and improving the performance of Monte Carlo Markov chains. *Statist. Sci.* **16** 340–350. MR1888449 https://doi.org/10.1214/ss/1015346319

[25] Murray, I., Ghahramani, Z. and MacKay, D.J. (2006). MCMC for doubly-intractable distributions. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence* 359–366.

[26] Nicholls, G.K., Fox, C. and Watt, A.M. (2012). Coupled MCMC with a randomized acceptance probability. Preprint. Available at arXiv:1205.6857.

[27] Park, J. and Haran, M. (2018). Bayesian inference in the presence of intractable normalizing functions. *J. Amer. Statist. Assoc.* **113** 1372–1390. MR3862364 https://doi.org/10.1080/01621459.2018.1448824

[28] Peskun, P.H. (1973). Optimum Monte-Carlo sampling using Markov chains. *Biometrika* **60** 607–612. MR0362823 https://doi.org/10.1093/biomet/60.3.607

[29] Potter, G.E., Handcock, M.S., Longini, I.M. Jr. and Halloran, M.E. (2012). Estimating within-school contact networks to understand influenza transmission. *Ann. Appl. Stat.* **6** 1–26. MR2951527 https://doi.org/10.1214/11-AOAS505

[30] Roberts, G.O. and Rosenthal, J.S. (1997). Geometric ergodicity and hybrid Markov chains. *Electron. Commun. Probab.* **2** 13–25. MR1448322 https://doi.org/10.1214/ECP.v2-981

[31] Roberts, G.O. and Rosenthal, J.S. (2008). Variance bounding Markov chains. *Ann. Appl. Probab.* **18** 1201–1214. MR2418242 https://doi.org/10.1214/07-AAP486

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

[33] Robins, G., Pattison, P., Kalish, Y. and Lusher, D. (2007). An introduction to exponential random graph ($p*$) models for social networks. *Soc. Netw.* **29** 173–191.

[34] Rosenthal, J.S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **90** 558–566. MR1340509

[35] Rudolf, D. and Ullrich, M. (2013). Positivity of hit-and-run and related algorithms. *Electron. Commun. Probab.* **18** no. 49, 8. MR3078012 https://doi.org/10.1214/ECP.v18-2507

[36] Tierney, L. (1998). A note on Metropolis-Hastings kernels for general state spaces. *Ann. Appl. Probab.* **8** 1–9. MR1620401 https://doi.org/10.1214/aoap/1027961031

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

[38] Younes, L. (1988). Estimation and annealing for Gibbsian fields. *Ann. Inst. Henri Poincaré Probab. Stat.* **24** 269–294. MR0953120

Bernoulli **28**(3), 2022, 1961–1985
https://doi.org/10.3150/21-BEJ1407

Pathwise large deviations for white noise chaos expansions

ALEXANDRE PANNIER[a]

Department of Mathematics, Imperial College London. [a]*a.pannier17@imperial.ac.uk*

We consider a family of continuous processes $\{X^\varepsilon\}_{\varepsilon>0}$ which are measurable with respect to a white noise measure, take values in the space of continuous functions $C([0,1]^d:\mathbb{R})$, and have the Wiener chaos expansion

$$X^\varepsilon = \sum_{n=0}^{\infty} \varepsilon^n I_n\big(f_n^\varepsilon\big).$$

We provide sufficient conditions for the large deviations principle of $\{X^\varepsilon\}_{\varepsilon>0}$ to hold in $C([0,1]^d:\mathbb{R})$, thereby refreshing a problem left open by Pérez–Abreu (1993) in the Brownian motion case. The proof is based on the weak convergence approach to large deviations: it involves demonstrating the convergence in distribution of certain perturbations of the original process, and thus the main difficulties lie in analysing and controlling the perturbed multiple stochastic integrals. Moreover, adopting this representation offers a new perspective on pathwise large deviations and induces a variety of applications thereof.

Keywords: Large deviations; Wiener chaos; white noise measure; weak convergence; multiple stochastic integrals; Malliavin calculus

References

[1] Azmoodeh, E. and Nourdin, I. (2019). Almost sure limit theorems on Wiener chaos: The non-central case. *Electron. Commun. Probab.* **24** Paper No. 9, 12. MR3916341 https://doi.org/10.1214/19-ECP212
[2] Bai, S. and Taqqu, M.S. (2020). Limit theorems for long-memory flows on Wiener chaos. *Bernoulli* **26** 1473–1503. MR4058375 https://doi.org/10.3150/19-BEJ1168
[3] Bell, D., Bolaños, R. and Nualart, D. (2020). Limit theorems for singular Skorohod integrals. *Theory Probab. Math. Statist.*
[4] Bessaih, H. and Millet, A. (2009). Large deviation principle and inviscid shell models. *Electron. J. Probab.* **14** 2551–2579. MR2570011 https://doi.org/10.1214/EJP.v14-719
[5] Biagini, F., Hu, Y., Øksendal, B. and Zhang, T. (2008). *Stochastic Calculus for Fractional Brownian Motion and Applications*. *Probability and Its Applications (New York)*. London: Springer. MR2387368 https://doi.org/10.1007/978-1-84628-797-8
[6] Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd ed. *Wiley Series in Probability and Statistics: Probability and Statistics*. New York: Wiley. MR1700749 https://doi.org/10.1002/9780470316962
[7] Breuer, P. and Major, P. (1983). Central limit theorems for nonlinear functionals of Gaussian fields. *J. Multivariate Anal.* **13** 425–441. MR0716933 https://doi.org/10.1016/0047-259X(83)90019-2
[8] Budhiraja, A. and Dupuis, P. (2000). A variational representation for positive functionals of infinite dimensional Brownian motion. *Probab. Math. Statist.* **20** 39–61. MR1785237
[9] Budhiraja, A. and Dupuis, P. (2019). *Analysis and Approximation of Rare Events: Representations and Weak Convergence Methods*. *Probability Theory and Stochastic Modelling* **94**. New York: Springer. MR3967100 https://doi.org/10.1007/978-1-4939-9579-0
[10] Budhiraja, A., Dupuis, P. and Maroulas, V. (2008). Large deviations for infinite dimensional stochastic dynamical systems. *Ann. Probab.* **36** 1390–1420. MR2435853 https://doi.org/10.1214/07-AOP362

1350-7265

[11] Budhiraja, A. and Song, X. (2020). Large deviation principles for stochastic dynamical systems with a fractional Brownian noise. Preprint, arXiv:2006.07683.

[12] Campese, S., Nourdin, I. and Nualart, D. (2020). Continuous Breuer-Major theorem: Tightness and nonstationarity. *Ann. Probab.* **48** 147–177. MR4079433 https://doi.org/10.1214/19-AOP1357

[13] Dalang, R., Khoshnevisan, D., Mueller, C., Nualart, D. and Xiao, Y. (2009). *A Minicourse on Stochastic Partial Differential Equations*. *Lecture Notes in Math.* **1962**. Berlin: Springer. MR1500166

[14] Delgado-Vences, F., Nualart, D. and Zheng, G. (2020). A central limit theorem for the stochastic wave equation with fractional noise. *Ann. Inst. Henri Poincaré Probab. Stat.* **56** 3020–3042. MR4164864 https://doi.org/10.1214/20-AIHP1069

[15] Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. *Applications of Mathematics (New York)* **38**. New York: Springer. MR1619036 https://doi.org/10.1007/978-1-4612-5320-4

[16] Deuschel, J.-D. and Stroock, D.W. (1989). *Large Deviations*. *Pure and Applied Mathematics* **137**. Boston, MA: Academic Press. MR0997938

[17] Dupuis, P. and Ellis, R.S. (1997). *A Weak Convergence Approach to the Theory of Large Deviations*. *Wiley Series in Probability and Statistics: Probability and Statistics*. New York: Wiley. MR1431744 https://doi.org/10.1002/9781118165904

[18] Dupuis, P. and Spiliopoulos, K. (2012). Large deviations for multiscale diffusion via weak convergence methods. *Stochastic Process. Appl.* **122** 1947–1987. MR2914778 https://doi.org/10.1016/j.spa.2011.12.006

[19] Eichelsbacher, P. and Thäle, C. (2015). Malliavin-Stein method for variance-gamma approximation on Wiener space. *Electron. J. Probab.* **20** Paper No. 123, 28. MR3425543 https://doi.org/10.1214/EJP.v20-4136

[20] Freidlin, M.I. and Wentzell, A.D. (1984). *Random Perturbations of Dynamical Systems*. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **260**. New York: Springer. MR0722136 https://doi.org/10.1007/978-1-4684-0176-9

[21] Ganguly, A. (2018). Large deviation principle for stochastic integrals and stochastic differential equations driven by infinite-dimensional semimartingales. *Stochastic Process. Appl.* **128** 2179–2227. MR3804791 https://doi.org/10.1016/j.spa.2017.09.011

[22] Garcia, J. (2008). A large deviation principle for stochastic integrals. *J. Theoret. Probab.* **21** 476–501. MR2391257 https://doi.org/10.1007/s10959-007-0136-4

[23] Harnett, D. and Nualart, D. (2014). Central limit theorem for an iterated integral with respect to fBm with $H > 1/2$. *Stochastics* **86** 187–202. MR3180032 https://doi.org/10.1080/17442508.2013.774403

[24] Huang, J., Nualart, D. and Viitasaari, L. (2020). A central limit theorem for the stochastic heat equation. *Stochastic Process. Appl.* **130** 7170–7184. MR4167203 https://doi.org/10.1016/j.spa.2020.07.010

[25] Ibragimov, I.A. (1983). Conditions for smoothness of trajectories of random functions. *Teor. Veroyatn. Primen.* **28** 229–250. MR0700208

[26] Jacquier, A. and Pannier, A. (2020). Large and moderate deviations for stochastic Volterra systems. Preprint, arXiv2004.10571.

[27] Ledoux, M. (1990). A note on large deviations for Wiener chaos. In *Séminaire de Probabilités, XXIV, 1988/89*. *Lecture Notes in Math.* **1426** 1–14. Berlin: Springer. MR1071528 https://doi.org/10.1007/BFb0083753

[28] Mayer-Wolf, E., Nualart, D. and Pérez-Abreu, V. (1992). Large deviations for multiple Wiener–Itô integral processes. In *Séminaire de Probabilités, XXVI*. *Lecture Notes in Math.* **1526** 11–31. Berlin: Springer. MR1231980 https://doi.org/10.1007/BFb0084307

[29] Mellouk, M. and Millet, A. (1998). Large deviations for stochastic flows and anticipating SDEs in Besov–Orlicz spaces. *Stoch. Stoch. Rep.* **63** 267–302. MR1658085

[30] Millet, A., Nualart, D. and Sanz, M. (1991). Composition of large deviation principles and applications. In *Stochastic Analysis* 383–395. Boston, MA: Academic Press. MR1119840

[31] Millet, A., Nualart, D. and Sanz, M. (1992). Large deviations for a class of anticipating stochastic differential equations. *Ann. Probab.* **20** 1902–1931. MR1188048

[32] Morse, M.R. and Spiliopoulos, K. (2017). Moderate deviations for systems of slow-fast diffusions. *Asymptot. Anal.* **105** 97–135. MR3729828 https://doi.org/10.3233/asy-171434

[33] Neuman, E. (2018). Pathwise uniqueness of the stochastic heat equation with spatially inhomogeneous white noise. *Ann. Probab.* **46** 3090–3187. MR3857853 https://doi.org/10.1214/17-AOP1239

[34] Nourdin, I. and Nualart, D. (2010). Central limit theorems for multiple Skorokhod integrals. *J. Theoret. Probab.* **23** 39–64. MR2591903 https://doi.org/10.1007/s10959-009-0258-y
[35] Nourdin, I., Nualart, D. and Poly, G. (2013). Absolute continuity and convergence of densities for random vectors on Wiener chaos. *Electron. J. Probab.* **18** no. 22, 19. MR3035750 https://doi.org/10.1214/EJP.v18-2181
[36] Nourdin, I. and Peccati, G. (2009). Stein's method on Wiener chaos. *Probab. Theory Related Fields* **145** 75–118. MR2520122 https://doi.org/10.1007/s00440-008-0162-x
[37] Nourdin, I. and Poly, G. (2012). Convergence in law in the second Wiener/Wigner chaos. *Electron. Commun. Probab.* **17** no. 36, 12. MR2970700 https://doi.org/10.1214/ecp.v17-2023
[38] Nualart, D. (2006). *The Malliavin Calculus and Related Topics*, 2nd ed. *Probability and Its Applications (New York)*. Berlin: Springer. MR2200233
[39] Nualart, D. and Peccati, G. (2005). Central limit theorems for sequences of multiple stochastic integrals. *Ann. Probab.* **33** 177–193. MR2118863 https://doi.org/10.1214/009117904000000621
[40] Nualart, D. and Zheng, G. (2020). Averaging Gaussian functionals. *Electron. J. Probab.* **25** Paper No. 48, 54. MR4092767 https://doi.org/10.1214/20-ejp453
[41] Pérez-Abreu, V. (1993). Chaos expansions: A review. *Resehans, Journal of IME-USP* **1** 335–359.
[42] Pérez-Abreu, V. and Tudor, C. (1994). Large deviations for a class of chaos expansions. *J. Theoret. Probab.* **7** 757–765. MR1295537 https://doi.org/10.1007/BF02214370
[43] Salins, M., Budhiraja, A. and Dupuis, P. (2019). Uniform large deviation principles for Banach space valued stochastic evolution equations. *Trans. Amer. Math. Soc.* **372** 8363–8421. MR4029700 https://doi.org/10.1090/tran/7872
[44] Freidlin, M.I. and Wentzell, A.D. (1970). On small random perturbations of dynamical systems. *Russian Math. Surveys* **25**.
[45] Walsh, J.B. (1986). An introduction to stochastic partial differential equations. In *École D'été de Probabilités de Saint-Flour, XIV—1984*. *Lecture Notes in Math.* **1180** 265–439. Berlin: Springer. MR0876085 https://doi.org/10.1007/BFb0074920
[46] Zhang, X. (2008). Euler schemes and large deviations for stochastic Volterra equations with singular kernels. *J. Differential Equations* **244** 2226–2250. MR2413840 https://doi.org/10.1016/j.jde.2008.02.019

Bernoulli **28**(3), 2022, 1986–2011
https://doi.org/10.3150/21-BEJ1408

An adaptive multiple-try Metropolis algorithm

SIMON FONTAINE[1,2,a] and MYLÈNE BÉDARD[1]

[1]*Département de mathématiques et de statistique, Université de Montréal, 2920 chemin de la Tour, Montréal, QC, Canada, H3T 1J4.*
[2]*Department of Statistics, University of Michigan, West Hall, 1085 South University, Ann Arbor, MI, USA, 48109.*
[a]*simfont@umich.edu*

Markov chain Monte Carlo (MCMC) methods, specifically samplers based on random walks, often have difficulty handling target distributions with complex geometry such as multi-modality. We propose an adaptive multiple-try Metropolis algorithm designed to tackle such problems by combining the flexibility of multiple-proposal samplers with the user-friendliness and optimality of adaptive algorithms. We prove the ergodicity of the resulting Markov chain with respect to the target distribution using common techniques in the adaptive MCMC literature. In a Bayesian model for loss of heterozygosity in cancer cells, we find that our method outperforms traditional adaptive samplers, non-adaptive multiple-try Metropolis samplers, and various more sophisticated competing methods.

Keywords: Adaptive scaling; ergodicity; limit theorem; loss of heterozygosity; multiple candidates; random walk sampler; robust adaptation

References

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

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

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

Atchadé, Y.F. and Rosenthal, J.S. (2005). On adaptive Markov chain Monte Carlo algorithms. *Bernoulli* **11** 815–828. MR2172842 https://doi.org/10.3150/bj/1130077595

Bai, Y., Craiu, R.V. and Di Narzo, A.F. (2011). Divide and conquer: A mixture-based approach to regional adaptation for MCMC. *J. Comput. Graph. Statist.* **20** 63–79. Supplementary material available online. MR2816538 https://doi.org/10.1198/jcgs.2010.09035

Barrett, M.T., Galipeau, P.C., Sanchez, C.A., Emond, M.J. and Reid, B.J. (1996). Determination of the frequency of loss of heterozygosity in esophageal adenocarcinoma by cell sorting, whole genome amplification and microsatellite polymorphisms. *Oncogene* **12** 1873–1878.

Bédard, M. (2007). Weak convergence of Metropolis algorithms for non-i.i.d. target distributions. *Ann. Appl. Probab.* **17** 1222–1244. MR2344305 https://doi.org/10.1214/105051607000000096

Bédard, M. (2008a). Efficient sampling using Metropolis algorithms: Applications of optimal scaling results. *J. Comput. Graph. Statist.* **17** 312–332. MR2439962 https://doi.org/10.1198/108571108X319970

Bédard, M. (2008b). Optimal acceptance rates for Metropolis algorithms: Moving beyond 0.234. *Stochastic Process. Appl.* **118** 2198–2222. MR2474348 https://doi.org/10.1016/j.spa.2007.12.005

Bédard, M., Douc, R. and Moulines, E. (2012). Scaling analysis of multiple-try MCMC methods. *Stochastic Process. Appl.* **122** 758–786. MR2891436 https://doi.org/10.1016/j.spa.2011.11.004

Bédard, M., Douc, R. and Moulines, E. (2014). Scaling analysis of delayed rejection MCMC methods. *Methodol. Comput. Appl. Probab.* **16** 811–838. MR3270597 https://doi.org/10.1007/s11009-013-9326-y

Bédard, M. and Rosenthal, J.S. (2008). Optimal scaling of Metropolis algorithms: Heading toward general target distributions. *Canad. J. Statist.* **36** 483–503. MR2532248 https://doi.org/10.1002/cjs.5550360401

Beskos, A., Roberts, G. and Stuart, A. (2009). Optimal scalings for local Metropolis-Hastings chains on nonproduct targets in high dimensions. *Ann. Appl. Probab.* **19** 863–898. MR2537193 https://doi.org/10.1214/08-AAP563

Beskos, A., Pillai, N., Roberts, G., Sanz-Serna, J.-M. and Stuart, A. (2013). Optimal tuning of the hybrid Monte Carlo algorithm. *Bernoulli* **19** 1501–1534. MR3129023 https://doi.org/10.3150/12-BEJ414

Breyer, L.A., Piccioni, M. and Scarlatti, S. (2004). Optimal scaling of MaLa for nonlinear regression. *Ann. Appl. Probab.* **14** 1479–1505. MR2071431 https://doi.org/10.1214/105051604000000369

Breyer, L.A. and Roberts, G.O. (2000). From Metropolis to diffusions: Gibbs states and optimal scaling. *Stochastic Process. Appl.* **90** 181–206. MR1794535 https://doi.org/10.1016/S0304-4149(00)00041-7

Casarin, R., Craiu, R. and Leisen, F. (2013). Interacting multiple try algorithms with different proposal distributions. *Stat. Comput.* **23** 185–200. MR3016937 https://doi.org/10.1007/s11222-011-9301-9

Chen, H.F., Lei, G. and Gao, A.J. (1988). Convergence and robustness of the Robbins-Monro algorithm truncated at randomly varying bounds. *Stochastic Process. Appl.* **27** 217–231. MR0931029 https://doi.org/10.1016/0304-4149(87)90039-1

Craiu, R.V. and Lemieux, C. (2007). Acceleration of the Multiple-Try Metropolis algorithm using antithetic and stratified sampling. *Stat. Comput.* **17** 109–120. MR2380640 https://doi.org/10.1007/s11222-006-9009-4

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

Craiu, R.V., Gray, L., Łatuszyński, K., Madras, N., Roberts, G.O. and Rosenthal, J.S. (2015). Stability of adversarial Markov chains, with an application to adaptive MCMC algorithms. *Ann. Appl. Probab.* **25** 3592–3623. MR3404645 https://doi.org/10.1214/14-AAP1083

Desai, M. (2000). Mixture Models for Genetic Changes in Cancer Cells. Ph.D. thesis. Univ. Washington.

Fontaine, S. (2019). MCMC adaptatifs à essais multiples. Master's thesis. Department of Mathematics and Statistics, Univ. Montreal.

Fontaine, S. and Bédard, M. (2022). Supplement to "An adaptive multiple-try Metropolis algorithm." https://doi.org/10.3150/21-BEJ1408SUPPA, https://doi.org/10.3150/21-BEJ1408SUPPB, https://doi.org/10.3150/21-BEJ1408SUPPC

Gelman, A., Roberts, G.O. and Gilks, W.R. (1996). Efficient Metropolis jumping rules. In *Bayesian Statistics, 5 (Alicante, 1994). Oxford Sci. Publ.* 599–607. New York: Oxford Univ. Press. MR1425429

Grenon-Godbout, N. and Bédard, M. (2021). On-line partitioning of the sample space in the regional adaptive algorithm. *Canad. J. Statist.* **49** 238–261. MR4267920 https://doi.org/10.1002/cjs.11562

Haario, H., Saksman, E. and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli* **7** 223–242. MR1828504 https://doi.org/10.2307/3318737

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109. MR3363437 https://doi.org/10.1093/biomet/57.1.97

Liu, J.S., Liang, F. and Wong, W.H. (2000). The multiple-try method and local optimization in Metropolis sampling. *J. Amer. Statist. Assoc.* **95** 121–134. MR1803145 https://doi.org/10.2307/2669532

Martino, L. and Read, J. (2013). On the flexibility of the design of multiple try Metropolis schemes. *Comput. Statist.* **28** 2797–2823. MR3141364 https://doi.org/10.1007/s00180-013-0429-2

Martino, L., Casarin, R., Leisen, F. and Luengo, D. (2018). Adaptive independent sticky MCMC algorithms. *EURASIP J. Adv. Signal Process.* **2018** 5.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1092.

Meyn, S. and Tweedie, R.L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge: Cambridge Univ. Press. With a prologue by Peter W. Glynn. MR2509253 https://doi.org/10.1017/CBO9780511626630

Pillai, N.S., Stuart, A.M. and Thiéry, A.H. (2012). Optimal scaling and diffusion limits for the Langevin algorithm in high dimensions. *Ann. Appl. Probab.* **22** 2320–2356. MR3024970 https://doi.org/10.1214/11-AAP828

Robbins, H. and Monro, S. (1951). A stochastic approximation method. *Ann. Math. Stat.* **22** 400–407. MR0042668 https://doi.org/10.1214/aoms/1177729586

Robert, C.P. and Casella, G. (2004). *Monte Carlo Statistical Methods*, 2nd ed. *Springer Texts in Statistics*. New York: Springer. MR2080278 https://doi.org/10.1007/978-1-4757-4145-2

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

Roberts, G.O. and Rosenthal, J.S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 255–268. MR1625691 https://doi.org/10.1111/1467-9868.00123

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

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

Rosenthal, J.S. and Yang, J. (2018). Ergodicity of combocontinuous adaptive MCMC algorithms. *Methodol. Comput. Appl. Probab.* **20** 535–551. MR3803934 https://doi.org/10.1007/s11009-017-9574-3

Saksman, E. and Vihola, M. (2010). On the ergodicity of the adaptive Metropolis algorithm on unbounded domains. *Ann. Appl. Probab.* **20** 2178–2203. MR2759732 https://doi.org/10.1214/10-AAP682

Sejdinovic, D., Strathmann, H., Garcia, M.L., Andrieu, C. and Gretton, A. (2014). Kernel adaptive Metropolis-Hastings. In *International Conference on Machine Learning* 1665–1673.

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

Sherlock, C. and Roberts, G. (2009). Optimal scaling of the random walk Metropolis on elliptically symmetric unimodal targets. *Bernoulli* **15** 774–798. MR2555199 https://doi.org/10.3150/08-BEJ176

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.* **22** 1701–1762. With discussion and a rejoinder by the author. MR1329166 https://doi.org/10.1214/aos/1176325750

Tran, M.-N., Pitt, M.K. and Kohn, R. (2016). Adaptive Metropolis-Hastings sampling using reversible dependent mixture proposals. *Stat. Comput.* **26** 361–381. MR3439379 https://doi.org/10.1007/s11222-014-9509-6

VanDerwerken, D.N. and Schmidler, S.C. (2013). Parallel Markov chain Monte Carlo. ArXiv Preprint. Available at arXiv:1312.7479.

Vats, D., Flegal, J.M. and Jones, G.L. (2019). Multivariate output analysis for Markov chain Monte Carlo. *Biometrika* **106** 321–337. MR3949306 https://doi.org/10.1093/biomet/asz002

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

Vihola, M. (2012). Robust adaptive Metropolis algorithm with coerced acceptance rate. *Stat. Comput.* **22** 997–1008. MR2950080 https://doi.org/10.1007/s11222-011-9269-5

Warnes, G.R. (2001). The Normal Kernel Coupler: An Adaptive Markov Chain Monte Carlo Method for Efficiently Sampling from Multi-modal Distributions. Ph.D. thesis. Univ. Washington.

Yang, J., Levi, E., Craiu, R.V. and Rosenthal, J.S. (2019). Adaptive component-wise multiple-try Metropolis sampling. *J. Comput. Graph. Statist.* **28** 276–289. MR3974879 https://doi.org/10.1080/10618600.2018.1513365

Bernoulli **28**(3), 2022, 2012–2033
https://doi.org/10.3150/21-BEJ1409

Exact convergence analysis of the independent Metropolis-Hastings algorithms

GUANYANG WANG[a]

Department of Statistics, Rutgers University, Piscataway, NJ 08854, USA . [a]*guanyang.wang@rutgers.edu*

A well-known difficult problem regarding Metropolis-Hastings algorithms is to get sharp bounds on their convergence rates. Moreover, a fundamental but often overlooked problem in Markov chain theory is to study the convergence rates for different initializations. In this paper, we study the two issues mentioned above of the Independent Metropolis-Hastings (IMH) algorithms on both general and discrete state spaces. We derive the exact convergence rate and prove that the IMH algorithm's different deterministic initializations have the same convergence rate. We get the exact convergence speed for IMH algorithms on general state spaces.

Keywords: Independent Metropolis-Hastings; Markov chain Monte Carlo; exact convergence rate

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] 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
[3] 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
[4] Atchadé, Y.F. and Perron, F. (2007). On the geometric ergodicity of Metropolis-Hastings algorithms. *Statistics* **41** 77–84. MR2303970 https://doi.org/10.1080/10485250601033214
[5] Bayer, D. and Diaconis, P. (1992). Trailing the dovetail shuffle to its lair. *Ann. Appl. Probab.* **2** 294–313. MR1161056
[6] Corcoran, J.N. and Tweedie, R.L. (2002). Perfect sampling from independent Metropolis-Hastings chains. *J. Statist. Plann. Inference* **104** 297–314. MR1906013 https://doi.org/10.1016/S0378-3758(01)00243-9
[7] Diaconis, P., Khare, K. and Saloff-Coste, L. (2008). Gibbs sampling, exponential families and orthogonal polynomials. *Statist. Sci.* **23** 151–178. MR2446500 https://doi.org/10.1214/07-STS252
[8] Diaconis, P. and Saloff-Coste, L. (1993). Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.* **3** 696–730. MR1233621
[9] Diaconis, P. and Saloff-Coste, L. (1996). Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.* **6** 695–750. MR1410112 https://doi.org/10.1214/aoap/1034968224
[10] Diaconis, P. and Shahshahani, M. (1981). Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete* **57** 159–179. MR0626813 https://doi.org/10.1007/BF00535487
[11] Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** 398–409. MR1141740
[12] Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **6** 721–741.
[13] Geyer, C.J. (1992). Practical Markov chain Monte Carlo. *Statist. Sci.* 473–483.
[14] Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109. MR3363437 https://doi.org/10.1093/biomet/57.1.97
[15] Holden, L., Hauge, R. and Holden, M. (2009). Adaptive independent Metropolis-Hastings. *Ann. Appl. Probab.* **19** 395–413. MR2498682 https://doi.org/10.1214/08-AAP545

[16] Jones, G.L. and Comment, A.A.J. (2008). Gibbs sampling, exponential families, and orthogonal polynomials. *Statist. Sci.* **23** 183–186.

[17] Jones, G.L. and Hobert, J.P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statist. Sci.* **16** 312–334. MR1888447 https://doi.org/10.1214/ss/1015346317

[18] Lee, C. and Neal, P. (2018). Optimal scaling of the independence sampler: Theory and practice. *Bernoulli* **24** 1636–1652. MR3757511 https://doi.org/10.3150/16-BEJ908

[19] Levin, D.A. and Peres, Y. (2017). *Markov Chains and Mixing Times*. Providence, RI: Amer. Math. Soc. MR3726904 https://doi.org/10.1090/mbk/107

[20] Liu, J.S. (1996). Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Stat. Comput.* **6** 113–119.

[21] Lubetzky, E. and Sly, A. (2021). Fast initial conditions for Glauber dynamics. *Probab. Theory Related Fields* **181** 647–667. MR4341083 https://doi.org/10.1007/s00440-020-01015-3

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

[23] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1092.

[24] Meyn, S.P. and Tweedie, R.L. (2012). *Markov Chains and Stochastic Stability*. Berlin: Springer.

[25] Middleton, L., Deligiannidis, G., Doucet, A. and Jacob, P.E. (2019). Unbiased smoothing using particle independent Metropolis-Hastings. In *The 22nd International Conference on Artificial Intelligence and Statistics* 2378–2387. PMLR.

[26] Murdoch, D.J. and Green, P.J. (1998). Exact sampling from a continuous state space. *Scand. J. Stat.* **25** 483–502. MR1650023 https://doi.org/10.1111/1467-9469.00116

[27] Persi Diaconis, Khare, K. and Saloff-Coste, L. (2008). Gibbs sampling, exponential families and orthogonal polynomials. *Statist. Sci.* **23** 196–200.

[28] Propp, J.G. and Wilson, D.B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures Algorithms* **9** 223–252.

[29] Qin, Q. and Hobert, J.P. (2021). On the limitations of single-step drift and minorization in Markov chain convergence analysis. *Ann. Appl. Probab.* **31** 1633–1659. MR4312841 https://doi.org/10.1214/20-aap1628

[30] Roberts, G.O. and Rosenthal, J.S. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv.* **1** 20–71. MR2095565 https://doi.org/10.1214/154957804100000024

[31] Roberts, G.O. and Rosenthal, J.S. (2011). Quantitative non-geometric convergence bounds for independence samplers. *Methodol. Comput. Appl. Probab.* **13** 391–403. MR2788864 https://doi.org/10.1007/s11009-009-9157-z

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

[33] Rosenthal, J.S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **90** 558–566. MR1340509

[34] Smith, R.L. and Tierney, L. (1996). Exact transition probabilities for the independence Metropolis sampler.

[35] Tan, Z. (2006). Monte Carlo integration with acceptance-rejection. *J. Comput. Graph. Statist.* **15** 735–752. MR2291270 https://doi.org/10.1198/106186006X142681

[36] Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.* **22** 1701–1762. MR1329166 https://doi.org/10.1214/aos/1176325750

[37] Wang, A.Q., Pollock, M., Roberts, G.O. and Steinsaltz, D. (2021). Regeneration-enriched Markov processes with application to Monte Carlo. *Ann. Appl. Probab.* **31** 703–735. MR4254493 https://doi.org/10.1214/20-aap1602

Bernoulli **28**(3), 2022, 2034–2066
https://doi.org/10.3150/21-BEJ1410

Locally polynomial Hilbertian additive regression

JEONG MIN JEON[1,a], YOUNG KYUNG LEE[2,b], ENNO MAMMEN[3,c] and BYEONG U. PARK[4,d]

[1]*KU Leuven, Leuven, Belgium.* [a]*jeongmin.jeon@kuleuven.be*
[2]*Kangwon National University, Chuncheon, South Korea.* [b]*youngklee@kangwon.ac.kr*
[3]*Universität Heidelberg, Heidelberg, Germany.* [c]*mammen@math.uni-heidelberg.de*
[4]*Seoul National University, Seoul, South Korea.* [d]*bupark@snu.ac.kr*

In this paper a new additive regression technique is developed for response variables that take values in general Hilbert spaces. The proposed method is based on the idea of smooth backfitting that has been developed mainly for real-valued responses. The local polynomial smoothing device is adopted, which renders various advantages of the technique evidenced in the classical univariate kernel regression with real-valued responses. It is demonstrated that the new technique eliminates many limitations which existing methods are subject to. In contrast to the existing techniques, the proposed approach is equipped with the estimation of the derivatives as well as the regression function itself, and provides options to make the estimated regression function free from boundary effects and possess oracle properties. A comprehensive theory is presented for the proposed method, which includes the rates of convergence in various modes and the asymptotic distributions of the estimators. The efficiency of the proposed method is also demonstrated via simulation study and is illustrated through real data applications.

Keywords: Additive model; Hilbert space; local polynomial smoothing; smooth backfitting; non-Euclidean data

References

[1] 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
[2] Bosq, D. (2000). *Linear Processes in Function Spaces*. *Lecture Notes in Statistics* **149**. New York: Springer. MR1783138 https://doi.org/10.1007/978-1-4612-1154-9
[3] Buja, A., Hastie, T. and Tibshirani, R. (1989). Linear smoothers and additive models. *Ann. Statist.* **17** 453–555. MR0994249 https://doi.org/10.1214/aos/1176347115
[4] Cohn, D.L. (2013). *Measure Theory*, 2nd ed. *Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks]*. New York: Birkhäuser/Springer. MR3098996 https://doi.org/10.1007/978-1-4614-6956-8
[5] Egozcue, J.J. and Pawlowsky-Glahn, V. (2019). Compositional data: The sample space and its structure. *TEST* **28** 599–638. MR3992128 https://doi.org/10.1007/s11749-019-00670-6
[6] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. *Monographs on Statistics and Applied Probability* **66**. London: CRC Press. MR1383587
[7] 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
[8] 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
[9] Hron, K., Menafoglio, A., Templ, M., Hruzová, K. and Filzmoser, P. (2016). Simplicial principal component analysis for density functions in Bayes spaces. *Comput. Statist. Data Anal.* **94** 330–350. MR3412829 https://doi.org/10.1016/j.csda.2015.07.007
[10] Jayasumana, S., Salzmann, M., Li, H. and Harandi, M. (2014). A framework for shape analysis via Hilbert space embedding. Available at arXiv:1412.417v1.

1350-7265

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

[12] Jeon, J.M., Lee, Y.K., Mammen, E. and Park, B.U. (2022). Supplement to "Locally polynomial Hilbertian additive regression." https://doi.org/10.3150/21-BEJ1410SUPP

[13] Kokoszka, P., Miao, H., Petersen, A. and Shang, H.L. (2019). Forecasting of density functions with an application to cross-sectional and intraday returns. *Int. J. Forecast.* **35** 1304–1317.

[14] 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

[15] 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

[16] 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

[17] Linton, O. and Nielsen, J.P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82** 93–100. MR1332841 https://doi.org/10.1093/biomet/82.1.93

[18] 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

[19] 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

[20] Mammen, E. and Sperlich, S. (2021). Backfitting tests in generalised structured models. *Biometrika*. in print.

[21] McNeil, J., Farris, M.S., Ruan, Y., Merry, H., Lynch, B.M., Matthews, C.E., Courneya, K.S. and Friedenreich, C.M. (2018). Effects of prescribed aerobic exercise volume on physical activity and sedentary time in postmenopausal women: A randomized controlled trial. *Int. J. Behav. Nutr. Phys. Act.* **15** 27. https://doi.org/10.1186/s12966-018-0659-3

[22] Nielsen, J.P. and Sperlich, S. (2005). Smooth backfitting in practice. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 43–61. MR2136638 https://doi.org/10.1111/j.1467-9868.2005.00487.x

[23] Pate, R.R., Stevens, J., Webber, L.S., Dowda, M., Murray, D.M., Young, D.R. and Going, S. (2009). Age-related change in physical activity in adolescent girls. *J. Adolesc. Health* **44** 275–282.

[24] Pawlowsky-Glahn, V. and Buccianti, A. (2011). *Compositional Data Analysis: Theory and Applications*. Chichester: Wiley.

[25] 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

[26] Petersen, A. and Müller, H.-G. (2019). Wasserstein covariance for multiple random densities. *Biometrika* **44** 183–218.

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

[28] Roca-Pardiñas, J. and Sperlich, S. (2010). Feasible estimation in generalized structured models. *Stat. Comput.* **20** 367–379. MR2725394 https://doi.org/10.1007/s11222-009-9130-2

[29] Steeves, J.A., Tudor-Locke, C., Murphy, R.A., King, G.A., Fitzhugh, E.C. and Harris, T.B. (2015). Classification of occupational activity categories using accelerometry: NHANES 2003-2004. *Int. J. Behav. Nutr. Phys. Act.* **12** 89. https://doi.org/10.1186/s12966-015-0235-z

[30] Talská, R., Menafoglio, A., Machalová, J., Hron, K. and Fišerová, E. (2018). Compositional regression with functional response. *Comput. Statist. Data Anal.* **123** 66–85.

[31] Wahid, A., Manek, N., Nichols, M., Kelly, P., Foster, C., Webster, P., Kaur, A., Smith, C.F., Wilkins, E., Rayner, M., Roberts, N. and Scarborough, P. (2016). Quantifying the association between physical activity and cardiovascular disease and diabetes: A systematic review and meta-analysis. *J. Am. Heart Assoc.* **5** e002495.

[32] Wang, J.-L., Chiou, J.-M. and Müller, H.-G. (2016). Functional data analysis. *Annu. Rev. Stat. Appl.* **3** 257–295.

[33] Wondergem, R., Pisters, M.F., Heijmans, M.W., Wouters, E.J.M., de Bie, R.A., Veenhof, C. and Visser-Meily, J.M.A. (2020). Movement behavior remains stable in stroke survivors within the first two months after returning home. *PLoS ONE* **15** e0229587.

[34] 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

[35] 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

Bernoulli **28**(3), 2022, 2067–2094
https://doi.org/10.3150/21-BEJ1411

Degenerate competing three-particle systems

TOMOYUKI ICHIBA[1,a] and IOANNIS KARATZAS[2,b]

[1]*Department of Statistics and Applied Probability, South Hall, University of California, Santa Barbara, CA 93106, USA.* [a]*ichiba@pstat.ucsb.edu*
[2]*Department of Mathematics, Columbia University, New York, NY 10027, USA.* [b]*ik1@columbia.edu*

We study systems of three interacting particles, in which drifts and variances are assigned by rank. These systems are degenerate: the variances corresponding to one or two ranks can vanish, so the corresponding ranked motions become ballistic rather than diffusive. Depending on which ranks are allowed to "go ballistic" the systems exhibit markedly different behavior, which we study here in some detail. Also studied are stability properties for the resulting planar process of gaps between successive ranks.

Keywords: Competing particle systems; local times; reflected planar Brownian motion; triple collisions; structure of filtrations

References

[1] Akahori, J., Izumi, M. and Watanabe, S. (2009). Noises, stochastic flows and E_0-semigroups [MR2354290]. In *Selected Papers on Probability and Statistics*. *Amer. Math. Soc. Transl. Ser. 2* **227** 1–23. Providence, RI: Amer. Math. Soc. MR2553243 https://doi.org/10.1090/trans2/227/01

[2] Azéma, J., Kaplan-Duflo, M. and Revuz, D. (1967). Mesure invariante sur les classes récurrentes des processus de Markov. *Z. Wahrsch. Verw. Gebiete* **8** 157–181. MR0222955 https://doi.org/10.1007/BF00531519

[3] Banner, A.D. and Ghomrasni, R. (2008). Local times of ranked continuous semimartingales. *Stochastic Process. Appl.* **118** 1244–1253. MR2428716 https://doi.org/10.1016/j.spa.2007.08.001

[4] Bass, R.F. and Pardoux, É. (1987). Uniqueness for diffusions with piecewise constant coefficients. *Probab. Theory Related Fields* **76** 557–572. MR0917679 https://doi.org/10.1007/BF00960074

[5] Bruggeman, C. and Sarantsev, A. (2018). Multiple collisions in systems of competing Brownian particles. *Bernoulli* **24** 156–201. MR3706753 https://doi.org/10.3150/16-BEJ869

[6] Burdzy, K., Kang, W. and Ramanan, K. (2009). The Skorokhod problem in a time-dependent interval. *Stochastic Process. Appl.* **119** 428–452. MR02493998 https://doi.org/10.1016/j.spa.2008.03.001

[7] Chernyĭ, A.S. (2001). On strong and weak uniqueness for stochastic differential equations. *Teor. Veroyatn. Primen.* **46** 483–497. MR1978664 https://doi.org/10.1137/S0040585X97979093

[8] Dembo, A., Shkolnikov, M., Varadhan, S.R.S. and Zeitouni, O. (2016). Large deviations for diffusions interacting through their ranks. *Comm. Pure Appl. Math.* **69** 1259–1313. MR3503022 https://doi.org/10.1002/cpa.21640

[9] Douc, R., Fort, G. and Guillin, A. (2009). Subgeometric rates of convergence of f-ergodic strong Markov processes. *Stochastic Process. Appl.* **119** 897–923. MR2499863 https://doi.org/10.1016/j.spa.2008.03.007

[10] Dupuis, P. and Williams, R.J. (1994). Lyapunov functions for semimartingale reflecting Brownian motions. *Ann. Probab.* **22** 680–702. MR1288127

[11] Engelbert, H.J. (1991). On the theorem of T. Yamada and S. Watanabe. *Stoch. Stoch. Rep.* **36** 205–216. MR1128494 https://doi.org/10.1080/17442509108833718

[12] Fernholz, E.R. (2010). A question regarding Brownian paths. Technical Report, INTECH Investment Management. Princeton, New Jersey.

[13] Fernholz, E.R. (2011). Time reversal for an $n = 3$ degenerate system. Technical Report, INTECH Investment Management. Princeton, New Jersey.

[14] Fernholz, E.R., Ichiba, T. and Karatzas, I. (2013). Two Brownian particles with rank-based characteristics and skew-elastic collisions. *Stochastic Process. Appl.* **123** 2999–3026. MR3062434 https://doi.org/10.1016/j.spa.2013.03.019

[15] Fernholz, E.R., Ichiba, T., Karatzas, I. and Prokaj, V. (2013). Planar diffusions with rank-based characteristics and perturbed Tanaka equations. *Probab. Theory Related Fields* **156** 343–374. MR3055262 https://doi.org/10.1007/s00440-012-0430-7
[16] Harrison, J.M. (2013). *Brownian Models of Performance and Control*. Cambridge: Cambridge Univ. Press. MR3157450 https://doi.org/10.1017/CBO9781139087698
[17] Harrison, J.M. and Reiman, M.I. (1981). Reflected Brownian motion on an orthant. *Ann. Probab.* **9** 302–308. MR0606992
[18] Harrison, J.M. and Williams, R.J. (1987). Brownian models of open queueing networks with homogeneous customer populations. *Stochastics* **22** 77–115. MR0912049 https://doi.org/10.1080/17442508708833469
[19] Harrison, J.M. and Williams, R.J. (1987). Multidimensional reflected Brownian motions having exponential stationary distributions. *Ann. Probab.* **15** 115–137. MR0877593
[20] Has'minskiĭ, R.Z. (1960). Ergodic properties of recurrent diffusion processes and stabilization of the solution of the Cauchy problem for parabolic equations. *Teor. Veroyatn. Primen.* **5** 196–214. MR0133871
[21] Hobson, D.G. and Rogers, L.C.G. (1993). Recurrence and transience of reflecting Brownian motion in the quadrant. *Math. Proc. Cambridge Philos. Soc.* **113** 387–399. MR1198420 https://doi.org/10.1017/S0305004100076040
[22] Ichiba, T. and Karatzas, I. (2010). On collisions of Brownian particles. *Ann. Appl. Probab.* **20** 951–977. MR2680554 https://doi.org/10.1214/09-AAP641
[23] Ichiba, T. and Karatzas, I. (2014). Skew-unfolding the Skorokhod reflection of a continuous semimartingale. In *Stochastic Analysis and Applications 2014*. *Springer Proc. Math. Stat.* **100** 349–376. Cham: Springer. MR3332719 https://doi.org/10.1007/978-3-319-11292-3_13
[24] Ichiba, T. and Karatzas, I. (2020). Degenerate Competing Three-Particle Systems. arXiv:2006.04970 [v2].
[25] Ichiba, T., Karatzas, I., Prokaj, V. and Yan, M. (2018). Stochastic integral equations for Walsh semimartingales. *Ann. Inst. Henri Poincaré Probab. Stat.* **54** 726–756. MR3795064 https://doi.org/10.1214/16-AIHP819
[26] Ichiba, T., Karatzas, I. and Shkolnikov, M. (2013). Strong solutions of stochastic equations with rank-based coefficients. *Probab. Theory Related Fields* **156** 229–248. MR3055258 https://doi.org/10.1007/s00440-012-0426-3
[27] Ichiba, T., Papathanakos, V., Banner, A., Karatzas, I. and Fernholz, R. (2011). Hybrid atlas models. *Ann. Appl. Probab.* **21** 609–644. MR2807968 https://doi.org/10.1214/10-AAP706
[28] Jourdain, B. and Reygner, J. (2013). Propogation of chaos for rank-based interacting diffusions and long time behaviour of a scalar quasilinear parabolic equation. *Stoch. Partial Differ. Equ. Anal. Comput.* **1** 455–506. MR3327514 https://doi.org/10.1007/s40072-013-0014-2
[29] Karatzas, I., Pal, S. and Shkolnikov, M. (2016). Systems of Brownian particles with asymmetric collisions. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 323–354. MR3449305 https://doi.org/10.1214/14-AIHP646
[30] Kella, O. and Ramasubramanian, S. (2012). Asymptotic irrelevance of initial conditions for Skorohod reflection mapping on the nonnegative orthant. *Math. Oper. Res.* **37** 301–312. MR2931282 https://doi.org/10.1287/moor.1120.0538
[31] Kolli, P. and Shkolnikov, M. (2018). SPDE limit of the global fluctuations in rank-based models. *Ann. Probab.* **46** 1042–1069. MR3773380 https://doi.org/10.1214/17-AOP1200
[32] Kruk, L., Lehoczky, J., Ramanan, K. and Shreve, S. (2007). An explicit formula for the Skorokhod map on $[0,a]$. *Ann. Probab.* **35** 1740–1768. MR2349573 https://doi.org/10.1214/009117906000000890
[33] Le Jan, Y. and Raimond, O. (2004). Flows, coalescence and noise. *Ann. Probab.* **32** 1247–1315. MR2060298 https://doi.org/10.1214/009117904000000207
[34] Le Jan, Y. and Raimond, O. (2004). Sticky flows on the circle and their noises. *Probab. Theory Related Fields* **129** 63–82. MR2052863 https://doi.org/10.1007/s00440-003-0324-9
[35] Meyn, S.P. and Tweedie, R.L. (1993). Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.* **25** 518–548. MR1234295 https://doi.org/10.2307/1427522
[36] Pal, S. and Pitman, J. (2008). One-dimensional Brownian particle systems with rank-dependent drifts. *Ann. Appl. Probab.* **18** 2179–2207. MR2473654 https://doi.org/10.1214/08-AAP516
[37] Prokaj, V. (2009). Unfolding the Skorohod reflection of a semimartingale. *Statist. Probab. Lett.* **79** 534–536. MR2494646 https://doi.org/10.1016/j.spl.2008.09.029
[38] Reiman, M.I. and Williams, R.J. (1988). A boundary property of semimartingale reflecting Brownian motions. *Probab. Theory Related Fields* **77** 87–97. MR0921820 https://doi.org/10.1007/BF01848132

[39] Revuz, D. and Yor, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **293**. Berlin: Springer. MR1725357 https://doi.org/10.1007/978-3-662-06400-9

[40] Sarantsev, A. (2015). Triple and simultaneous collisions of competing Brownian particles. *Electron. J. Probab.* **20** no. 29, 28. MR3325099 https://doi.org/10.1214/EJP.v20-3279

[41] Sarantsev, A. (2017). Reflected Brownian motion in a convex polyhedral cone: Tail estimates for the stationary distribution. *J. Theoret. Probab.* **30** 1200–1223. MR3687255 https://doi.org/10.1007/s10959-016-0674-8

[42] Stroock, D.W. and Varadhan, S.R.S. (1979). *Multidimensional Diffusion Processes*. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **233**. Berlin: Springer. MR0532498

[43] Tsirelson, B. (1997). Triple points: From non-Brownian filtrations to harmonic measures. *Geom. Funct. Anal.* **7** 1096–1142. MR1487755 https://doi.org/10.1007/s000390050038

[44] Varadhan, S.R.S. and Williams, R.J. (1985). Brownian motion in a wedge with oblique reflection. *Comm. Pure Appl. Math.* **38** 405–443. MR0792398 https://doi.org/10.1002/cpa.3160380405

[45] Watanabe, S. (2000). The stochastic flow and the noise associated to Tanaka's stochastic differential equation. *Ukraïn. Mat. Zh.* **52** 1176–1193. MR1816931 https://doi.org/10.1023/A:1010319800544

[46] Williams, R.J. (1995). Semimartingale reflecting Brownian motions in the orthant. In *Stochastic Networks*. *IMA Vol. Math. Appl.* **71** 125–137. New York: Springer. MR1381009 https://doi.org/10.1007/978-1-4757-2418-9_7

Bernoulli **28**(3), 2022, 2095–2121
https://doi.org/10.3150/21-BEJ1414

Degrees of freedom for off-the-grid sparse estimation

CLARICE POON[1,a] and GABRIEL PEYRÉ[2,b]

[1]*Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK.* [a]*cmhsp20@bath.ac.uk*
[2]*CNRS and DMA, PSL University, Ecole Normale Supérieure, 45 rue d'Ulm, F-75230 PARIS cedex 05, France.*
[b]*gabriel.peyre@ens.fr*

The degrees of freedom quantify the number of parameters of an estimator and is central to various risk minimization procedures. Its computation and properties are well understood when dealing with finite dimensional linear models, possibly regularized using sparsity through the celebrated Lasso estimator. However, for some applications (such as super-resolution in imaging or training multi-layer perceptrons with a single hidden layer) it makes sense to rather consider "continuous" methods. For these applications, these methods avoid the discretization of the parameter space and lead to more efficient numerical solvers. Training these continuous models with a sparsity inducing prior can be achieved by solving a convex optimization problem over the infinite dimensional space of measures, which is often called the Beurling Lasso (Blasso), and is the continuous counterpart of the Lasso. Previous works (*Ann. Statist.* **35** (2007) 2173–2192; *Statist. Sinica* **23** (2013) 809–828) show that the size of the smallest solution support is an unbiased estimator of the degrees of freedom of the Lasso. Our main contribution is a proof of a continuous counterpart to this result for the Blasso. In contrast to the Lasso, our new formula shows that the empirical degrees of freedom of the Blasso is not proportional to the size of the support. Our second contribution is a detailed study of the case of sampling Fourier coefficients in 1D, which corresponds to a super-resolution problem. We show that our formula for the unbiased estimation of the degrees of freedom is valid outside of a set of measure zero of observations, which in turn justifies its use to compute an unbiased estimator of the prediction risk using the Stein Unbiased Risk Estimator (SURE). We also report numerical results for both the case of Fourier sampling and the learning of a multilayers perceptron with a single hidden layer.

Keywords: Degrees of freedom; Lasso; Beurling lasso; off-the-grid

References

[1] Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*. *Springer Series in Statistics. Perspectives in Statistics*. New York: Springer.

[2] Azaïs, J.-M., de Castro, Y. and Gamboa, F. (2015). Spike detection from inaccurate samplings. *Appl. Comput. Harmon. Anal.* **38** 177–195. MR3303671 https://doi.org/10.1016/j.acha.2014.03.004

[3] Bach, F. (2017). Breaking the curse of dimensionality with convex neutral networks. *J. Mach. Learn. Res.* **18** Paper No. 19, 53. MR3634886

[4] Boyd, N., Hastie, T., Boyd, S., Recht, B. and Jordan, M.I. (2017). Saturating splines and feature selection. *J. Mach. Learn. Res.* **18** Paper No. 197, 32. MR3827085

[5] Boyd, N., Schiebinger, G. and Recht, B. (2017). The alternating descent conditional gradient method for sparse inverse problems. *SIAM J. Optim.* **27** 616–639. MR3634995 https://doi.org/10.1137/15M1035793

[6] Bredies, K. and Pikkarainen, H.K. (2013). Inverse problems in spaces of measures. *ESAIM Control Optim. Calc. Var.* **19** 190–218. MR3023066 https://doi.org/10.1051/cocv/2011205

[7] Candès, E.J. and Fernandez-Granda, C. (2013). Super-resolution from noisy data. *J. Fourier Anal. Appl.* **19** 1229–1254. MR3132912 https://doi.org/10.1007/s00041-013-9292-3

[8] Candès, E.J. and Fernandez-Granda, C. (2014). Towards a mathematical theory of super-resolution. *Comm. Pure Appl. Math.* **67** 906–956. MR3193963 https://doi.org/10.1002/cpa.21455

[9] Candès, E.J., Sing-Long, C.A. and Trzasko, J.D. (2013). Unbiased risk estimates for singular value thresholding and spectral estimators. *IEEE Trans. Signal Process.* **61** 4643–4657. MR3105401 https://doi.org/10.1109/TSP.2013.2270464
[10] Chen, X., Lin, Q. and Sen, B. (2020). On degrees of freedom of projection estimators with applications to multivariate nonparametric regression. *J. Amer. Statist. Assoc.* **115** 173–186. MR4078455 https://doi.org/10.1080/01621459.2018.1537917
[11] Chizat, L. (2019). Sparse optimization on measures with over-parameterized gradient descen. arXiv:1907.10300.
[12] Chizat, L. and Bach, F. (2018). On the global convergence of gradient descent for over-parameterized models using optimal transport. In *Advances in Neural Information Processing Systems* 3036–3046.
[13] Coste, M. (1999). An Introduction to O-minimal Geometry.
[14] Coste, M. (2000). An Introduction to Semialgebraic Geometry. Citeseer.
[15] de Castro, Y. and Gamboa, F. (2012). Exact reconstruction using Beurling minimal extrapolation. *J. Math. Anal. Appl.* **395** 336–354. MR2943626 https://doi.org/10.1016/j.jmaa.2012.05.011
[16] Denoyelle, Q., Duval, V., Peyré, G. and Soubies, E. (2020). The sliding Frank-Wolfe algorithm and its application to super-resolution microscopy. *Inverse Probl.* **36** 014001, 42. MR4040984 https://doi.org/10.1088/1361-6420/ab2a29
[17] Donoho, D.L. and Johnstone, I.M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90** 1200–1224. MR1379464
[18] Dossal, C., Kachour, M., Fadili, M.J., Peyré, G. and Chesneau, C. (2013). The degrees of freedom of the Lasso for general design matrix. *Statist. Sinica* **23** 809–828. MR3086657
[19] Duval, V. and Peyré, G. (2015). Exact support recovery for sparse spikes deconvolution. *Found. Comput. Math.* **15** 1315–1355. MR3394712 https://doi.org/10.1007/s10208-014-9228-6
[20] Duval, V. and Peyré, G. (2017). Sparse regularization on thin grids I: The Lasso. *Inverse Probl.* **33** 055008, 29. MR3628904 https://doi.org/10.1088/1361-6420/aa5e12
[21] Efron, B. (1986). How biased is the apparent error rate of a prediction rule? *J. Amer. Statist. Assoc.* **81** 461–470. MR0845884
[22] Efron, B. (2004). The estimation of prediction error: Covariance penalties and cross-validation. *J. Amer. Statist. Assoc.* **99** 619–642. MR2090899 https://doi.org/10.1198/016214504000000692
[23] Eldar, Y.C. (2009). Generalized SURE for exponential families: Applications to regularization. *IEEE Trans. Signal Process.* **57** 471–481. MR2603376 https://doi.org/10.1109/TSP.2008.2008212
[24] Fisher, S.D. and Jerome, J.W. (1975). Spline solutions to L^1 extremal problems in one and several variables. *J. Approx. Theory* **13** 73–83. MR0361577 https://doi.org/10.1016/0021-9045(75)90016-7
[25] Foucart, S. and Rauhut, H. (2013). *A Mathematical Introduction to Compressive Sensing*. *Applied and Numerical Harmonic Analysis*. New York: Birkhäuser/Springer. MR3100033 https://doi.org/10.1007/978-0-8176-4948-7
[26] Giryes, R., Elad, M. and Eldar, Y.C. (2011). The projected GSURE for automatic parameter tuning in iterative shrinkage methods. *Appl. Comput. Harmon. Anal.* **30** 407–422. MR2784572 https://doi.org/10.1016/j.acha.2010.11.005
[27] Golub, G.H., Heath, M. and Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* **21** 215–223. MR0533250 https://doi.org/10.2307/1268518
[28] Hudson, H.M. (1978). A natural identity for exponential families with applications in multiparameter estimation. *Ann. Statist.* **6** 473–484. MR0467991
[29] Hwang, J.T. (1982). Improving upon standard estimators in discrete exponential families with applications to Poisson and negative binomial cases. *Ann. Statist.* **10** 857–867. MR0663437
[30] Johnstone, I.M. (1999). Wavelet shrinkage for correlated data and inverse problems: Adaptivity results. *Statist. Sinica* **9** 51–83. MR1678881
[31] Kato, K. (2009). On the degrees of freedom in shrinkage estimation. *J. Multivariate Anal.* **100** 1338–1352. MR2514133 https://doi.org/10.1016/j.jmva.2008.12.002
[32] Künsch, H.R. (1994). Robust priors for smoothing and image restoration. *Ann. Inst. Statist. Math.* **46** 1–19. MR1272743 https://doi.org/10.1007/BF00773588
[33] Liu, D.C. and Nocedal, J. (1989). On the limited memory BFGS method for large scale optimization. *Math. Program.* **45** 503–528. MR1038245 https://doi.org/10.1007/BF01589116

[34] Lorentz, R.A. (2000). Multivariate Hermite interpolation by algebraic polynomials: A survey *J. Comput. Appl. Math.* **122** 167–201.
[35] Mallows, C.L. (1973). Some comments on c_p. *Technometrics* **15** 661–675.
[36] Mammen, E. and van de Geer, S. (1997). Locally adaptive regression splines. *Ann. Statist.* **25** 387–413. MR1429931 https://doi.org/10.1214/aos/1034276635
[37] Meyer, M. and Woodroofe, M. (2000). On the degrees of freedom in shape-restricted regression. *Ann. Statist.* **28** 1083–1104. MR1810920 https://doi.org/10.1214/aos/1015956708
[38] Mikkelsen, F.R. and Hansen, N.R. (2018). Degrees of freedom for piecewise Lipschitz estimators. *Ann. Inst. Henri Poincaré Probab. Stat.* **54** 819–841. MR3795067 https://doi.org/10.1214/17-AIHP822
[39] Poon, C. and Peyré, G. (2022). Supplement to "Degrees of freedom for off-the-grid sparse estimation." https://doi.org/10.3150/21-BEJ1414SUPP
[40] Poon, C., Keriven, N. and Peyré, G. (2018). The geometry of off-the-grid compressed sensing. Preprint, arXiv:1802.08464.
[41] Ramani, S., Blu, T. and Unser, M. (2008). Monte-Carlo sure: A black-box optimization of regularization parameters for general denoising algorithms. *IEEE Trans. Image Process.* **17** 1540–1554. MR2517100 https://doi.org/10.1109/TIP.2008.2001404
[42] Rosset, S., Swirszcz, G., Srebro, N. and Zhu, J. (2007). ℓ_1 regularization in infinite dimensional feature spaces. In *Learning Theory*. *Lecture Notes in Computer Science* **4539** 544–558. Berlin: Springer. MR2397611 https://doi.org/10.1007/978-3-540-72927-3_39
[43] Rosset, S., Zhu, J. and Hastie, T. (2003/04). Boosting as a regularized path to a maximum margin classifier. *J. Mach. Learn. Res.* **5** 941–973. MR2248005
[44] Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464. MR0468014
[45] Stein, C.M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. MR0630098
[46] Tibshirani, R.J. (2014). Adaptive piecewise polynomial estimation via trend filtering. *Ann. Statist.* **42** 285–323. MR3189487 https://doi.org/10.1214/13-AOS1189
[47] Tibshirani, R.J. (2015). Degrees of freedom and model search. *Statist. Sinica* **25** 1265–1296. MR3410308
[48] Tibshirani, R.J. and Rosset, S. (2019). Excess optimism: How biased is the apparent error of an estimator tuned by SURE? *J. Amer. Statist. Assoc.* **114** 697–712. MR3963173 https://doi.org/10.1080/01621459.2018.1429276
[49] Tibshirani, R.J. and Taylor, J. (2012). Degrees of freedom in lasso problems. *Ann. Statist.* **40** 1198–1232. MR2985948 https://doi.org/10.1214/12-AOS1003
[50] Vaiter, S., Deledalle, C., Fadili, J., Peyré, G. and Dossal, C. (2017). The degrees of freedom of partly smooth regularizers. *Ann. Inst. Statist. Math.* **69** 791–832. MR3671639 https://doi.org/10.1007/s10463-016-0563-z
[51] Wilkie, A.J. (1996). Model completeness results for expansions of the ordered field of real numbers by restricted Pfaffian functions and the exponential function. *J. Amer. Math. Soc.* **9** 1051–1094. MR1398816 https://doi.org/10.1090/S0894-0347-96-00216-0
[52] Zou, H., Hastie, T. and Tibshirani, R. (2007). On the "degrees of freedom" of the lasso. *Ann. Statist.* **35** 2173–2192. MR2363967 https://doi.org/10.1214/009053607000000127
[53] Zou, H. and Yuan, M. (2008). Regularized simultaneous model selection in multiple quantiles regression. *Comput. Statist. Data Anal.* **52** 5296–5304. MR2526595 https://doi.org/10.1016/j.csda.2008.05.013